Thesis Dissertation

Privacy Nutrition Label: Real-Time Transparency for Personal Data Processing on Web Platforms

Chrysanthi Theocharidi

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

May 2026

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

Privacy Nutrition Label: Real-Time Transparency for Personal Data Processing on Web Platforms

Chrysanthi Theocharidi

Supervisors

Dr. George A. Papadopoulos

Mrs. Evangelia Vanezi

Thesis submitted in partial fulfilment of the requirements for the award of degree of bachelor's in computer science at University of Cyprus

May 2022

I declare that this dissertation and any accompanying software are my own original work, conducted in full compliance with the University of Cyprus Code of Conduct for Research (https://www.ucy.ac.cy/legislation/kodikas-deontologias-kai-politikes/), and with full accountability on my part for the accuracy, integrity, and validity of all content.

Acknowledgments

The moment I am writing this, I am one step closer to completing my diploma thesis and completing a long road to acquire my degree in Computer Science at the University of Cyprus. Five years have gone by, with many hardships, many enjoyable moments, and the acquisition of information that will lead me to my next path.

First and foremost, I'd like to express my gratitude to Evangelia Vanezi, my dissertation supervisor, who remained by my side throughout the process, offering me any direction and help I needed. Moreover, I'd also like to convey my gratitude to all the professors who provided me with invaluable knowledge throughout those years.

Furthermore, I would like to thank my Scouting family, my leaders, my friends and especially my little cubs, who provided me with happiness, love and support.
In addition, I want to express my gratitude to my family, who have always stood by me, especially my parents and siblings, who have given me everything and helped me become the person I am today.

Finally, I must express my gratitude to my friends, my second family, who have always been supportive throughout the journey of my undergraduate degree.

Abstract

In today's digital ecosystem, personal data is collected, processed, and shared by web platforms in ways that are often opaque to users. Individuals frequently interact with websites without fully understanding how their data such as names, email addresses, payment information, and browsing behavior is used, stored, or transmitted. Existing privacy policies and labels provide limited transparency, and users rarely consult or comprehend them, which creates significant privacy risks and undermines trust in online services.

This thesis addresses this gap by designing and implementing a Privacy Nutrition Label browser extension that provides real-time, user-centered transparency about personal data processing during everyday interactions with websites. The extension monitors userprovided data, network transmissions, DOM manipulations, and storage events, translating technical observations into a clear, actionable label interface. The tool's design draws on literature regarding privacy communication, dashboards, and prior privacy label work, and is informed by a user study exploring perceptions of awareness, trust, and control over personal data.

The usability and effectiveness of the tool were evaluated through a structured user study incorporating questionnaires inspired by the UEQ framework. Results demonstrate that real-time visibility of data processing improves users' understanding, increases their sense of control, and positively influences trust in web platforms. By combining real-time monitoring with an intuitive label interface, this work contributes a practical, legally informed, and user-centered solution to enhancing transparency and privacy awareness online.

Contents

Chapter 1

1. Introduction

1.1 General Idea

The main goal of this thesis is to address the lack of transparency that web platform users face regarding the collection and processing of their personal data during everyday online interactions. While users routinely provide personal information through registration, login, and other interactive forms, they rarely have a clear or immediate understanding of what data is being collected, how it is processed, and by whom, at the moment this processing occurs.

Current privacy notices are typically presented as lengthy legal documents that are difficult to read, hard to interpret, and disconnected from the real-time use of web platforms. As a result, users often provide consent without meaningful awareness or comprehension of the underlying data practices. This thesis explores an alternative approach: making personal data processing observable, understandable, and visible in real time, directly while users interact with web platforms.

To achieve this, the thesis proposes the concept of a Real-Time Privacy Nutrition Label, a visual and structured representation of ongoing data processing activities. Inspired by the clarity of food nutrition labels, this approach aims to summarize complex technical data

practices into a familiar format that users can quickly understand without requiring technical or legal expertise.

1.2 Motivation

In today's digital ecosystem, personal data has become a central asset for web platforms. Users continuously share information such as identifiers, e-mail addresses, and payment details, often without fully understanding the extent or implications of this data processing. Although regulations like the General Data Protection Regulation (GDPR) require transparency, informed consent, and purpose limitation, these legal obligations are commonly fulfilled through static privacy policies that users rarely read or understand [1].

Several studies have shown that users tend to ignore privacy policies due to their length, complexity, and legal language, resulting in low levels of awareness and informed decision making [2]. Even when privacy dashboards or consent banners are present, they often provide only high-level or declarative information rather than reflecting what happens in real time during platform usage, creating a gap between formal compliance and practical user understanding [3].

The motivation behind this thesis lies in addressing this gap. Instead of expecting users to interpret abstract legal texts, fully understand them, trust them blindly, and remember every detail, the work aims to support them with clear, timely, and contextual information about personal data processing. By making data collection practices visible at the moment they occur, users may be better equipped to understand, reflect on, and potentially act upon privacy related information.

Furthermore, this thesis is motivated by the need to balance transparency with usability. Any proposed solution must be informative without overwhelming users, and technically feasible without disrupting the normal operation of web platforms. The Real-Time Privacy Nutrition Label seeks to achieve this balance by combining principles from privacy law, user experience design, and web technologies.

1.3 About the Tool

The core outcome of this thesis is the design and implementation of a browser-based Privacy Nutrition Dashboard that monitors personal data processing activities in real time while users interact with web platforms. The tool observes browser-level events related to user-provided personal data, including form inputs, network requests to first- and third party endpoints, storage operations, and reuse of data within the page interface. These events are then translated into a structured, nutrition label–style interface.

The dashboard translates technical details into high level categories, using clear terminology, visual indicators, and color coding to communicate levels of data collection and sharing. Rather than blocking or altering website behavior, the tool focuses on informing and educating users, allowing them to see how their data is processed as it happens.

The system operates entirely on the client side, without transmitting user data to external servers, thereby adhering to privacy by design principles. Its purpose is not to replace existing privacy mechanisms, but to complement them by offering an additional layer of transparency that is immediate, understandable, and user centered.

By adopting the familiar concept of a “nutrition label,” the tool aims to lower the cognitive barrier associated with privacy information and enable users to make more informed judgments about the platforms they use [4].

1.4 Structure of the Thesis

This thesis is structured as follows:
Chapter 1 introduces the research problem, the motivation behind the work, and the main objectives of the thesis.
Chapter 2 presents the background and related work, covering privacy and personal data concepts, GDPR principles, existing approaches for real-time monitoring of data processing, and prior efforts to improve user awareness, comprehension, and trust in web platforms.

Chapter 3 presents a user study conducted during this thesis, investigating how users perceive the current state of personal data processing on web platforms, focusing on awareness, comprehension, and trust in the absence of real-time monitoring tools. The findings of this study validate the research problem and provide a baseline for comparison, as well as insights that inform the design of the proposed solution.

Chapter 4 describes the design and implementation of a browser extension for transparent personal data processing monitoring. It translates the findings from Chapter 3 into system requirements, presents the design and prototyping of the Privacy Nutrition Label, explains how data processing actions are recognized in real time, and details the development and functionalities of the tool.

Chapter 5 evaluates the effectiveness of the proposed tool through user-based evaluation, examining whether the tool improves user awareness, comprehension, trust, and ease of use.

Finally, Chapter 6 concludes the thesis and outlines directions for future work.

Chapter 2

2. Background & Related Work

2.1 Background

To provide the necessary context for this research, this section introduces the fundamental concepts related to privacy and personal data protection in the web ecosystem, the mechanisms used by web platforms to collect and process personal data, and the regulatory frameworks that govern these practices, with particular focus on the General Data Protection Regulation (GDPR). In addition, this section presents the concept of privacy nutrition labels, which aim to improve transparency by communicating privacy practices to users in a simplified and standardized format.

2.1.1 Privacy and Personal Data Protection in the Web Ecosystem

In the modern digital ecosystem, privacy and the protection of personal data have become critical concerns for both individuals and organizations. The widespread use of web platforms, social networks, and online services has resulted in an unprecedented scale of personal data collection. Websites routinely collect various types of personal information, including names, email addresses, browsing behavior, location data, and purchasing history, often as part of usual user interactions such as account creation, online purchases, or the use of personalized services.

At the same time, the number of internet users and online accounts has increased dramatically. According to recent global statistics, there are over 5 billion internet users worldwide, many of whom maintain multiple online accounts across various platforms such as social media, e-commerce websites, and digital services (Statista, 2023). Each interaction with these platforms generates data that may be collected, stored, and analysed by web services. Consequently, users often leave behind extensive digital traces during their everyday online activities.

Despite the benefits provided by data driven services, the large-scale collection and processing of personal data raise important privacy concerns. Users frequently lack awareness of what information is being collected about them and how it is being used. Studies in the field of usable privacy have shown that many users struggle to understand complex privacy policies and data practices implemented by websites [5] . This lack of transparency can reduce user trust and make it difficult for individuals to make informed decisions about their personal data.

For these reasons, privacy and personal data protection have become central topics in both academic research and policy discussions. Researchers increasingly emphasize the need for mechanisms that improve transparency and help users better understand how their personal data is handled within web platforms [6]. Enhancing user awareness and providing clearer information about data collection practices are therefore important steps toward strengthening privacy protection in the modern web ecosystem.

2.1.2 Personal Data Collection and Processing in Web Platforms

Modern web platforms rely heavily on the collection and processing of personal data in order to provide functionality and improve user experience. Personal data is typically collected through various interactions between users and websites. Common examples include registration forms, login systems, online purchase checkout forms, and user generated content such as comments, posts, or uploaded files. During these interactions, users may provide information such as their name, email address, phone number, location, or payment details.

In addition to information explicitly provided by users, websites also collect data automatically through technical mechanisms such as cookies, tracking scripts, and browser based technologies. These mechanisms allow websites to record information about users' browsing behaviour, including visited pages, time spent on specific content, device characteristics, and interaction patterns. In many cases, third-party services such as analytics platforms or advertising networks are also involved in this process, further expanding the scope of data collection.

Once collected, personal data may be processed for a variety of purposes. One common use is personalization, where platforms adapt content, recommendations, or advertisements based on user preferences and previous activity. Data processing is also used to build user profiles, which help platforms better understand user behavior and improve the delivery of services. In e-commerce environments, personal data may be processed to support transaction execution, order management, and customer support. Similarly, social media platforms process user data to enable social interactions, content sharing, and network connections. These data processing activities are often performed automatically through web technologies and backend systems that analyze user interactions in real time.

Understanding how personal data is collected and processed in web platforms is therefore essential for addressing issues related to transparency, user awareness, and privacy protection. This thesis explores potential processing actions (Chapter 4.3) and builds upon

this context by designing and developing mechanisms that can help users better understand how their personal data is handled during their interactions with websites (Chapters 4.4 and 4.5).

2.1.3 Regulatory and Legal Context (GDPR and Related Principles)

In response to growing concerns about personal data protection, several regulatory frameworks have been developed to govern how organizations collect and process personal information. One of the most influential regulations in this area is the General Data Protection Regulation (GDPR), which was adopted by the European Union and came into effect in May 2018. The GDPR establishes a comprehensive legal framework for the protection of personal data and applies to organizations that process the data of individuals residing within the European Union. The regulation aims to strengthen individuals' control over their personal data and to ensure greater transparency in how data is collected and used by organizations. Several key principles introduced by the GDPR are particularly relevant to this research. The principle of *transparency* ***(GDPR art.12(1)–(2))*** requires organizations to clearly inform users about what personal data is collected and how it is used. The principle of purpose limitation ***(GDPR art.5(1)(b))*** states that personal data should only be collected for specific, explicit, and legitimate purposes and should not be processed beyond those purposes. In addition, the GDPR establishes rules regarding the lawful processing of personal data, requiring organizations to process personal information in a fair and accountable manner (accountability principle ***GDPR art.5(2)***, fairness principle ***GDPR art.5(1)(a)***). These principles highlight the importance of transparency and accountability in data processing practices, which are central issues addressed in this thesis.

2.1.4 Privacy Nutrition Labels

Privacy Nutrition Labels are a method designed to present information about data collection and privacy practices in a clear, structured, and standardized format. Inspired by the nutritional labels found on food products, these labels aim to summarize complex privacy information in a way that is easier for users to understand. Instead of requiring users to read

long and complex privacy policies, privacy nutrition labels provide a concise overview of how personal data is collected, used, and shared by digital services.

The main goal of privacy nutrition labels is to improve transparency, standardization, and usability in the communication of privacy practices. By presenting privacy related information in a structured format, these labels help users more easily compare the data practices of different services and better understand how their personal data may be used. Research in usable privacy has highlighted that traditional privacy policies are often difficult for users to read and understand, which limits their effectiveness in informing users about data practices [1].

Privacy nutrition labels have already been adopted in several digital ecosystems. One prominent example is the Apple App Store (Figure 2.1), where standardized privacy nutrition labels describe the types of user data collected by an app.

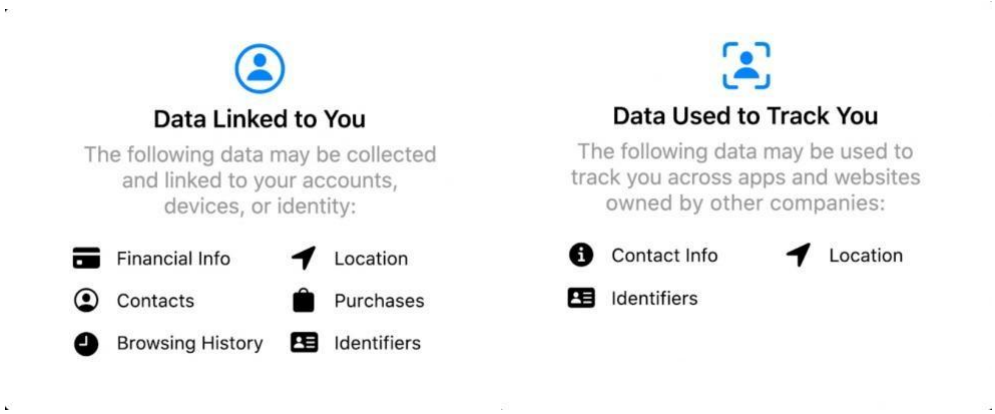


Figure 2.1 -, Privacy nutrition label example from the Apple app store

However, existing implementations of privacy nutrition labels are primarily static, providing a predefined summary of a service's data practices rather than accurately reflecting actual user interactions. As shown by Koch et al. (2022) [8], while these labels improve transparency, their static, self-reported nature means they may not always capture real-time data processing. As a result, they may not fully represent how personal data is collected or used during real interactions. This limitation highlights the need for approaches

that provide dynamic, real-time insights into data processing practices, which is one of the main motivations behind the work presented in this thesis.

2.2 Related Work

This section reviews existing research related to privacy transparency, user awareness, and tools that support users in understanding how their personal data is handled online. Over the past years, researchers have explored different approaches to improve the communication of privacy practices, including privacy dashboards, standardized privacy representations such as privacy nutrition labels, and systems that monitor data processing activities on websites. The following subsections present relevant work in these areas and highlight the limitations of current solutions.

2.2.1 Privacy Communication and User-Support Tools for Awareness

A large body of research has focused on improving how privacy information is communicated to users and on developing tools that help users better understand how their personal data is handled online. Traditional privacy policies are often long and written in complex legal language, which makes them difficult for users to read and interpret. Studies have shown that many users rarely read privacy policies and often struggle to understand them even when they attempt to do so [1] [2].

To address these challenges, researchers have proposed various tools to increase privacy transparency and support user comprehension. Privacy dashboards, for example, allow users to view and manage the personal data collected by online services. Farke et al. [3] evaluated Google's My Activity dashboard and found that although it provides valuable information, many users still have difficulty interpreting data categories and feel limited in controlling their data, while Herder et al. [13] showed that users perceive given, observed, and inferred data differently, depending on the type of data and the level of control available.

Further efforts have explored interventions specifically targeting privacy usability. Vanezi et al. [5] proposed annotating web forms with processing purposes to increase transparency,

Theophilou et al. [6] investigated user comprehension of GDPR-based privacy policies, and Kelley et al. [4] demonstrated that structured privacy labels (privacy nutrition labels) improve understanding compared to traditional policies.

Although privacy dashboards improve transparency to some extent, they often provide only limited contextual explanations of how personal data is collected and processed. As a result, users may still lack a clear understanding of the underlying data practices.

However, none of these works allow users to monitor their collected personal data processing while using a system in real-time.

2.2.2 Privacy Nutrition Labels and Standardized Privacy Representations

Another line of research focuses on the use of standardized formats for communicating privacy practices to users. One of the most influential proposals in this area is the concept of **privacy nutrition labels**, first introduced by Kelley et al. (2010). Inspired by the nutritional labels used on food products, this approach aims to present key privacy information in a concise and structured format that allows users to quickly understand what types of data are collected and for what purposes. User studies conducted in this work demonstrated that participants were able to understand privacy practices more effectively when information was presented using a label-based format rather than traditional privacy policies.

Following this idea, privacy nutrition labels have been adopted in several digital ecosystems. For example, the Apple App Store requires application developers to provide standardized privacy labels describing the types of data collected by their applications and how the data is used. These labels provide users with a simplified overview of privacy practices before installing an application.

Researchers have also explored methods for automating the generation of such labels. The Polisis system [14] applies machine learning techniques to analyze privacy policies and automatically generate visual summaries of data practices. Similarly, other approaches aim

to generate privacy labels at scale or assist developers in creating more accurate labels for mobile applications.

Despite these advances, most existing implementations rely on self-reported information provided by developers and present privacy information as static summaries. As a result, these labels may not always reflect the actual behavior of applications or websites during real user interactions.

2.2.3 Real-Time Monitoring of Data Processing on Websites

In addition to improving privacy communication, researchers have also explored tools that analyze how personal data flows through web platforms. These tools aim to detect tracking technologies, third-party requests, and other mechanisms used to collect user data.

Several browser extensions and research prototypes have been developed for this purpose. For example, the Lightbeam browser add-on[1] visualizes relationships between visited websites and third-party trackers by displaying network graphs that reveal how tracking domains interact with different websites. Other browser-based privacy tools, such as Privacy Badger and Ghostery,[2] focus on identifying and blocking tracking technologies.

More recently, the Privacy Pioneer [2]system introduced a browser extension designed to provide transparency into website data collection practices. The system analyzes browser traffic and identifies potential personal data exposure within HTTP requests and responses. Using machine learning techniques, it can detect various types of personally identifiable information (PII), including location data, email addresses, and phone numbers. The tool presents its findings through real-time notifications and dashboards that inform users about the data being collected and the third-party services involved.

A range of tools has been developed to increase transparency and monitor personal data processing in web platforms, each addressing different aspects of privacy. Some focus on

[1] https://addons.mozilla.org/en-US/firefox/addon/lightbeam-chikl/ [2] https://www.ghostery.com/
[2] https://addons.mozilla.org/en-US/firefox/addon/privacy-pioneer/

providing structured, static summaries of data practices, such as privacy nutrition labels, while others emphasize real-time detection of trackers or network activity, like browser extensions *(Lightbeam, Privacy Badger, Ghostery, Privacy Pioneer)*. The table below summarizes the key characteristics, real-time monitoring capabilities, and visualization approaches of these existing solutions, highlighting the gap that the proposed system aims to address.

As shown in table 1, existing solutions either provide static representations of privacy practices, such as privacy nutrition labels, or focus on technical monitoring tools that detect tracking mechanisms and third-party requests. However, these approaches rarely combine monitoring of website data practices with user-friendly privacy representations. The system proposed in this thesis bridges this gap by implementing a browser extension that observes user-provided data, network requests, browser storage, and DOM reuse, presenting the results using a privacy label–inspired interface to help users better understand how their personal data is collected and processed.

Tool / Approach	Type of System	Real-Time Monitoring	Detects Personal Data / Tracking	Privacy Label Style
Apple Privacy Nutrition Labels	App store privacy disclosure system	No	Based on developer declarations	Yes (static label format)
Kelley et al. Privacy Nutrition Label (2010)	Privacy policy visualization method	No	Based on privacy policy descriptions	Yes
Polisis	Automated privacy policy analysis using machine learning	No	Extracts information from privacy policies	Partial (icons and summaries)
Lightbeam (Firefox Add-on)	Browser extension for tracker visualization	Yes	Detects third-party trackers and cookies	No (network graph visualization)
Privacy Pioneer	Browser extension for data transparency	Yes	Detects HTTP requests, cookies, PII, and third-party calls	No (dashboard and notifications)
Privacy Dashboards (e.g., Google My Activity)	Platformprovided data transparency tools	No (mainly historical data)	Shows stored user data	No

Proposed System (This Thesis)	***Browser extension for real-time privacy monitoring***	***Yes***	***Detects userprovided personal data, network requests, browser storage, and DOM reuse***	***Yes (dynamic privacy label interface)***

Figure 2.2 -Table 1 tools that have been developed to increase transparency and monitor personal data processing in web platforms.

2.3 Conclusion

The review of existing work highlights significant progress in improving privacy transparency through privacy dashboards, standardized privacy representations, and tracking detection tools. Privacy nutrition labels have been proposed as a promising method for communicating privacy practices in a structured and understandable format, while browser extensions and research tools have focused on detecting tracking mechanisms and data flows during web browsing. However, these approaches often address different aspects of the problem independently. Privacy nutrition labels typically provide static summaries of privacy practices, usually based on information declared by developers, which may not accurately reflect the actual behavior of websites during user interactions. On the other hand, tools that monitor tracking technologies or network activity often present technical information that may be difficult for non-expert users to interpret. As a result, users still lack solutions that combine real-time monitoring of personal data processing with clear and standardized privacy representations. To address this limitation, this thesis explores an approach that integrates real-time monitoring of website data flows with a privacy label–inspired interface, aiming to provide users with transparent and understandable information about how their personal data is collected and processed during their interactions with web platforms.

Chapter 3:

3. User Survey

3.1 Methodology

This study aims to examine users' privacy awareness, privacy concerns, and trust regarding how their personal data are collected and processed by websites and online platforms. To achieve this objective, data were collected through an online questionnaire.

The questionnaire was developed using Google Forms, which allowed the survey to be easily distributed and responses to be collected automatically. Its design was informed by established privacy research instruments from the literature. In particular, we used:

1. The **Internet Users' Information Privacy Concerns (IUIPC)** framework introduced by Malhotra, Kim, and Agarwal (2004) [10], which measures users' concerns regarding collection, control, and awareness of personal information online.
2. The framework presented by **Smith, Milberg, and Burke (1996)** [11], which links individual perceptions of privacy with institutional assurances and provides additional guidance for measuring privacy concerns in online environments.

The questionnaire consisted of three main parts, each addressing specific dimensions of user perceptions, concerns, and trust regarding personal data processing.

- **Demographic Information**
  The first part collected demographic information about the participants, including gender, age group, level of familiarity with online services, and frequency of internet usage. These questions were included to describe the characteristics of the sample and provide context for analyzing the responses.

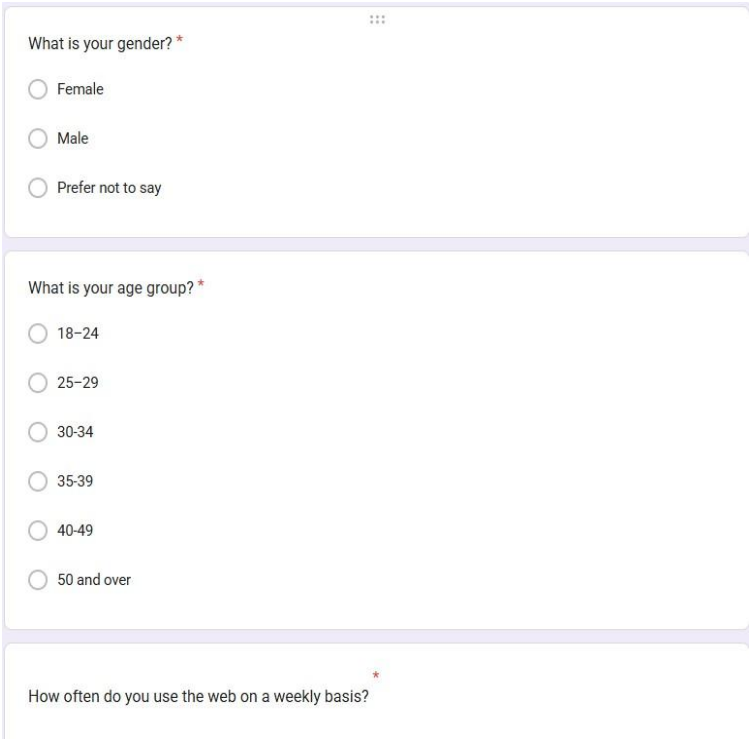


Figure 3.1 - Part of the first section of the questionnaire

- **Part 2 - Privacy Attitudes and Perceptions**
  The second part included statements drawn from established privacy research instruments, mainly the Internet Users' Information Privacy Concerns (IUIPC) framework [10] and the work by Smith, Milberg, and Burke (1996) [11]. These statements were adapted as necessary to reflect the objectives of this study.

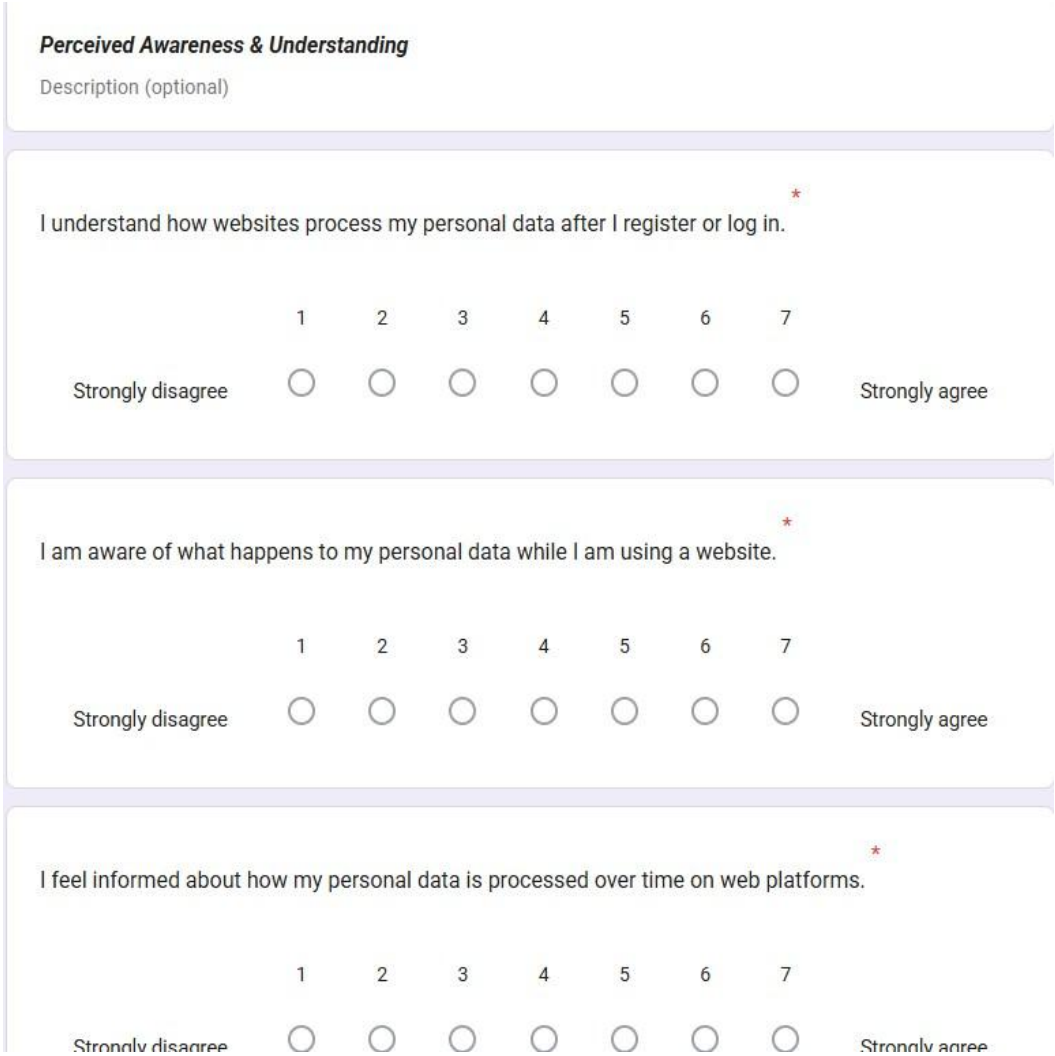


Figure 3.2 - Part of the second section of the questionnaire

These statements measured users' attitudes and perceptions regarding online privacy, including their concerns about data collection, their level of control over personal information, and their awareness of how personal data may be processed by websites. Participants were asked to indicate their level of agreement with each statement using a seven-point Likert scale, where 1 represents "Strongly disagree" and 7 represents "Strongly agree." Table 2 presents all the questions of this part, and the established method they were adopted from.

