ABSTRACT

This research was conducted in the context of the University of Cyprus, Computer Science Master's Dissertation. The research's purpose was to define and address the gap between Privacy Policies comprehension by end-users and their 'Blind' Consent. In order to achieve this, research was conducted in terms of Privacy in Software Engineering and Usable Privacy. This dissertation initially provides a Systemization of Knowledge through comparative tables reviewing the literature and finding common ground between research papers. After that, the need to create a software tool arose in order to assess user's understanding in various types of online platforms through an easy-to-use visual form. This was done through the OhKéy platform, where existing web platforms' Privacy Policies were gathered, the GDPR part was extracted, anonymized and grouped in small 'clickable boxes' based on their GDPR context. Users could select each one to assess their text in terms of comprehensibility based on a set of pre-defined terms using a one-to five Likert scale rating. 1098 annotations were gathered in total. This thesis , towards achieving the initial objectives, analyzed in depth the negative annotations. The result showed crucial need for Enhancing privacy policies with elements that will address the issues recognized. Therefore the AlRíght Policy Generation tool links Policy Makers and end-users by introducing a panel for Policy Makers and Web Developers to insert the Privacy Policy text and along with other functionalities like summaries, important question answering and media, to create an Enhanced Privacy Policy form. The outcome could be downloaded by the Policy Makers and Web developers in order to be used in their actual platforms. The tool's purpose was to address 'Ambiguity', 'Vagueness', 'Language Difficulty', 'Verbosity' 'Suspiciousness' and 'Complexity' by providing users with many options through the policy reading procedure. Very positive results of the AlRíght tool were gathered in two different surveys regarding User Experience (with Policy makers and Web developers) and End-users evaluation.

ADDRESSING 'BLIND' CONSENT : A STUDY ON USERS PERCEPTION ON PRIVACY POLICIES AND A TOOL TO ENHANCE THEIR UNDERSTANDING OF THEM.

Ioanna Theophilou

A Thesis

Submitted in Partial Fulfilment of the

Requirements for the Degree of

Master of Science

at the

University of Cyprus

Recommended for Acceptance

by the Department of Computer Science

May 2024

APPROVAL PAGE

Master of Science Thesis

ADDRESSING 'BLIND' CONSENT : A STUDY ON USERS PERCEPTION ON PRIVACY POLICIES AND A TOOL TO ENHANCE THEIR UNDERSTANDING OF THEM.

Presented by

Ioanna Theophilou

Research Supervisor

Professor George A. Papadopoulos

Committee Member

Associate Professor Georgia M. Kapitsaki

Committee Member

Dr. Christos Mettouris

University of Cyprus

May, 2024

ACKNOWLEDGEMENTS

With the help of God, I have successfully completed this master's Dissertation. I wouldn't be able to do so if it wasn't for the strength that He gave me. I would firstly like to thank my supervisor, Professor George A. Papadopoulos for giving me this opportunity in working with this amazing topic. A very big thank you goes to my co-supervisor Evangelia Vanezi for her unconditional support, guidance and mentoring throughout this thesis. I would also like to thank my mother for her support and my father who I'm sure was praying for me from up above. I would finally want to thank my partner for his emotional support and all my friends.

CREDITS

Credits to my co-supervisor Evangelia Vanezi for her creative ideas of this thesis and her invaluable contribution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 : Introduction

Introduction

1.1 Motivation

The GDPR stands for General Data Protection Regulation, and it is applied since May 25, 2018. It defines rules regarding personal data collection, storage and processing [1]. Privacy policies are written documents that are made as an agreement between users and platforms on how their data are going to be used, processed etc. This is very important in Software Engineering since in most platforms, user registration and therefore user login, are crucial. So why is personal data privacy protection important in information systems? Most people use web applications, mobile applications and others that save personal data of users for personalization purposes [2]. User registration often requires user's email address, name, surname, and other information that is considered personal data. Most platforms do not evaluate Privacy in the very first steps of the Software Engineering Process which leads to crucial changes and modifications in the future. To be exact, the Software Engineering Process according to [3] starts with the 'Requirements Engineering' phase and continues with the 'Analysis'. Then follows the 'Design' , the 'Development' and lastly the 'Validation' and 'Deployment' phases.

Privacy by Design means "data protection through technology design" according to [4]. Although this term is almost 54 years old, there is no exact definition for it. Some say that Privacy by Design means including the notion of Privacy from the early stages of the Software Engineering Life Cycle. This is what is believed to be the best practice in order to emphasize its importance and create awareness around the topic.

Privacy policy documents though, are very long and not user-friendly because they contain a lot of text and information that may confuse the reader or even overwhelm them. According to

[5], only one out of four of the readers read the privacy policies and they spent about a minute reading it. This is where the notion of Usability comes in handy. According to [6] "usability of a product is a function of the particular user or class of users being studied, the task they perform, and environment in which they work".

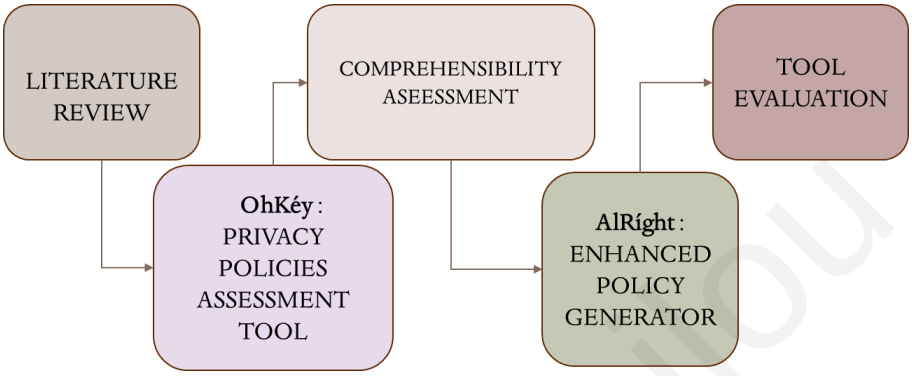
Combining Privacy and Usability is a new-coming field that involves observing different types of content visualizations, User Experience, Human Computer Interaction and others. It requires a combination of research in order to be classified correctly. In this dissertation, we examine comprehensibility of Privacy Policies. A survey is conducted in order to assess comprehension and a tool is then designed and developed as the proposed solution of this assessment.

1.2 The research problem

The research question is split intro three parts. The first question to be answered is : Do users understand the contents of Privacy Policies ? The second question is : Can we assist Policy Makers and Software Engineers in creating comprehensible Privacy Policies ? And lastly: Can we enhance users' comprehension ? The purpose was to find ways in addressing this blind consent the users do in making them aware of what the contents of Privacy Policies are. The way to do this will be described below. The overall purpose was to assess the comprehension of users and with the analysis to propose a way in answering this main question. In this thesis, we use the term 'Blind' consent to describe the situation where users do not read the privacy policies [5] and they immediately consent to them.

1.3 Methodology

Figure 1 Thesis Methodology



As described above, the purpose is to study users' comprehension on Privacy Policies. The first step in achieving this was to create an easy-to-use tool that includes many real policies and lets users assess the policy texts in terms of understanding. This is described in Chapter 3. After this, a thorough comprehension analysis was needed to analyze the users needs and general perceptions of Privacy Policies in order to propose the solution to the problem. A solution is proposed, through a tool that derives Enhanced Privacy Policies made to address comprehension problems extracted in step two.

1.4 Thesis Structure

In the second chapter of the thesis a literature review is presented. The literature review is split into two sections. The literature review regarding software engineering and privacy and the literature review regarding privacy and usability. Terminologies are also provided along with comparison tables discussing the critical keywords for each research paper. On the next Chapter, chapter 3, the OhKéy Privacy Policies assessment tool is presented. The purpose of this tool is the comprehensibility assessment of various types of privacy policies, coming from different platform types. The analysis of the results is presented in Chapter 4 that is the Comprehensibility Analysis. The results showed the crucial need for a solution. Therefore in chapter 5 this thesis proposes the AlRíght enhanced policies generator tool. Lastly, in Chapter 6 there is a discussion of the tool's evaluation in terms of usefulness and User Experience. After that the conclusions are presented in Chapter 7.

Chapter 2 : Background Study

Background Study

2.1 Introduction

In this section related work is presented. This review includes thorough examination of existing research. Analyzing the existing landscape allows for recognition of areas that require innovation, seeking out gaps for exploration and offering opportunities for original contributions.

The literature review is separated into two parts. The first part explores the related work on Privacy and Software engineering in general. The second part explores the notion of Privacy in terms of Usability.

2.2 The notion of Privacy and Privacy in ICT

ICT refers to information and communication technology. The concept of privacy in the context of ICT is based on the way data is transmitted, stored and processed in the system. When we refer to privacy terms or privacy regulations, we mean the strict rules for the transfer, transmission and processing of data. For example, the GDPR (General Data Protection Regulation) legislation. Privacy policies or privacy rules are rules that must be followed by systems to ensure that data is protected, and privacy rules are followed.

By privacy in general we mean ensuring how personal data is managed, regulated, collected and used. The concept of security relates to ensuring the confidentiality, integrity and availability of data. Although the two concepts are different, they are closely related, as privacy is an important part of data security. In addition, the concept of privacy includes other concepts,

such as non-connectivity, transparency and accessibility, which contribute to ensuring data security .

2.3 Related work on Privacy and Software Engineering

2.3.1 Relevant Terminology

This chapter presents relevant notions found in the bibliography and their definitions.

Table 1 Terminology found in literature about Privacy and Software Engineering

Notion	Explanation	Bibliography
Privacy by Design	It is an engineering and strategic approach committed to selectively and sustainably minimizing the privacy risks of IT systems through technical and governance controls.	[7]
Privacy Engineering	The field of engineering that provides ways through techniques, methodology and tools to ensure a system with good levels of privacy.	[8]
Privacy-preserving systems	Users use them to protect their personal data.	[9]
Software Development	This is a set of computer science tasks that focus on the creation, design and maintenance of software	[9]
Systematic Mapping Study (SMS)	The process of identifying, classifying and reviewing previously published materials that address a specific research topic.	[9]
GDPR.	General Data Protection Regulation	[9]
Data Minimization	A data controller must limit the collection of personal information to what is strictly necessary to achieve a specific purpose.	[9]
Privacy-Enhanced Systems	Technologies that incorporate key data protection principles by reducing the use of personal data, increasing data security and empowering people.	[9]
Informational Privacy	The ability to maintain control over the use and distribution of one's personal information.	[9]
Privacy Preserving Software Systems	It helps to ensure that none of the parties involved has access to each other's data.	[9]

Human-Centered Privacy	Privacy with a focus on human needs.	[10]
Privacy Software Application	Software designed to protect the privacy of its users.	[10]
Privacy Knowledge Base	The knowledge base where the key privacy factors are defined is the basis for supporting developers' decisions in privacy software development.	[10]
Usable Security and Privacy	The action intended to prevent risks to the security and confidentiality of users arising from the interaction of humans (users) with computer systems.	[11]
Privacy Compliance	It states how organisations (regardless of their industry) meet the regulatory and legal requirements for collecting, processing, and maintaining personal information.	[12]
Domain Specific Language	A programming language that targets a specific kind of problem, rather than a general-purpose programming language that targets any kind of software problem.	[12]
Architecture Design	This is a field that focuses on satisfying requirements and desires for the construction of living spaces, using specific tools and, above all, creativity.	[13]
Software as a Service	The software is accessed online via a subscription, rather than being purchased and installed on separate computers.	[14]
Privacy Policies Update	Changing privacy according to the environment.	[14]

2.3.2 Main points of Literature review

According to [7], legal definitions are necessary. The concept of 'Privacy by Design' is a methodology that aims to incorporate the concept of privacy from the initial stages of designing a software system. The ultimate goal is for software engineers to consider the concept of privacy from the initial stages of designing a software [7]. In [7] the authors tried to see and analyze the concept of 'Privacy by Design (PbD) ' in several software implementations. Through their literature review, it appeared that most of the articles they studied about 'PbD', the legislation concerning 'data minimization' is the one that is discussed the most in them. 'Data Minimization'

is the GDPR principle defining that the data controller of the system, is required to limit its data collection. That is, the data that is necessary for the implementation of the system, i.e. for a specific purpose and only that purpose, must be collected [7].

