Empowering Users To Reclaim Their Online Data Privacy

Written by: Sruthi Rangarajan, Brooke Simon, and Nikhil Suresh

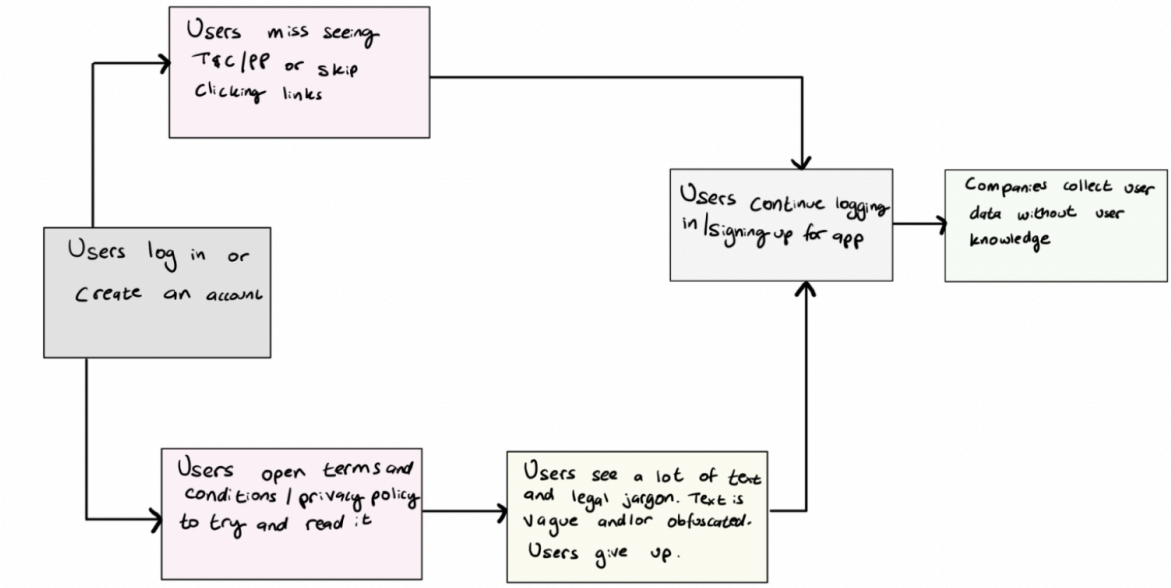
No matter what application you are signing up for, in today's day and age there will most always be some terms and policies that you implicitly agree to just by making an account. These policies often give the company expansive access and control over user data, and yet based on our research most users never even look at them. These privacy policies are typically filled with dense, legal jargon, making it difficult for people to fully grasp how their personal data is being stored and used. This lack of transparency often results in a sense of ignorance, where users blindly accept terms without truly understanding their implications. Through our user research, we determined that the primary issue is the disconnect between the legalistic language of these documents and the average user's need for clarity. Because the policies are so long and the language so obscure, understanding how their policies impact user data and user privacy would take an absurdly long time for people without legal or technical backgrounds. Motivated by this lack of accessible information, we set out to design a privacy policy application that simplifies these long, obfuscated documents. By breaking them down into clear digestible summaries and privacy ratings, we empower users to make more informed decisions about their online presence.

Although privacy policies are prevalent in apps and websites of many types, we chose to focus on major social media applications such as Instagram, Tik Tok, Facebook, and Twitter. This was motivated by the vast quantities of data that users upload and consume on these platforms, often without thinking about the implications. This data includes interacting with other users and their content as well as posting their own content. In addition to this, social media companies are well known for their exploitation of user data for their own gain. Many lawsuits have been brought against these major companies for their misuse of user data, such as a lawsuit against Facebook in which user data was used by third party analytics companies without user permission. Another lawsuit brought against Twitter by the Federal Trade Commission prosecuted Twitter for misusing user data, which they claimed was being collected for security reasons. In reality, the FTC says that it was exploiting this data for commercial use. By understanding what they agree to when signing up for these companies, users will have a more robust knowledge of the dangers they risk when using social media. Our hope is that by providing clear, accessible summaries of these privacy policies, users will be able to make informed decisions about their digital presence, understand how their data is being used, and take proactive steps to protect their privacy.

Before beginning prototyping and design of our project, we worked with users to gain a better understanding of their needs and desires when it came to social media privacy policies. The goal of this research was to find out if current users care about their data privacy, determine the optimal way to facilitate their understanding of the privacy policies, and determine what aspects of data privacy are most important to people. In order to conduct this research we recruited adults who use social media across varying ages, gender identities, and social

backgrounds. We also aimed to include people with varying use of social media, ranging from frequent users to infrequent users.