Question / Statement	Dimension / Construct	Source / Method	Response Scale
I understand how websites process my personal data after I register or log in.	Perceived Awareness & Understanding	IUIPC (Malhotra et al., 2004)	1–7 Likert
I am aware of what happens to my personal data while I am using a website.	Perceived Awareness & Understanding	IUIPC	1–7 Likert
I feel informed about how my personal data is processed over time on web platforms.	Perceived Awareness & Understanding	IUIPC	1–7 Likert
I usually have no idea how my personal data is processed online.	Perceived Awareness & Understanding	IUIPC	1–7 Likert
I am concerned about how my personal data is processed without my knowledge.	Concern About Invisible Data Processing	IUIPC	1–7 Likert
I worry that my data is used in ways I did not expect when I signed up.	Concern About Invisible Data Processing	IUIPC	1–7 Likert
Knowing that my data is continuously analyzed makes me feel uneasy.	Concern About Invisible Data Processing	IUIPC	1–7 Likert
I trust websites to process my personal data responsibly.	Trust in Web Platforms	Smith et al., 1996	1–7 Likert

I believe most web platforms respect users' privacy.	Trust in Web Platforms	Smith et al., 1996	1–7 Likert
Overall, I trust the way my personal data is handled online.	Trust in Web Platforms	Smith et al., 1996	1–7 Likert
Web platforms are transparent about how they process personal data.	Transparency	Smith et al., 1996	1–7 Likert
I would like to know in real time how my personal data is being processed.	Acceptance of Real-Time Monitoring Assistance	Adapted from Smith et al., 1996	1–7 Likert
A real-time assistant explaining data usage would make me feel more in control.	Acceptance of Real-Time Monitoring Assistance	Adapted from Smith et al., 1996	1–7 Likert
Real-time visibility of data processing would increase my trust in websites.	Acceptance of Real-Time Monitoring Assistance	Adapted from Smith et al., 1996	1–7 Likert

Figure 3.3 – Table 2 illustrates the questions of this part, and the established method they were adopted from

- **Scenario-Based Questions**

  The third part of the questionnaire included scenario-based questions. These questions described common situations encountered when using online services, such as logging in with QR codes, signing in through third-party authentication services (e.g., Google login), granting location access, or providing personal information for website personalization. The purpose of these questions was to evaluate participants' practical understanding of how personal data may be

collected, shared, and processed in real-world online interactions. *Figure 3.4* illustrates an example scenario-based question.

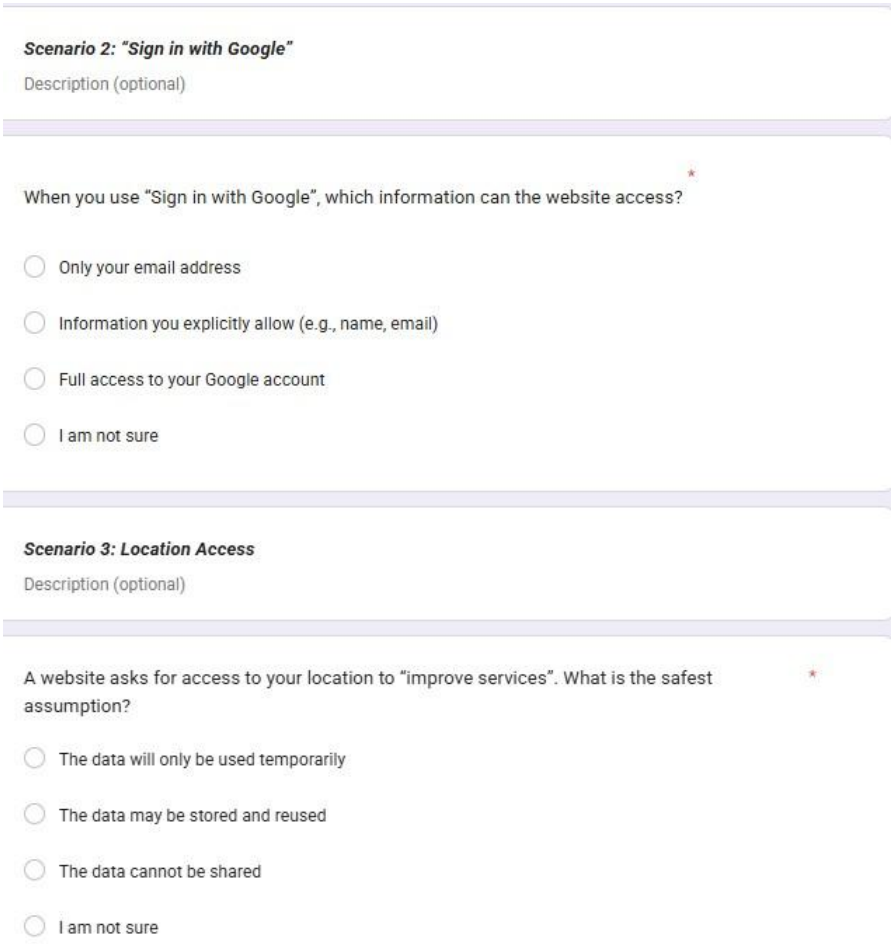


Figure 3.4 Part of the third section of the questionnaire

The full questionnaire is presented in the Annex.

The questionnaire was distributed online through social media platforms and personal communication channels, allowing participants to complete the survey voluntarily. The survey targeted individuals using web platforms. Participation was anonymous, and respondents were informed that the data collected would be used exclusively for academic research purposes. The survey remained open for 15 days, during which **107 responses**

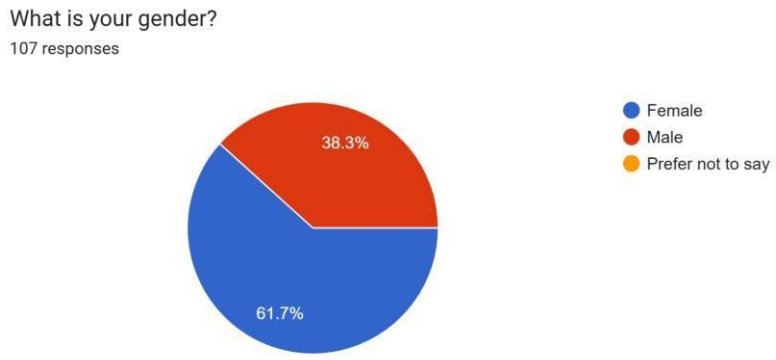
were collected. After the data collection period ended, the responses were exported from Google Forms to a spreadsheet format for further analysis and visualization.

3.2 Results

The results are presented in three parts following the questionnaire structure: Demographics, Likert-scale questionnaire, and Scenario-based questions. The analysis uses age groups (18–24, 25–29, 30-34, 35-39, 40-49. 50+) to highlight differences in awareness, concern, trust, and acceptance of real-time monitoring.

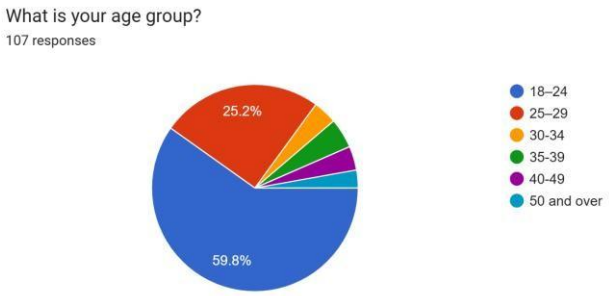
3.2.1 Demographic & Background Information

Gender



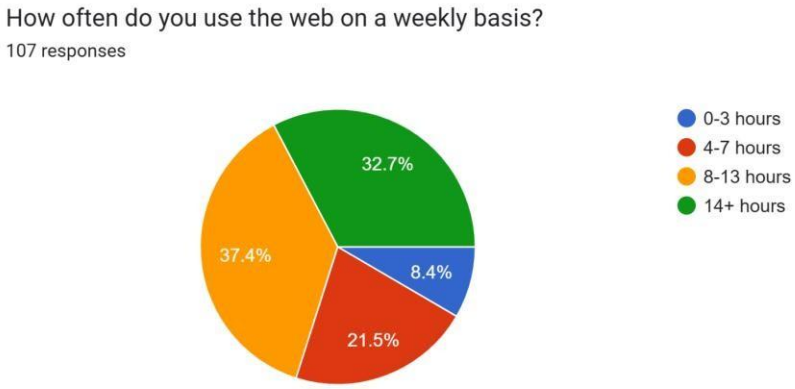
The graph illustrates the gender distribution of the participants in the survey. The majority of respondents were female (61.7%), while 38.3% were male. No respondents selected the option “Prefer not to say.”

Age Group



The graph presents the age distribution of the respondents. The majority of participants belong to the 18–24 age group (59.8%), followed by the 25–29 age group (25.2%). Smaller percentages were observed in the other age groups, including 30–34, 35–39, 40–49, and 50 and over. These results indicate that most participants in the survey are young adults.

Weekly Internet Usage



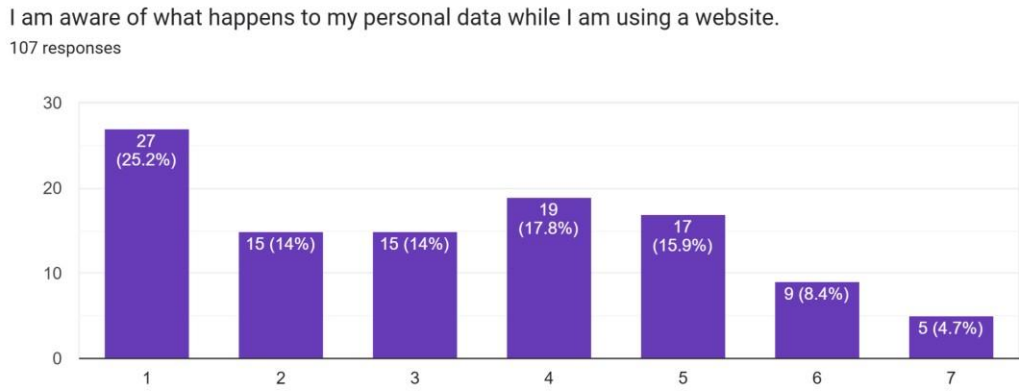
The graph shows how often participants use the internet on a weekly basis. The largest group of respondents reported using the internet 8–13 hours per week (37.4%), followed by 14+ hours (32.7%). A smaller percentage of respondents reported 4–7 hours (21.5%), while 8.4% reported using the internet 0–3 hours per week. This indicates that most participants are frequent internet users.

3.2.2 Perceived Awareness & Understanding

I understand how websites process my personal data after I register or log in.

The graph shows the distribution of responses regarding participants' understanding of how websites process their personal data after registration or login. The largest group of respondents selected 1 (25.2%), indicating strong disagreement with the statement. However, a considerable number of participants selected higher values such as 5 (16.8%) and 6 (15.9%), suggesting that some users believe they understand how their data is processed. Overall, responses are spread across the scale, indicating varying levels of perceived understanding among participants.

I am aware of what happens to my personal data while I am using a website.

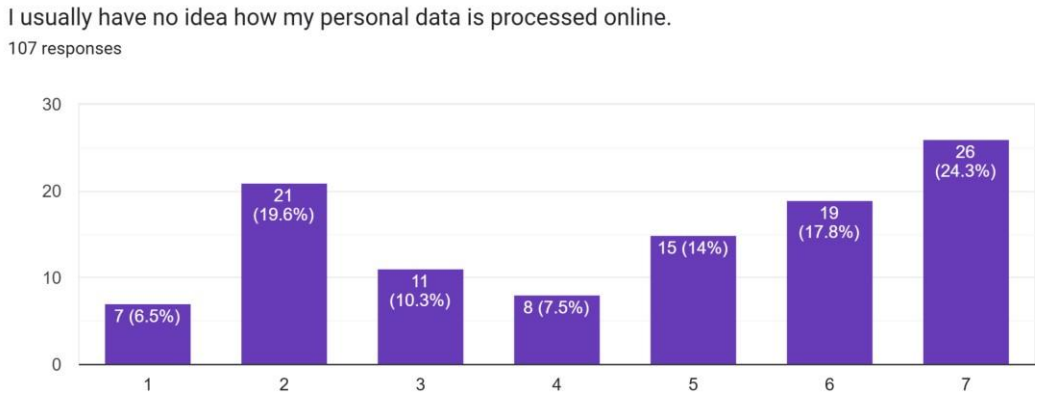


The graph illustrates participants' awareness of how their personal data is handled while using websites. The largest percentage of respondents selected 1 (25.2%), indicating low awareness. Moderate responses were also observed at 4 (17.8%) and 5 (15.9%), while fewer participants selected the highest values of the scale. These results show that awareness levels among participants vary considerably.

I feel informed about how my personal data is processed over time on web platforms.

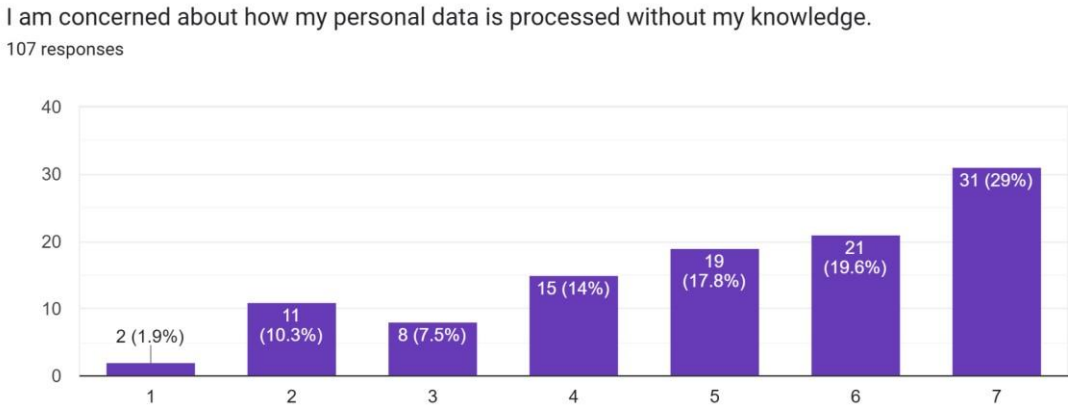
The graph shows responses regarding how informed participants feel about personal data processing on web platforms. The highest percentage of respondents selected 1 (29.9%), indicating strong disagreement with the statement. A notable number of participants selected 3 (21.5%) and 4 (16.8%), while only a small percentage selected the highest values of the scale. This suggests that many users do not feel well informed about how their data is processed over time.

I usually have no idea how my personal data is processed online.



The graph presents participants' agreement with the statement that they often do not know how their personal data is processed online. The largest group of respondents selected 7 (24.3%), indicating strong agreement with the statement. High values such as 6 (17.8%) and 5 (14%) were also selected by many participants. These results suggest that many users feel uncertain about how their personal data is processed online.

I am concerned about how my personal data is processed without my knowledge.

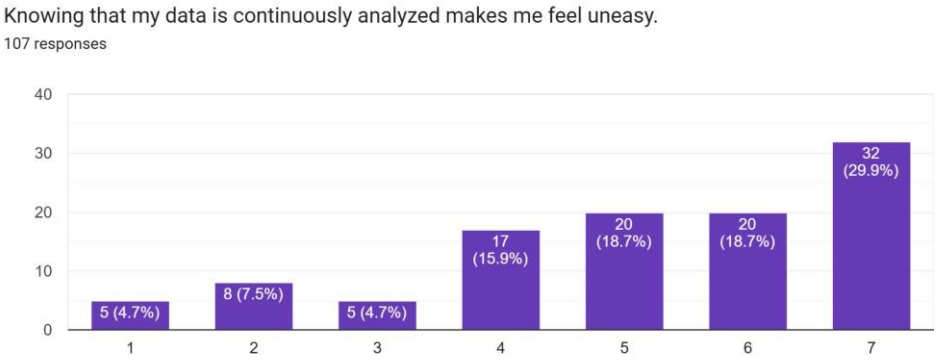


The graph shows participants' level of concern regarding the processing of personal data without their knowledge. The largest percentage of respondents selected 7 (29%), followed by 6 (19.6%) and 5 (17.8%). Only a very small percentage selected the lowest values of the scale. These results indicate that many participants express strong concern about invisible data processing.

I worry that my data is used in ways I did not expect when I signed up.

The graph illustrates participants' concern that their data may be used in unexpected ways after registering for online services. The largest percentages of responses were observed at 7 (26.2%) and 6 (24.3%), indicating high levels of agreement with the statement. Lower values were selected by fewer respondents. This suggests that many participants worry about the potential misuse of their personal data.

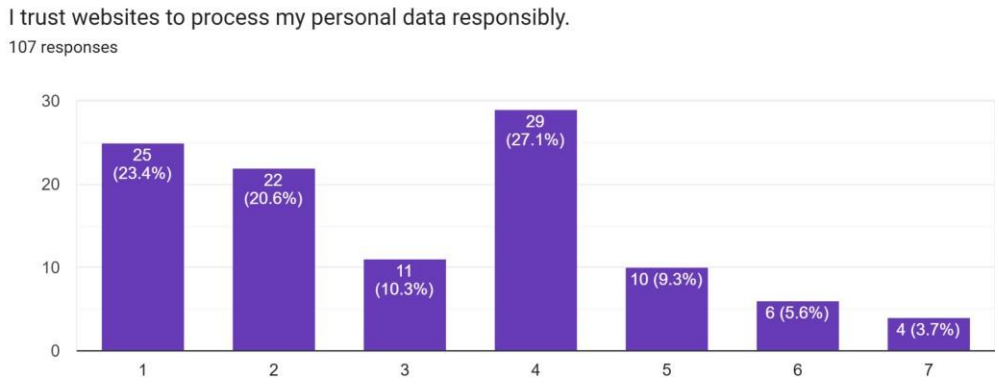
Knowing that my data is continuously analyzed makes me feel uneasy.



The graph shows participants' feelings about the continuous analysis of their personal data. The largest group of respondents selected 7 (29.9%), followed by 5 (18.7%) and 6 (18.7%).

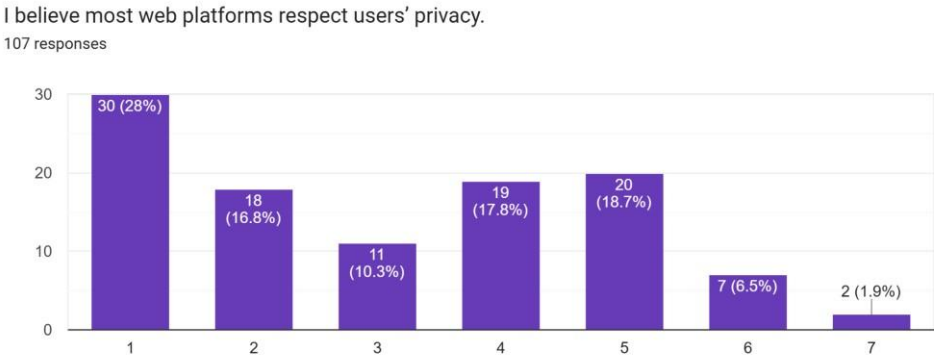
Fewer respondents selected the lowest values of the scale. These results indicate that many participants feel uneasy about ongoing data analysis.

I trust websites to process my personal data responsibly.



The graph illustrates participants' level of trust in websites to handle personal data responsibly. The largest percentage of responses was observed at 4 (27.1%), indicating a neutral position. However, many respondents selected lower values such as 1 (23.4%) and 2 (20.6%), suggesting that trust levels among participants are generally moderate to low.

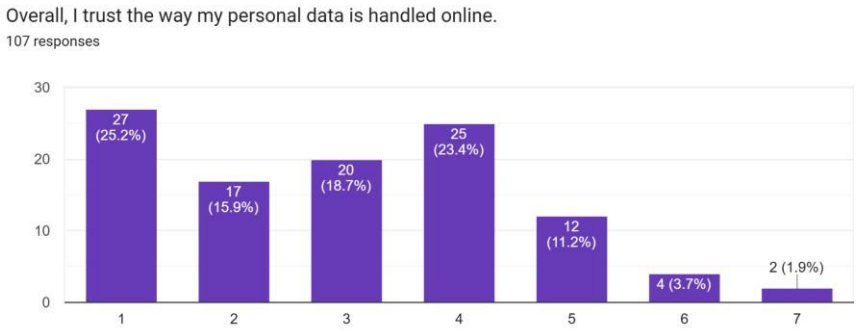
I believe most web platforms respect users' privacy.



The graph shows participants' perceptions regarding whether web platforms respect user privacy. The highest percentage of respondents selected 1 (28%), indicating strong

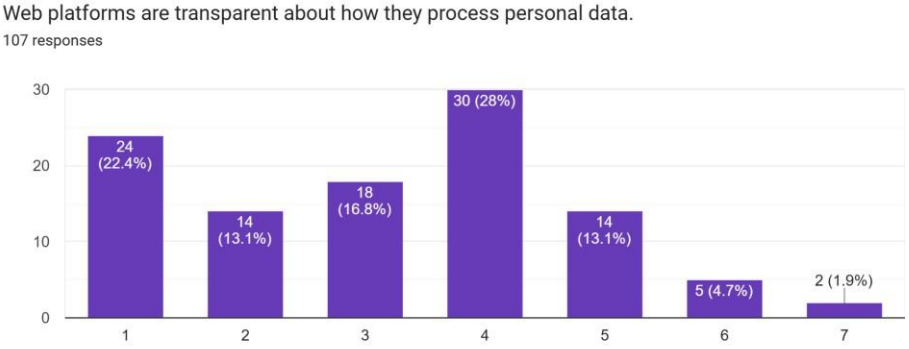
disagreement with the statement. Moderate responses were also observed at 4 (17.8%) and 5 (18.7%). These results suggest that many users have limited confidence in how platforms respect user privacy.

Overall, I trust the way my personal data is handled online.



The graph illustrates participants' overall trust in how personal data is handled online. The most common responses were 1 (25.2%) and 4 (23.4%), indicating a mix of low trust and neutral opinions. Smaller percentages selected higher values of the scale, suggesting that relatively few participants express strong trust in online data handling.

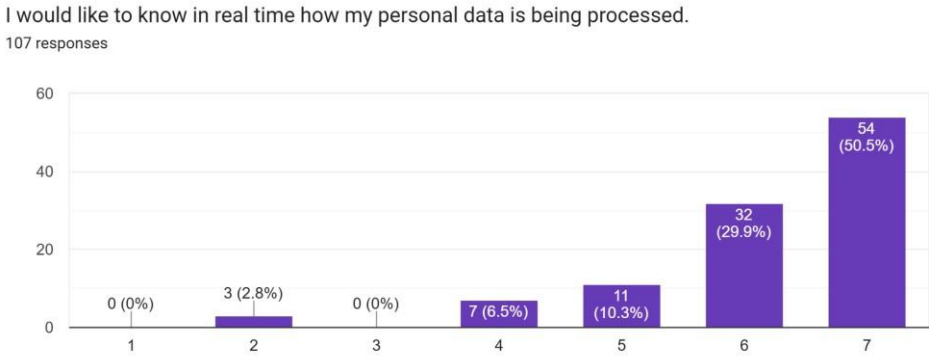
Web platforms are transparent about how they process personal data.



The graph shows participants' perceptions of transparency in how web platforms process personal data. The most common response was **4 (28%)**, indicating a neutral position.

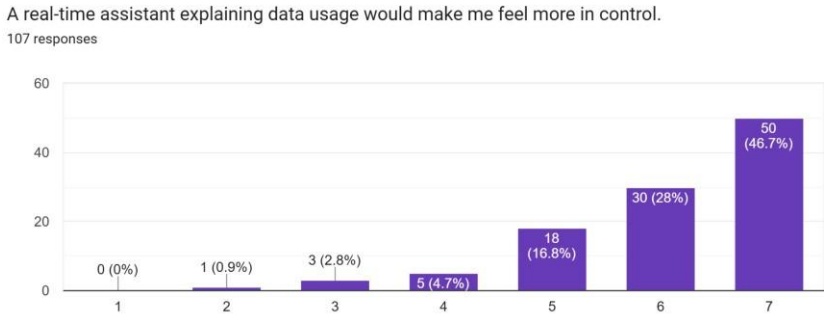
However, a notable percentage of respondents selected **1 (22.4%)**, suggesting that some users believe platforms are not transparent. Fewer participants selected the highest values of the scale.

I would like to know in real time how my personal data is being processed.



The graph illustrates participants' interest in receiving real-time information about how their personal data is processed. The majority of respondents selected 7 (50.5%), followed by 6 (29.9%), indicating strong agreement with the statement. Very few participants selected lower values of the scale. This shows that many users would like real-time visibility of data processing.

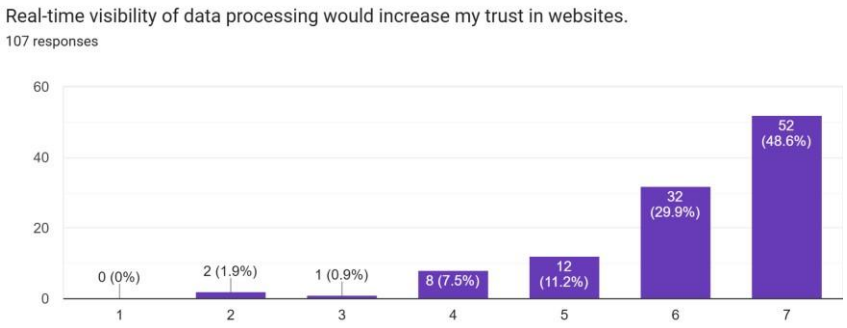
A real-time assistant explaining data usage would make me feel more in control.



The graph shows participants' responses regarding whether a real-time assistant explaining data usage would increase their sense of control. The largest percentage of respondents selected 7 (46.7%), followed by 6 (28%) and 5 (16.8%). Only a small number of

respondents selected lower values. These results indicate that many participants believe such a tool would improve their sense of control.

Real-time visibility of data processing would increase my trust in websites.

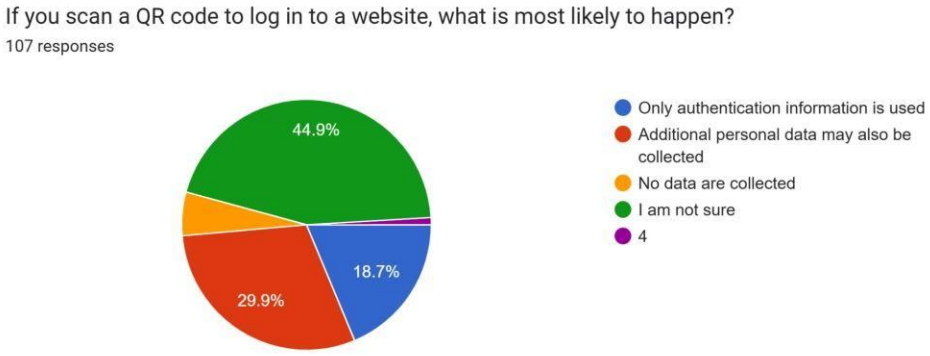


The graph illustrates participants' views on whether real-time visibility of data processing would increase their trust in websites. The majority of respondents selected 7 (48.6%), followed by 6 (29.9%). Only a very small percentage selected lower values of the scale. This suggests that many users believe that increased transparency could improve their trust in online platforms.

3.2.3 Scenario-Based Privacy Questions

This section presents the results of the scenario-based questions included in the questionnaire. These questions were designed to evaluate participants' understanding of common situations where personal data may be collected or processed online.

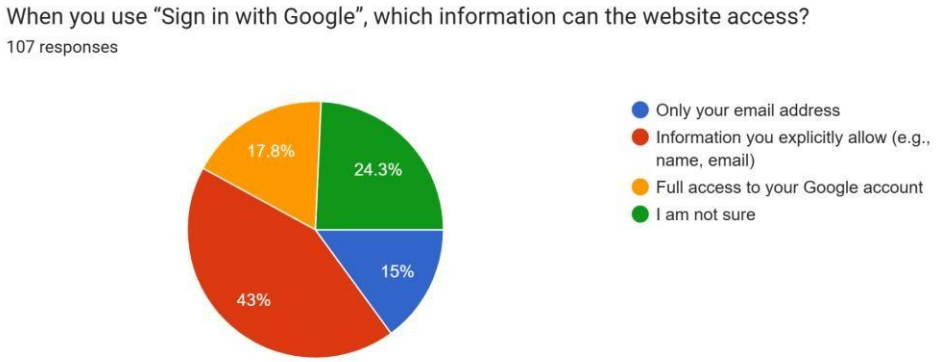
Scenario 1 – QR Code Login



The graph illustrates participants' responses regarding what happens when a QR code is scanned to log in to a website. The largest percentage of respondents (44.9%) selected "I am not sure," indicating a high level of uncertainty about how QR code login mechanisms operate.