More generally on the concept of 'PbD'. It is a concept that has not yet matured, although it was first introduced in 1990 [7]. The concept of 'Privacy' is the one that enables individuals to manage the data whose use may in the future identify them [7]. Many people confuse the concept of privacy with the concept of security [7]. The concept of Security deals with 'CIA'. Confidentiality, completeness and availability. Whereas, the concept of privacy deals with 'CIA' but also with the concept of transparency and accessibility. [7] concludes that the most important thing is indeed the addition of privacy in all stages of software technology as well as in organizational processes and government systems.

[8], on the other hand, sees the issue from a more general point of view, since it states that software technology in general is a combination of services. The academic literature concerning 'privacy engineering' often overlooks some parameters related to specific implementations in order to discover something that can be used everywhere. Most proposed architectures take a specific architecture for granted, as well as other methodologies that have been put into practice in other systems. [8] emphasizes that it is at the discretion of the software engineer not to use these practices. The same applies in the field of research. According to [8] there are very significant limitations to attempts to implement technologies that promote the notion of privacy. It is these limitations that make up the concept of 'Privacy Engineering'. Applications of this can be seen in [9], which has introduced the tactic of 'POSD' which stands for 'Privacy Oriented Software Development'. 'POSD' can find the 'sensitivities' of a software and arrange appropriate patterns for that software.

[15] has conducted a literature review on the concept of 'PbD' and concludes that, although the concept is necessary, it is difficult for software engineers to learn and put it into practice in their

systems. However, there are implementations that show that over time this concept is becoming more and more accessible in the field of software engineering. [10] has implemented a graphical tool through which the decision-making process in privacy-focused software systems is facilitated.

According to [10], the complexity of systems is increasing and therefore the possibility of cyber-attacks and data leaks is also increasing. This paper presents the 'Privacy Knowledge Base (PKB)' and the VIS-PRISE tool. With the tool, software engineers can integrate the concept of privacy and security into their systems in all phases of the software lifecycle and development. The privacy model includes seven rules that define a privacy environment with each rule describing specific actions and constraints. These constraints are: 'Proactive not Reactive, Privacy as a default setting, privacy embedded into design , full functionality , end- to-end security, visibility and transparency, respect for user privacy' [10]. There are 2 strategies of 'PbD'. The first strategy is 'Privacy by Architecture'. The purpose of this strategy is minimization, transparency, separation and abstractness [10]. The second strategy is 'Privacy by Policy'. The purpose of this strategy is information and control [10]. [10] also presents 'Privacy Patterns'. Patterns are methodologies that use scenarios whereby personal data is exposed. It identifies these scenarios and finds solutions which if reused can prevent data leakage and ensure privacy.

Many implementations have been made for privacy applications in software engineering. [11] has created a game that provides a training environment/tool for good software code practices that incorporate the concept of privacy. Another article concludes that there is a lack of models, processes and tools to facilitate 'PbD' in the software development process. This shortage has become increasingly important due to the requirements of GDPR. [16] presents a literature review, concluding that privacy patterns should be the first step of testing, so that they can be applied if appropriate to the system before creating a new one.

[17] also discusses the 2 strategies: architecture-based privacy and policy-based privacy. It has developed a three-layer model for controlling system operations with respect to privacy, based on user behavior. Privacy practices in software engineering systems are also found in [12]where the 'PRIAM' model was implemented. This model is a GDPR metamodel implemented with 'DSL'. This metamodel can confirm GDPR compliance and asserts privacy enforcement. The way people use technology can help create better systems that protect people's privacy [13]. [13] suggests that this process should be done through specific tools that understand the concept of privacy and incorporate insights from society.

Many other applications have also been made in the field of 'Cloud Computing'. An automatic privacy policy customization scheme has been implemented based on user requirements [14]. It is concluded here that the concept of 'Software as a Service' does not protect the privacy of users. Also, there is still the uncertainty whether the services provided by a system actually satisfy its policies. An algorithm is proposed which has been compared with Tableau's algorithm for satisfiability. [18]proposes the STP Chain. A tool that controls the privacy of systems based on 'blockchains' which thereby avoids malicious user behaviors in 'Crowdsourced Software Engineering (CSE)'. Finally, [19]provides an analysis of several different approaches and tactics applied in industry to maintain data privacy in 'Data-driven' systems in the software engineering domain. He implemented a tool that monitors data usage in order to identify unauthorized access to these systems.

2.3.3 Literature Review Comparison

2.3.3.1 Methodology

The methodology followed to compare the survey results is presented through a table in this section. The table compares the literature review conducted on the concept of Privacy in the Software Technology Sector.

Initially, by reading the literature, some "labels" have been selected for comparison purposes. "Labels" are keywords that may appear in more than one article and characterize the content of these articles. Then all the "labels" have been collected so that a list of them can be created to be presented as common features in the comparison table. In the comparison below 20 labels have been selected.

The left column of the comparison table shows the "labels" (vertically) and the relevant articles are shown horizontally. A '✓' symbol is shown if the article is tagged, and a 'X' symbol if the article is not tagged.

2.3.3.2 Results

Table 2 Comparative Table in literature review on Privacy in Software Engineering

Labels	**Papers**												
	[7]	[8]	[9]	[15]	[10]	[11]	[16]	[17]	[12]	[13]	[14]	[18]	[19]
Literature Review	×	✓	×	✓	✓	×	✓	×	×	×	×	×	×
Proposed solutions	×	✓	×	×	×	×	×	×	×	×	×	×	×
Privacy engineering	×	✓	×	×	×	×	×	✓	×	×	×	×	×
Privacy by Design	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×
Privacy enhancing	×	✓	×	×	×	×	×	×	×	×	×	×	×
GDPR	✓	✓	×	×	×	×	✓	×	✓	×	×	×	✓
Security	✓		×	×	✓	×	×	×	×	×	×	×	×
Tool implementation	×	×	×	×	✓	✓	×	×	✓		✓	✓	✓
Privacy by architecture	×	×	×	×	✓	×	×	×	×	×	×	×	×
Privacy Strategies	×	×	✓	×	✓	×	×	✓	×	×	×	×	×
Privacy Pattern	×	×	✓	×	✓	×	✓	×	×	×	×	×	×
Privacy Compliance	×	×	✓	×	×	×	✓	×	✓	×	×	×	×
Sociology	×	×		×	×	×	×	×	×	✓	×	×	×
Privacy Oriented Software (POSD)	×	×	✓	×	×	×	×	×	×	×	×	×	✓
Software as a Service (SaaS)	×	×	×	×	×	×	×	×	×	×	✓	×	×
Knowledge Base / Formal language / Algorithm Proposal	×	×	×	×	×	×	×	×	×	×	✓	×	×

Policy update	×	×	×	×	×	×	×	×	×	×	✓	×	×
Changing Environment	×	×	×	×	×	×	×	×	×	×	✓	×	×
Crowdsourced Software Engineering	×	×	×	×	×	×	×	×	×	×	×	✓	×
Blockchain	×	×	×	×	×	×	×	×	×	×	×	✓	×

It can be seen that the concept of Privacy by Design is a very important concept that is still being studied and has not yet matured into practice. Many articles are trying to address it but there is no article that covers all 20 labels.

2.3.4 Conclusions on Privacy and Software Engineering

In conclusion, the concepts of Privacy and Security are often confused. But they have different meanings. Privacy refers to the right of protecting Personally Identifiable Information, while Security refers to the protection of systems and data from unauthorized access. Most software engineers, do not have the experience or even the knowledge to fully understand the details and regulations for privacy and therefore cannot identify to avoid features that violate it. Therefore, it is necessary to incorporate Privacy by Design (PbD) approaches in software development. This ensures that software engineers take the concept of privacy into account from the early stages of design. PbD approaches include Privacy by Architecture, where privacy is built into the software architecture, and by Policy, as policies are created to ensure privacy is protected during the software development lifecycle.

The General Data Protection Regulation (GDPR) is a comprehensive privacy law that is widely used to protect personal information. However, GDPR compliance should not be limited to written policies and documents. There must be practical limitations on software systems that process personal information, such as 'privacy-preserving data collection and processing', 'data minimization' and secure storage and transmission. Many tools have been developed to integrate privacy into software development, such as Privacy Impact Assessments, privacy- enhancing

technologies and privacy engineering methodologies. However, these tools are not widely known or used and more efforts are needed.

2.4 Related work on Privacy and Usability

2.4.1 Relevant Terminology

Table 3 Terminology found in literature about Privacy and Usability

Notion	Explanation	Bibliography
Personally Identifiable Information (PII)	Information that can be used to personally identify an entity or individual	[20]
Agile methodologies	Its a software development project management approach where the tasks are processed into phases for efficiency	[21]
Machine Learning	Machine learning is a field in computing where programmers use large data to train a model in order to solve/generalize something.	[22]
Heuristics	It's a problem solving strategy that may not provide the best solution.	[22]
Privacy Calculus	Its a formal language definition used to guarantee legislation compliance.	[23]
Artificial Intelligence (AI)	It the field in computing where the human intelligence is simulated through computing devices, algorithms, and computer systems.	[24]

2.4.2 Main points of Literature review

As mentioned above, privacy policy documents are long and not easy to comprehend [20]. They are legal documents that express the way an organization uses the personally identifiable information of their clientele. The authors here suggested a publicly available browser extension called PrivacyCheck v2. By answering 20 questions the browser automatically summarizes any privacy policy. The questions are based on the GDPR. The browser extension is a tool for analysing competitors that identifies the leading rivals with the most robust privacy policies in the same industry sector. It brings awareness to the user for the purpose of making informed decisions in terms of selecting the services they use. They also state that the usage of Personally

Identifiable Information (PII) over the Internet has become a major privacy concern. The authors' goal here was to bring awareness to the common user of how their personal data are being processed and therefore enforcing them to make more considerate decisions.

The author in [21] suggests the notion of Practicable Privacy. That is privacy practices that are considered to be usable, acceptable and appropriable. The author also states that the digital privacy and security has focused on well developed countries only, despite the fact that the 80% of the global population are in countries that are not well developed. They presented computers and the internet to a group of supervisors in Nepal, emphasizing the socio-political factors that shaped the design landscape concerning digital privacy.

Kostova and others in [8] as seen in the previous section have pointed out that academic literature does not emphasize in the conditions of software production and therefore their results can be generalized easily. They have also mentioned that the notions of Software Architectures are not taken into account in recent proposed solutions in achieving Privacy by Desing. These then tend to propose solutions that are not in parallel with current practices. They have identified important limitations in the approaches that research is being done in terms of designing and evaluating privacy enhancing technologies. They have also outlined necessary research and actions to realign research with practical application, changes that are a prerequisite for incorporating academic findings on privacy into standard software engineering practices. The authors believe that literature on privacy technology and engineering are increasing exponentially. Therefore they state that is it very difficult to translate to current-day software systems. Lastly they want to emphasize that current privacy research is not compatible with the many common service architectures, including agile methodologies.