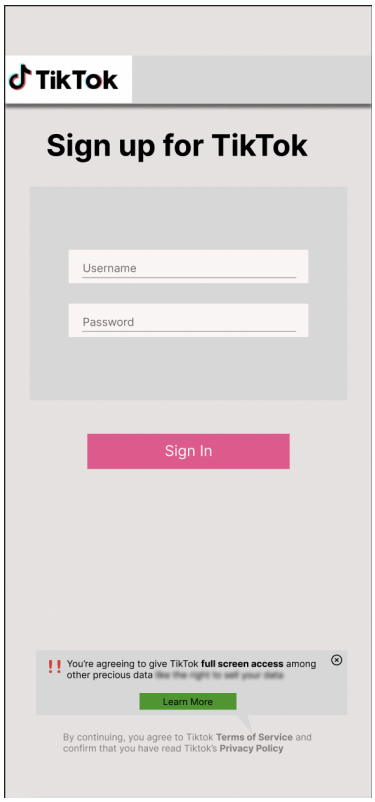
To answer our research questions, we interviewed ten participants either on zoom or in person and conducted four think-aloud studies. During the semi-structured interviews, participants described their prior knowledge, opinions, and interactions with privacy policies on social media websites. They also discussed their social media use, current levels of concern surrounding their data privacy, and if they would be interested in learning more about their data privacy. During the think-aloud studies, participants read through a privacy policy (Meta or Tik Tok) and vocalized their inner monologue while reading through it. We have included our process map below, which identifies the two scenarios that users go through, while signing up for a social media website.

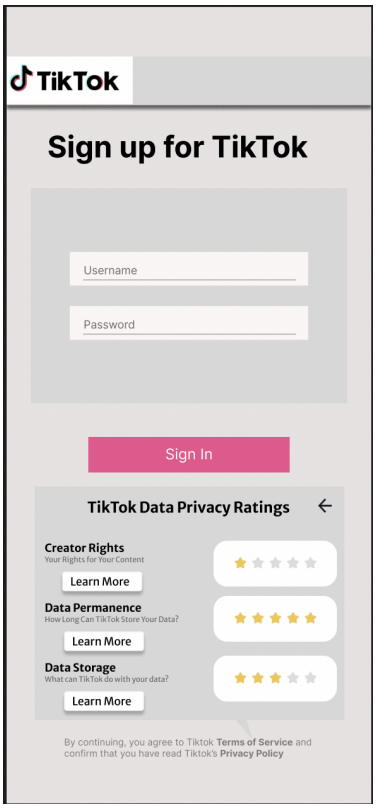


Upon analyzing the results of this study we found that users can be categorized based on three main areas: the extent to which they care about data privacy, their willingness to change their actions, and their knowledge about data privacy. From these we developed two initial personas. The first is users who care somewhat about data privacy and have some knowledge, but are unwilling to change their use of social media platforms. Several of these kinds of users mentioned that they were more likely to care about sensitive data such as medical and financial data. The second persona we identified is users who care about the impact of these policies, but want to be more informed before deciding if they would change their behavior. This analysis of our user research further motivates our project, as we feel confident that presenting privacy policies in a more digestible format will enable these users to make more informed decisions. Feedback from a crit workshop helped us refine our goal further. Many people were proponents of the idea, but mentioned that users may still decide to use social media apps. Given this, they

may not care enough about the privacy policy to read it. Ultimately, we realized that while users may keep using the platforms, having more knowledge empowers them to make informed decisions about what to share or withhold. Grounded by this refined motivation, we have also been able to iterate designing a system that facilitates user engagement and understanding with more clarity on what matters to users.

In order to refine our design goals, we tested a paper prototype with the help of several peers. The paper prototype was designed as a popup that would appear on an iphone screen when users arrived at the sign-up page for a social media platform. The popup invited users to view a summary of the privacy policy, and then showed them ratings out of five for three different categories: creator rights, data permanence, and data storage. The ratings were rendered as stars similar to the format used when reviewing a product, as seen in the prototype. The users also had a button they could click to 'learn more' about each category. We then refined our paper prototype by making a mockup on Figma, and asking users to test that as well. Here are some pictures from our Figma mockup.





Testers stated that they liked the pop-up format and appreciated that the rating system quickly captured if the platform was 'good' or 'bad' in a visual manner, but mentioned that they were confused about what it meant to have one star versus two, three, four, or five. Along with this they mentioned that a star rating system may not be the best visual choice, and suggested a scale of some kind instead. They also voiced concerns about the use of large language models

and their capacity to hallucinate or misinterpret text. From this feedback we were able to craft two main design goals for our final system.