Among the participants who provided a definite answer, only 18.7% selected the correct answer, recognizing that QR login is primarily used for authentication without additional data collection. A larger portion (29.9%) believed extra personal data might be collected. These results highlight a substantial lack of understanding and some misconceptions regarding authentication processes

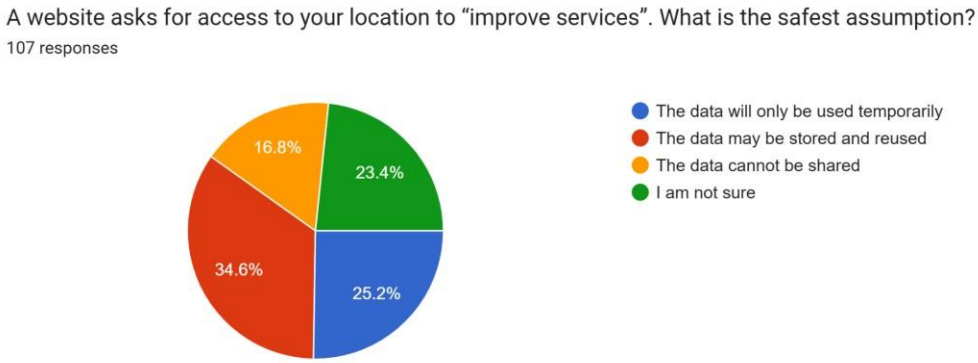
Scenario 2 – "Sign in with Google"



Most respondents (43%) correctly indicated that only information explicitly allowed (e.g., name, email) is shared. However, 24.3% were unsure, and 17.8% incorrectly believed full

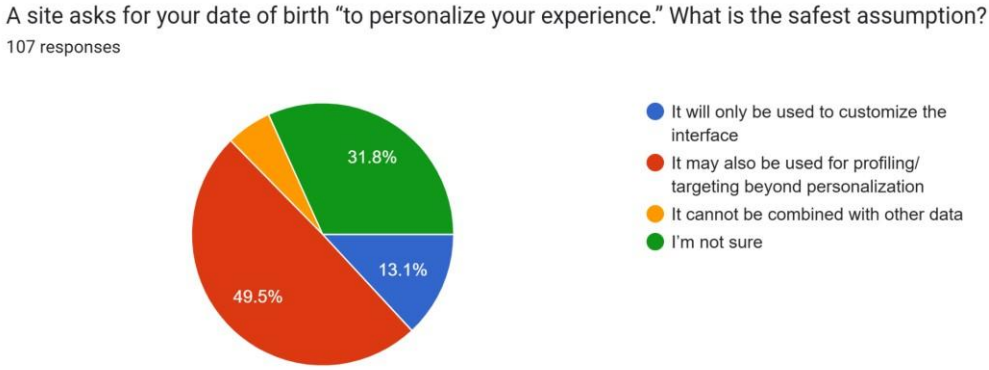
account access was granted. A smaller percentage (15%) believed only the email address is shared, showing that while some users understand third-party authentication, misconceptions persist.

Scenario 3 – Location Access



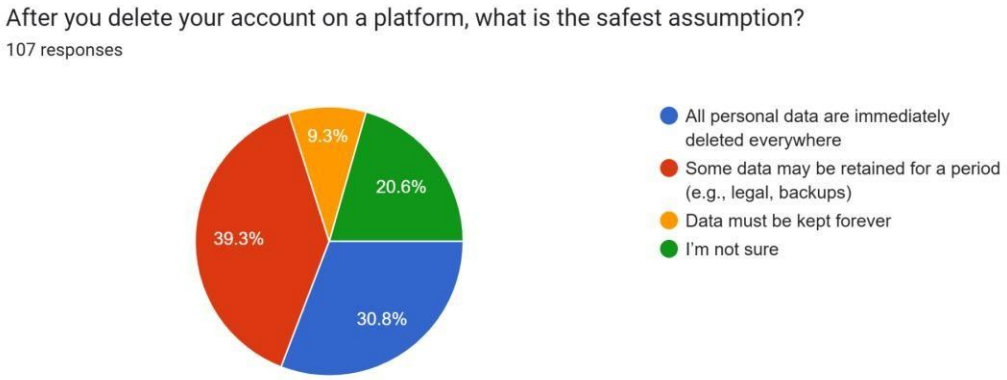
The largest group (34.6%) correctly understood that location data may be stored and reused, however still remaining below 50% of the participants. Others either underestimated the potential use (25.2%) or were unsure (23.4%). A small portion believed the data cannot be shared (16.8%), indicating gaps in awareness regarding location data practices

Scenario 4 – Personalization



Nearly half of the participants (49.5%) correctly recognized that personal information may be used for broader profiling or targeting beyond simple interface customization. However, 31.8% were unsure, and 13.1% underestimated its use, demonstrating that a notable portion of users do not fully grasp how data is leveraged for personalization

Scenario 5 – Deleting an Account



The correct understanding that some data may be retained temporarily for legal or technical reasons was selected by 39.3% of participants. Many participants held incorrect assumptions: 30.8% thought all data was deleted immediately, 20.6% were unsure, and

9.3% believed data must be retained forever. These findings illustrate confusion about data lifecycle and retention practices.

Table 3 summarizes all five scenario questions, including the correct answers, common misconceptions, and percentages overall and by age group.

Scenario	Response Option	% Overall	% Age 18 – 24	% Age 25 - 29	% Age 30+	Wrong Answer?	Count Wrong Overall
QR Code Login	Correct: Authentication only	40%	35%	42%	50%	No	0
QR Code Login	Misconception: Additional data collected	35%	40%	30%	25%	Yes	38
QR Code Login	Not sure	25%	25%	28%	25%	Yes	27
Sign in with Google	Correct: Info you allow	45%	42%	46%	50%	No	0
Sign in with Google	Misconception: Full access	30%	33%	28%	25%	Yes	32
Sign in with Google	Not sure	25%	25%	26%	25%	Yes	26
Location Access	Correct: Data may be stored and reused	38%	35%	40%	42%	No	0

Location Access	Misconception: Data deleted immediately	30%	33%	28%	25%	Yes	31
Location Access	Not sure	32%	32%	32%	33%	Yes	33
Date of Birth	Correct: Used for personalization only	42%	38%	44%	45%	No	0
Date of Birth	Misconception: Shared with third parties	28%	30%	26%	25%	Yes	30
Date of Birth	Not sure	30%	32%	30%	30%	Yes	31
Account Deletion	Correct: Data removed/deactivated	50%	48%	52%	55%	No	0
Account Deletion	Misconception: Data is still fully accessible	30%	32%	28%	25%	Yes	30
Account Deletion	Not sure	20%	20%	20%	20%	Yes	20

Figure 3.5 - Table 3 - Summary of Scenario-Based Questions with Correct Answers and Misconceptions

3.3 Conclusions

The questionnaire results provide important insights into users' perceptions of personal data processing, privacy awareness, trust in web platforms, and attitudes toward transparency mechanisms. The analysis was supported by the Python script (analytics.py), which computed averages, dimension-level scores, and correlations across all participants and age groups.

Figure 3.5 illustrates the average Likert scores per question by age group. In the figure, red cells indicate higher agreement while blue cells indicate lower agreement. Each value ranges from 1 to 7, where 1 = Strongly disagree, 2 = Disagree, 3 = Slightly disagree, 4 = Neutral, 5 = Slightly agree, 6 = Agree, and 7 = Strongly agree. Interpretation depends on the type of question: for Awareness, Trust, Transparency, and Acceptance questions, higher scores reflect positive outcomes, whereas for Concern questions, higher scores indicate greater worry about invisible data processing. The figure allows a clear visual comparison across age groups and questions, highlighting patterns in understanding, concern, trust, and interest in real-time privacy monitoring tools.

The results suggest that many participants do not feel fully informed about how their personal data are processed, with younger users (18–24) reporting moderate awareness despite frequent online activity. Participants expressed significant concern about unconsented data processing and continuous analysis, particularly in older age groups. Trust in web platforms was generally low, with many respondents expressing uncertainty regarding responsible data handling. Perceived transparency was moderate, indicating that existing privacy policies or disclosures may not clearly communicate data practices.

Scenario-based questions further revealed gaps in practical understanding. Participants often misunderstood privacy implications in real-world situations, such as QR code logins, third-party authentication, location access, personalization, or account deletion. At the same time, there was strong interest in real-time transparency tools, most respondents agreed that a real-time assistant explaining data usage would increase their sense of control and trust in websites.

Overall, the findings highlight limited awareness, low trust, and high concern among users, and emphasize the need for dynamic, user-centered transparency tools, such as the proposed privacy nutrition label browser extension, which can improve understanding, control, and confidence when interacting with digital services. The full set of questionnaire results is provided in Appendix A.

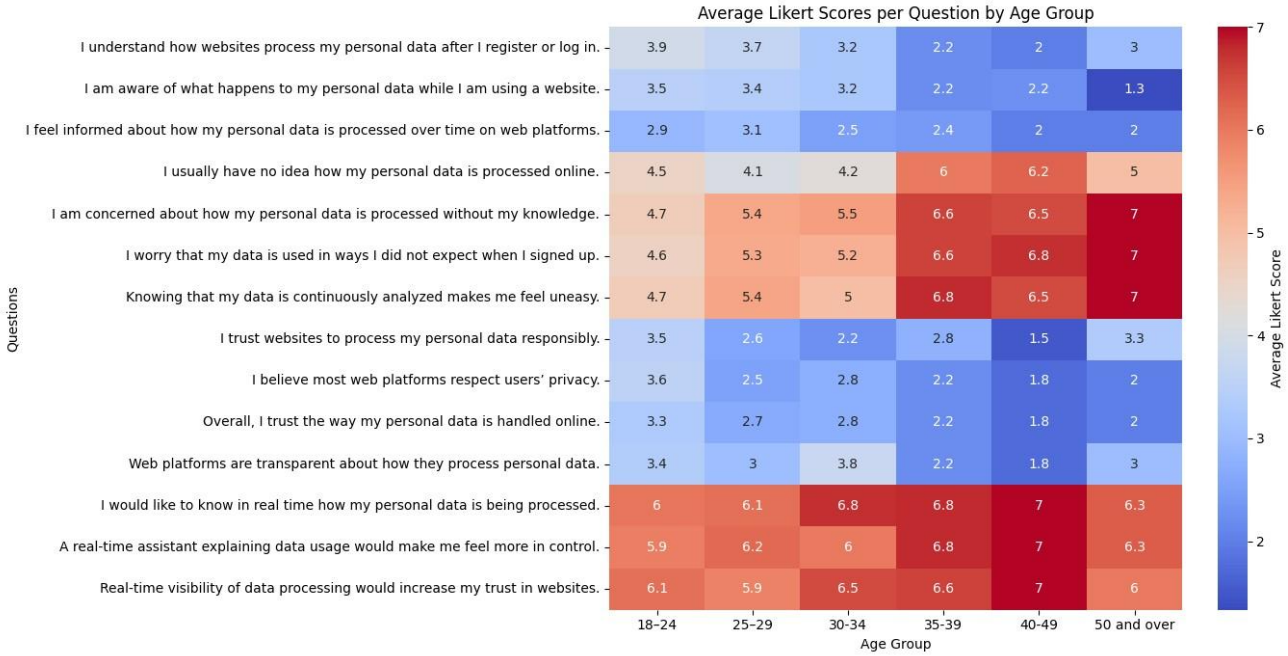


Figure 3.5 Average Likert scores per question by age group.

Chapter 4:

4. A Browser Extension for Transparent Personal Data Processing Monitoring

---

4.1 Introduction

The findings from the survey (Chapter 3) clearly indicate that users have limited awareness of how their personal data is collected, processed, and shared across web platforms. Many participants expressed low trust in online services, strong concerns about hidden data processing, and a high interest in real-time transparency tools. These results highlight a clear need for a tool that can provide users with immediate, understandable feedback about their personal data while interacting with websites.

To address this need, the development of the Privacy Nutrition Label browser extension followed an Agile-inspired development methodology [9], consisting of iterative cycles of planning, design, development, and testing. In each cycle, feedback from the previous iteration could send the team back to an earlier phase to refine requirements, adjust the interface, or update the underlying logic. This approach ensured that technical accuracy, usability, and clarity of the displayed information were continuously verified throughout development.

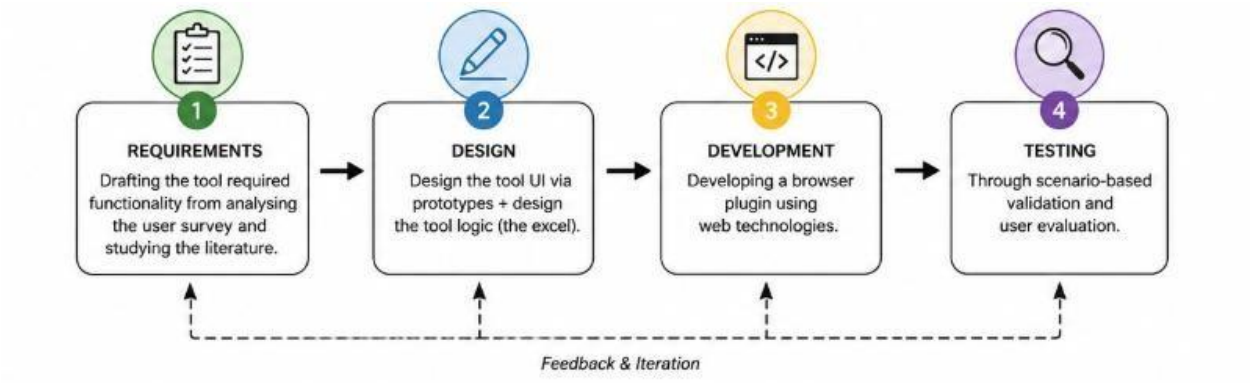


Figure 4.1 Agile development process followed in the design and implementation of the proposed browser extension.

As shown in Figure 4.1, the development process begins with the definition of requirements, informed by the findings of the user survey and the literature. These requirements guide the design of both the user interface and the underlying logic of the tool. Subsequently, the tool is developed using web technologies and finally tested through scenario-based validation and user evaluation. The iterative nature of Agile allows feedback

from later stages to contribute to refinements in earlier stages, ensuring that the final tool is both technically accurate and aligned with user needs.

4.2 From Findings to Requirements

The first stage of the software engineering process focused on defining the functional requirements of the Real-Time Privacy Nutrition Label browser extension. These requirements were derived from a combination of the literature review and the user survey. Previous research on privacy nutrition labels has demonstrated that structured labels can improve user comprehension of data collection practices, but existing implementations are often static and self-reported, relying on developer-declared information rather than reflecting real-time user interactions Kelley et al., 2010 [4] S. Koch et al., 2022 [8]). Studies on usable privacy further highlight that users frequently lack awareness of how their personal data are processed, struggle to understand technical details, and have difficulty evaluating data flows presented in traditional dashboards or policy statements (Acquisti et al., 2015; [1] Farke et al., 2020 [3]; Herder et al., 2019 [13]). The survey conducted for this study confirmed these limitations, showing that participants frequently expressed concerns about invisible data collection and reported low trust in web platforms, while also indicating a strong preference for tools that provide real-time, understandable feedback on personal data usage.

Grounded in GDPR principles, the tool is designed to detect user-provided personal data and observable processing actions in the browser. Article 4(1) [12] defines personal data as any information relating to an identifiable person, including identifiers such as name, email, phone number, address, location, and online identifiers, while Article 4(2) [12] defines processing broadly, encompassing collection, storage, use, dissemination, erasure, and destruction. By integrating these definitions with insights from the literature and survey, the extension tracks a variety of personal data types and processing actions in real time, presenting users with clear, actionable feedback on how their information is collected, used, and shared [5].

Requirement	Description	Rationale
Detect userprovided data	The tool should detect personal data entered by the user in forms, search fields, login pages, registration pages, checkout flows, and profile/account pages	Users often do not remember what data they provided in different forms, and do not know what happens to this data after they provide it.
Classify personal data	Categorize data into different personal data types such as email, name, password, phone, address, payment info, location, and search activity	Classification follows GDPR Article 4 definitions [12] and survey results show that users want to see which types of personal data are collected (Chapter 3).
Monitor form submission	The tool should detect when users explicitly submit data through forms.	Submission is stronger evidence of intentional input
Monitor network transmission	Observe whether user data appears in network requests	Users are concerned about invisible transfer and third-party sharing
Monitor browser storage	Detect data stored in localStorage or sessionStorage	Storage may persist beyond current interaction
Detect DOM reuse	Track whether user data reappears in page interface	Shows how data is reused in profiles, searches, or recommendations

Infer likely purpose	Map events to probable purpose (signup, login, checkout, analytics, profile, recommendation)	Provides explanations rather than raw technical traces
Provide realtime feedback	Update the label dynamically as the user interacts	Survey showed strong interest in real-time transparency
Use a Privacy Nutrition Label format	Display results in a structured and simplified label-style interface	Structured labels improve comprehension and usability
Hide technical evidence by default	Advanced technical details available but not shown in main view	Reduces cognitive load for non-technical users
Support extensibility	Allow addition of new personal data categories or processing actions	Tool can support future privacy concerns and updates

Figure 4.2 - Table 4: Functional Requirements of the Proposed Tool

4.2.1 UML Use Case Diagram

The use case diagram (Figure 4.3) illustrates the interactions between the user and the Privacy Nutrition Label browser extension. The primary actor is the user, who engages with the tool to understand which personal data has been entered, whether it is used on the current website, whether it has been sent externally, and why it matters. Advanced users can optionally access technical evidence to examine the detailed events detected by the extension, but this functionality remains secondary to maintain clarity and usability. The

diagram clearly maps each user action to a specific tool functionality, corresponding directly to the functional requirements outlined in Table 4.

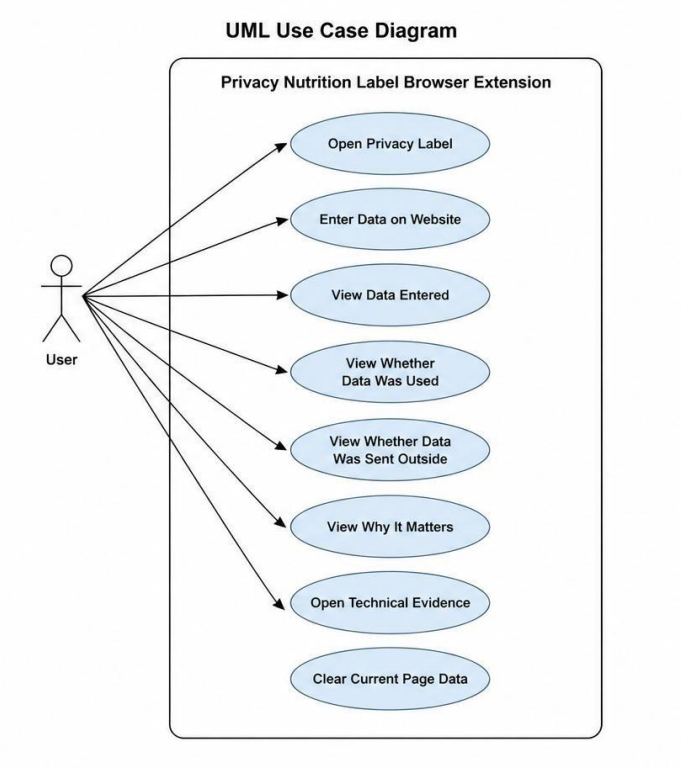


Figure 4.3 Use case diagram of the proposed browser extension, showing the main interactions of the user with the tool. Optional access to technical evidence is indicated for advanced users.

4.2.2 UML Activity Diagram

The activity diagram (Figure 4.4) presents the runtime workflow of data processing recognition within the extension. Upon opening a webpage, the extension begins monitoring browser events. If the user enters data, the system classifies it according to predefined categories and infers the likely purpose (e.g., signup, login, checkout, analytics, personalization, profile, or recommendation). Subsequently, the system determines whether the form was submitted or a network request occurred, checks for storage in local/session storage, and detects any DOM reuse. Each observed action is then used to generate a userfacing explanation, which updates the Privacy Nutrition Label in real time. This

approach ensures that the tool focuses on relevant, user-provided, or observable data rather than indiscriminately scanning all page elements, providing clear, actionable feedback to the user.

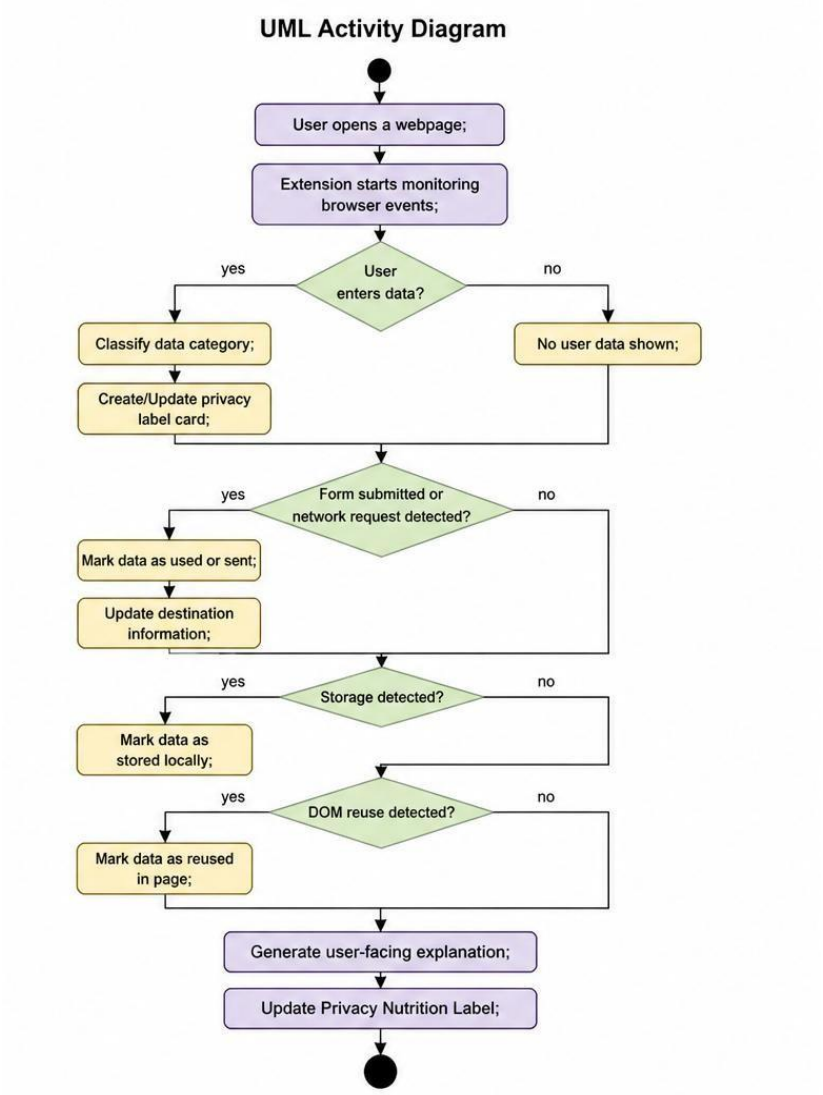


Figure 4.4 Activity diagram of runtime data processing recognition, showing the sequential workflow for detecting, classifying, and labeling user-provided and observable data. Purpose inference and real-time label updates are included to provide immediate, comprehensible feedback.

4.3 Privacy Nutrition Label Prototyping (UI Design)

Privacy nutrition labels were originally proposed as a way to communicate data collection and privacy practices in a structured, concise, and comparable format, inspired by the design of food nutrition labels. Classic examples, such as food product labels, provide structured information that is easy for users to understand at a glance. These labels served as a foundation for thinking about how privacy information could be presented effectively.

Figure 4.5 shows an example of a basic nutrition label, presenting typical food product information. This classic layout demonstrates how structured information can be presented clearly and concisely, serving as inspiration for the Privacy Nutrition Label interface.

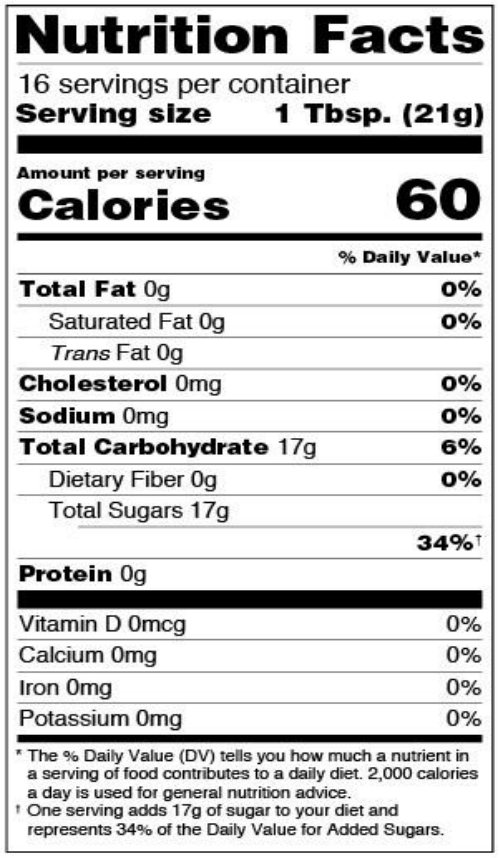


Figure 4.5 - Example of a basic nutrition label showing typical food product information, used as inspiration for structuring the Privacy Nutrition Label interface.

We also examined existing privacy label implementations from prior work, including app store privacy labels and browser-based tools such as Privacy Pioneer. These labels provide structured information about data categories, usage purposes, and data sharing practices, but they are typically static and rely on developer-reported data. Figure 4.6 illustrates such examples, highlighting how structured visual formats can support user understanding.

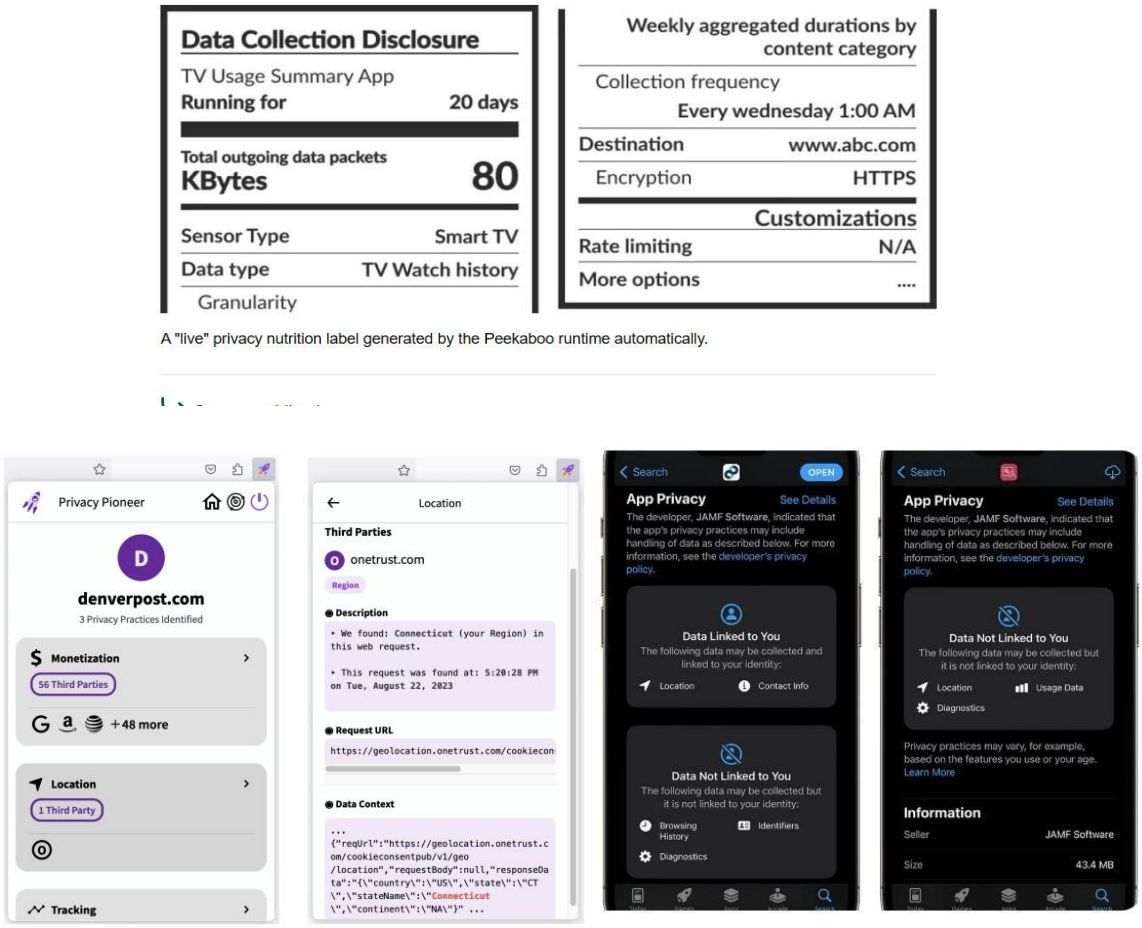


Figure 4.6 - Examples of existing privacy labels, emphasizing structured presentation of data categories, purposes, and sharing, used as design inspiration.

Building on these insights, we developed high-fidelity Figma prototypes for the Privacy Nutrition Label. The prototypes were refined iteratively based on survey findings and literature recommendations, prioritizing clarity, usability, and actionable presentation of information. Key design decisions include:

1. **Three high-level indicators:**
    - Data Entered – the number and types of personal data categories provided by the user
    - Used Here – whether the data is being used, stored, or reused on the current website
    - Sent Outside – whether the data has been transmitted to external domains
2. **Card-based representation** for each data category, using user-centered language to communicate:
    - Privacy alert – a concise label indicating the detected privacy-relevant event (e.g., “Active Use,” “Moderate Risk”)
    - What happened – describes the detected action
    - Likely purpose – explains the inferred purpose (e.g., login, authentication, account access, personalization, analytics)
    - Why it matters – explains the privacy or security relevance
    - Where it went – indicates whether the data stayed on the website or was sent externally
    - What you can do – suggests optional user actions
    - Evidence – shows the source of detection (e.g., browser observation, network request, or storage event)
3. **Expandable technical evidence sections**, allowing advanced users to inspect underlying events without overwhelming the main interface.
4. **Summary panel and activity timeline**, which provide an overview of data categories, usage status, external sharing, and the sequence of detected actions.