Diving more into the online world, it is known that website privacy policies let users avoid some collections of personal information [22] . Bannihatti et al have found that the instructions on how to do so, are buried 'deep in their text' expressing that users most often get tired of trying

to find the way to opt-out that they eventually give up. They have created a browser extension that automatically detects the opt-out choices and presents them to the user. They do that by combining many heuristics using machine leaning. The browser extension is made to present the options available. On another approach, Salgado et al [25] are positive in the thought of enhancing privacy policy interfaces in order to bring trust to the common user in terms of legislation. The technical ability for that was also using heuristics, since they suggested 'six usable privacy heuristics' along with usable privacy guidelines called "pugs" (pug#) . These, can actually identify problems in privacy policy interfaces. Sigmund in [23] on the other hand combined privacy calculus to verify determinants of reading privacy policy statements. He did that because he strongly states that users 'disregard' privacy policies, although they are made to mitigate their privacy concerns.

What will happen when theory becomes practice? The study in [26] analyzes developers' approaches to privacy tasks in software development, finding a focus on regulatory compliance and confidentiality in 119 Stack Overflow answers. It suggests promoting overlooked strategies through tools and enhancing support for managing third-party data practices.

In terms of visualization, Reinhardt in [27] states a huge question. Privacy policies are made to enable users make 'informed decisions' as mentioned before, on the other hand though, they lack of usability so users just don't read them. They also state that although visualizing the representation of the privacy policies may be beneficiary, they are not used in practice. So they created design principles along with a Visual Interactive Privacy Policy, derived from the Privacy Policy Nutrition Label, enhanced with control features and additional interactive components.

Last but not least, in Johansen's PhD thesis, the paper [24] presents a Usable Privacy Cube that represents the model to 'support evaluations of privacy'. There are three axes of variability captures. These axes are "rights of the data subjects, privacy principles, and usable privacy

criteria.” Johansen believes in two privacy perspectives. The one of the data subjects and the other of the controllers and processors. The criteria suggested are made to measure the level of usability. They measure effectiveness, efficiency, and satisfaction, considering both the objective and perceived usability outcomes. This involves evaluating accuracy, completeness, resource utilization (such as time, effort, and financial resources), and satisfaction ratings. Ultimately, the UP Cube is envisioned to serve as the foundation for a novel certification approach capable of appraising the usability of privacy, thereby benefiting everyday users.

2.4.3 Literature Review Comparison

2.4.3.1 Methodology

The same methodology used in 2.3.3.1 is used here as well. Specifically, the methodology followed to compare the survey results is presented through a table in this section. The table compares the literature review conducted on the concept of Privacy and Usability.

Initially, by reading the literature, some "labels" have been selected for comparison purposes. "Tags" are keywords that may appear in more than one article and characterize the content of these articles. Then all the "tags" have been collected so that a list of them can be created to be presented as common features in the comparison table. In the comparison below 'tags' have been selected.

The left column of the comparison table shows the "labels" (vertically) and the relevant articles are shown horizontally. A '✓' symbol is shown if the article is tagged, and a 'X' symbol if the article is not tagged.

2.4.3.2 Results

Table 4 Comparative Table in literature review on Privacy and Usability

Labels	**Papers**								
	[20]	[21]	[8]	[22]	[25]	[23]	[26]	[27]	[24]
Usability	✓	✓	✓	✓	✓	✓	✓	✓	✓
Visual Representations	×	×	×	×	×	×	×	✓	×
Practicable Privacy	×	✓	×	×	×	×	×	×	×
Digital Privacy	×	✓	×	×	×	×	×	×	×
Socio-political	×	✓	×	×	×	×	×	×	×
Software engineering	×	×	✓	×	×	×	×	×	×
Software Architectures	×	×	✓	×	×	×	×	×	×
Software by Design	×	×	✓	×	×	×	×	×	×
Privacy Engineering	×	×	✓	×	×	×	×	×	×
Privacy Enhancing Technologies	✓	×	✓	×	✓	×	×	×	×
Agile Methods	×	×	✓	×	×	×	×	×	×
Privacy Policy Comprehension	✓	×	×	×	×	×	×	×	×
Browser Extension	✓	×	×	✓	×	×	×	×	×
Policy Summary	✓	×	×	×	×	×	×	×	×
Personally Identifiable Information (PII)	✓	×	×	×	×	×	×	×	×
Opt-out user Choices	×	×	×	✓	×	×	×	×	×
Heuristics	×	×	×	✓	✓	×	×	×	×
Machine Learning	×	×	×	✓	✓	×	×	×	×
AI	×	×	×	×	×	×	×	×	✓
Privacy Design Guidelines	×	×	×	×	✓	×	×	✓	✓
Interactive Privacy Policy	×	×	×	×	×	×	×	✓	×
Legislation compliance	×	×	×	×	✓	×	×	×	✓
Readability	×	×	×	×	×	✓	×	×	×
Privacy Calculus	×	×	×	×	×	✓	×	×	×
Privacy Tasks	×	×	×	×	×	×	✓	×	×

2.4.4 Conclusions on Privacy and Usability

The review presented above shows that there are many approaches in the field and confirms that this field is emerging. Many authors expressed their concerns in putting into practice

visualization techniques in presenting privacy policies. They believe that even if there are solutions, no one will adopt them. To counterargue, research also shows that people are not aware at all. Not only of what they can do, and how they can do it (e.g. changing their privacy settings) , but they don't know the seriousness of the situation (e.g. if their personal data are being exchanged through platforms online). So, the following question arises : How can they adopt something they don't know exists?

From the review, we can extract that the reason why people have no awareness about the topic is that there is no easy / user-friendly way for them to learn more about it. The suggested solution in this dissertation is to make Privacy Policies easy to understand with visualization techniques in a way to achieve the level of awareness needed. After that the everyday user can make more informed decisions.

Chapter 3 : OhKéy – Privacy Policies Assessment Tool

OhKéy – Privacy Policies Assessment Tool

3.1 Introduction / About the tool

3.1.1 Motivation

The purpose of this comprehensibility assessment is to evaluate the understanding of people, that may or may not have a legal background, of the privacy policy since it is well known that they are long documents of text and not user friendly. The results are analysed in further chapters and a solution is also proposed. The tool created is a Privacy Policy annotation tool that fits the purpose described above.

The tool has three main pages. The 'Home' page, the 'Survey' page and the 'Annotations' page. The home page describes the purpose and asks for the user's consent in this survey. The 'Survey' page asks demographical questions about the user of the tool that will remain anonymous when submitted. The 'Annotations' page presents the user with the three types of Online platforms and the user must choose one in order to assess one of its privacy policies. Each type has three policies available.

Each participant was asked to assess one Privacy Policy of each type. It was mentioned that it would be best to assess as many policies as possible. Each submission assesses one privacy policy. To assess another, the participant should press the 'Home' button on the navigation bar and return to the home page in order to take the survey again.

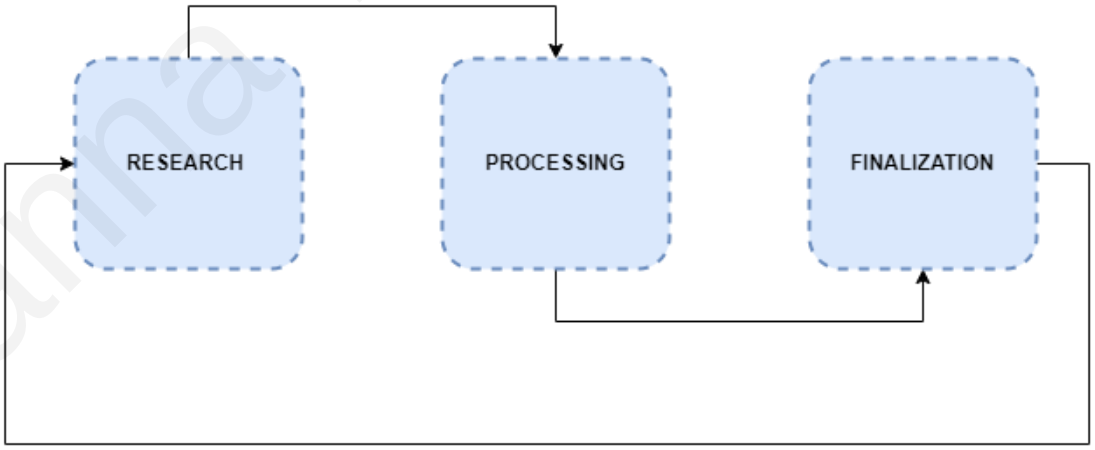
3.1.2 The need

This platform is crucial in order to analyze the comprehensibility of such legal documents in an interactive way. In the literature it was mentioned that most people find privacy policies difficult to read because they are very long. For the purpose of this comprehensibility assessment, its crucial to provide the user with an interactive and engaging tool in order to make the reading process less uninteresting.

3.1.3 The general methodology

The purpose is to present the user with original privacy policies, so that they can review and assess actual content. For the tool, two types of content were crucial. The actual privacy policies and the annotations / tags that the users would assess their contents with. The general methodology is the same for both requirements described above and can be seen in the figure below.

Figure 2 OhKéy tool - General methodology followed



As it can be seen in the figure, for each type of content in the platform the above methodology was followed. There was firstly some research, for example finding relevant bibliography about text annotation as long as finding which privacy policy platforms to include in the final tool. After that, there was always a processing procedure. For example, creating pairs of tags in order to present them in a Likert scale to be assessed more easily. Another example of processing was the anonymization needed for the actual policies of platforms. These will be further discussed

in the next section in more detail. Lastly the finalization where the methodology reaches the goal for the contents' outcomes, with a possibility of change, hence the arrow redirecting the whole methodology to the beginning.

3.2 Requirements Engineering

3.2.1 Tool Specification

The tool specification will be visually presented in the table below.

Table 5 OhKéy tool - specifications

Specification	Purpose	Details
Home page	The homepage is necessary to present the purpose of the tool, and to briefly describe the contents of it.	- Description of tool and its purpose - Consent form. - "START" button that redirects the user to the first page of the tool
Survey page	The tool does not have a user registration and user login since its purpose is to conduct an anonymous survey. Although some demographics will be collected anonymously for the purposes of the survey. Due to the fact that the tool assesses Privacy Policies, it is necessary to know whether the user studied or is studying Law. This will let us know about their background and therefore the result will be clearer.	- Age range ✓ "18-25" ✓ "26-40" ✓ "41-50" ✓ -"60+" - Gender ✓ "Male" ✓ "Female" ✓ "Other" ✓ "Prefer not to say" - Do you have a legal background? ✓ "Yes" ✓ "No"
Annotation Page	After the participant presses SUBMIT on the previous page, they will be redirected to the Annotation page, where they will be asked to 'Select the type of Platform' they would like to assess.	- Buttons for each type of platform that redirects them to the corresponding Policy Library.
Policy Libraries	There are three Policy Libraries that correspond to the different types of platforms assessed by the users. 1. Social Media Platforms 2. E-commerce Platforms 3. Online Advertising Platforms	- Each Library will have three buttons. - Each button will correspond to a real platform's Privacy Policy - When pressed, the user will be redirected to the chosen Privacy Policy in order to assess it. - Details in the Privacy Policy that may be used to identify which platform is assessed will be removed.