The first design goal is to employ visualizations as much as possible to represent the impact of the policy on data privacy. All testers noted that they liked the visualization provided by the stars and hoped to see that iterated upon in future designs. Because many users largely only care about data privacy to a minimal extent, making the design of the popup as visually engaging as possible draws people in and captures attention. It also makes it easy to understand from just a cursory glance without having to spend too much time on the popup.

The second design goal is to make the information as concise and accessible as possible to help users understand the information being presented quickly and easily. Our user studies suggested that the legal language of the policies was a pain point, so simplifying and shortening it will benefit those users. This was also evidenced in our prototype testing, where our testers noted that having less text made them more likely to look at the popup.

Our system is designed to automatically pull out, analyze, and present the summaries of the privacy policies to users in a comprehensible and accessible format. When the user visits the sign-up page of a social media website, our backend instantly scrapes the privacy policy of the website and passes it through Gemini Lite, a large language model capable of understanding and summarizing complex text, which in this case would be the privacy policy. On the first pass, the model compares the scraped policy with respect to three key categories – medical/financial information access, data storage, and data sharing with third-party providers – and generates a structured summary regarding what exactly is explicitly stated by the policy. This ensures that our system accurately captures the most critical information from the original document in the respective categories. Once this summary is generated, we pass these summaries through Gemini Lite again, but with a different prompt. Rather than summarizing the policy, the second pass extracts the most alarming or essential facts. This is the one sentence that users most need to be aware of. Along with this, our software provides a 1 to 5 privacy risk rating, indicating how alarming the policy is. This two-step process ensures the summaries are actionable and informative, enabling users to immediately understand the risks of registering on a given platform.

Once processed through the large language models, this information is sent back to our front end to be displayed on a fully developed interface that is designed to allow the key points to be easily identifiable at a glance. The highlight of this popup is to visually differentiate between different sections of the privacy policy into clearly labeled sections. For example, users are able to see categories like "Access to Medical/Financial Data" which explains how the site handles sensitive health or financial-related information, or "Data Storage" which explains for how long a user's data is stored and whether it can be deleted. Under each category, there is a brief but informative summary in simple language to help users understand better. In order to further facilitate the information being easily comprehensible at a glance, we employed a color-coded rating scale that graphically illustrates the level of concern for each of the policy areas.

This rating system utilizes a red-to-green scale, where red signifies high-risk or objectionable policies, yellow/orange indicates moderately objectionable content, and green indicates user-friendly or transparent policies. With this visual cue, users can determine at a glance which sections of the privacy policy are most objectionable without having to read through the whole document. This helps users instantly recognize the degree of privacy implications without needing to trudge through dense legal jargon. Further, each category also has a "Learn More" button, which allows users to expand the section to read a more detailed explanation if they are interested in knowing more. This way, even though the summaries are short and easy to read, those who want more background information can choose to open and read it without being bogged down. One of the most important aspects of our system is the way it ensures user contribution.

In contrast to typical privacy policies that are buried in the terms and conditions of a website, our summaries are displayed in a persistent popup on the sign-up page itself. The popup remains on the screen until the user actively closes it, so at least they have to see the privacy information before signing up. The majority of individuals ignore privacy policies completely because they are long and difficult to read, but by surfacing clear, concise, and relevant information at the exact time when they are making the decision, our system radically improves transparency.

To evaluate our system, we focused on two key questions: first, whether the popup effectively educates people about data privacy in a dynamic and engaging way, and second, whether users care enough about the information for the system to influence their behavior. For the first question, we examined whether the information was well-organized, relevant, and critical, as well as whether the system allowed users to choose the level of depth they wanted when engaging with the content. For the second question, we explored whether users found the information compelling enough to impact their decision-making, if they preferred this format over traditional privacy policies, and whether they would use the system across different types of websites.

To answer these questions, we conducted user studies that combined qualitative and quantitative analyses. Participants were asked to interact with the popup, explore its features, and provide feedback through semi-structured interviews. We performed thematic analysis on user responses, categorizing feedback into key themes. In addition to qualitative results, we also tracked engagement data by logging the amount of time users spent on various sections of the popup, including the "Learn More" buttons and total time spent on the system. Using a Python script, we analyzed this data to identify trends, correlations, and engagement patterns. Through statistical analysis, we examined whether user behavior followed predictable patterns and explored factors that influenced how long participants engaged with different sections of the popup.