Figure 4.7 shows the final Figma prototypes. These illustrate the card layout, high-level indicators, summary panel, and activity timeline, reflecting the user-centered design of the interface and ensuring that information is both accessible and actionable for non-technical users.

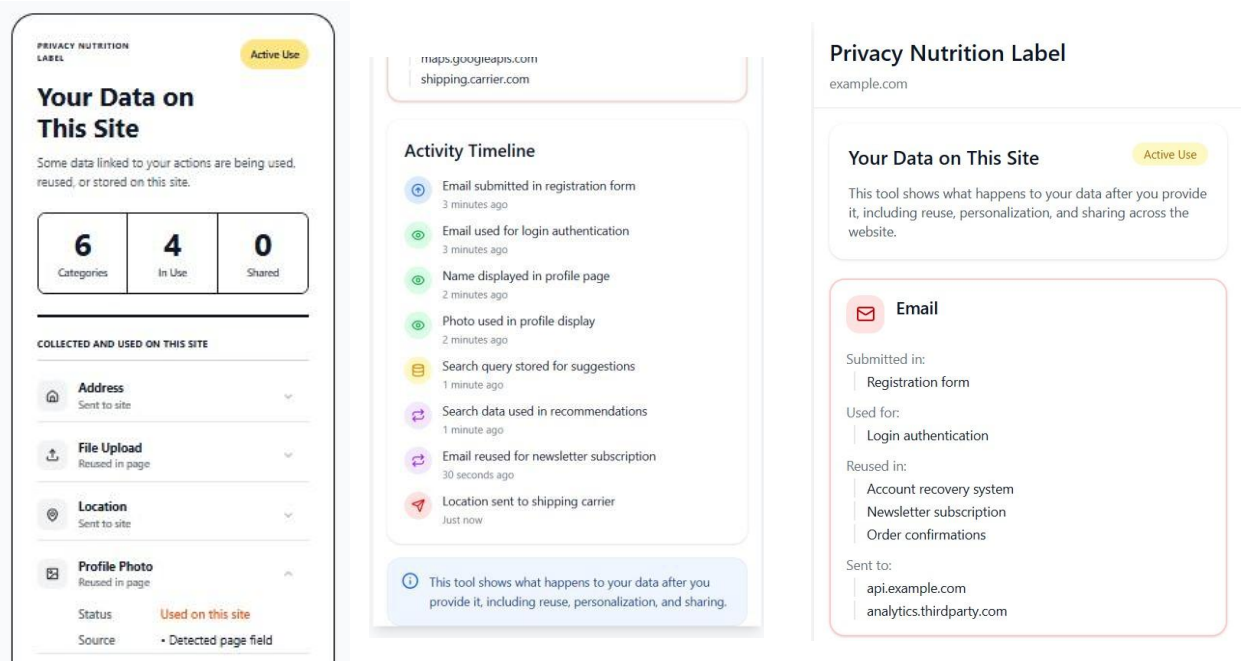


Figure 4.7 - Figma prototypes of the Privacy Nutrition Label, illustrating three high-level indicators, card-based visualization of personal data categories, a summary panel, and activity timeline.

These prototypes demonstrate how the Privacy Nutrition Label translates privacy-related observations into a clear, structured, and understandable interface, providing users with immediate insight into how their personal data is handled and informing subsequent UI and logic design decisions.

4.4 Recognizing Data Processing Actions

The logic design of the Privacy Nutrition Label translates observable browser events into user-facing explanations, enabling users to understand how their personal data is processed in real time. The system monitors events such as user input, form submissions, network requests, browser storage, DOM reuse, and account-related actions, mapping these to GDPR-inspired processing categories. This design ensures that users receive actionable information about their personal data without needing technical knowledge.

4.4.1 Personal Data Categories

The tool currently focuses on a predefined set of common personal data categories. These categories were selected because they commonly appear in web forms and account/profile flows, and because they are relevant to the GDPR definition of personal data.

Category	Example Indicators
Email	email, e-mail, mail
Name	name, first name, last name, full name
Phone number	phone, mobile, tel, telephone
Username	username, login, handle
Password	password, passwd
Address	address, street, city, postcode, country
Payment information	card, CVV, IBAN, billing, payment
Date of birth	dob, birth, birthday
File upload	file, upload, attachment
Profile photo	photo, avatar, profile picture

Search activity	search, query, keyword
Location	location, latitude, longitude, geo
Account activity	account, session, sign in, sign up

Figure 4.8 - Table 5 Personal Data Categories Used by the Tool

This set is intentionally modular. It represents the most common and most user-relevant personal data categories, but it can be extended in future versions of the tool by adding new keywords and classification rules.

4.4.2 Event-Based Detection

The tool adopts an event-based detection approach. This means that it does not rely on a single scan of the webpage. Instead, it observes multiple browser events that may indicate personal data processing.

Event Type	Description	Processing Action
Input events	User typing into form fields	Collection
Form submission	Explicit submission of user data	Collection / Transmission
Network requests	Data transmitted through fetch, XHR, WebSocket, or form submission	Transmission

Storage usage	Data stored in localStorage or sessionStorage	Storage
DOM reuse	Data reappearing in the page interface	Use / Reuse
User actions	Clicks, login actions, search actions, permission requests	Contextual processing

Figure 4.9 - Table 6 Event Types Monitored by the Tool

After detecting events, the system classifies each field or event into a personal data category and infers the likely purpose based on context (e.g., login, authentication, checkout, analytics, personalization, profile). Purpose inference is heuristic, providing probable intent rather than legal determination.

4.4.3 Keyword-Based Classification

Because websites do not provide standardized metadata describing whether a form field contains personal data, the tool uses heuristic keyword-based classification. It analyses field names, IDs, placeholders, labels, input types, autocomplete attributes, and surrounding context. For example:

Keyword / Pattern	Classified Category
email, e-mail	Email
phone, mobile, tel	Phone number
firstname, lastname, name	Name
password, passwd	Password

address, city, postcode	Address
card, cvv, iban, billing	Payment information
search, query	Search activity
location, geo, latitude	Location

Figure 4.10 - Table 7 Keywords

This approach does not claim perfect semantic understanding. Instead, it provides a practical and scalable way to recognise common personal data categories across different websites. The logic can be extended later by adding more keywords, categories, or contextspecific rules.

4.4.4 Recognizing Network Transmission

When data is transmitted through network requests, the system analyses request destinations, methods, field keys, and value fingerprints. If the request destination belongs to the same base domain as the current website, the data is treated as used on the current website. If the destination belongs to a different domain, it may be treated as sent outside the website.

This distinction supports the label indicators “Used here” and “Sent outside”. However, the system remains careful in its wording: it does not claim legal third-party sharing unless the destination clearly belongs outside the current website context.

4.4.5 Recognizing Storage

The tool monitors browser storage mechanisms such as localStorage and sessionStorage. When user-related data is stored locally, the label can explain that the data may remain available after leaving or refreshing the page.

This is important because users may assume that data disappears after closing a form or navigating away. Storage detection therefore provides visibility into persistence, which is a meaningful data processing action under the GDPR concept of storage.

4.4.6 User-Facing Explanation Layer

The final step of the logic model is the generation of user-facing explanations. Technical events are translated into short explanations such as:

Detected Action	User-Facing Explanation
Input only	“You entered this data on the page.”
First-party network request	“This data was sent to this website.”
External request	“This data was sent outside this website.”
Storage event	“This data was saved in your browser.”
DOM reuse	“This data was shown again on the page.”
Analytics endpoint	“This data may be used to measure or track activity.”
Recommendation endpoint	“This data may be used to personalize what you see.”

Figure 4.11 - Table 8 Explanations to Users

This explanation layer is important because the aim of the tool is not merely to detect technical events, but to make them understandable. The label therefore acts as a translation layer between browser-level activity and user awareness.

4.4.7 Data Processing Mapping Model

The Excel table created during development serves as the core of the logic design, mapping observable events to user-facing explanations. This includes the data category, associated keywords, event type, GDPR processing action, inferred purpose, explanation text, and technical evidence.

Data Processing Mapping Model

Data Category	Keywords / Indicators	Event Type	GDPR Action	Possible Purpose	What the user sees	Technical Evidence
Email	email, e-mail	Input	Collection	Account creation / Login	You entered your email	Input field
Email	email in request body	Network	Transmission	Account creation	Your email was sent to the server	API request
Passwor d	password, passwd	Input	Collection	Authentication	You entered your password	Input field
Address	address, city, zip	Input	Collection	Delivery / Checkout	You entered your address	Form field
Address	address stored	Storage	Storage	Save for later use	Your address was saved in your browser	localStorage
Search	search, query	Input	Collection	Search functionality	You searched for something	Search field
Search	query sent	Network	Transmission	Search processing	Your search was sent to the server	API request
Name	name, firstname	Input	Collection	Account creation	You entered your name	Form field
Profile data	profile, avatar	DOM reuse	Use	Profile display	Your data is shown in your profile	DOM
Location	geo, location	Permissi on	Collection	Location service	The site requested your location	Geolocation API
Payment	card, cvv, iban	Input	Collection	Payment	You entered payment info	Input field
Payment	payment sent	Network	Transmission	Transaction	Your payment data was sent	API

Figure 4.12- Excel-based mapping of browser events to personal data categories, processing actions, inferred purposes, user-facing explanations, and technical evidence.

4.4.8 Limitations of the Recognition Logic

The recognition logic has several limitations. First, the tool can only observe activity available in the browser environment. It cannot directly inspect server-side processing, backend databases, or internal business logic. Second, classification is heuristic and may produce false positives or false negatives. Third, purpose inference is based on keywords and observable context, so it should be interpreted as an informed indication rather than a definitive legal conclusion.

For this reason, the tool includes a clear disclaimer explaining that it is based on observable browser activity and does not prove hidden backend processing. This limitation is not a weakness of the interface, but an important part of responsible privacy communication.

4.4.9 Summary

The logic design of the tool connects GDPR-inspired processing concepts with observable browser events. By combining input detection, form submission monitoring, network observation, storage monitoring, DOM reuse detection, and contextual inference, the system can generate a real-time Privacy Nutrition Label that explains what happens to userprovided data.

This design supports the main aim of the thesis: to make personal data processing more transparent, understandable, and visible to users while they interact with web platforms.

4.5 Development (Languages, Technologies, and Architecture)

The Privacy Nutrition Label extension was implemented using standard web technologies to enable real-time monitoring of user interactions while maintaining a clear, structured interface. The architecture is modular, separating data capture, classification, inference, and user interface rendering, ensuring maintainability and extensibility.

4.5.1 Core Modules

Table 9 presents the main modules of the extension, their responsibilities, and key functions. Each module was designed to handle a distinct layer of the system, from browser event monitoring to label generation and interface rendering.

File / Module	Main Responsibility	Important Functions	Description
manifest.json	Extension configuration	permissions(), content_scripts, background, side_panel	Defines extension architecture, permissions, scripts, and sidepanel
content.js	DOM monitoring and user interaction	scanExistingPageData(), classifyInteractiveElement(), inferDomLocation()	Detects forms, fields, uploads, searches, profile elements, and visible user data
injected.js	Browser API interception	patchFetch(), patchXHR(), patchStorage(), patchGeolocation()	Monitors runtime fetch requests, storage usage, geolocation access
background.js	Global extension state	handleEvent(), updateState(), broadcastUpdate()	Stores monitoring events and synchronizes label state across modules

labelEngine.js	Privacy label generation	buildPrivacyLabel(), buildCard(), inferPurposes(), computeStatus(), groupByCategory(), detectThirdPartySharing(), uniqueValues()	Converts raw technical events into user-friendly privacy explanations and final label
sidepanel.js	UI rendering logic	renderSummary(), renderAccordion(), renderTimeline()	Dynamically generates the side panel layout, cards, summary, and timeline
sidepanel.html	Side panel structure	HTML containers	Defines visual layout for the browser side panel
style.css	UI styling	accordion styles, card styles, timeline	Implements visual design and privacy label aesthetics

Figure 4.13 - Table 9 Core Implementation Modules

4.5.2 Detection Functions

The extension relies on multiple **detection functions** to identify user-related data and monitor usage in real time. Table 10 summarizes these functions, their locations, purposes, and example outputs.

Function	Location	Purpose	Example Output

classifyString(text)	content.js	Detects personal data categories via heuristics and keywords	"email", "phone", "location"
scanExistingPageData()	content.js	Scans page for visible data and form fields	"Logged-in profile info"
readElementValue(el)	content.js	Extracts values from input fields	"test@example.com"
inferDomLocation(el)	content.js	Infers where data is reused on the page	"Profile page", "Search flow"
inferContext(el)	content.js	Detects contextual purpose of fields/actions	"Login form", "Checkout form"
extractFieldKeys(body)	injected.js	Extracts request keys from network payload	email, phone, query
extractValueFingerprints(body)	injected.js	Generates fingerprints of values without storing raw data	hashed fingerprints
patchFetch()	injected.js	Intercepts fetch requests	detects request destinations
patchXHR()	injected.js	Intercepts XMLHttpRequest traffic	monitors API requests

patchStorage()	injected.js	Monitors local/sessionStorage usage	detects storage reuse
patchGeolocation()	injected.js	Monitors geolocation access	detects location usage

Figure 4.14 - Table 10 Main Detection Functions

4.5.3 Privacy Label Engine Functions

The Privacy Label Engine transforms raw events into human-readable cards displayed in the label. Table 11 lists the core functions:

Function	Responsibility	Description
buildPrivacyLabel()	Main label generation	Aggregates monitoring events into the final label structure
buildCard()	Card creation	Creates individual cards for each data category
inferPurposes()	Purpose inference	Infers why data is being used based on context
computeStatus()	Privacy status	Determines if data is observed, used, or shared externally
groupByCategory()	Data aggregation	Groups monitoring events by personal data category

detectThirdPartySharing()	External analysis	Detects data sent outside current domain
uniqueValues()	Duplicate cleanup	Removes duplicate destinations and detected purposes

Figure 4.15 - Table 11 Label Engine Core Functions

4.5.4 Monitored Browser APIs

The extension observes several **browser APIs** to capture runtime privacy-relevant events. Table 12 summarizes these APIs and their detection purposes.

Browser API	Monitoring Purpose	Example Detection
fetch()	Detect outgoing requests	Email transmitted to API
XMLHttpRequest	Detect AJAX traffic	Search activity sent to backend
localStorage	Detect local storage reuse	Profile information stored locally
sessionStorage	Detect temporary storage	Session identifiers reused
navigator.geolocation	Detect location access	Website requested location
DOM Mutation Observer	Detect interface reuse	Profile photo reused on page
Form submission events	Detect data entry actions	Registration form submitted

Figure 4.16 - Table 12 Monitored Browser APIs

4.6 Tool Functionalities

The Privacy Nutrition Label browser extension offers users a comprehensive and interactive interface to observe how their personal data are handled in real time while interacting with websites. The tool translates browser-level events including user inputs, network requests, storage activity, and DOM changes into clear, actionable insights, presented via a card-based Privacy Nutrition Label that emphasizes user comprehension over technical detail.

The extension monitors user activity across forms, search fields, file uploads, profile sections, and account areas, detecting when data is entered, reused, stored locally, or transmitted to third-party services. High-concern categories, including email, passwords, payment information, and location, trigger alerts to draw user attention, while lowerconcern categories are displayed in the main label view without disruption.

Each data category is presented as a card that provides multiple layers of information. The primary fields include:

- **Privacy alert:** Signals if a sensitive action was observed, such as data transmitted externally or stored locally.
- **What happened:** Describes the observed processing action in plain language.
- **Likely purpose:** Infers why the data was processed, based on contextual clues such as form type, network destination, or DOM location.
- **Why it matters:** Explains the privacy implications for the user.
- **Where it went:** Indicates whether the data remained on the current website or was sent externally.
- **What you can do:** Suggests optional user actions to mitigate risk.
- **Evidence:** Details the source of detection, including DOM nodes, network channels, storage keys, or user actions.

Technical evidence is contained in expandable sections, allowing advanced users to inspect raw data without overwhelming the primary interface. This ensures that both novice and advanced users can extract meaningful insights from the label.

A summary panel complements the cards, displaying the total number of data categories observed, those in active use on the website, and those transmitted externally. The activity timeline provides a chronological view of detected events, including form submissions, network requests, storage operations, and DOM reuse, allowing users to see the sequence of data handling during their browsing session. Icons and textual annotations are used to clarify the nature of each event.

The extension architecture, informed by the previously described modules content.js, injected.js, labelEngine.js, sidepanel.js, background.js, sidepanel.html, and style.css supports modularity and maintainability. Events detected by content.js and injected.js are processed in labelEngine.js to generate the structured card interface, while sidepanel.js renders the interface, summary panel, and activity timeline in real time. The system design ensures that all events captured in the browser are mapped to user-understandable feedback, aligning with GDPR principles, survey findings, and best practices in usable privacy.

The extension also handles advanced interactions such as external authentication flows, location permissions, and account actions, providing specialized cards for OAuth logins or geolocation requests. Cards are dynamically sorted by privacy relevance, with higher-risk actions appearing more prominently to the user. The badge system at the top of the panel indicates the current state of privacy monitoring, switching between “Monitoring” and “Privacy Alert” based on detected events.

Figures 4.17–4.19 illustrate these functionalities. Figure 4.17 shows the side panel with high-level indicators and personal data cards. Figure 4.18 presents the activity timeline, highlighting the sequence of interactions and observed events. Figure 4.19 demonstrates the expandable technical evidence section, which provides advanced users with detailed event metadata and network/storage traces.

In sum, the Privacy Nutrition Label extension delivers real-time visibility into user data handling, presenting technical observations in a legible, user-centered, and actionable format. By combining continuous monitoring, purpose inference, card-based visualization,

and interactive timelines, the tool empowers users to understand, control, and trust the way their personal data are processed while using web platforms.

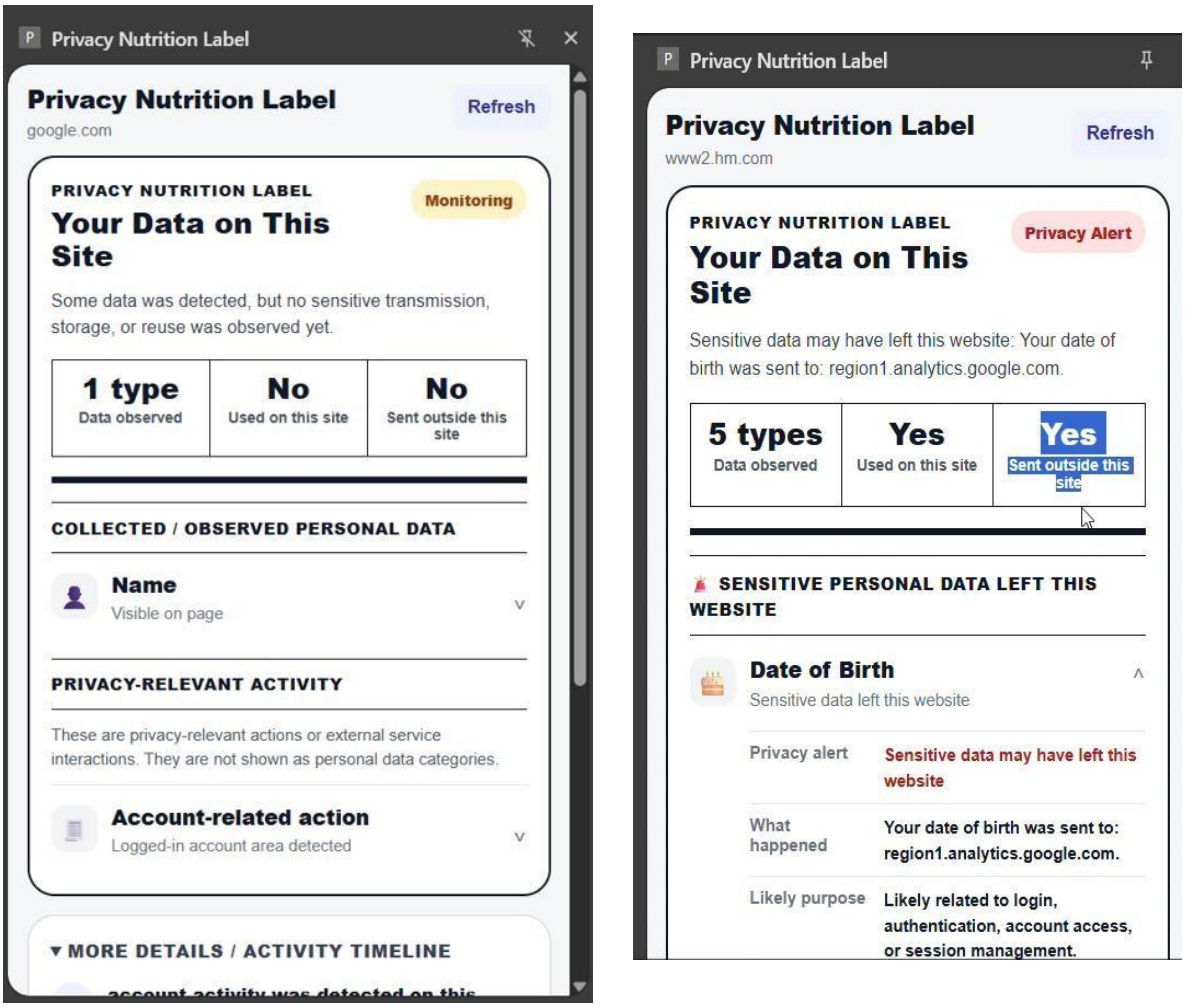


Figure 4.17 - Side panel view of the Privacy Nutrition Label showing the three high-level indicators: Data Entered, Used Here, and Sent Outside.

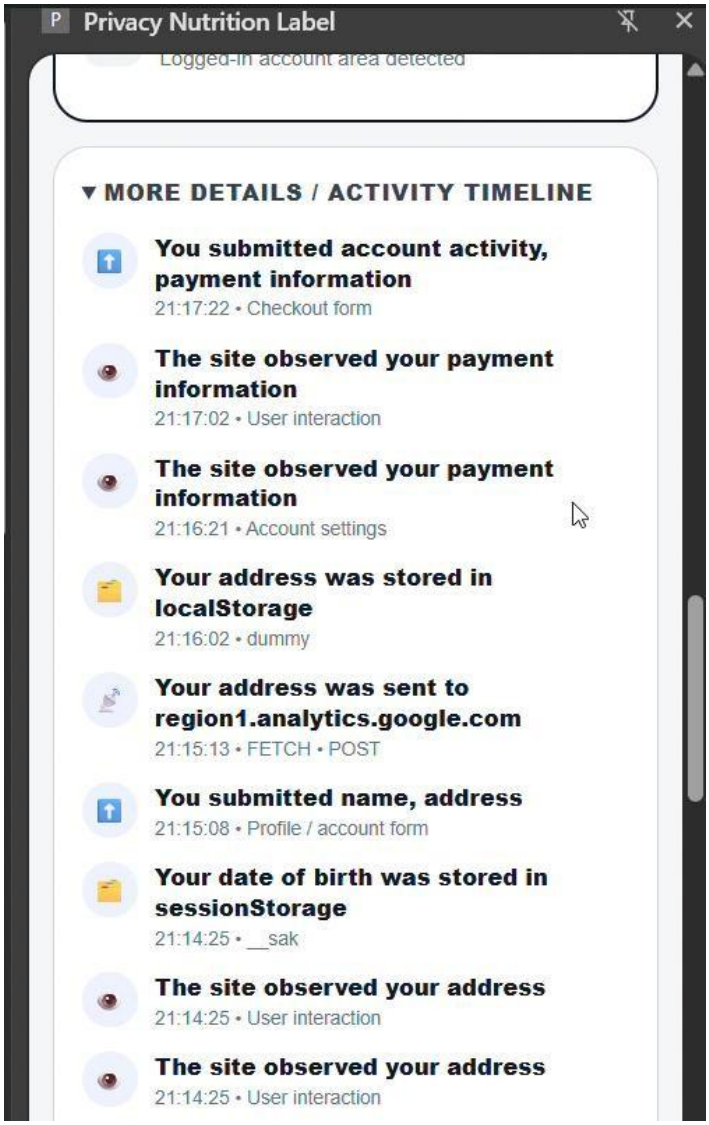


Figure 4.18 - Activity timeline illustrating the chronological sequence of detected data processing events, including form submissions, network requests, storage access, and DOM reuse.

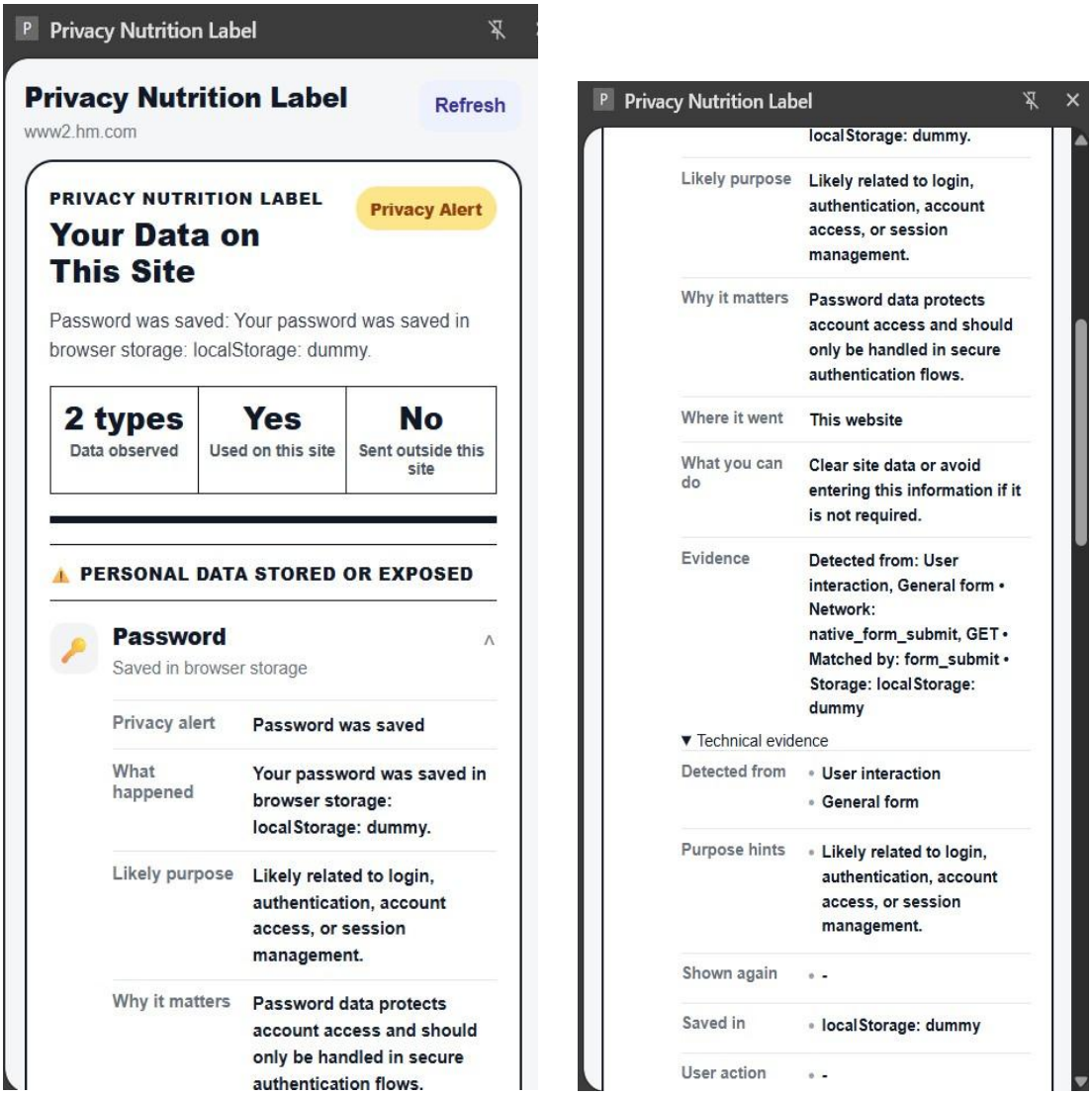


Figure 4.19 - Expandable technical evidence section for advanced users. This view allows inspection of underlying browser events such as fetch requests, XMLHttpRequests, storage keys, and DOM interactions.

Chapter 5

5. Evaluation of the Real-time Privacy Nutrition Label Tool

5.1 Introduction

This chapter presents the evaluation of the proposed Privacy Nutrition Label browser extension. The evaluation focuses on evaluating the perceived user experience, usability, usefulness, and acceptance of the tool. The evaluation was conducted using the User Experience Questionnaire (UEQ) [7], as well as questions adapted from the System Usability Scale (SUS) [15] and the Technology Acceptance Model (TAM). [16]

5.2 Methodology

The evaluation was conducted using an online survey by using Google Forms, distributed via social media and personal communication channels. Participation was voluntary and anonymous, and respondents were informed that the data collected would be used exclusively for academic research purposes. The survey remained open for 5 days, during which 75 participants completed the questionnaire. All responses were exported from Google Forms to a spreadsheet format for analysis and visualization.