Policy Loading (Content)	The Policy Loading page is the page where the actual Privacy Policy of one platform type is presented. A procedure is needed to arrange the contents of this page, since actual Policies of online platforms are presented.	- The policy will be divided in 'clickable' boxes. - There will be a procedure where these boxes will be defined. - There will be a literature review where the annotating couples / tags will be defined. - In the backend procedure, there will be a way to determine whether the corresponding box is a header in the policy or a normal text.
Policy Loading (Technical Procedures)	The technical requirements for this page must be formed in a way that the assessment will be an easy-to-use step-by-step procedure, in a visually usable way.	- Clickable boxes of text ✓ When a text is pressed a pop-up will appear that will ask the user to start the assessment procedure - Assessment Procedure ✓ It will be a three-step procedure - Step 1: Select tags ✓ The user selects the tags they think are applicable for the pressed text and a Likert scale will appear - Step 2 : Rating using the Likert scale ✓ The user will be asked to rate the tags using a Likert scale ✓ The scale will be a color coded 'one – to - five' scale ✓ '1' means that the negative characteristics applies the most, and '5' means that the positive characteristic applies the most. - Step 3: Comments ✓ The user can leave any comments about the text they annotated in a new popup. ✓ Save Button - The assessment will be presented above the assessed text (Tags will be shown in a color-coded Likert scale manner) - Submit Button
Submission confirmation	This page is required to reassure the user that the assessment is completed successfully	Visual way to confirm the user that the assessment is being done and saved in the database.

According to the tool specifications described in the table above, it can be seen that in the tool there was a need for six different pages, each one with different characteristics and technical requirements. It also describes two critical procedures needed for the content of the platform. The first critical procedure is the Privacy Policies selection and anonymization and the second critical procedure is the selection of the annotating tags. These will be described in more detail in the section below.

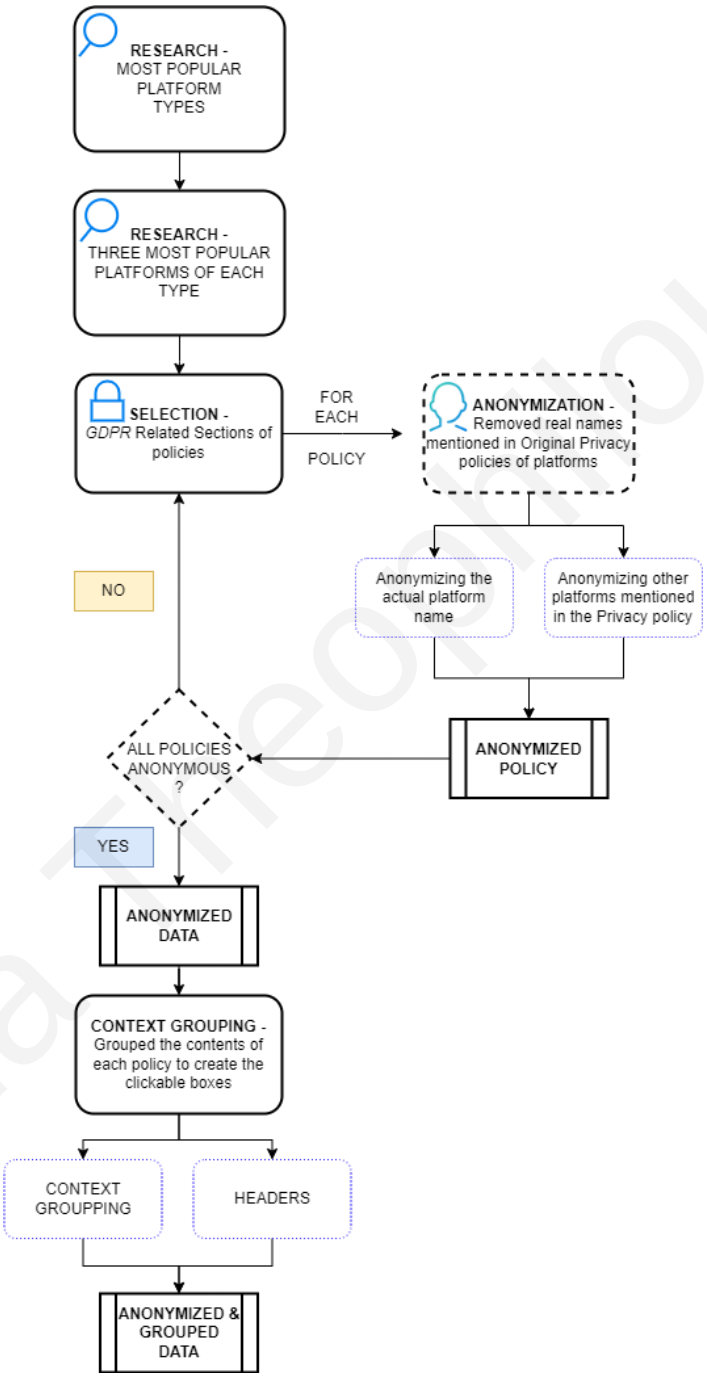
3.2.2 Content creation methodology

In this section the two content creation methodologies followed will be described. As mentioned above, the first methodology is about the Policies used in the platform. The second methodology followed is about the annotations used in the platform. The annotations are the words used to annotate the contents of the policies.

Before determining which policies will be used, there was a need to define which platforms will the Privacy Libraries include. According to [28] there are four types of platforms. E-commerce marketplaces , app stores , social media platforms and online advertising platforms. For the purposes of the platform three online platform types will be included. The one to be excluded are the apps because for this survey only online platforms will be analyzed.

For each type of platform, I selected its three most popular platforms to include in the corresponding Privacy Library. As described below, all platforms were anonymized before being available in the platform. The methodology involves literature review searching, gathering of data, processing of said data and grouping. More is shown in the figure below.

Figure 3 OhKéy - Privacy Policies Content Creation Methodology

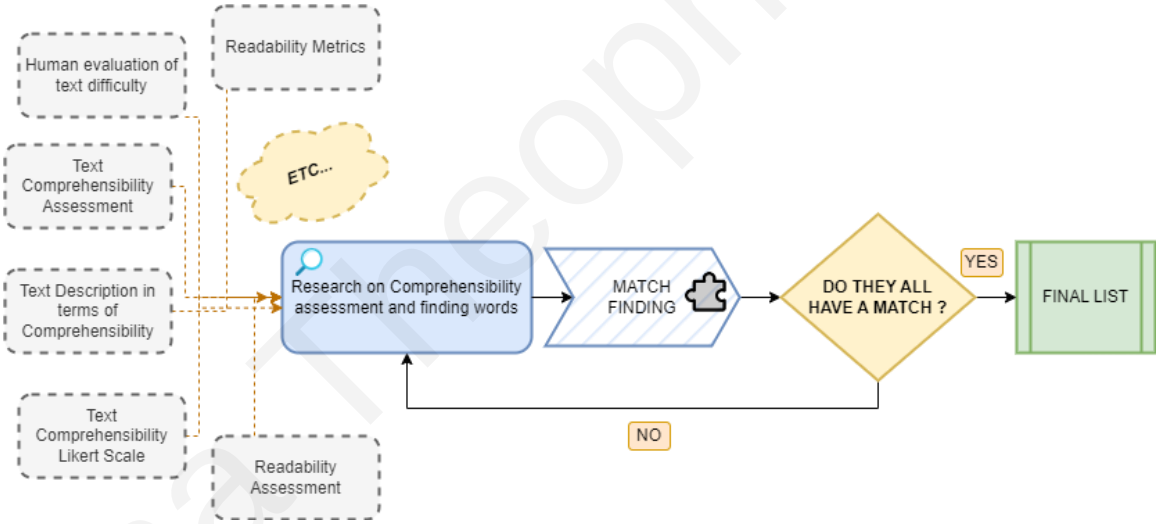


In the above figure, it can be seen that the first step of the methodology was indeed the selection of the most popular platform types and then the selection of their three most popular platforms. Then, there was a processing of these data since each platform included the official names of popular organizations, companies and others. Therefore, an anonymization process was crucial. For each Privacy Policy any names of companies or products were removed. After that, the anonymized data were grouped in terms of context relevance. Headers were also put in order to

distinguish the different parts of the policies. It is very important to mention that the contents of the Privacy Policies selected to be presented in the platform were only related to the different GDPR rights. After this procedure, the data were both grouped and anonymized correctly and they were imported in the database of the tool using CSV file uploading.

As mentioned above, the second crucial methodology is the one where all the annotations will be defined. The following figure shows this methodology.

Figure 4 OhKéy - Annotations Content Creation Methodology



The figure shows that there mainly was research on comprehensibility assessment in order to find words that could describe texts and paragraphs in general. There was an attempt to find an established list of annotations, but since similar work was not published, for the context of this thesis, a list of annotations is created by combining words from different literature. Some fields (not all) that were searched are shown in the figure for example: 'Text Comprehensibility assessment', 'Readability Metrics' and others. In the research it was also found that most text analysis is used to be rated with the Likert scale [29]. The Likert scale is a scale that rates something from one to five, one being the lowest. In order to use this in the content creation methodology, it was needed to define pairs of words where the left side of the pair would be the negative characteristic (most negative = rated with 1) and the right side of the pair would be the

positive characteristic (most positive = rated with 5). In order to achieve this, for each word found in the literature a pair was needed. In the figure this is referred to as a 'Match'.

According to [30] where they wanted to analyze Privacy Policies using Contextual Integrity annotations, the authors characterized Privacy Policy Text as '**clear'**, '**vague'** , '**abstract'** in terms of information transfer. The authors of [31] in their research of 'Text comprehensibility assessment for people with intellectual disabilities using a mobile application' they discovered that text can also be assessed in terms of language difficulty. Hence they stated that text can be assessed as '**plain language**', '**simple language'**, '**easy language'** etc. From this literature the following annotations were extracted : clear, vague, abstract, plain ,simple. It can be seen that some of them express something positive like 'clear' and some others express something negative like 'vague'. It can also be seen that none of the texts assess the trust that the user has in the read text, so in the list of the Final Annotations 'trust' should be added.

As it can be seen in the figure above, and as it was explained in the previous paragraph, the annotations can sometimes be negative or positive. Due to the fact that the annotations would be rated in the Likert scale, for the selected annotations, we defined its corresponding opposite pair. For negative text annotations we defined the positive and so forth. The final list of annotations is presented in the Table bellow, including what each tag would assess.

Table 6 OhKéy - Established List of Annotations

Annotation	**Details**
Ambiguous / Clear	Users can indicate whether the language is precise and unambiguous or if it leaves room for multiple interpretations
Vague / Informative	Users can state whether the text is informational or not indicating that there are or there are not gaps in the explanation.
Complex / Simple	This basic pair allows users to rate the complexity of the language used in the privacy policy.
Verbose / Concise	This pair helps in assessing whether the policy is straightforward and to the point or overly lengthy and filled with unnecessary detail

Suspicious / Trustworthy	This pair allows users to rate their trust in how their data is handled based on the policy's clarity and transparency.
Difficult language / Plain Language	This pair will be used by the users when they would like to assess whether the language used in the Privacy Policy is difficult to understand.

At this stage it is time to point out that the users can select the annotations they believe match with the text they have read. This is preferable because users may have many backgrounds. For example technical users will find simple texts referring to ‘caches’ ‘IP addresses’ etc. They can select as many pairs as they find applicable in each text. They can also annotate the titles. More about the platform will be explained in the next sections.