Our qualitative evaluation revealed several key insights into the efficacy of our system. Overall, people appreciated the information provided by the popup and the manner in which it was provided. Thirteen of the fifteen participants said that the text was easy to understand and

that the popup was visually appealing. People especially noted that having a rating displayed in a visual way helped draw them in: “the colorbar and rating was good for a high level overview”. People also noted that they liked the rating and visual color bar because they thought that people were unlikely to read the summaries or text, especially the longer ‘learn more’ summary.
In addition to this, fourteen out of fifteen participants said that they felt more informed about the privacy policy after reading our popup. However, it is important to note that many participants stated that this is because they had little to no prior knowledge about privacy policies.

Although the popup effectively educated people about privacy policy, seven of the fifteen people mentioned that they were confused about how the privacy policy ratings were being determined. It was unclear to them what “bad” versus “good” meant in this context: “The main thing [to be improved] would be making the scale at the bottom more clear.”. Some users also mentioned wanting references back to the original privacy policy so they could see more detail about the claims being made, or ensure consistency between the claims and the actual policy: “I wish that there was a direct quote from the privacy policy…”.

There was some debate amongst users as to whether or not the popup would cause them to change their actions towards social media. A common theme amongst users was that they were more likely to change their behavior when using ‘random’ apps and not the mainstream ones such as Instagram and Tik Tok. This sentiment was motivated by the idea of disposability and diversity of choice: “I would mainly use the extension on more random sites where if their data security is bad I’ll just not use it. I’m always going to use sites like instagram and twitter, big social media sites.”

Our quantitative data supported the claim made by evaluation participants that users were less likely to read the text, especially the ‘learn more’ text. Our analysis showed that on average users spent 24% less time on the third learn more section than the first. This suggests that while users initially engage deeply, their attention declines as they move through the content. A strong correlation ($r > 0.79$) was observed between time spent on "Learn More" sections and total time on the app, indicating that users who engaged with one section in depth were more likely to explore other areas as well

Cluster analysis revealed two distinct user groups: those who spent less than one minute engaging with the popup (skimmers) and those who spent over one minute exploring the content in greater depth. Regression analysis gave us an R^2 value of 0.94, which can indicate that the time spent before scrolling and time spent on individual sections were strong predictors of total engagement. These findings show that while some users engaged deeply with the content, others preferred a quick overview, reinforcing the importance of providing multiple levels of information depth.

Our hope is that although many users will still use big social media sites, they might change the way they interact with these sites by limiting the amount of content or kinds of content they post. Including concise summaries, visual scores, and expandable information allowed users to interact at their preferred level of depth, making privacy information more consumable. Users did feel more informed after using the popup and appreciated having control

over how much they read. That being said, our evaluations also suggest that while transparency is valued, it does not necessarily result in instant behavioral change. Some users were aware of the importance of privacy policies but said they would continue using these top platforms irrespective of the information provided. Users did mention that it would be more useful for evaluating websites that are known less, where these privacy concerns would influence their decisions more.

Looking ahead, we also plan to expand the functionality of the system in several key ways. One of our primary future directions is user customization, allowing individuals to select which specific aspects of a privacy policy they are most interested in learning about. By providing customizable filters, we can ensure that users receive insights that align with their personal privacy concerns. We also plan to test different LLM models to determine which one performs best in analyzing and summarizing privacy policies. While Gemini Lite has proven to be effective, evaluating other models, such as OpenAI's GPT4 or Claude 3.7 Sonnet, could help improve accuracy, reduce hallucinations, and enhance the validity and conciseness of summaries. Finally, we aim to expand beyond social media platforms and apply our system to a broader range of websites. Currently, our application works on major social media apps, like Instagram and Facebook, but privacy concerns extend far beyond these platforms. Users frequently create accounts for websites that do financial services, portals for job applications, and healthcare platforms, all of which have complex and often opaque data policies. By broadening our scope to include any website requiring account creation, we can provide valuable privacy insights in a wider range of digital interactions.

Through continued system improvement based on user feedback, improving personalization capabilities, optimizing AI model selection, and expanding the range of sites covered, we hope to make privacy policy transparency a standard feature of the internet. Through these future directions, our goal remains the same: to bridge the gap between complex legal policies and user awareness, empowering people to make more informed decisions about their digital privacy.

References:

https://www.nytimes.com/2023/04/20/business/facebook-settlement-apply.html
https://www.ftc.gov/legal-library/browse/cases-proceedings/2023062-twitter-inc-us-v