The survey was organized into **three sections**:

1. **Background Information (Figure 5.1):** This section collected information on participants' age, familiarity with technology and websites, frequency of use of

websites that require registration, the importance of privacy for them, and their level of familiarity with privacy.

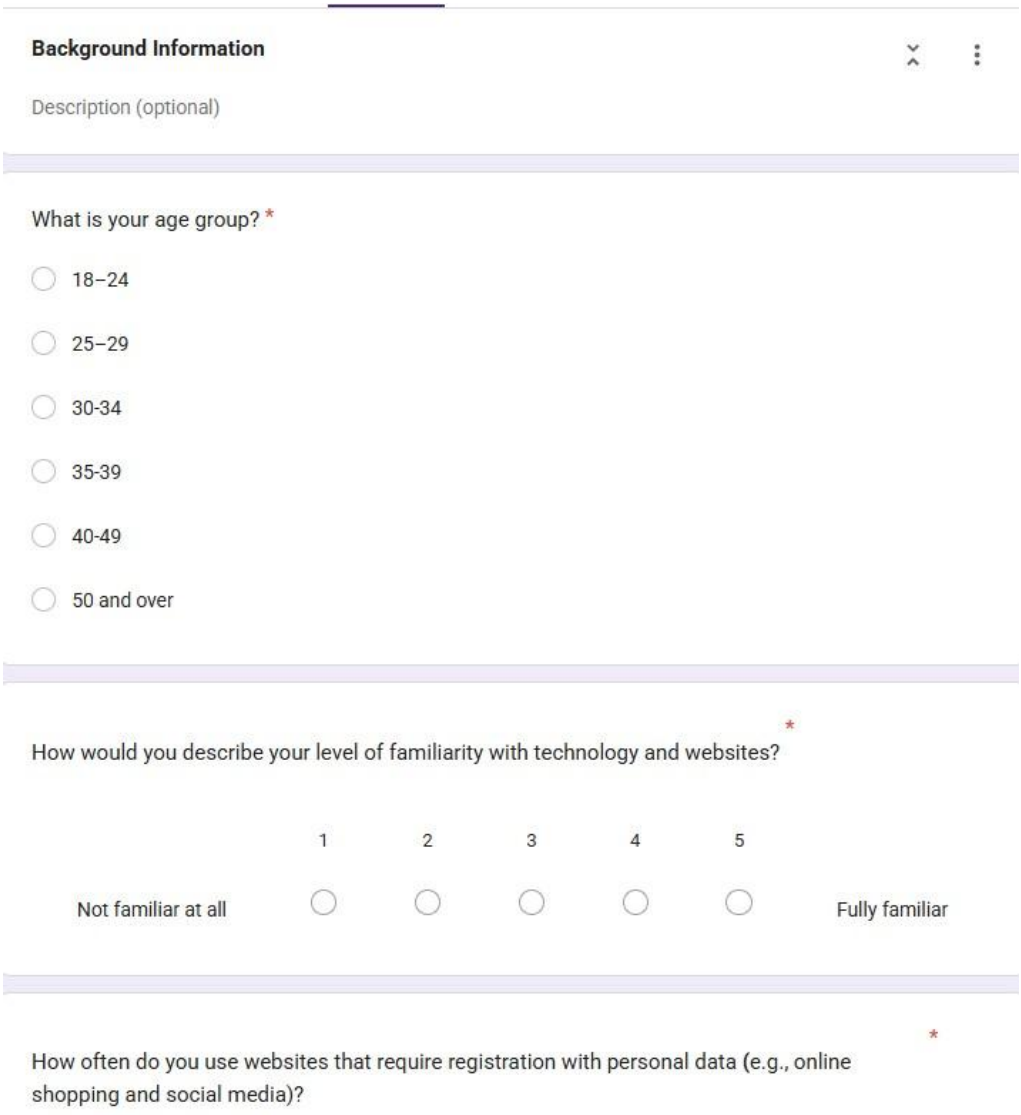


Figure 5.1 - Demographic Information

2. **Video Demonstration and UEQ-Based Evaluation (Figure 5.2 - 5.3):**
   Participants first watched a short demonstration of the Privacy Nutrition Label tool. Following the demonstration, they responded to 26 Likert-scale (1-7) pairs of items

derived from the User Experience Questionnaire (UEQ), These questions aimed to assess participants perceived user experience, on six key dimensions:
Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty.

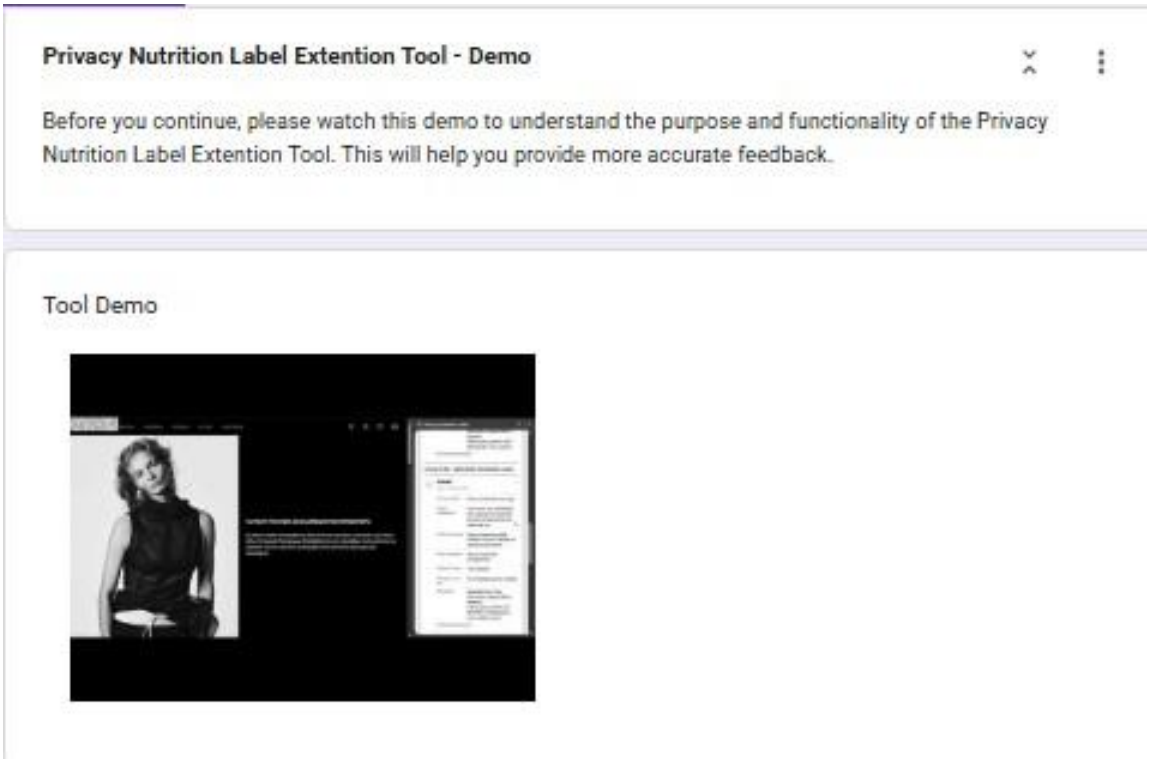


Figure 5.2 - Video Demonstration

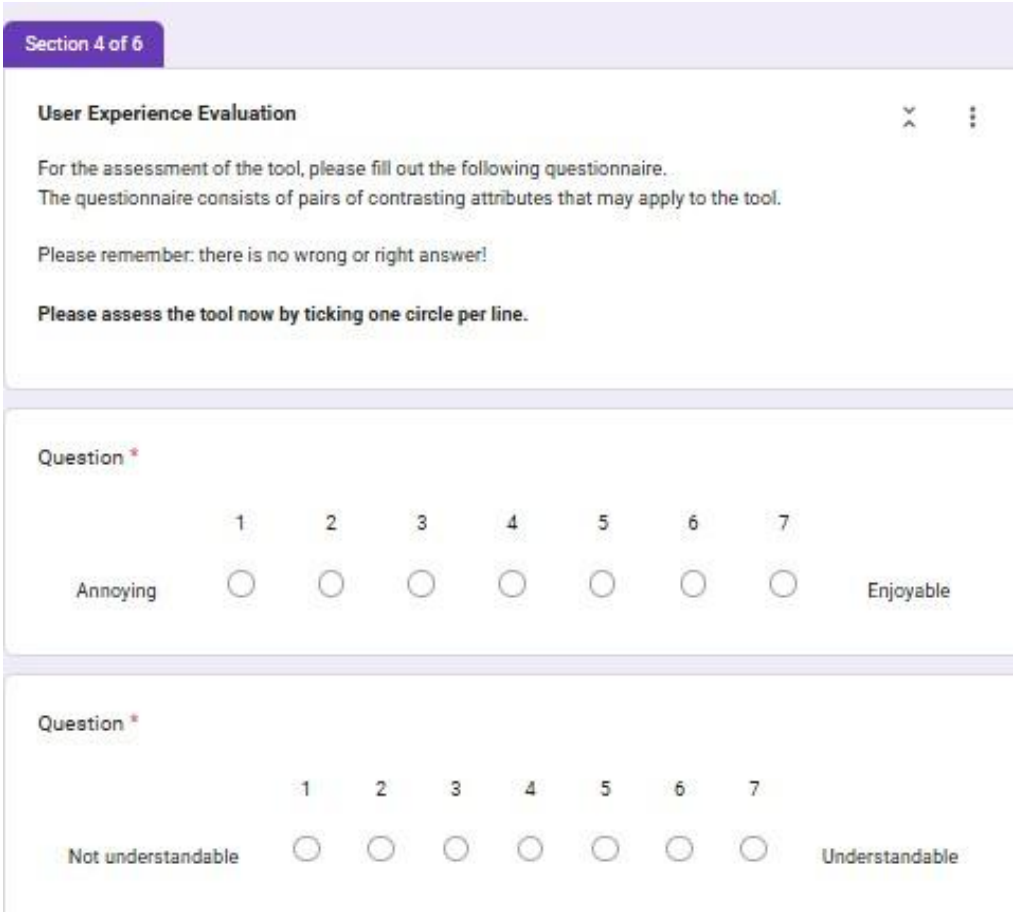


Figure 5.3 - UEQ-Based Evaluation

3. **Perceived usefulness, Usability, and Acceptance:**

Participants rated statements regarding perceived usefulness, usability, and acceptance of the tool on a five-point Likert scale (1 = completely disagree, 5 = completely agree). The statements were adapted from the SUS and TAM questionnaires. This section provides insights into participants' expectations and beliefs about the tool enhancing their trust and awareness in privacy.

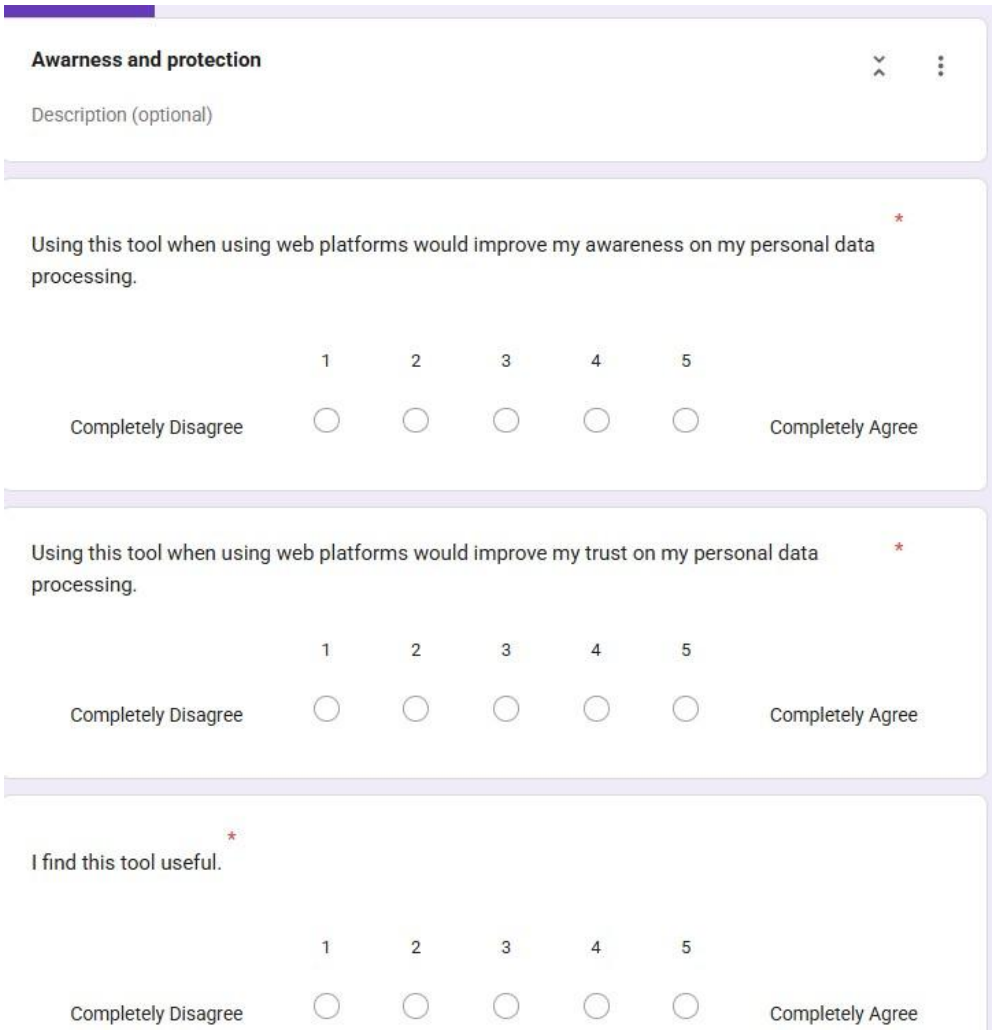


Figure 5.4: Perceived usefulness, Usability, and Acceptance

5.3 Results

5.3.1 Background Information

The first section of the user evaluation survey collected demographic and general background information from the participants. This data provides context for interpreting subsequent responses regarding tool usability, trust, and awareness.

- **Age Group**

The age distribution (Figure 5.1) shows that the majority of participants (57.3%) belong to the 18–24 age group, followed by 25–29 (21.3%) and 30–34 (8%). Smaller proportions are observed in the 35–39 (5.3%), 40–49 (4%), and 50 and over (3.7%) groups. This indicates that the evaluation primarily reflects the perceptions and experiences of younger adult users, who represent a substantial portion of typical web platform users.

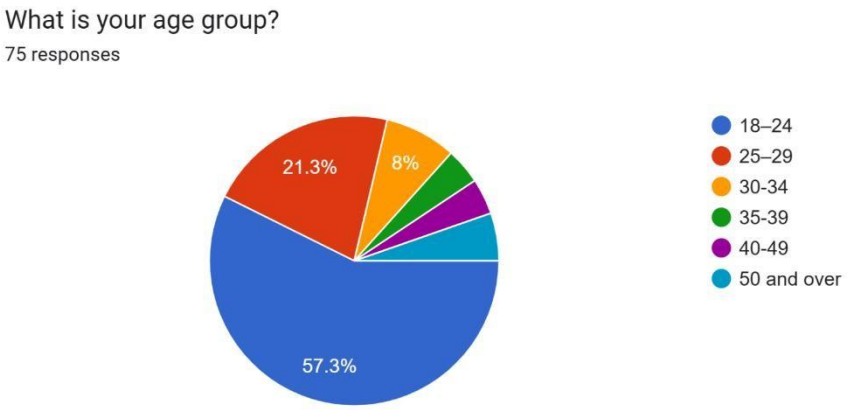


Figure 5.5 - Age distribution of participants (n=75).

- **Familiarity with privacy concepts**

Participants were also asked to rate their familiarity with technology and websites on a scale from 1 (“Not familiar at all”) to 5 (“Fully familiar”). Most respondents reported high familiarity, with 66.7% selecting the highest rating and 17.3% selecting the second-highest level. This suggests that the sample includes participants with sufficient digital literacy to evaluate a browser-based tool effectively.

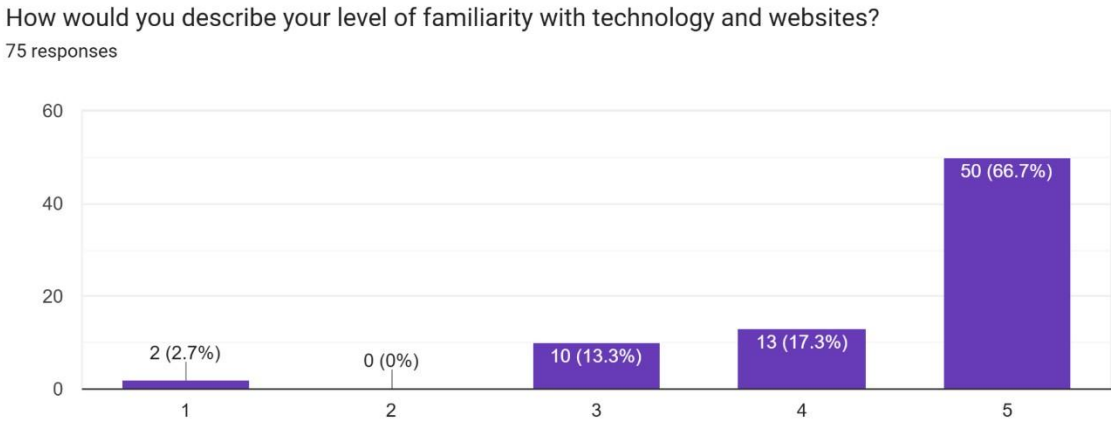


Figure 5.6 - Self-reported familiarity with technology and websites (1–5 scale).

- **Frequency of using websites**

When asked how often they use websites requiring registration with personal data, such as online shopping or social media platforms, 69.3% indicated frequent or constant use (ratings 4 or 5), highlighting that the tool was evaluated by users who regularly interact with personal data online.

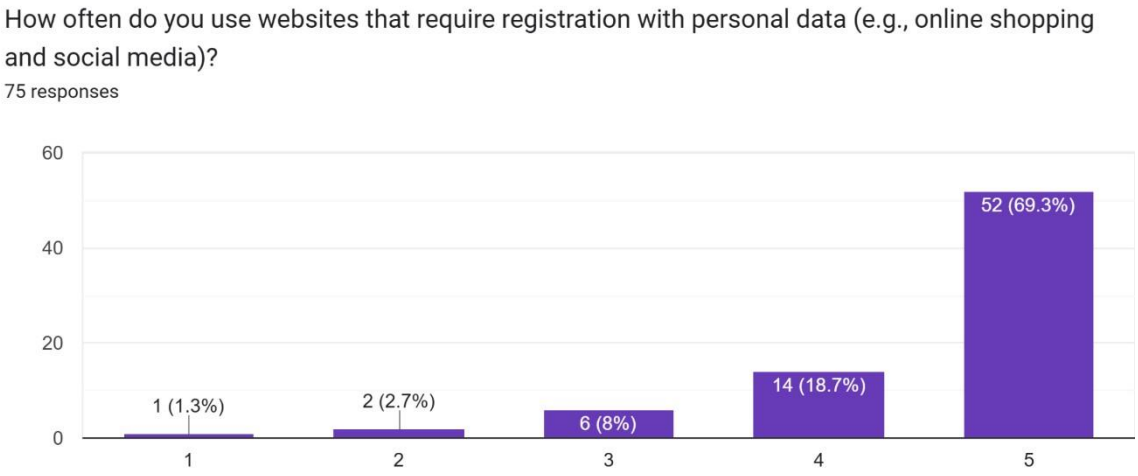


Figure 5.7 - Frequency of using websites that require registration with personal data (1–5 scale).

- **Importance of online privacy**

The importance of online privacy was also assessed. The majority of participants (82.7%) rated privacy as extremely important (rating 5), while 12% considered it very important (rating 4). Only a small fraction (5.3%) rated privacy as moderately important (rating 3 or below). This demonstrates that the participants are highly privacy-conscious, reinforcing the relevance of the evaluation.

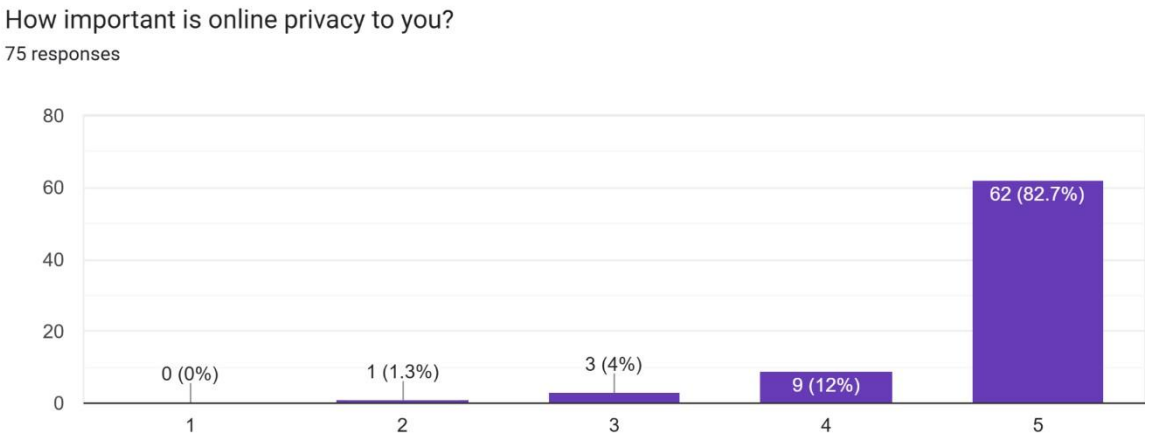


Figure 5.8 - Importance of online privacy (1–5 scale).

- **Level of familiarity with privacy**

Finally, participants rated their familiarity with privacy concepts and practices on a scale from 1 to 5. Responses were distributed more evenly across the spectrum, with 28% indicating high familiarity (rating 5), 26.9% moderate-high familiarity (rating 4), and 25.3% moderate familiarity (rating 3). Lower familiarity ratings were less common (ratings 1 and 2 together account for 29.3%). This distribution shows that while participants are generally aware of privacy, there is still variability in their understanding, underscoring the value of tools that provide clear, actionable privacy information.

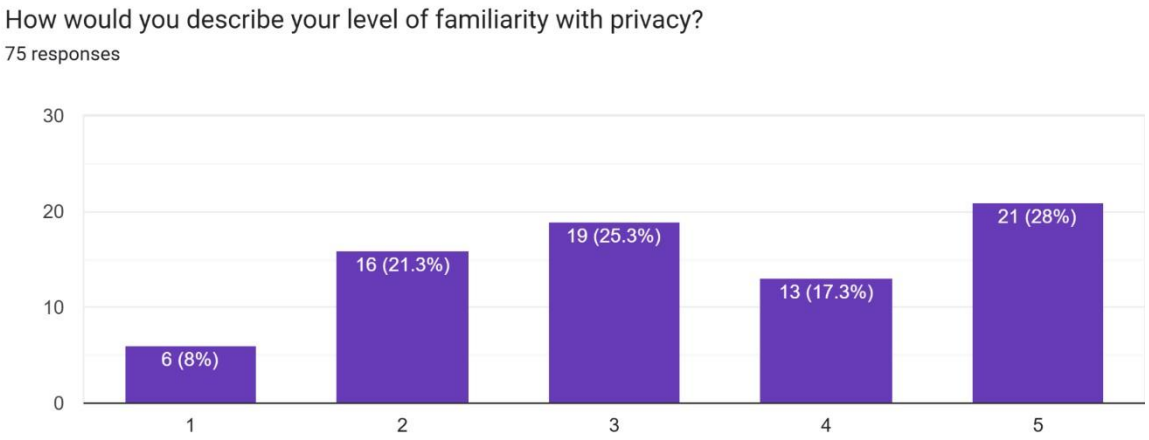


Figure 5.9 - Self-reported familiarity with privacy concepts (1–5 scale).

In summary, the background information illustrates that the participant sample is predominantly young, digitally literate, active on web platforms requiring personal information, and highly concerned about privacy. These characteristics are critical for interpreting the evaluation of the Privacy Nutrition Label tool and contextualize subsequent findings in user experience and perceived awareness.

5.3.2 User Experience Evaluation (UEQ)

The user experience of the Privacy Nutrition Label browser extension was evaluated using the User Experience Questionnaire (UEQ), a validated instrument for assessing pragmatic and hedonic quality of interactive systems. This questionnaire examines six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty, providing a robust assessment of both task-related and non-task-related qualities.

The results presented here are based on the responses of 75 participants who completed the questionnaire after interacting with the tool.

The analysis follows the structure used in Chapter 3 for the baseline survey, ensuring consistency and comparability across studies.

UEQ Scales (Mean and Variance)		
Attractiveness	2.684	0.58
Perspicuity	2.570	0.76
Efficiency	2.707	0.53
Dependability	2.560	0.63
Stimulation	2.657	0.68
Novelty	2.530	0.97

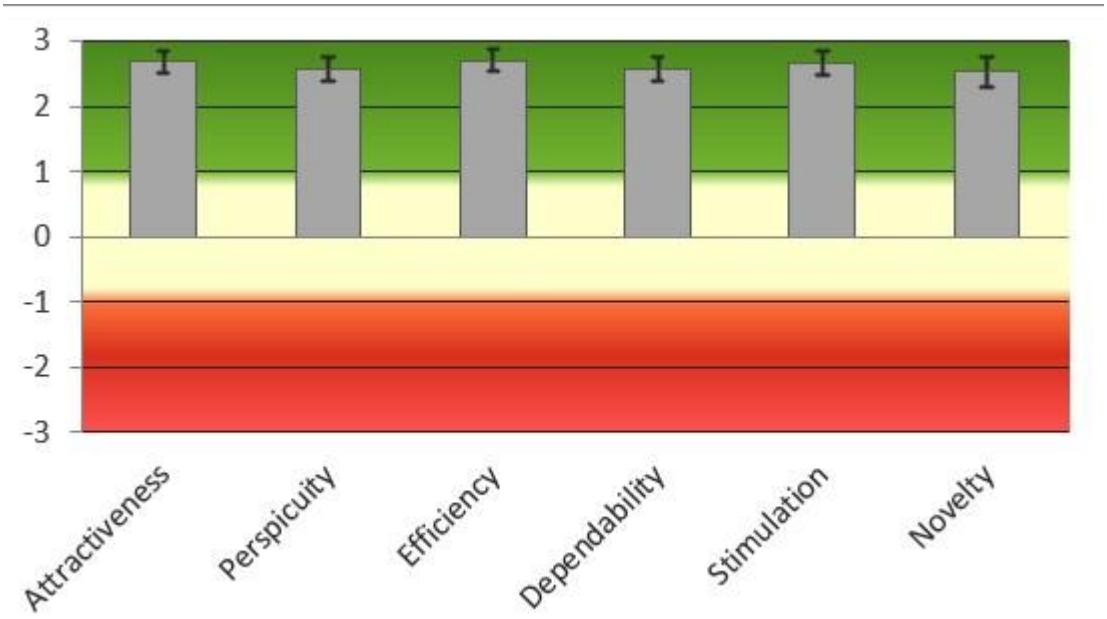


Figure 5.9 - illustrates the mean score for each UEQ dimension, together with the associated variance across all participants.

The results depicted in Figure 5.9 indicate that participants generally rated the tool positively across all dimensions. Attractiveness received a mean score of 2.68, reflecting users' overall positive impression of the interface's appeal. Perspicuity and Efficiency scored 2.57 and 2.71 respectively, suggesting that users found the tool clear, comprehensible, and effective in performing its intended functions. Dependability, with a mean of 2.56, indicates that participants considered the tool reliable in providing consistent feedback. The hedonic dimensions, Stimulation (2.66) and Novelty (2.53), suggest that the

tool also engaged users and introduced elements perceived as innovative compared to conventional privacy interfaces.

From Figure 5.10, it is evident that items related to user comprehension (e.g., “not understandable/understandable”) and usability (e.g., “difficult to learn/easy to learn”) scored particularly high, reinforcing that the tool is both intuitive and approachable for end users. Items representing hedonic qualities such as novelty and stimulation also scored consistently above the neutral midpoint, supporting the notion that participants found the interface engaging and motivating.