3.3 Technologies Used

3.3.1 HTML

HTML (Hypertext Markup Language) is the standard language used to create and design web pages. It provides the basic structure and content of a webpage, using tags to define elements such as headings, paragraphs, links, and images. For this thesis, the HTML version used is HTML 5 which is the latest. [32]

3.3.2 CSS

CSS stands for Cascading Style Sheets and used to enhance the presentation and visual appearance of HTML documents. In this dissertation the version used is CSS 3 which is the latest version of CSS. [33]

3.3.3 Bootstrap

Bootstrap is a front-end framework for developing responsive web applications. It provides pre-designed templates, components, and utilities that help developers create consistent and attractive layouts across different devices and screen sizes. [34]

3.3.4 JavaScript

JavaScript is a programming language commonly used for client-side scripting in web development. It enables interactive and dynamic features on web pages. In this thesis, all interactive actions were implemented with this programming language. [35]

3.3.5 PHP

PHP (Hypertext Preprocessor) is a server-side scripting language used for web development. It is used for communication with the server. It is widely used for building dynamic websites and web applications, handling tasks such as processing form data, interacting with databases, and generating dynamic content. [36]

3.3.6 MariaDB

MariaDB is an open-source relational database management system (RDBMS) and a fork of MySQL. The language used for this database is SQL. [37]

3.3.7 Apache

It's a widely used web server software. It powers a large percentage of websites on the internet and provides features such as HTTP request handling, virtual hosting, and security configurations. The version used is the Apache/2.4.6. [38]

3.3.8 PhpMyAdmin

It is a software tool that incorporates MySQL and MariaDB along with the Apache web server. In this tool we can visually see tables of the database, execute SQL and view the relationships among our database tables. The contents of the database can be downloaded within the tool in an SQL format a CSV format and many others. In this thesis implementation I used the CSV format to import Privacy Policies and content in general in the database. CSV is explained below [39].

3.3.9 CSV

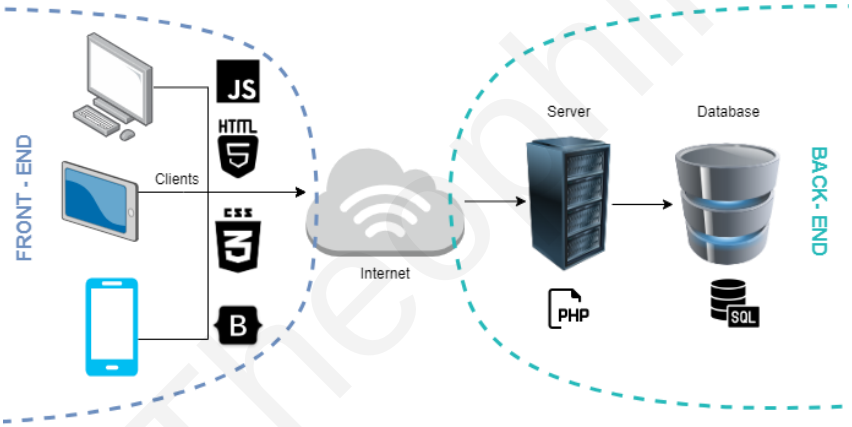
CSV (Comma-Separated Values) is a simple file format used to store data, such as spreadsheets and databases. It's an excel-like document. It consists of plain text data

organized in rows and columns, with each row representing a record and each column representing a field separated by commas or sometimes tabs. As mentioned in the section above, this format was used to import Privacy Policies in the database [40].

3.4 Architecture

3.4.1 Tool Architecture

Figure 5 OhKéy - Tool Architecture

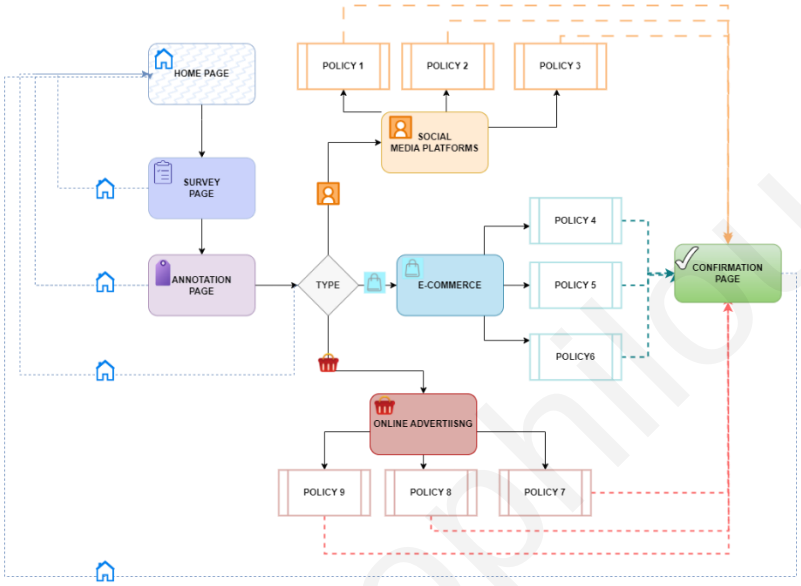


The figure above shows the general architecture of the tool. Just like in every web platform the architecture includes Front-end Components and Back-end Components. The Front -end components are 'what people see'. Clients are any user device that has access to the internet (e.g. phones, computers tablets etc.). The languages used for the Front-end are: JavaScript, HTML, CSS and the Bootstrap Framework.

The back-end components include the Server and the Database of the website with back-end languages being PHP and SQL respectively. As mentioned in the previous sub-chapter, the access and usage of SQL and the connectivity between the server and the Database was through the phpMyAdmin software tool.

3.4.2 Front-end Architecture

Figure 6 OhKéy - Front End Architecture



In the Front-end Architecture all the pages of the website are displayed. The architecture followed the exact specification defined in the previous sub-chapters. The goal here was to create a survey-like structure in order to replicate it, it was needed to create a sequential flow of actions for the user to follow. Meaning that the user should not have many options of how to proceed to the next steps since I was trying to replicate the step-by-step survey procedure.

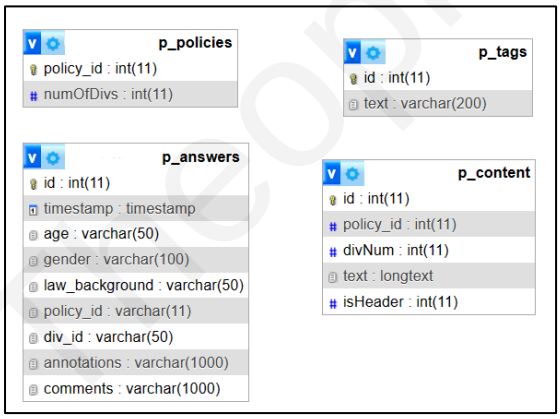
The first interaction that the user has with the tool is the home page, just like any other website. From the Home page the user could only move sequentially to the Survey page since the home page only provides a “START” button that navigates the user to the Survey page. The Survey page is also sequential since there is a small questionnaire as described in the Tool Specifications table, that requires only three questions. In order to move forward the user must press “SUBMIT” indicating that their answers are truthful and correct. After this page, the user can see three buttons in the Annotations page indicating the three types of platforms as described in the Content Creation Methodology above. The user then can navigate to one of the platform types, reassuring again the sequential walkthrough of actions. After that, the corresponding Policy Library will display its three options indicating three different policies. The user will then select one and start annotating it. After the annotation ends, the user can see an “OhKéy!”

button and press it when finished with the policy annotation. To annotate another policy the user must start from the beginning (i.e. the Home Page).

Finally, the figure shows that almost all pages have the option to go to the Home Page and not the previous page. This was purposely implemented in this manner, to replicate the option that the user has of deciding not to participate in the survey at any point while going through it.

3.4.3 Back-end Architecture / Database

Figure 7 OhKéy - Back End Architecture



The figure shows the Back-end Architecture where all the tables in the database are presented. It can be seen that all tables have their own identifiers. Starting with the `***p_policies***` table. This table is the one that defines how many policies there are. In this implementation there were nine privacy policies. As described in the Content Creation Methodology, each policy was anonymized and then divided into 'clickable boxes'. These clickable boxes in Web Design principles are called 'divisors' and are indicated with the HTML tag '*** '***. Hence in the table, there is a field called '***numOfDivs'*** in order to let the parameterized function that loads the content know, how many clickable boxes to load. The `**p_tags**` table is the one where the final list of annotations are stored. The final list of annotations and its explanation can also be found in the Content Creation Methodology in the same chapter.

The `***p_content***` table, replicates the contents of the Privacy Policy pages. These were imported in the database using CSV files. Each Policy has a unique identifier (the one that the table `***p_policies***` defines). The field `***divNum***` in this case is a number that increases in each row. Meaning that each row of the `***p_content***` table is a clickable box (a) of one Policy. For example if I have Policy 10 that has 50 clickable boxes, the row : " `***p_content***` - [***id*** : 1101 , ***policy_id*** : 10 , ***divNum*** : 51 , ***text***: Right to Erasure , ***isHeader***: 1] " would not be displayed. The `***text***` field is the one where actual content is stored, and the `***isHeader***` is a yes or no field defining whether the text is a header in the Privacy Policy.

Last but not least, the `***p_answers***` table, is the one that stores all user answers when submitted. The `***timestamp***` field is an automatic field that stores the date and time a submission was done. Since the survey is anonymous, each row is identified by a unique ID, in order to avoid collisions in the extreme scenario of two participants submitting the same time an annotation with all fields the same. The `***age***` , `***gender***` and `***law_background***` fields are the ones from the demographical survey (the 'Survey page') that was mentioned above. The `***policy_id***` is the foreign key that connects this table with the `***p_policies***` table. The field `***div_id***` is a string concatenation of the policy and the `***divNum***` described above. For example if someone annotated the 5th box of the 10th policy, the `***div_id***` field would be : "***policy:10_div:5***". The `***annotations***` are the pairs of words described above. For this field (continuing the previous example) a valid value would be: ***"Ambiguous / Clear: 2, Complex / Simple: 2,***" meaning that this user annotated the 5th box as ambiguous and complex (since the Likert scale rating is below 3 hence is low). Finally, the `**comments**` field stores the comments provided by the user. If there were no comments, the field stores ***'No comment provided'***. The table below summarizes the detailed description provided above:

Table 7 OhKéy - Back End Architecture Summary

Table	Field	Explanation
p_policies	policy_id	Unique identifier of policy
	numOfDivs	Number of clickable boxes in the privacy policy
p_tags	id	Unique identifier of tag
	text	Annotations content
p_content	id	Unique identifier
	p_policies	Corresponding policy ID
	divNum	Integer : a number that increases in each row and indicates the sequence of the content in each policy
	text	Actual policy content
	isHeader	Values {1,0}. Indicate whether the text is a header or not. If it is it will appear in bold letters in the page.
p_answers	id	Unique identifier
	timestamp	Automatic field that stores the date and time a submission was done.
	age	Values {“18-25”, “26-40” , “41-50”, “60+”} String: provides information about the age group (demographical data)
	gender	Values {“Male”, “Female” , “Other”, “Prefer not to say”} String: provides information about the gender of the participants (demographical data)
	law_background	Values {“Yes”, “No”} String : Provides information about whether the participants have a legal background
	policy_id	Foreign key : identifies the policy id of the corresponding annotation
	div_id	Its a string concatenation of the `***policy_id***` and the `***divNum***`
	comments	String: text that the users input. If the users did not, then the value is by default : “No comments provided”

3.5 User Manual

3.5.1 Home Page

Figure 8 OhKéy - Home page part 1/3.

Figure 8 shows what is the first interaction with the tool. The user firstly sees this page, displaying the name of the tool, and its purpose along with the navigation bar with the only option being the “HOME” which is the current page. The reason why the “HOME” is the only option in the navigation bar, is explained in the Front-end Architecture’s chapter. The platform is live and can be visited when clicking here.

Figure 9 OhKéy - Home page part 2/3.

Figure 9 shows the screen on the first scroll where the importance of the tool is described along with the purpose of the survey, in order to only conduct this survey with informed participants. The “START ANNOTATING” button is the one that when pressed will navigate the user to the next page in order to answer in some demographics.

Figure 10 OhKéy - Home page part 3/3

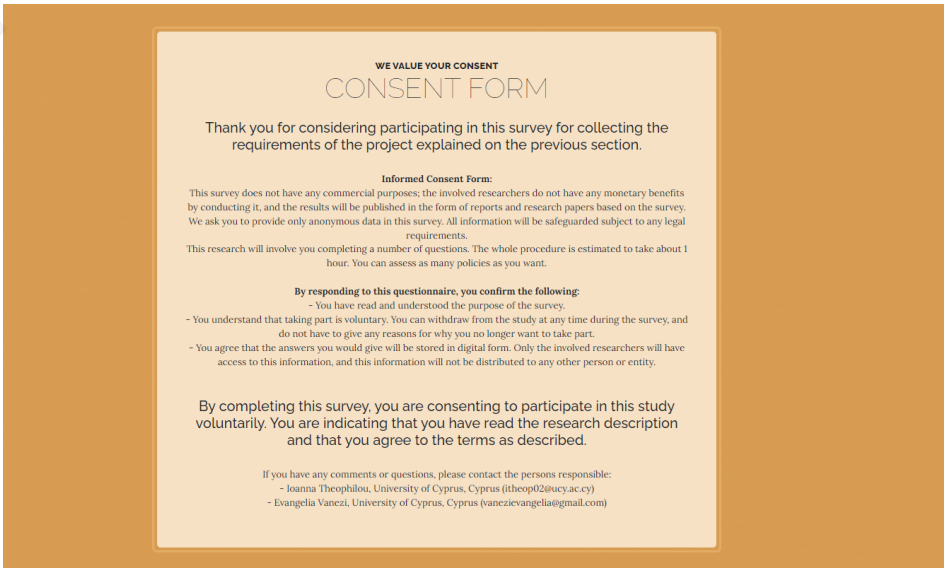


Figure 10 shows the Consent form, that is available at the end of the home page since it is divided into two parts. The explanation of the importance and the consent. This statement informs the users about the survey, how much time it will take to complete it, that the survey is anonymous and who to contact in case something concerns them. Since the Consent form is a written statement, participants who do answer the survey are immediately considered that they consent to its terms.

3.5.2 Survey Page

Figure 11 OhKéy - Survey page part 1/3

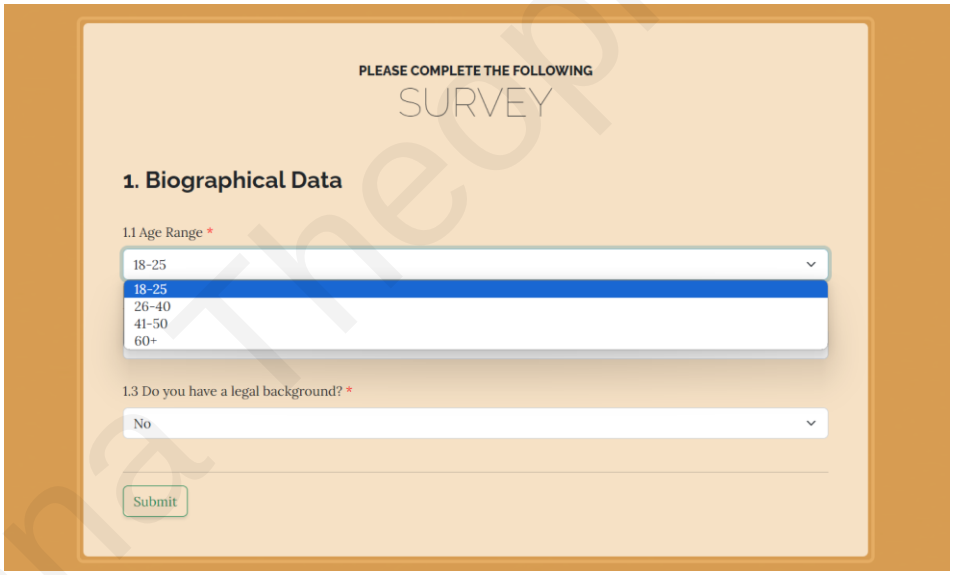


Figure 11 shows the first question of the Survey Page, where the user must select their age range. The possible answers to this question are shown in the figure.

Figure 12 OhKéy - Survey page part 2/3

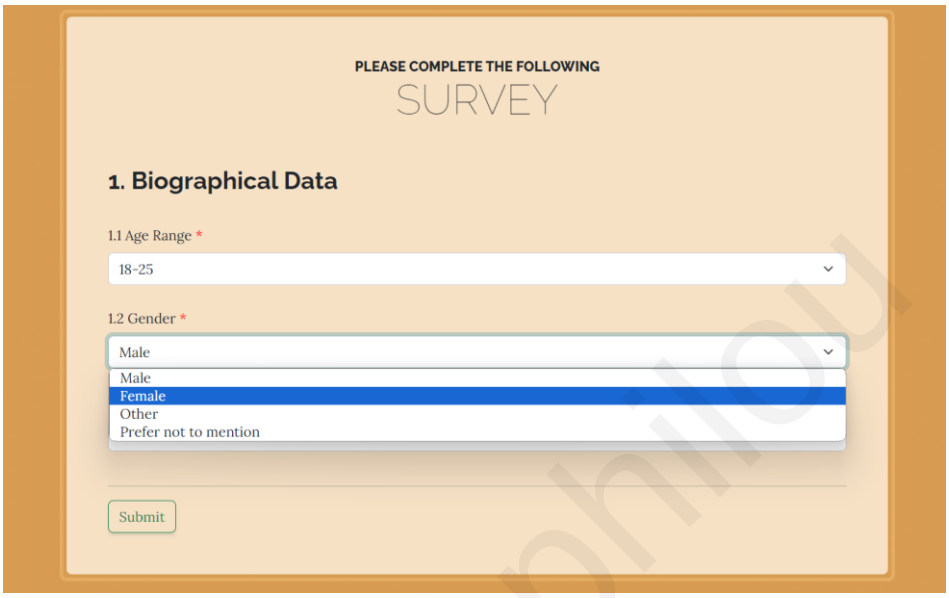


Figure 12 shows the possible answers for the second question to the biographical data survey, where users must state their gender.

Figure 13 OhKéy - Survey page part 3/3

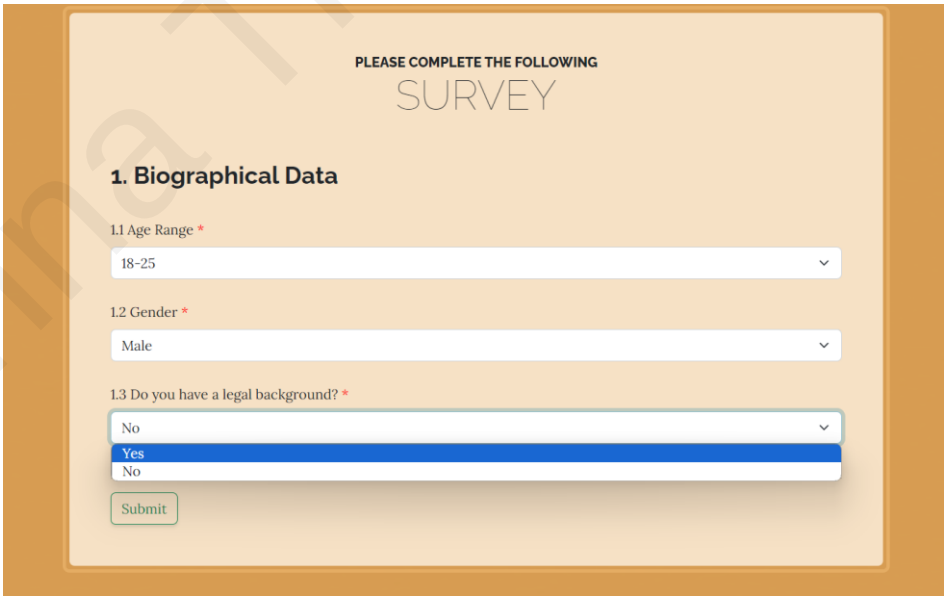
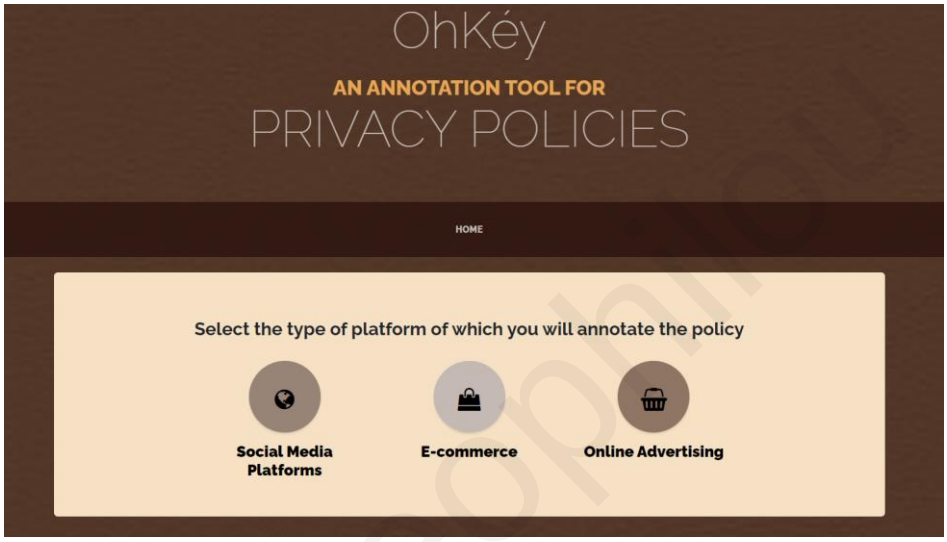


Figure 13 shows the last biographical data question in which the users must select whether they have a legal background. Having legal background means that they have studied law, or that they are currently studying law or even that they work in the field (lawyers etc.) . By pressing "SUBMIT" the user can navigate to the next page, where they are asked to select the type of the platform they want to assess.

3.5.3 Annotation Page

3.5.3.1 Platform Type selection

Figure 14 OhKéy - Annotation page / Types of Platforms



In the figure above the Annotations Page is shown. The user is presented with the three types of platforms: Software Media Platforms, E-commerce, Online Advertising. They must select one in order to go to the corresponding Policy Library.