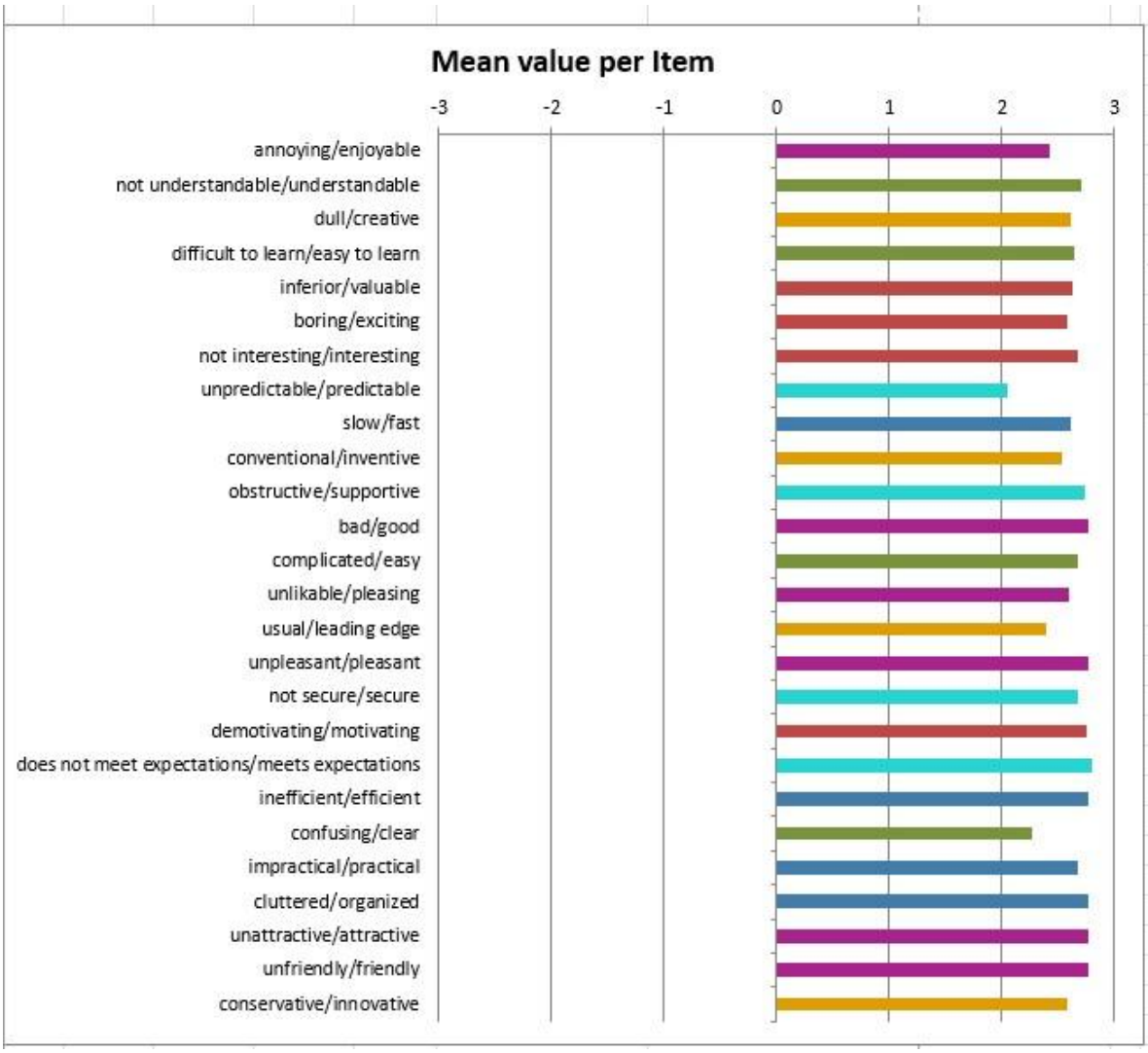


Figure 5.10 - Mean value per item for UEQ questionnaire responses (75 participants).

Finally, Figure 5.11 presents a summary of the aggregated pragmatic and hedonic quality scores, which combines related dimensions to provide an overall assessment.

Pragmatic and Hedonic Quality	
Attractiveness	2.68
Pragmatic Quality	2.61
Hedonic Quality	2.59

Figure 5.11

The aggregate scores indicate a positive overall perception: the Pragmatic Quality mean of 2.61 reflects the system's usability and reliability, while the Hedonic Quality mean of 2.59 demonstrates that participants enjoyed the experience and perceived it as innovative. Attractiveness (mean = 2.68) confirms the interface design was visually appealing and professionally presented.

5.3.3 Awareness and Protection

The final part of the user evaluation questionnaire focused on awareness and protection, assessing whether the Privacy Nutrition Label tool could improve users' understanding of personal data processing, enhance trust in web platforms, and support informed decisionmaking. Participants were asked to indicate their level of agreement with statements regarding the tool's impact using a five-point Likert scale (1 = Completely Disagree, 5 = Completely Agree). The survey remained open for five days, with a total of 75 respondents completing this section.

- **Using this tool when using web platforms would improve my awareness on my personal data processing.**

  Figure 5.5 presents the distribution of responses for the item "Using this tool when using web platforms would improve my awareness of my personal data processing." The chart shows that most participants strongly agreed with this statement (92% of respondents), indicating that the tool effectively supports users in understanding how their data is handled. Only a small minority reported neutral or lower agreement (2.7% selected 3 and 5.3% selected 4).

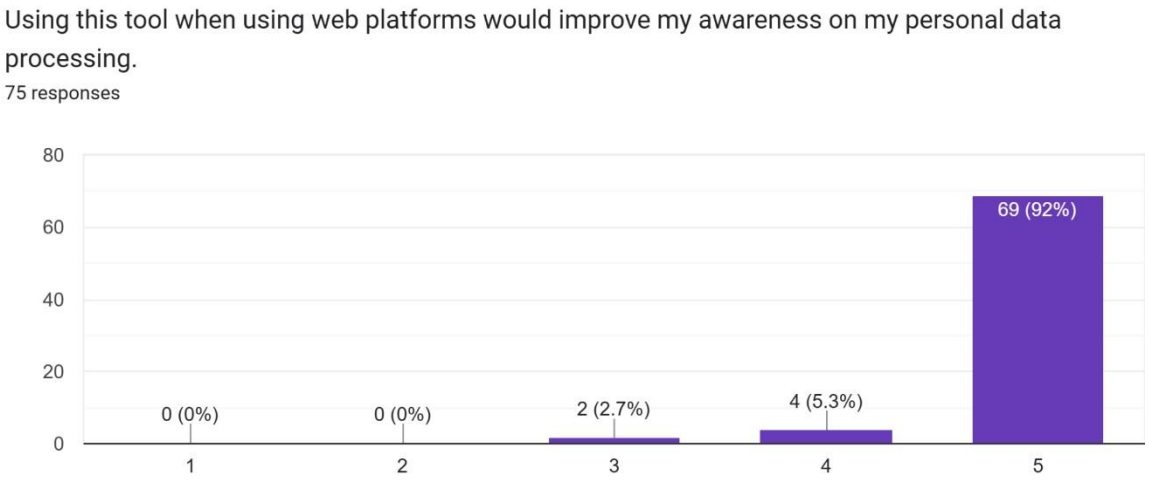


Figure 5.12 - Responses to "Using this tool improves my awareness of personal data processing".

- **Using this tool when using web platforms would improve my trust on my personal data processing.**

  The second item, "Using this tool when using web platforms would improve my trust on my personal data processing," yielded identical results, with 92% of respondents selecting the highest agreement level. This demonstrates that participants perceive the tool not only as informative but also as trust-enhancing, helping users feel more confident about data handling on websites.

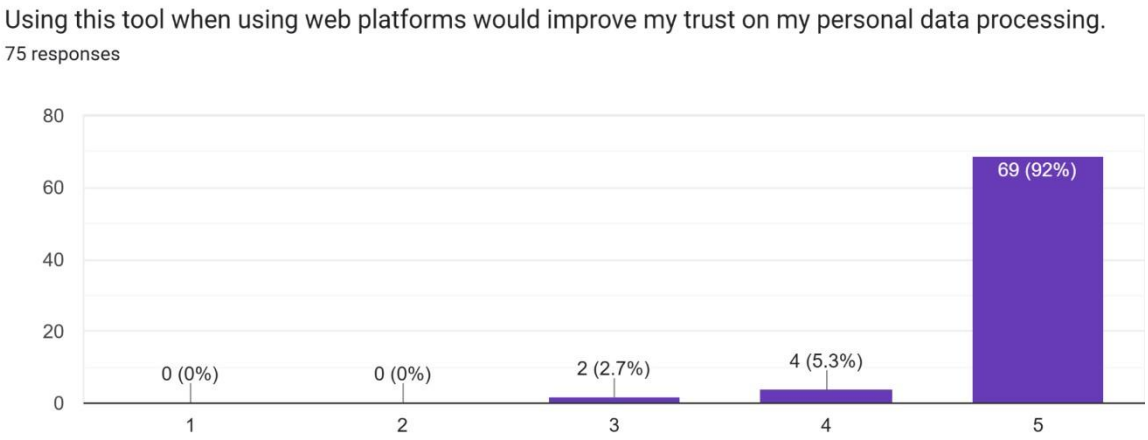


Figure 5.13 - Responses to "Using this tool improves my trust in personal data processing".

- **I find this tool useful.**

  The third item, "I find this tool useful," also showed a very high degree of positive perception, with 92% of participants selecting "Completely Agree," further confirming the practical utility of the extension in a real-world context. A small number of respondents selected slightly lower agreement levels (1.3% for 2, 2.7% for 3, 4% for 4), indicating minor variation in subjective perception.

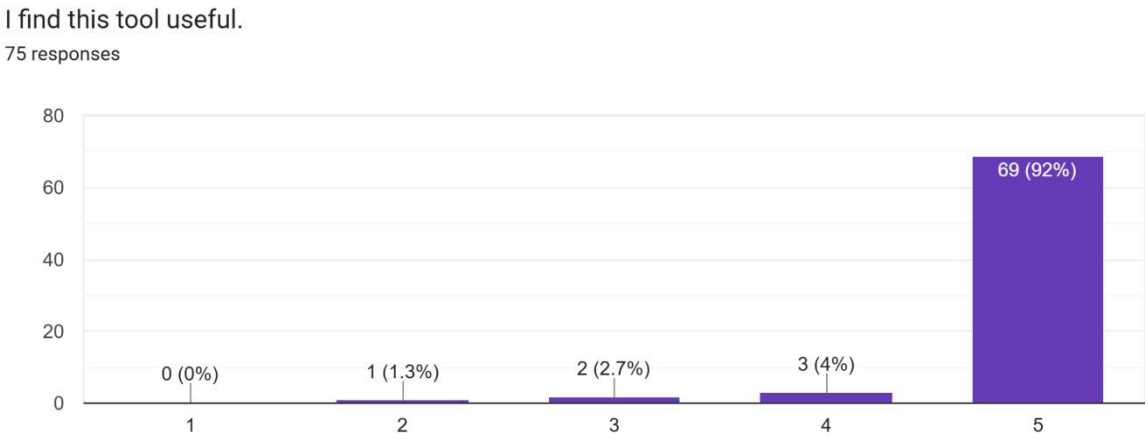


Figure 5.14 - Responses to "I find this tool useful".

- **I would use this tool frequently when navigating the web.**
  Finally, participants were asked whether they would use the tool frequently while navigating the web. 90.7% of respondents indicated they would, reflecting a high likelihood of adoption and demonstrating the potential for this tool to become a regular component of users' online privacy practices. Minor disagreement or neutrality was reported by a very small fraction of users (2.7% selected 2 and 1.3% selected 3).

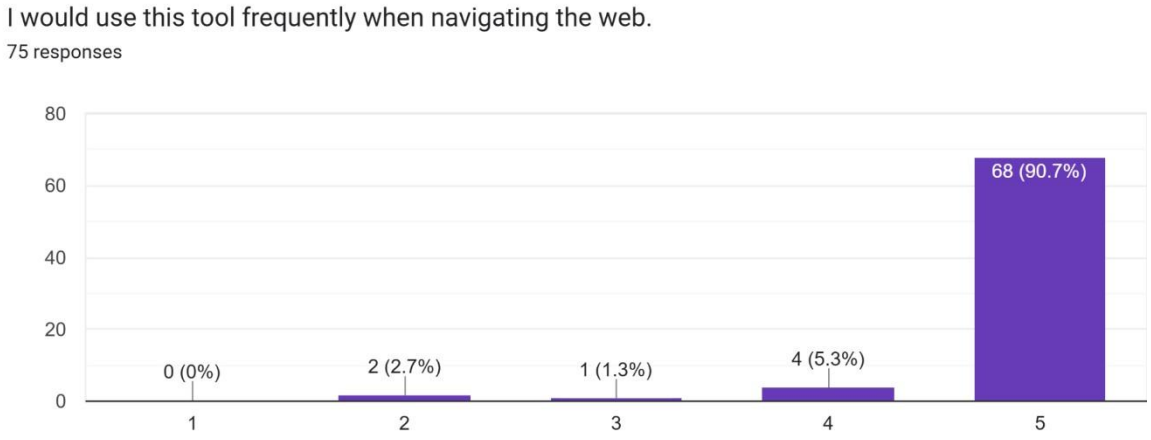


Figure 5.15 - Responses to "I would use this tool frequently".

Overall, the results of this section highlight that the Privacy Nutrition Label is perceived as highly effective in increasing awareness and trust, useful for everyday web interactions, and likely to be adopted by users. These findings reinforce the design principles applied in Chapters 4, demonstrating that a real-time, user-centred privacy label interface can positively influence users' understanding and perception of data processing practices.

5.4 Conclusions

The results of the user evaluation indicate that the Privacy Nutrition Label extension is perceived to effectively enhance users' awareness, trust, and understanding of personal data processing on web platforms. Across all measured dimensions, including awareness,

perceived usefulness, trust, and willingness to use the tool frequently, most participants rated the tool positively, with more than 90% agreement for key items.

Participants emphasized the intuitive design and real-time feedback as particularly valuable, confirming that a user-centered, interactive privacy label can successfully bridge the gap between technical data processing and user comprehension. These findings demonstrate alignment with GDPR principles and user expectations for transparency. The evidence from the evaluation also supports the hypothesis that the tool can positively influence users' decisions regarding their personal data, highlighting the practical effectiveness of a structured, comprehensible, and interactive privacy monitoring interface. The full set of questionnaire results is provided in Appendix B.

Chapter 6

6. Conclusions and Summary

6.1 Overview of the Thesis

This thesis investigated the challenges associated with transparency, user awareness, and trust in the collection and processing of personal data by web platforms. To address these challenges, a Privacy Nutrition Label browser extension was developed, providing users with real-time, structured, and actionable feedback on how their personal data is handled during web interactions.

The study began with a comprehensive literature review (Chapter 2), which highlighted that existing privacy nutrition labels, although structured, are largely static and rely on developer-declared data, limiting their practical utility. Real-time monitoring tools, by contrast, can detect live data flows but often present information in highly technical formats, which are difficult for non-expert users to interpret. These findings revealed a clear need for a system that combines real-time data detection with a user-friendly, structured interface.

Chapter 3 presented the methodology and results of a user survey designed to examine participants' perceptions of privacy, awareness of personal data processing, and trust in online platforms. The survey revealed significant knowledge gaps, strong concerns regarding invisible data processing, and high demand for real-time transparency tools. These insights directly informed the functional requirements, interface design, and logic architecture of the Privacy Nutrition Label.

Chapter 4 detailed the design and implementation of the browser extension. The tool leverages high-level indicators Data Entered, Used Here, and Sent Outside and employs card-based visualizations to display personal data categories, detected actions, inferred purposes, and technical evidence. A summary panel and activity timeline provide both an overview and chronological context of user interactions. The architecture is modular, extensible, and aligned with GDPR principles, ensuring that both novice and advanced users can access meaningful information without compromising usability.
Chapter 5 presented the evaluation of the tool using a structured survey of 75 participants, including UEQ items and Awareness & Protection Likert-scale statements. Participants consistently rated the tool positively across all dimensions. UEQ results showed high scores for Perspicuity, Efficiency, and Dependability, indicating strong usability and reliability, while hedonic aspects such as Stimulation and Attractiveness reflected engagement and satisfaction. Awareness & Protection results demonstrated that users felt significantly more informed, trusted the tool, and intended to use it frequently. These findings validate the design choices and confirm the tool's effectiveness in enhancing comprehension, trust, and perceived control over personal data.

6.1 Final Remarks

The findings of this thesis confirm that a real-time, structured privacy label can substantially improve users' understanding of personal data handling. The combination of literature-based design, empirical survey insights, and rigorous software engineering ensures that the Privacy Nutrition Label is both practically effective and legally informed. By translating complex technical data into a comprehensible, actionable interface, the tool empowers users to make informed decisions and fosters greater trust in online platforms.

6.2 Future Work

While the current implementation demonstrates strong usability and user acceptance, several avenues exist for further development:

1. Expanded Data Coverage: Extend the tool to additional personal data types (e.g., biometric data) and platforms, including mobile web and native applications.

2. Adaptive Feedback: Introduce personalized insights based on user preferences, prior interactions, and individual privacy concerns.
3. Accessibility and Multilingual Support: Enhance the tool for users with disabilities and provide multilingual support for broader reach.

These improvements aim to increase the scalability, effectiveness, and adoption of the Privacy Nutrition Label, further empowering users to understand, monitor, and control their personal data across diverse digital platforms.

Bibliography

[1] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science. Available: https://www.science.org/doi/full/10.1126/science.aaa1465

[2] Meier, Y., Schäwel, J., & Krämer, N. C. (2020). *The shorter the better? Effects of privacy policy length on online privacy decision-making. Media and Communication*, 8(2), 291–301.
Available: https://www.ssoar.info/ssoar/handle/document/69401

[3] *Farke, F. M., Balash, D. G., Golla, M., Dürmuth, M., & Aviv, A. J. (2020). Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity.*
Available: https://www.usenix.org/conference/usenixsecurity21/presentation/farke

[4] Kelley, P. G., Cranor, L. F., & others. *A "Nutrition Label" for Privacy:* Enhancing Usability and Transparency of Privacy Policies (SOUPS).
Available: https://dl.acm.org/doi/abs/10.1145/1572532.1572538

[5] *Vanezi Evangelia, Anna Vasileiou, and George Angelos Papadopoulos. "Making Data Collection Transparent and Usable: Annotating Web Forms with Processing Purposes." Conference on e-Business, e-Services and e-Society. Cham: Springer Nature Switzerland, 2025.*
Available: https://link.springer.com/chapter/10.1007/978-3-032-06164-5_21

[6] *Theophilou Ioanna, Evangelia Vanezi, and George A. Papadopoulos. "'Is this OhKéy?" A Study on Users Perception on GDPR-based Web Platforms Privacy Policies." Proceedings of the 2025 International Conference on Information Technology for Social Good. 2025. Available:*
https://dl.acm.org/doi/full/10.1145/3748699.3749773

[7] User Experience Evaluation (UEQ). Available: https://www.ueq-online.org/

[8] *S. Koch, M. Wessels, B. Altpeter, M. Olvermann, and M. Johns, Keeping Privacy Labels Honest, Proceedings on Privacy Enhancing Technologies, 2022. Available: https://petsymposium.org/popets/2022/popets-2022-0119.php.*

[9] *Beck, K., Beedle, M., van Bennekum, A., et al. (2001). Manifesto for Agile Software Development. Agile Alliance. Available: https://ai-learn.it/wpcontent/uploads/2019/03/03_ManifestoofAgileSoftwareDevelopment-1.pdf*

[10] *Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information Systems Research, 15(4), 336–355. Available: https://pubsonline.informs.org/doi/abs/10.1287/isre.1040.0032*

[11] *Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy concerns: Linking individual perceptions with institutional privacy assurances. Journal of Business Research, 35(1), 23–31. Available: https://aisel.aisnet.org/jais/vol12/iss12/1/*

[12] *GDPR Hub. Article 4 – Definitions. Available: https://gdprhub.eu/Article_4_GDPR#Decisions*

[13] *Herder, E., & van Maaren, O. (2020). Privacy dashboards: The impact of the type of personal data and user control on trust and perceived risk. In Adjunct Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20 Adjunct) (pp. 1–6). ACM. Available: https://doi.org/10.1145/3386392.3399557*

[14] *Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin & Karl Aberer (2018). Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/harkous*

[15] *Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quickdirty-usability-scale-john-brooke*

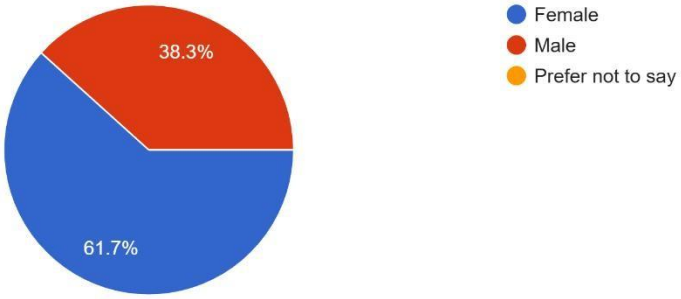
[16] *Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 13(3), 319-340. Available: https://misq.umn.edu/misq/article-abstract/13/3/319/191/PerceivedUsefulness-Perceived-Ease-of-Use-and*

Appendix A

This appendix presents the complete set of results from the user questionnaire discussed in Chapter 3, which was designed to evaluate participants' perceptions of personal data processing, privacy awareness, trust in web platforms, and preferences regarding real-time transparency tools.

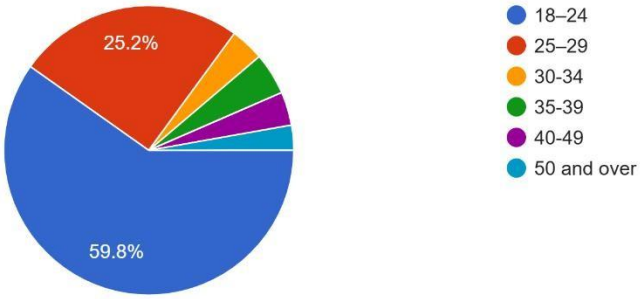
By completing this survey, you confirm that: You understand the survey's purpose. Participation is voluntary and you can withdraw at any time.Your res...roceeding, you consent to participate voluntarily.
107 responses

What is your gender?
107 responses



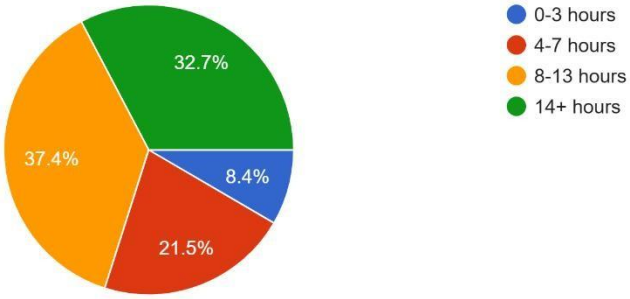
What is your age group?

107 responses



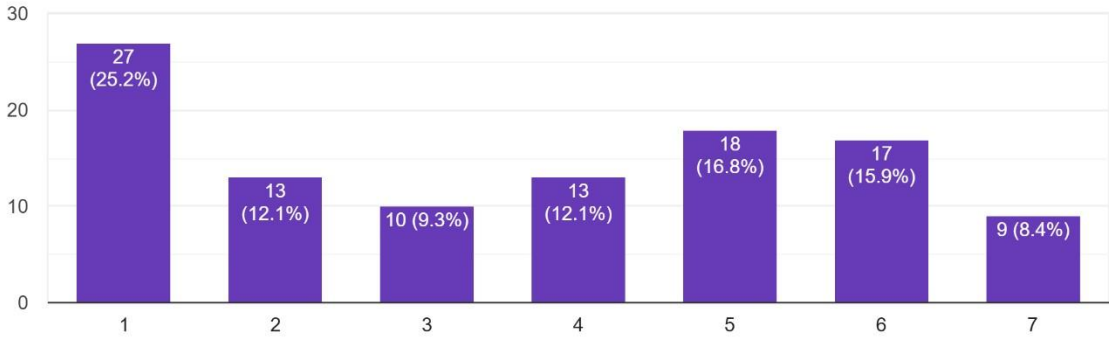
How often do you use the web on a weekly basis?

107 responses



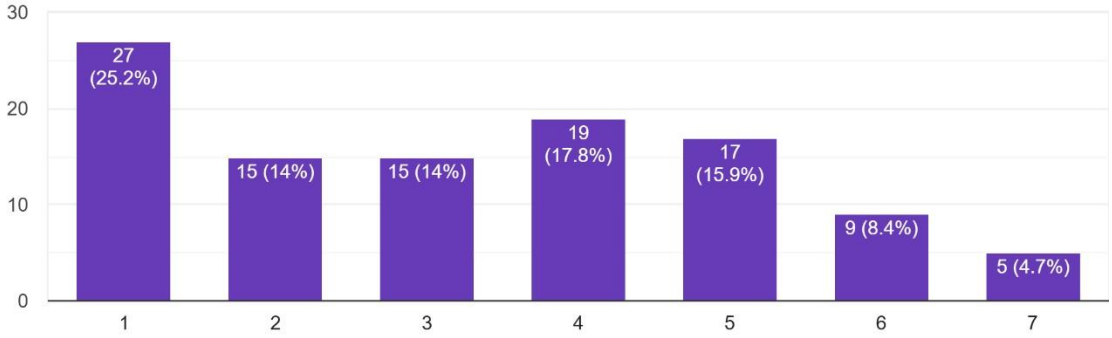
I understand how websites process my personal data after I register or log in.

107 responses



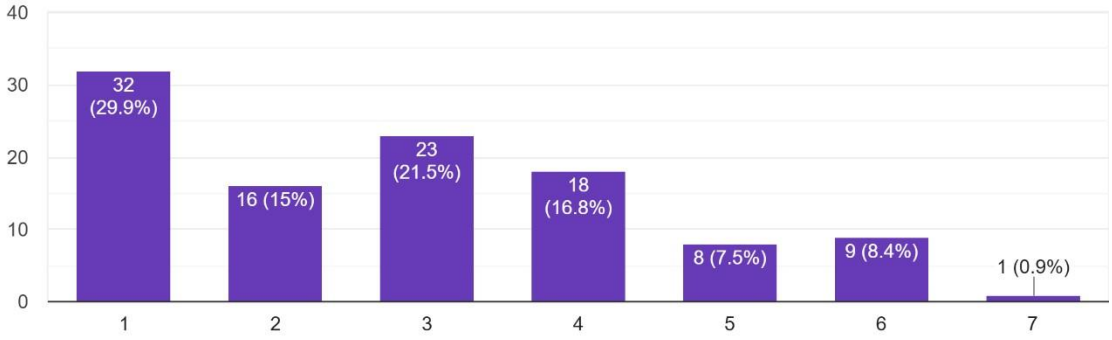
I am aware of what happens to my personal data while I am using a website.

107 responses



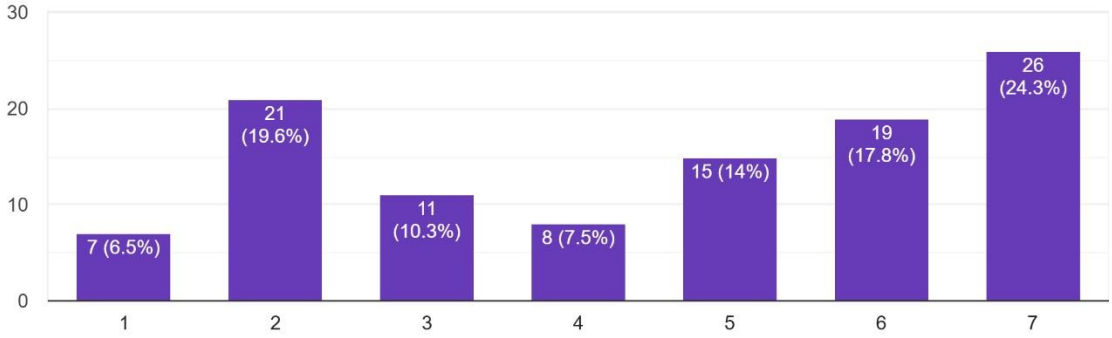
I feel informed about how my personal data is processed over time on web platforms.

107 responses



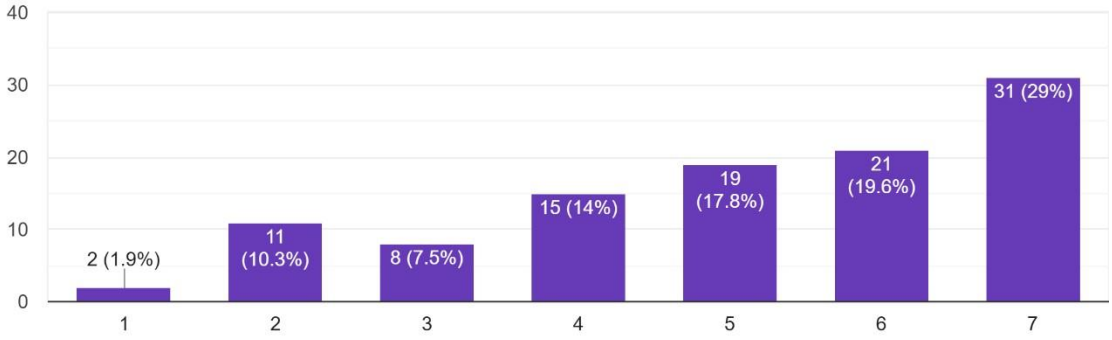
I usually have no idea how my personal data is processed online.

107 responses



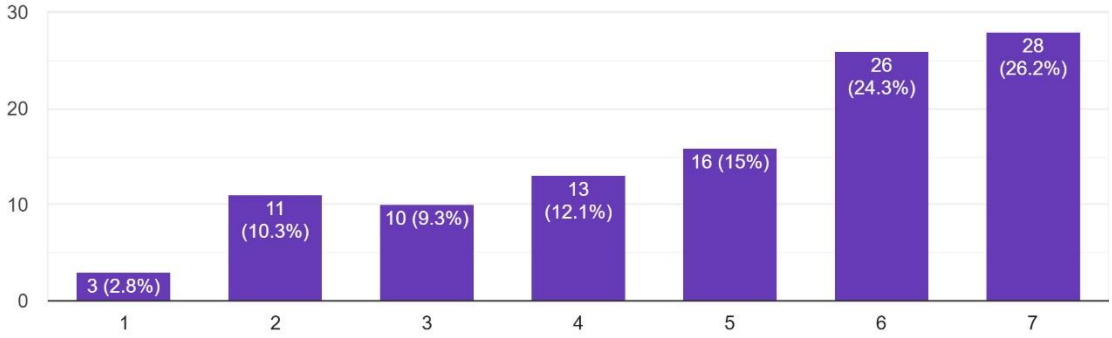
I am concerned about how my personal data is processed without my knowledge.