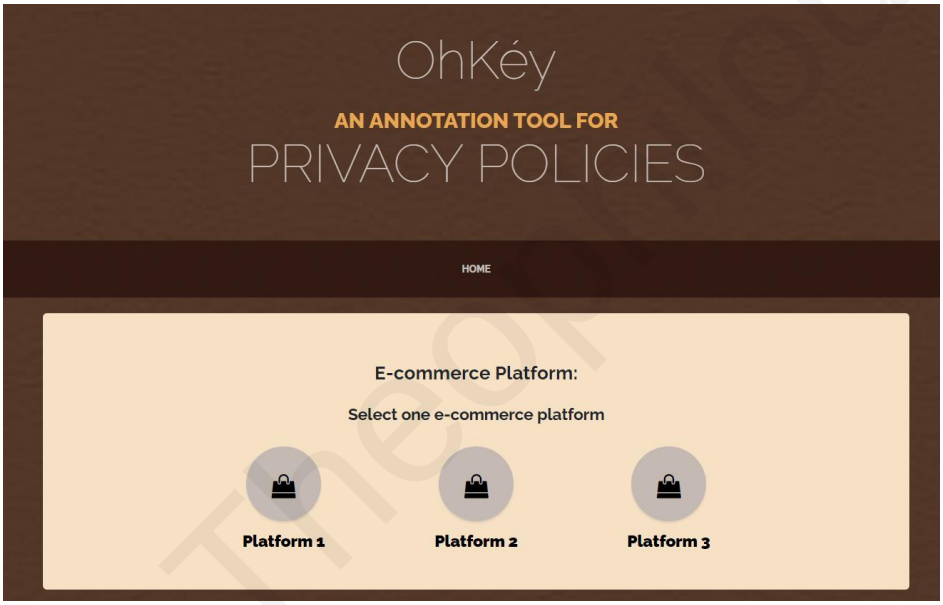
3.5.3.2 Social Media Platforms

Figure 15 OhKéy - Annotation page / Social Media Platforms

If the user selects ‘Social Media Platforms’ in the Annotation’s page, they will see these three options. These are actual Policies of Social Media Platforms, anonymized, grouped and only GDPR related.

3.5.3.3 E-commerce

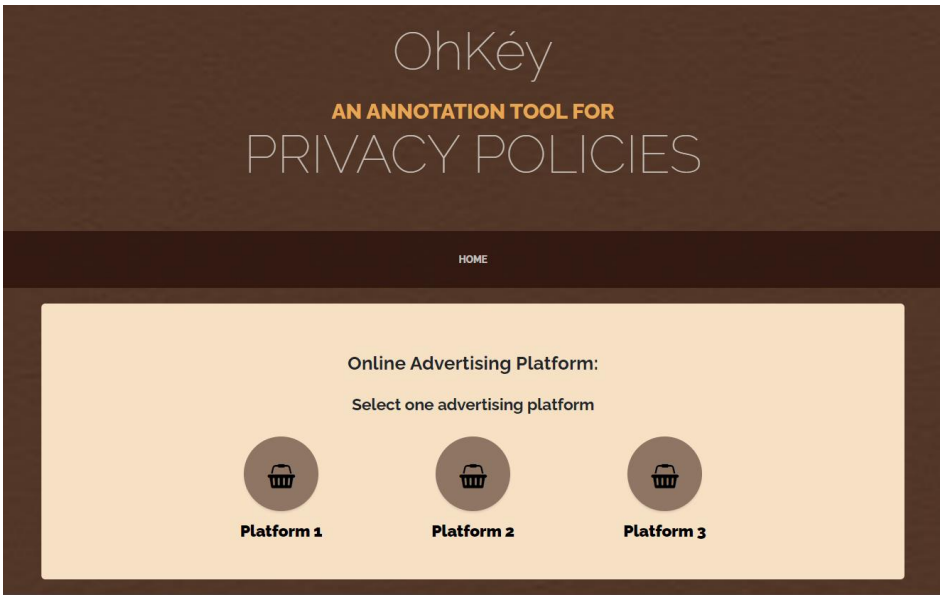
Figure 16 OhKéy - Annotation page / E-commerce Platforms



When the user selects ‘E-commerce’ in the Annotations page, they will be presented in three other policies of real E-commerce platforms.

3.5.3.4 Online Advertising

Figure 17 OhKéy - Annotation page / Online Advertising Platforms

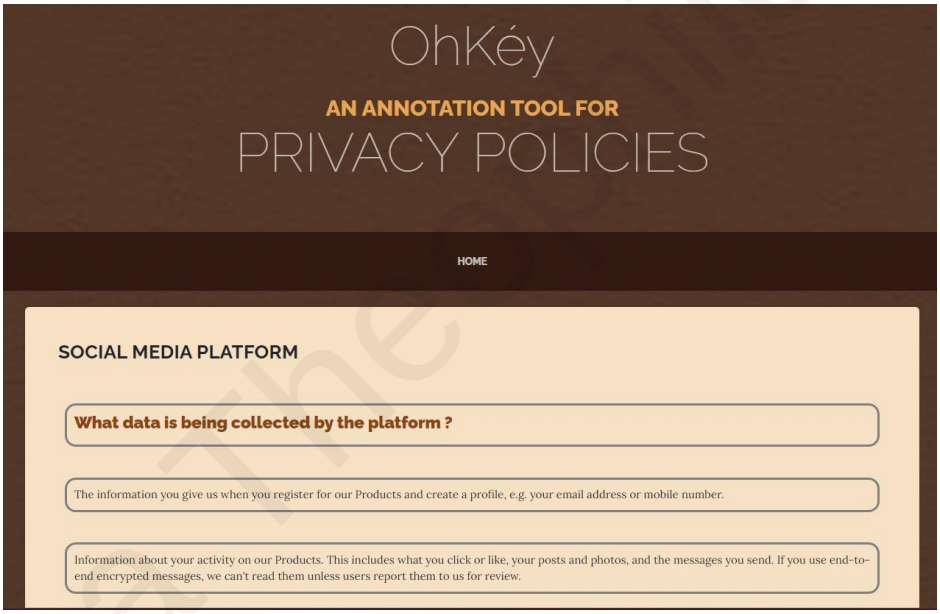


The option ‘Online Advertising’ bring the user to this Policies Library in which again they can select one policy to start annotating.

3.5.4 Survey Submission

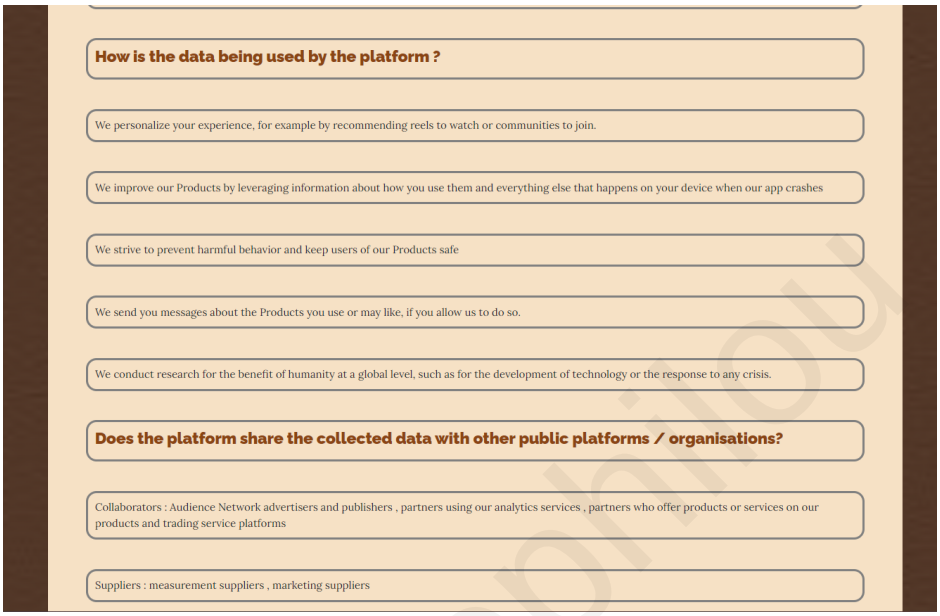
In order to demonstrate the functionalities of the platform, in the figures below, Policy 1 of the Social Media Platforms will be annotated.

Figure 18 OhKéy - Survey Submission Part 1/11



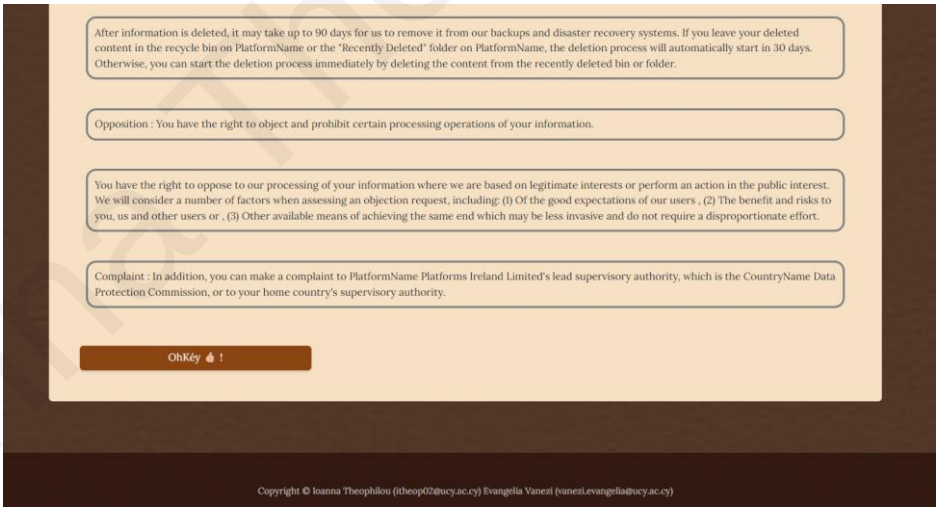
In previous chapters it was mentioned that each policy is grouped and categorized into ‘clickable’ boxes where any Company or Organization names are removed. These clickable boxes are the grey bordered text shown in Figure 18 and in all upcoming figures related to the Survey Submission. It was also mentioned in the Back-end Architecture, that headers are indicated in bold color. It can be seen that headers are shown indeed in bold brown color.

Figure 19 OhKéy - Survey Submission Part 2/11



Scrolling through the page, we can see the different boxes and headers.

Figure 20 OhKéy - Survey Submission Part 3/11



At the end of the page, there is an ‘OhKéy’ button indicating the final submission of the whole survey. Now, lets see the steps of the annotation one by one. The user firstly presses on a box of policy text. In the example, the user pressed on the box that explains the “Withdrawal of consent”.

Figure 21 OhKéy - Survey Submission Part 4/11

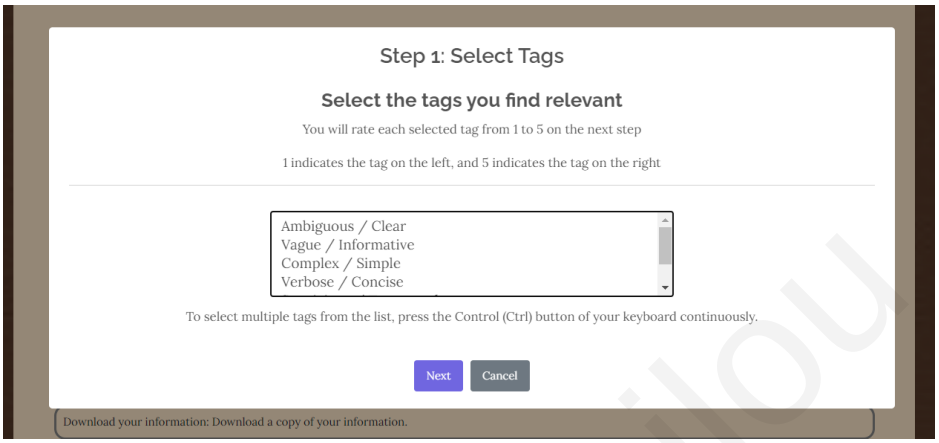
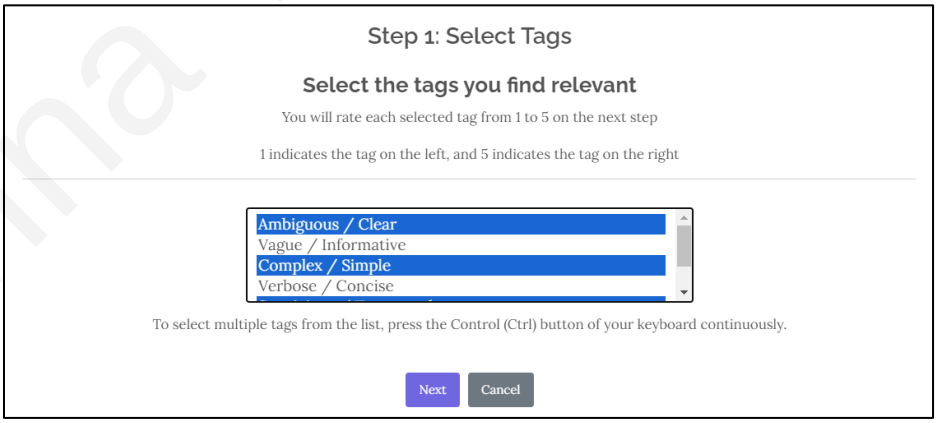