107 responses



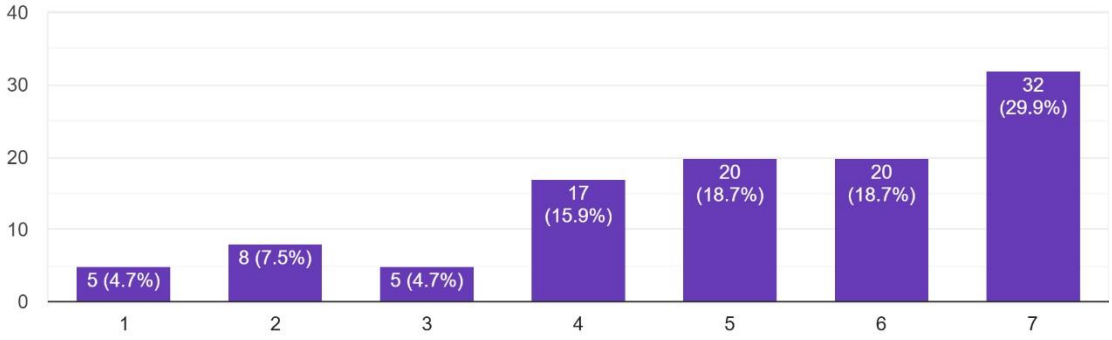
I worry that my data is used in ways I did not expect when I signed up.

107 responses



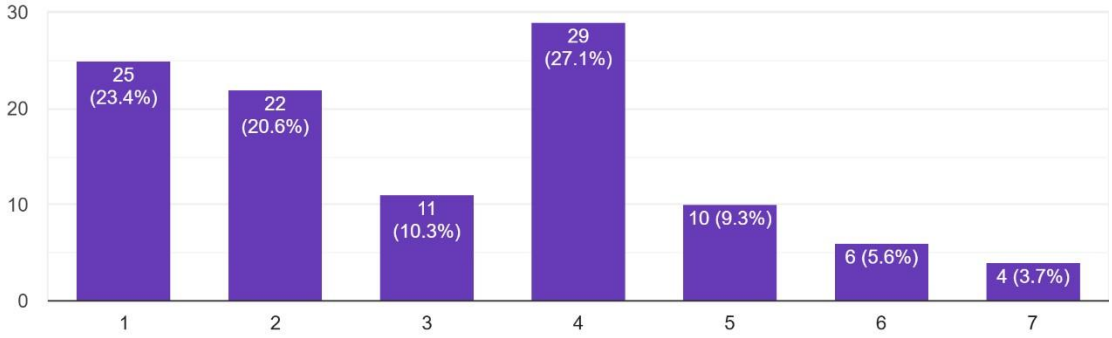
Knowing that my data is continuously analyzed makes me feel uneasy.

107 responses



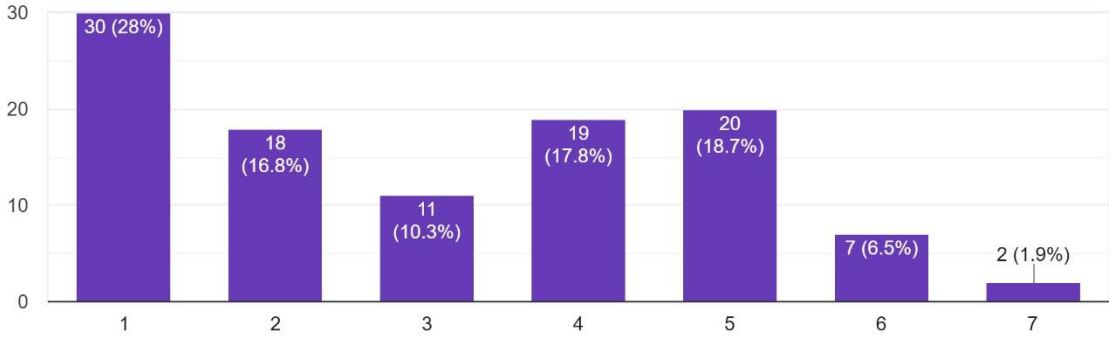
I trust websites to process my personal data responsibly.

107 responses



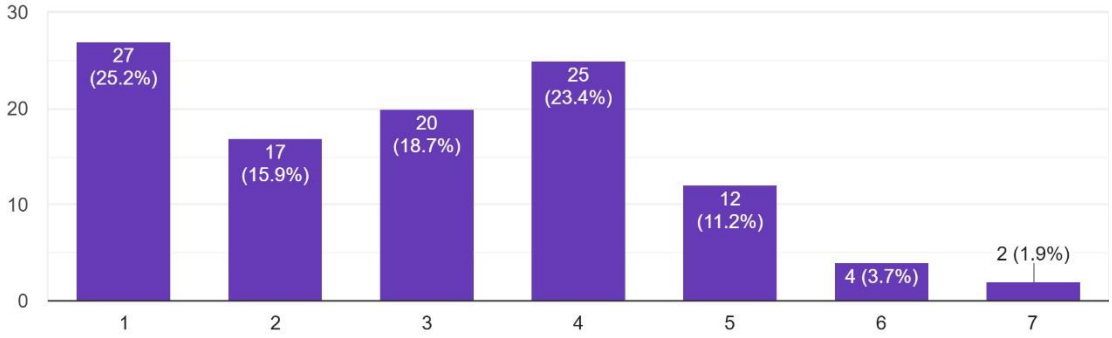
I believe most web platforms respect users' privacy.

107 responses



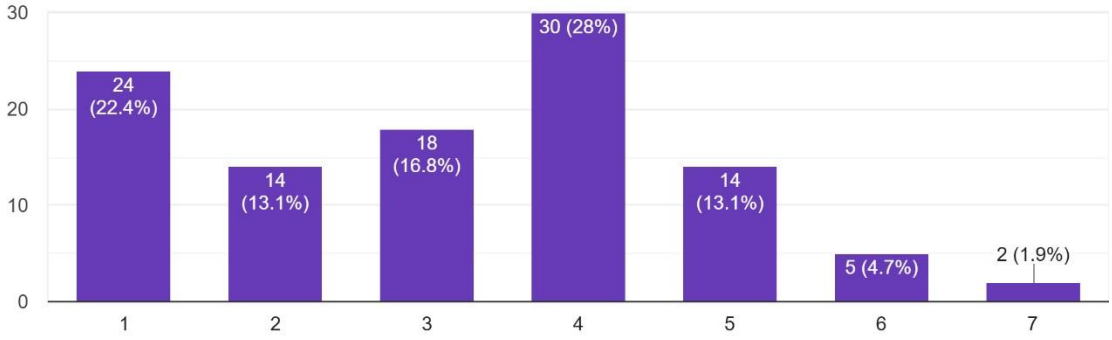
Overall, I trust the way my personal data is handled online.

107 responses



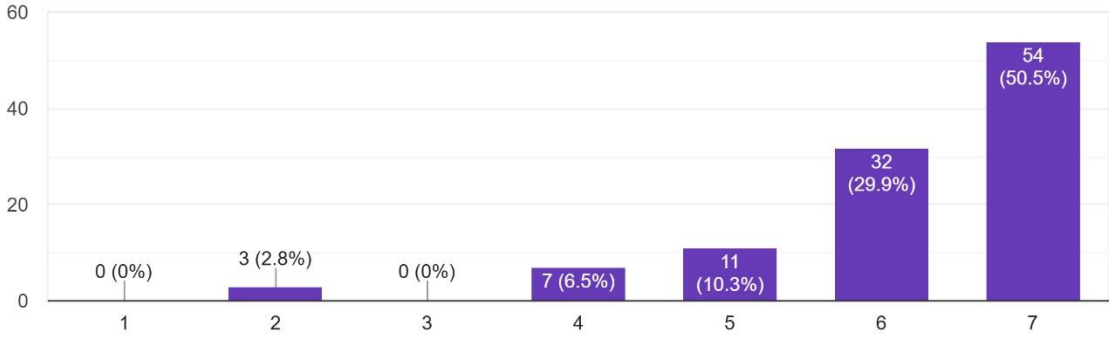
Web platforms are transparent about how they process personal data.

107 responses



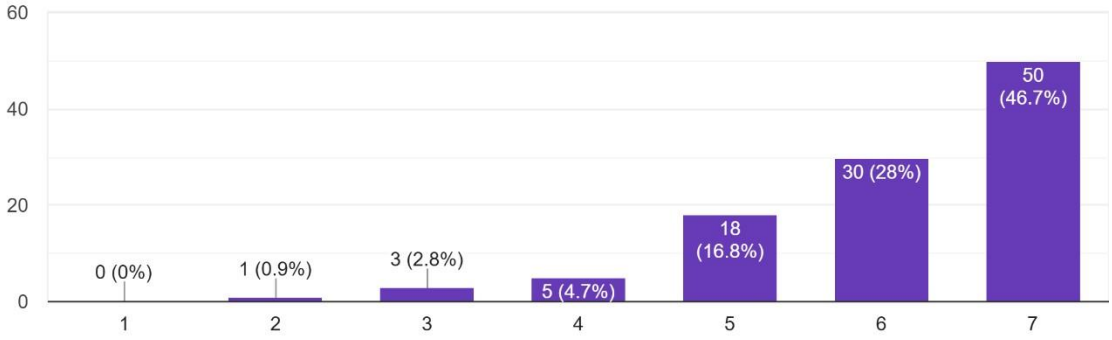
I would like to know in real time how my personal data is being processed.

107 responses



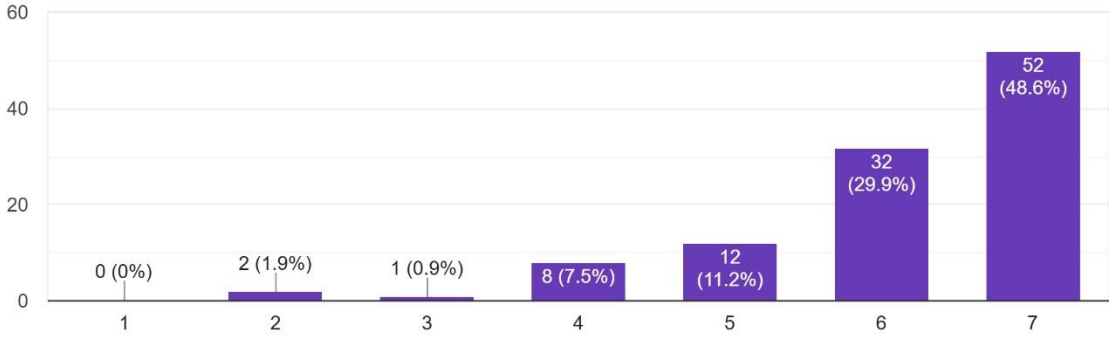
A real-time assistant explaining data usage would make me feel more in control.

107 responses



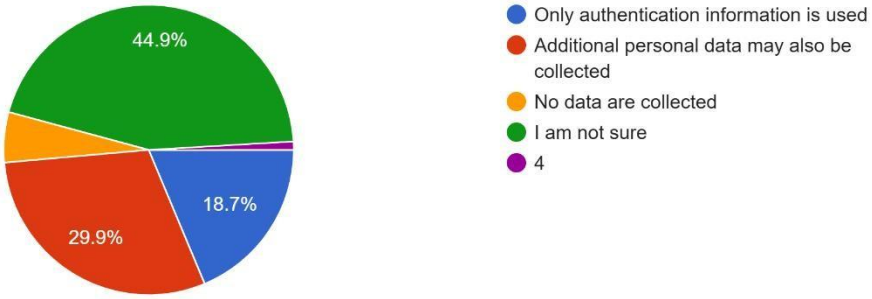
Real-time visibility of data processing would increase my trust in websites.

107 responses



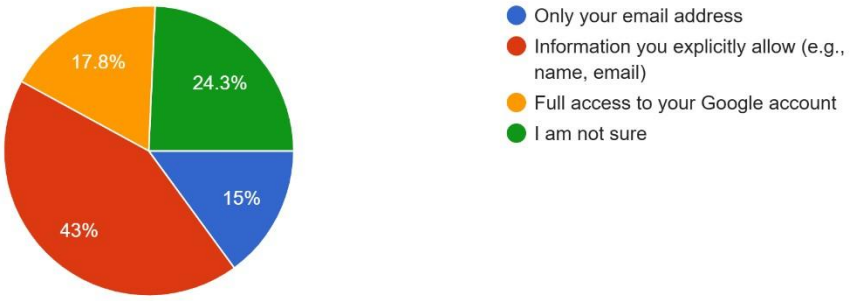
If you scan a QR code to log in to a website, what is most likely to happen?

107 responses



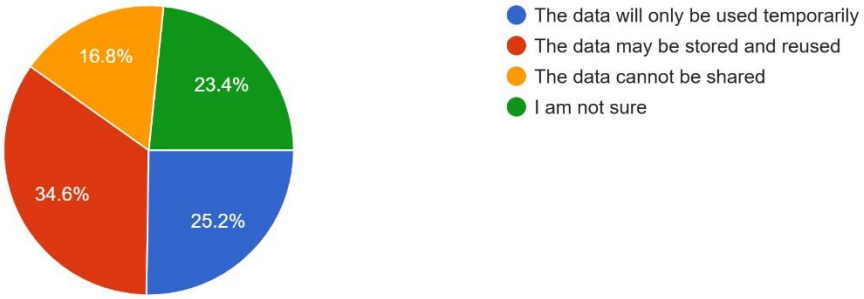
When you use “Sign in with Google”, which information can the website access?

107 responses



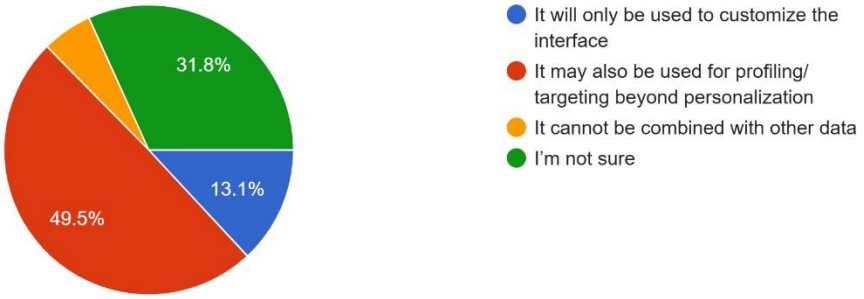
A website asks for access to your location to “improve services”. What is the safest assumption?

107 responses



A site asks for your date of birth “to personalize your experience.” What is the safest assumption?

107 responses



After you delete your account on a platform, what is the safest assumption?

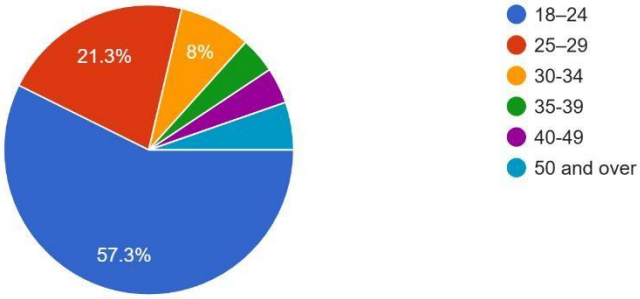
107 responses

Appendix B

This appendix includes the full questionnaire results used in Chapter 05 to evaluate the Privacy Nutrition Labels Tool's usability, visual appeal, and effectiveness. The survey was distributed to users as part of the final evaluation phase, aiming to collect feedback on their overall experience with the tool.

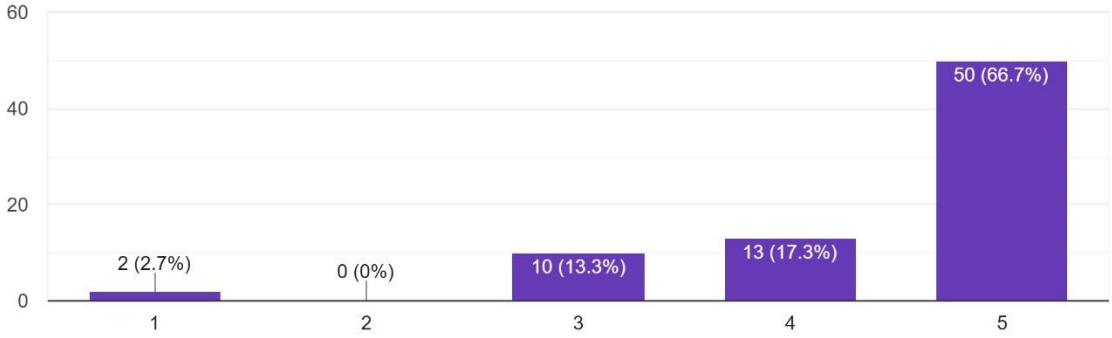
By completing this survey, you confirm that: You understand the survey's purpose. Participation is voluntary and you can withdraw at any time.Your res...roceeding, you consent to participate voluntarily.
75 responses

What is your age group?
75 responses



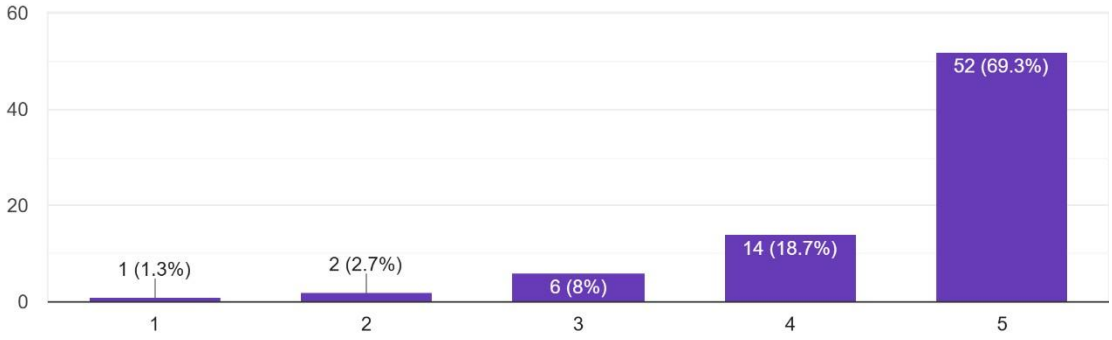
How would you describe your level of familiarity with technology and websites?

75 responses



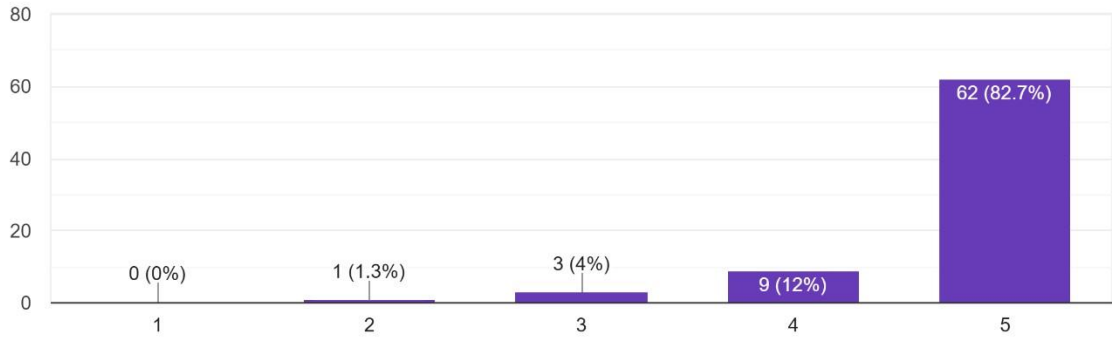
How often do you use websites that require registration with personal data (e.g., online shopping and social media)?

75 responses



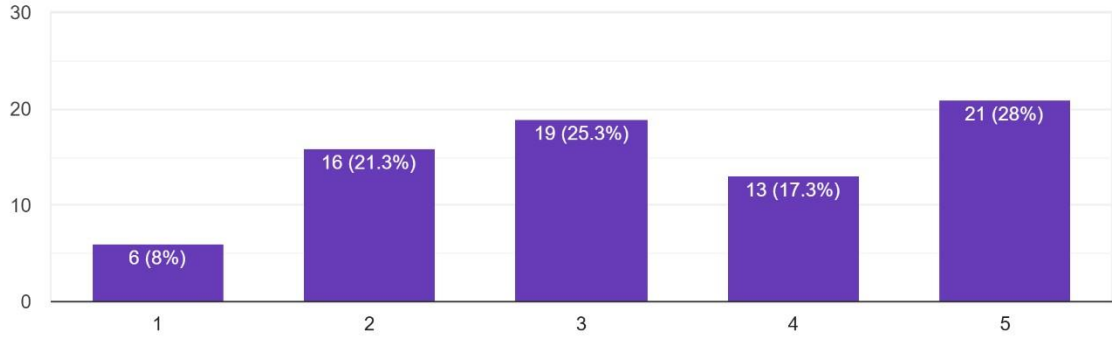
How important is online privacy to you?

75 responses

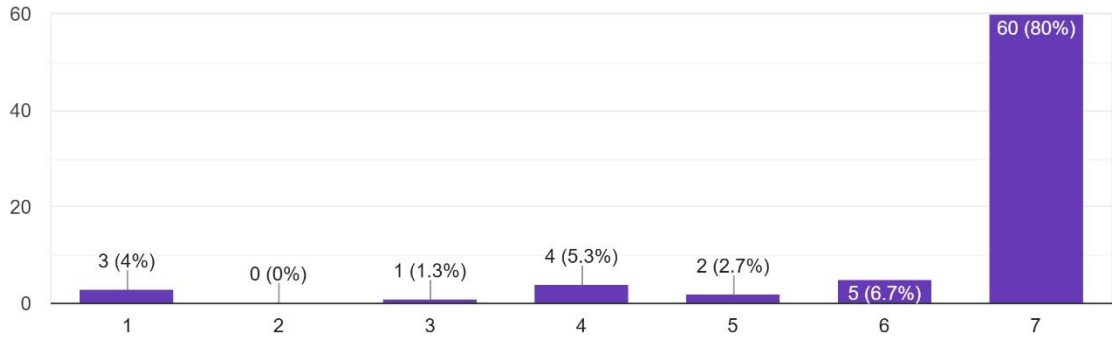


How would you describe your level of familiarity with privacy?

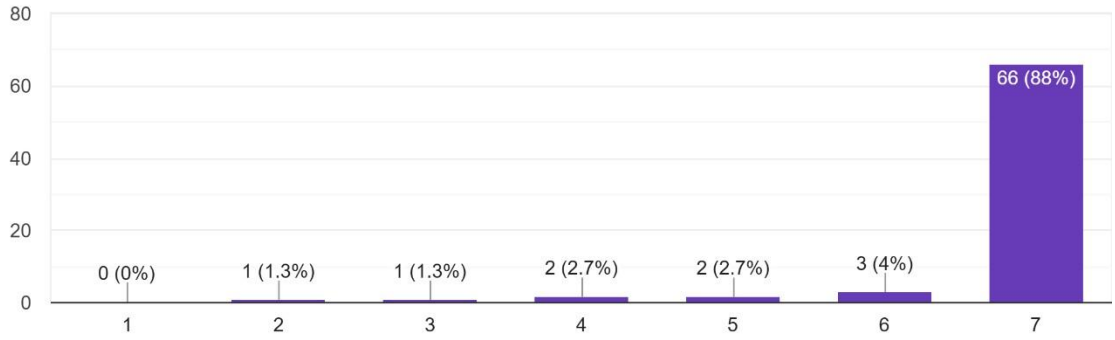
75 responses



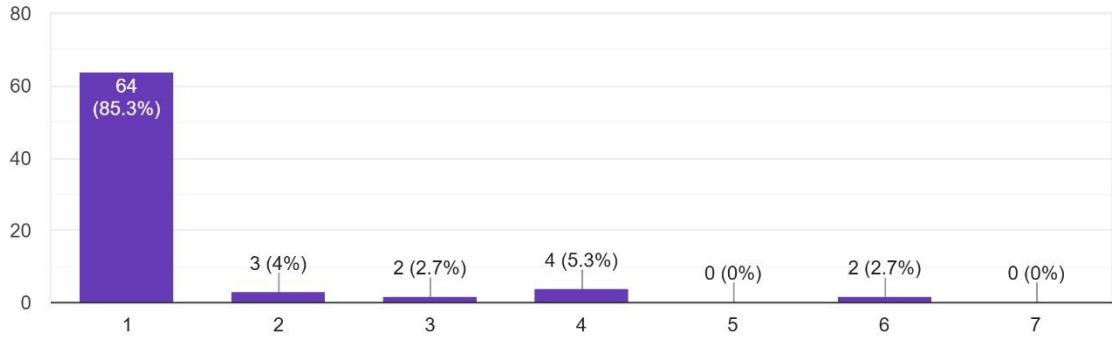
75 responses



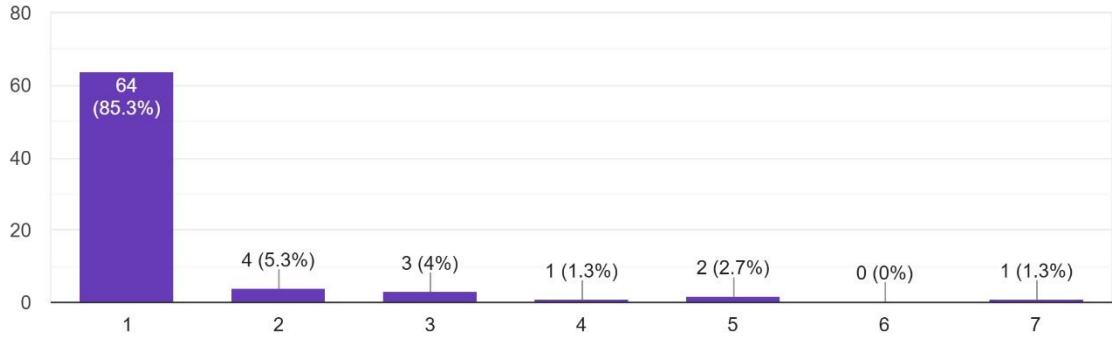
75 responses



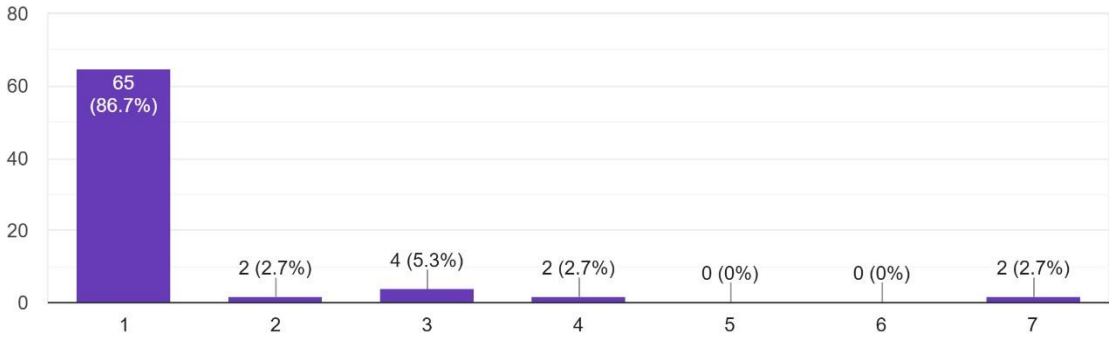
75 responses



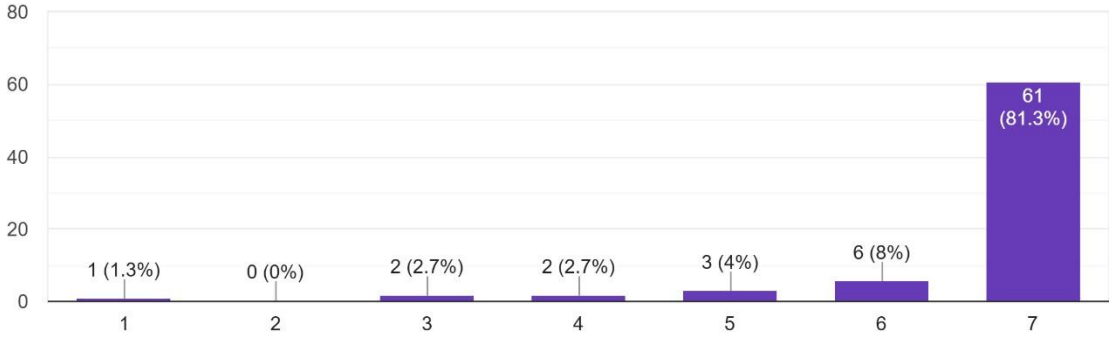
75 responses



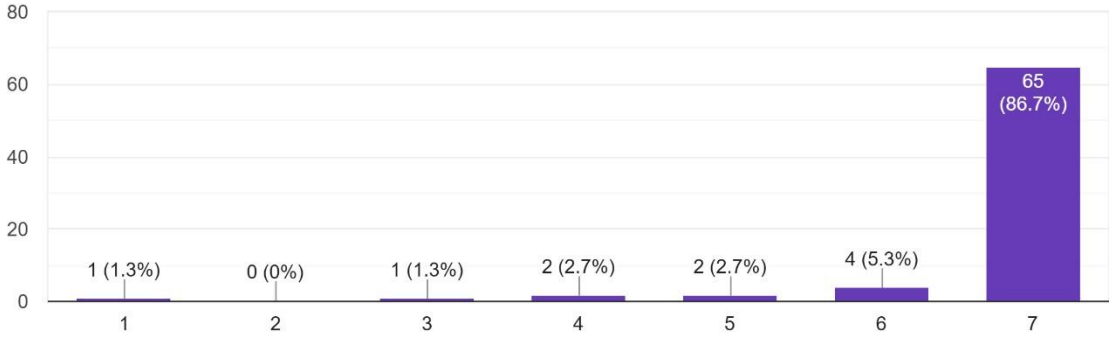
75 responses



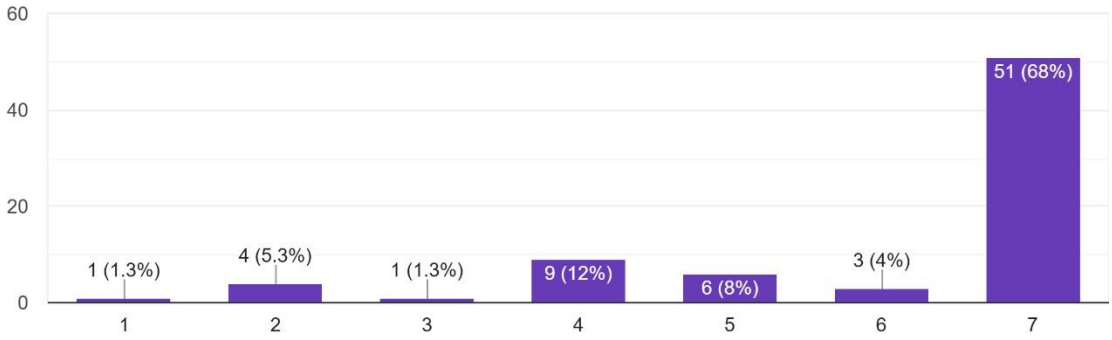
75 responses



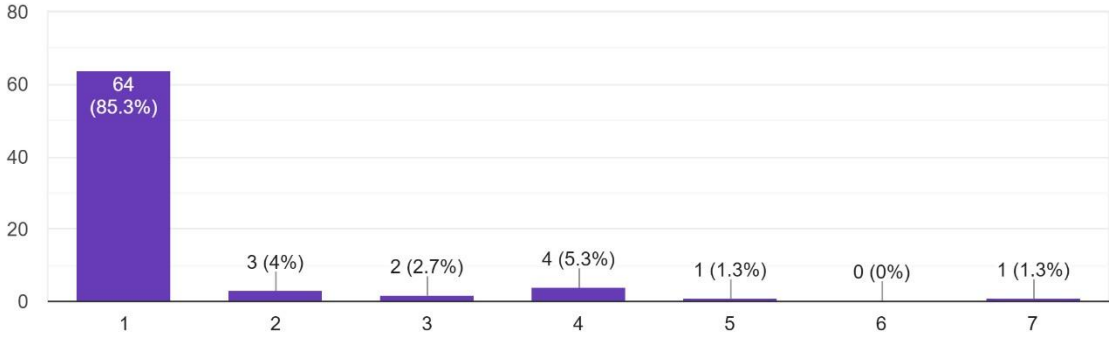
75 responses



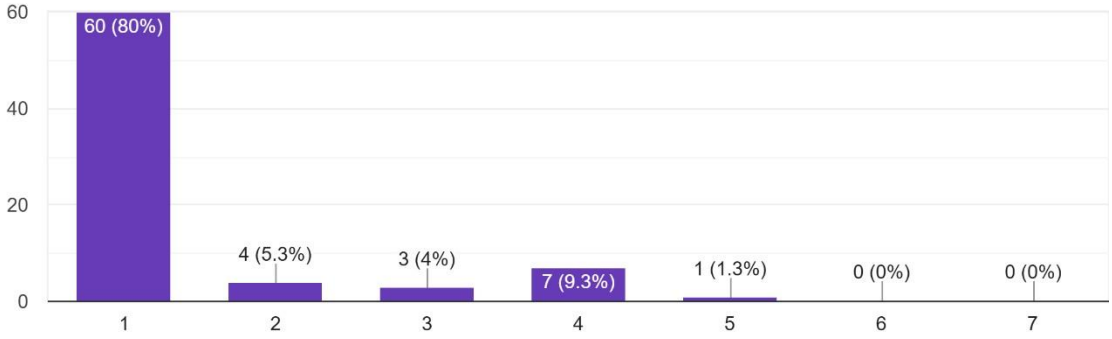
75 responses



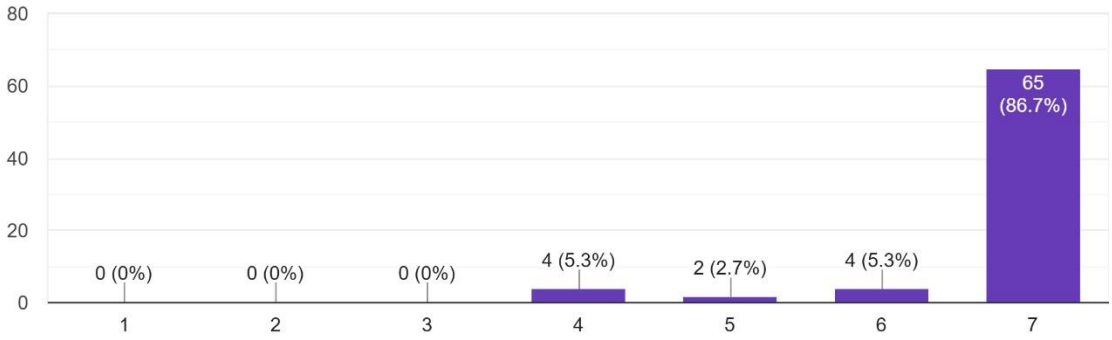
75 responses



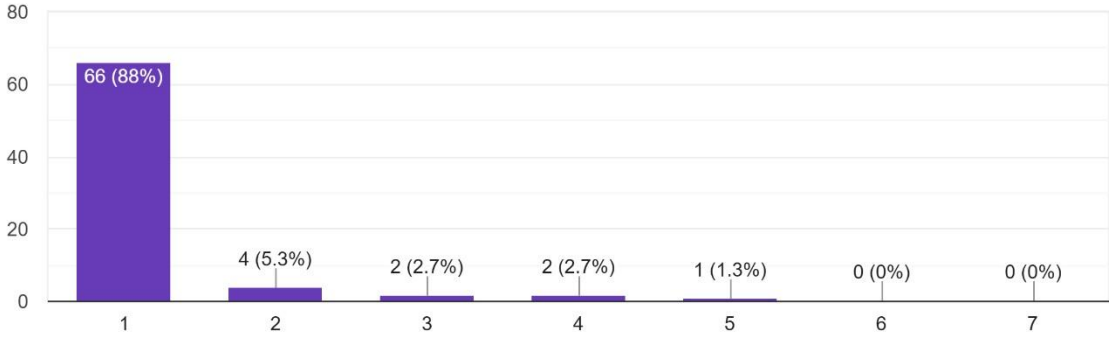
75 responses



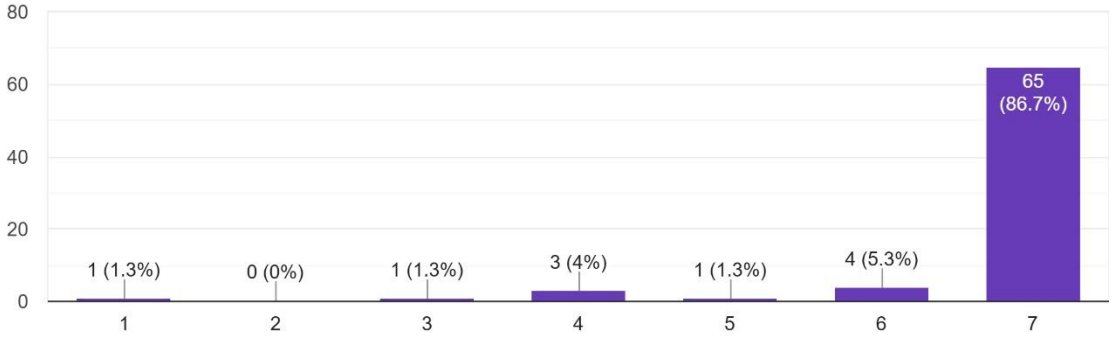
75 responses



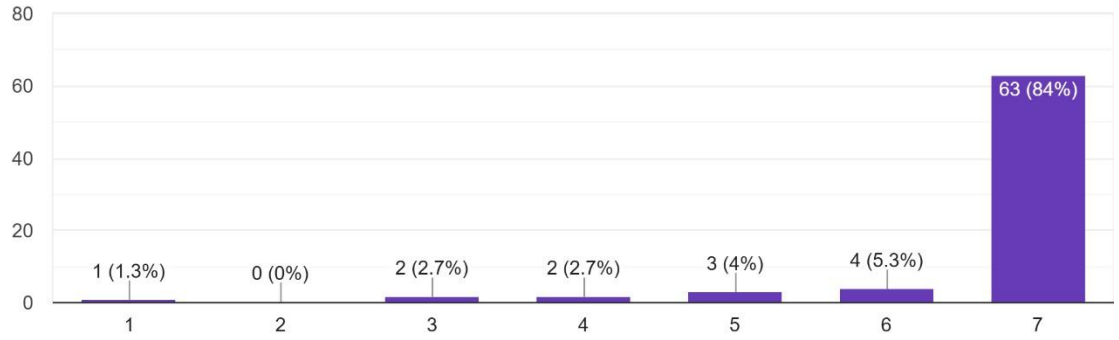
75 responses



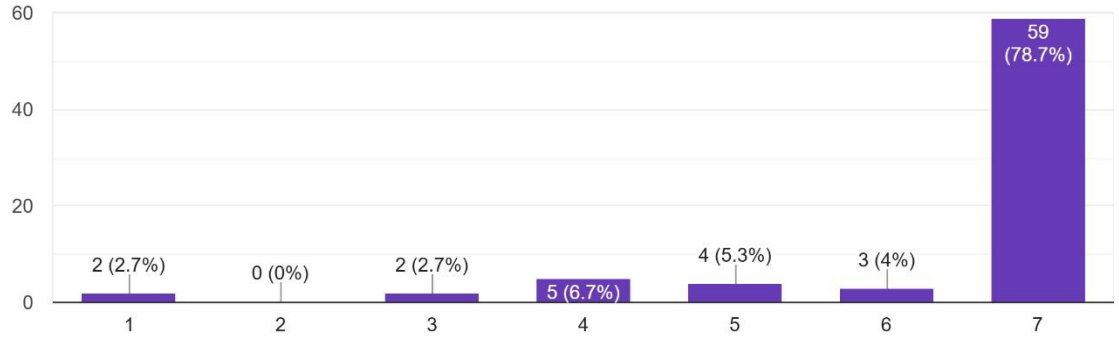
75 responses



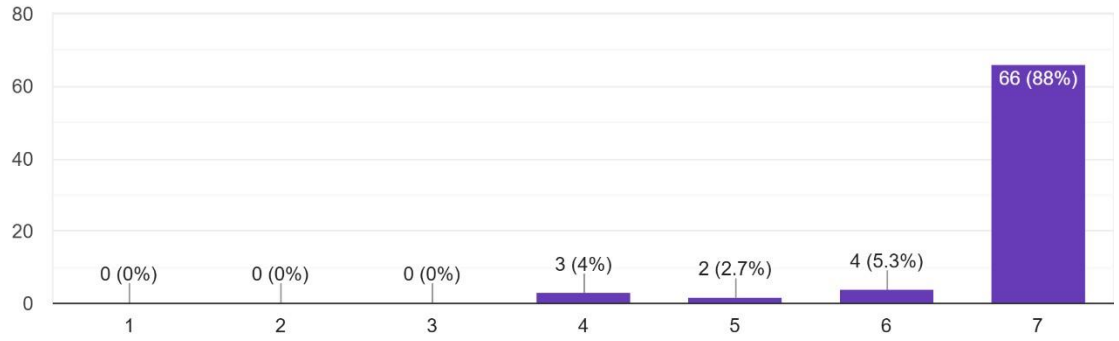
75 responses



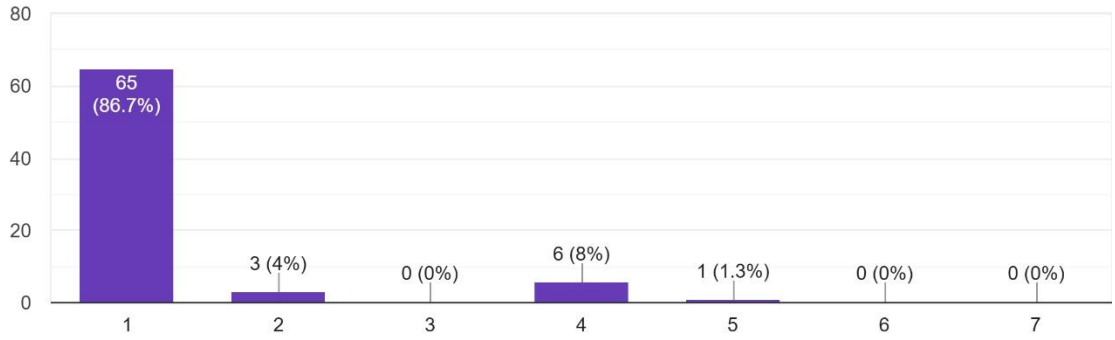
75 responses



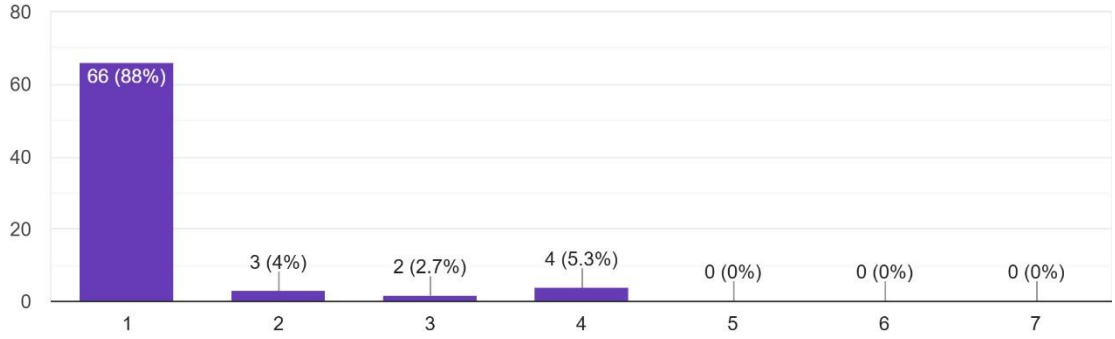
75 responses



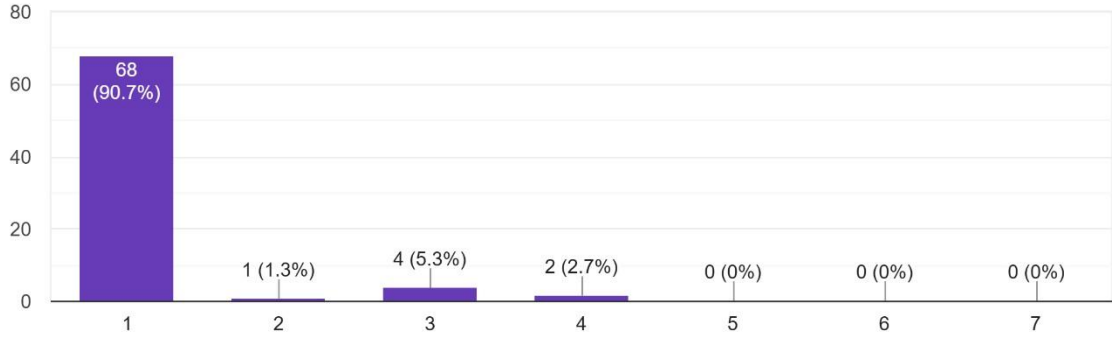
75 responses



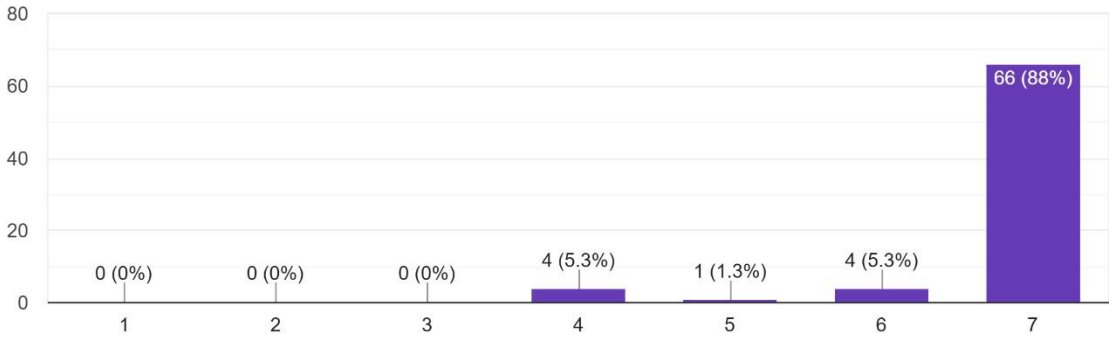
75 responses



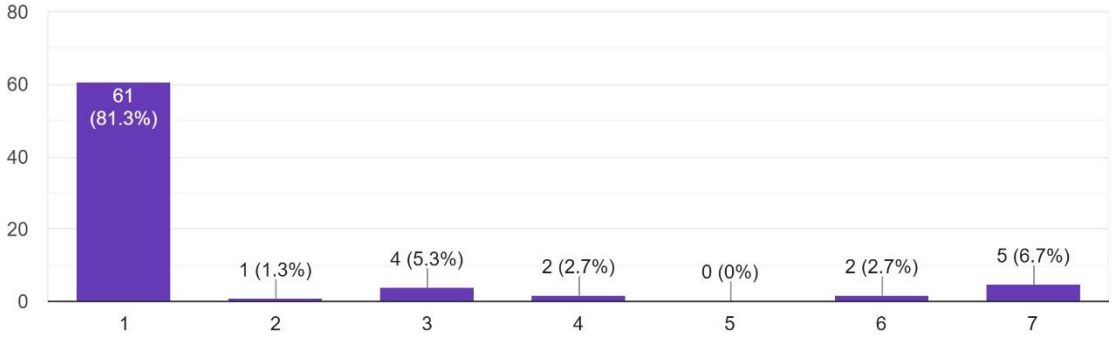
75 responses



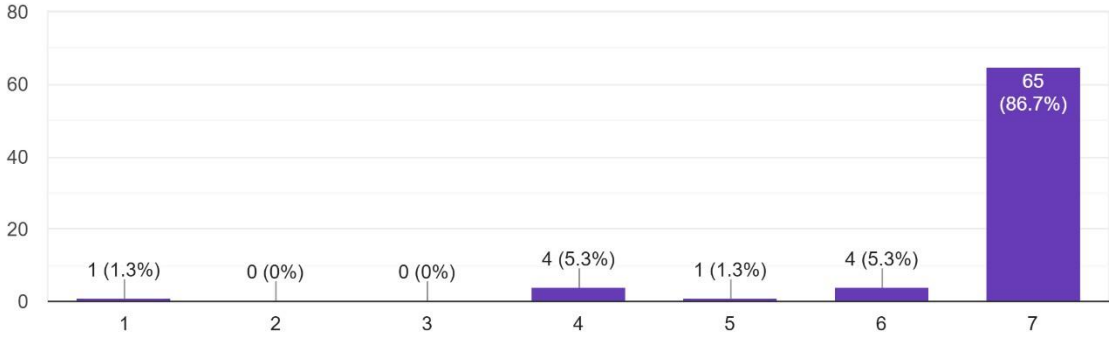
75 responses



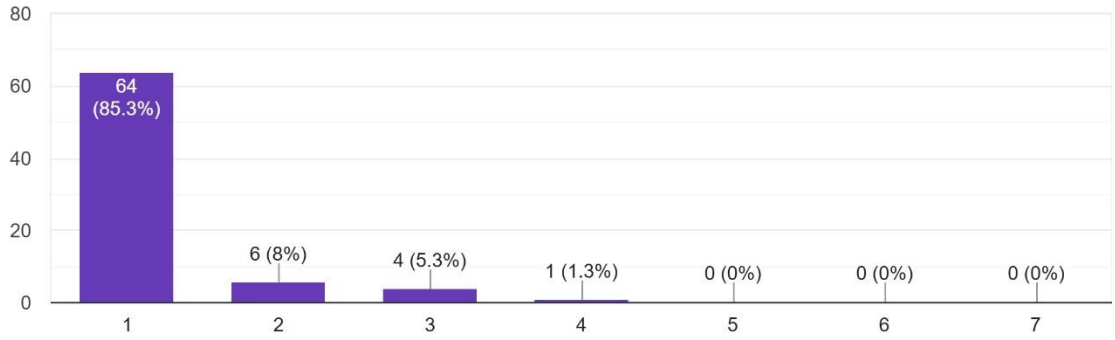
75 responses



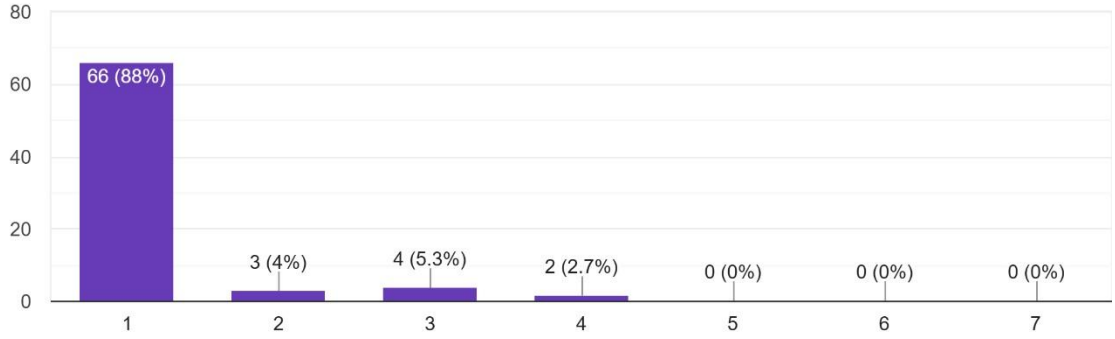
75 responses



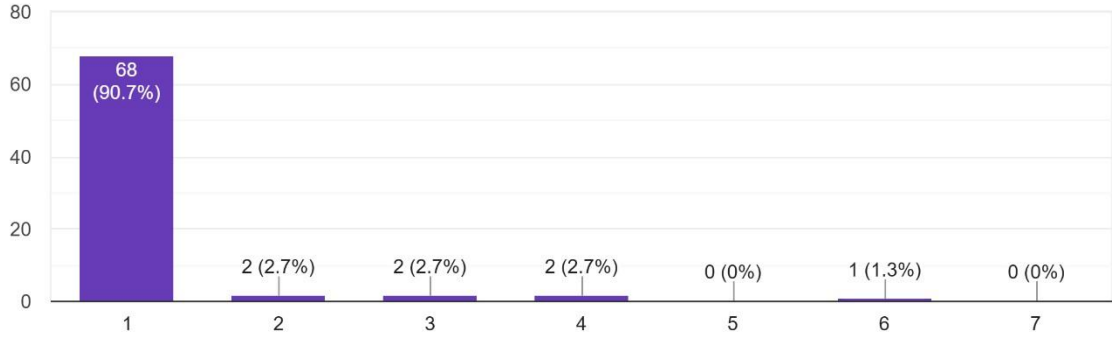
75 responses



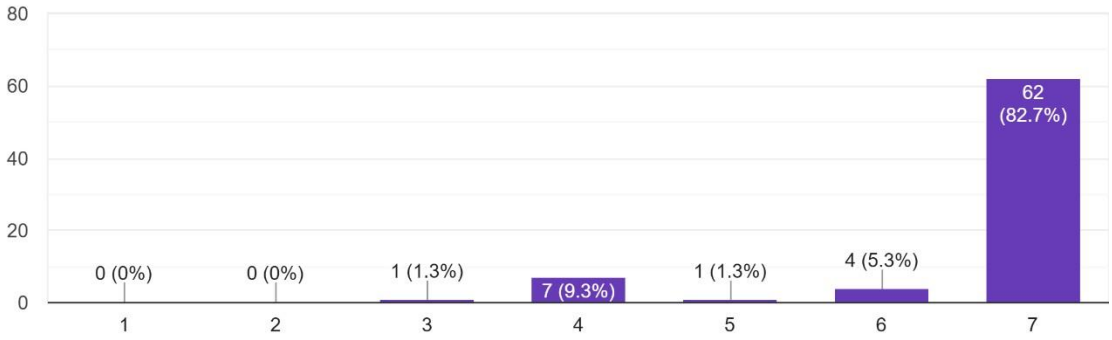
75 responses



75 responses

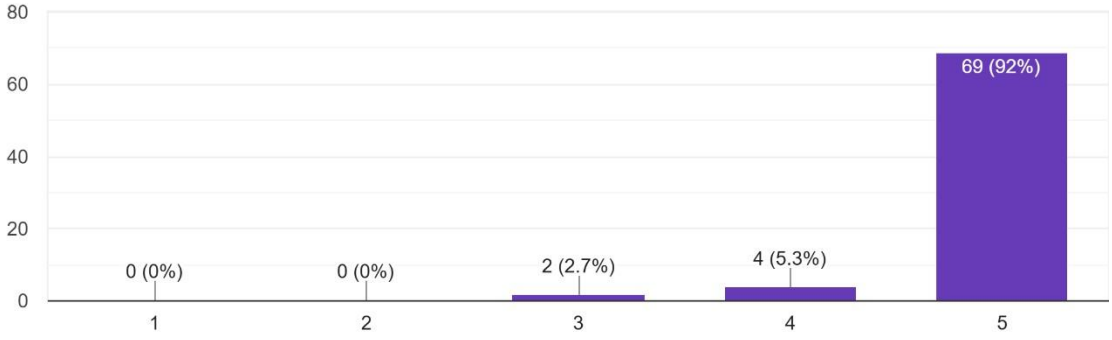


75 responses



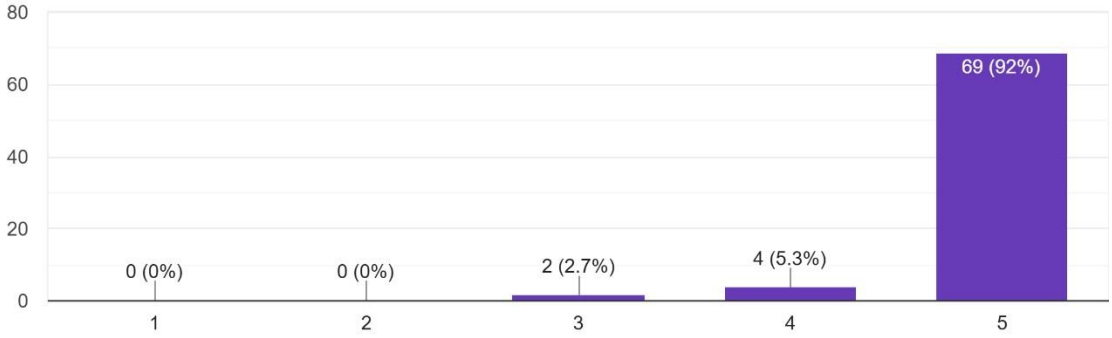
Using this tool when using web platforms would improve my awareness on my personal data processing.

75 responses



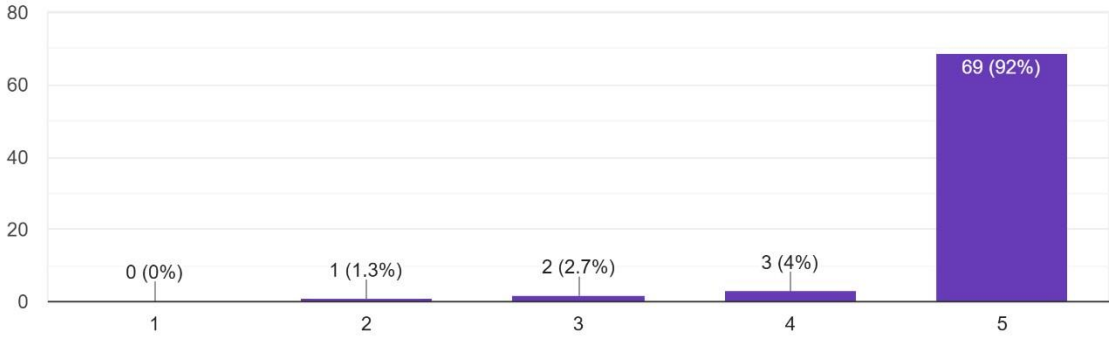
Using this tool when using web platforms would improve my trust on my personal data processing.

75 responses



I find this tool useful.

75 responses



I would use this tool frequently when navigating the web.

75 responses