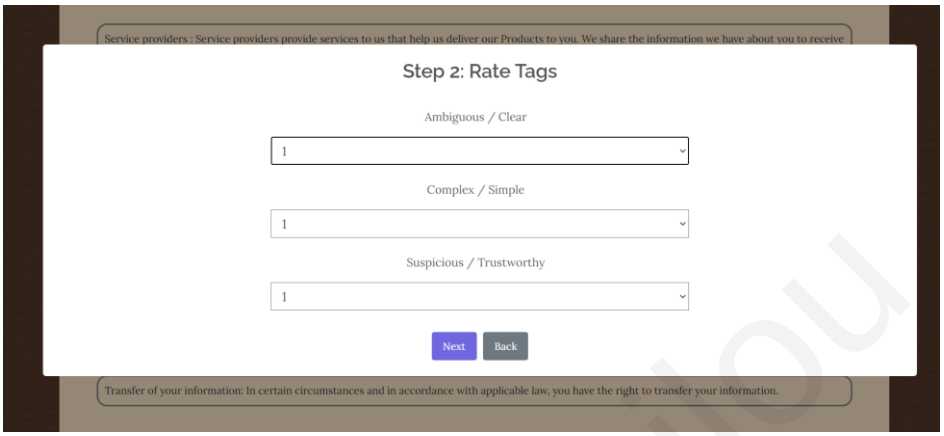
Figure 21 shows the popup that firstly appears during annotating a Policy Text. The popup says that the user must select the tags they believe are relevant to the Policy Text read. If they believe that more than one tags are applicable they shall press the CTRL button on their computer continuously and press with the computer mouse the next tag. It also informs the user that the selected tags will then be rated using the Likert scale ('You will rate each selected tag from 1 to 5 on the next step').

Figure 22 OhKéy - Survey Submission Part 5/11



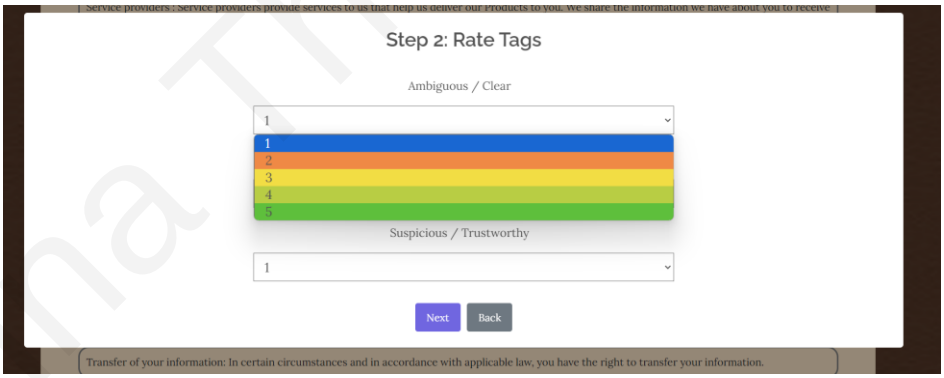
Taking a closer look at the demostrating example, where the user here selected three tags to rate for the specific text they read. The tags they selected, are 'Ambiguous / Clear' , 'Complex / Simple' and 'Suspicious / Trustworthy'. They then press 'NEXT'.

Figure 23 OhKéy - Survey Submission Part 6/11



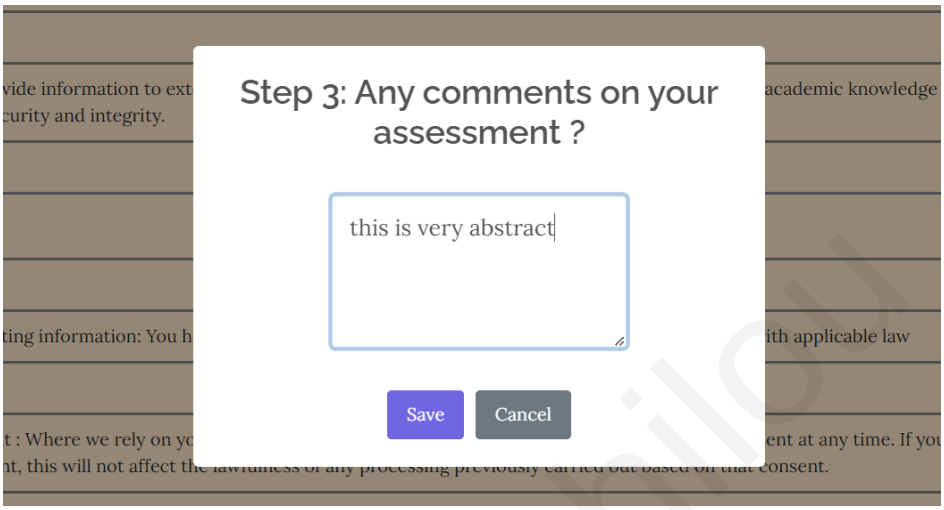
After pressing 'NEXT' on the first step of the annotation procedure, the second popup appears that asks the user to rate each pair on a scale of one to five where five is considered to be the most positive characteristic, and one is considered to be the most negative characteristic.

Figure 24 OhKéy - Survey Submission Part 7/11



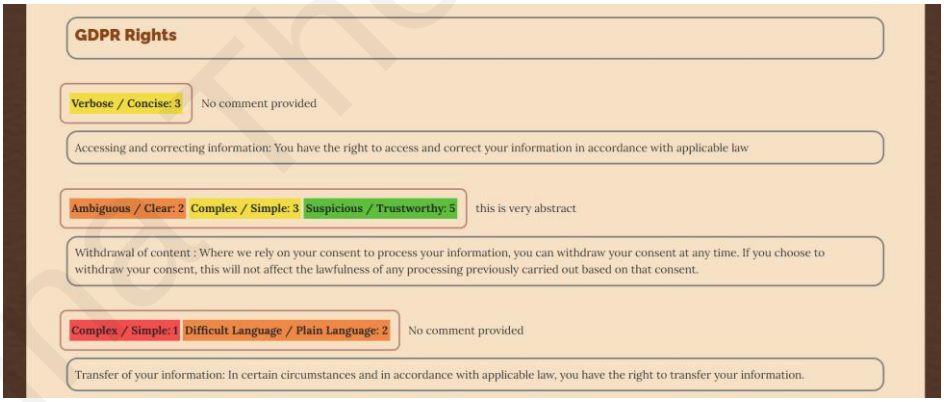
When the user pressed the drop-down menus under each annotation, the scale is presented in a colour coded manner.

Figure 25 OhKéy - Survey Submission Part 8/11



As mentioned in previous chapters, the final step of the annotation is to provide comments. Here the user stated that 'this is very abstact' and pressed 'SAVE'.

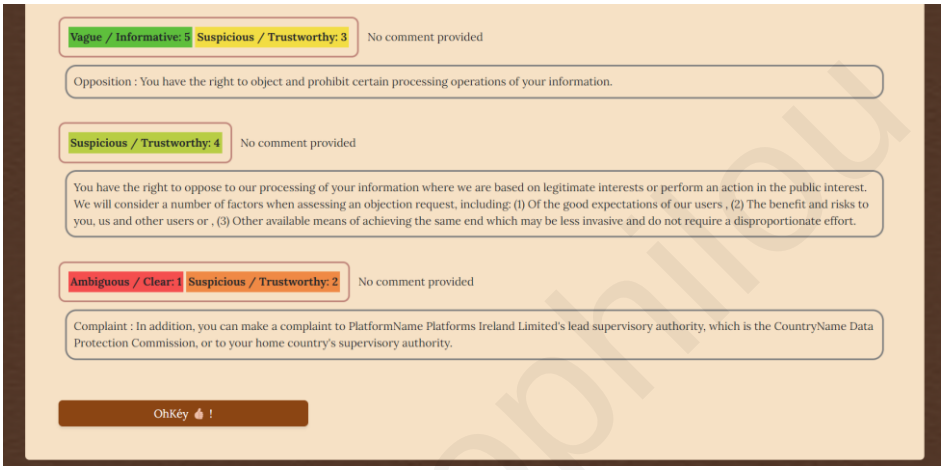
Figure 26 OhKéy - Survey Submission Part 9/11



The user in this specific demonstration annotated the text regarding the 'Withdrawal of consent'. It can be seen that all options selected in the second step of the annotation are presented above the annotated text, color-coded with the selected Likert scale rating. For example, the user in the second step selected 'Ambiguous / Clear' with the rating of 2. In Figure 24 it can be seen that the colour for this rating is orange. Hence the pair is highlighted with orange. It is also noted that rated with two means that is more Ambiguous than clear since its under 3 which indicates a neutral value. On the other hand, it seems that the specific text is 'Trustworthy' since the user annotated the text 'Suspicious / Trustwothy' with the highest score. The comment that the user provided in Figure 25, is presented in Figure 26 near the Annotations.

Important to mention that other annotations with no comments provided by the user are indicated with the ‘ No comment provided’ text.

Figure 27 OhKéy - Survey Submission Part 10 /11



The user can then press the final button indicating that the annotation presented to them is correct and shall be submitted.

Figure 28 OhKéy - Survey Submission Part 11/11

Finally, the confirmation page appears reassuring the user of the successful submission of their annotation. The user can annotate one Privacy Policy at a time, therefore they can press the ‘HOME’ button on the navigation menu in order to be redirect to the first page of the survey.

This page concludes the demonstration of the OhKéy Assessment tool. The survey’s results and methodologies are discussed in the next section.

Chapter 4 : Comprehensibility Assessment Analysis

Comprehensibility Assessment Analysis

4.1 Introduction

In this chapter, the results of the Comprehension Assessment conducted through the OhKéy platform will be presented. As described above, the OhKéy Privacy Policies assessment tool provides users with Policy Libraries of three types of platforms.

Each participant was asked to assess at least one policy of each platform. It is important to note that each submission assessed one Policy at a time. Participants were asked to choose from a list of pairs to annotate specific policy texts. The pairs were words that could describe texts, paragraphs in terms of comprehension, readability and language difficulty. Each pair contains opposing words (=words that describe the same thing positively and negatively). After selecting one or more pairs for a specific policy text, they were asked to rate these pairs using a Likert scale of one to five (five meaning that they believe the most positive characteristic is applicable in the selected policy text).

The analysis of this procedure and the outcomes of this survey will be analyzed in detail in the following sub-sections.

4.2 Purpose

The general purpose of this Comprehensibility Assessment is to identify gaps and propose a way to solve them in regard to users' understanding of Privacy Policies. As mentioned in the literature review, it was proven that most users do not read what they consent to. Here we are trying to find whether this blind consent is due to the contents of the privacy policies in terms

of language usage and lengthiness (e.g. many legal terminologies , many technical terminologies, vagueness and verbosity etc.).

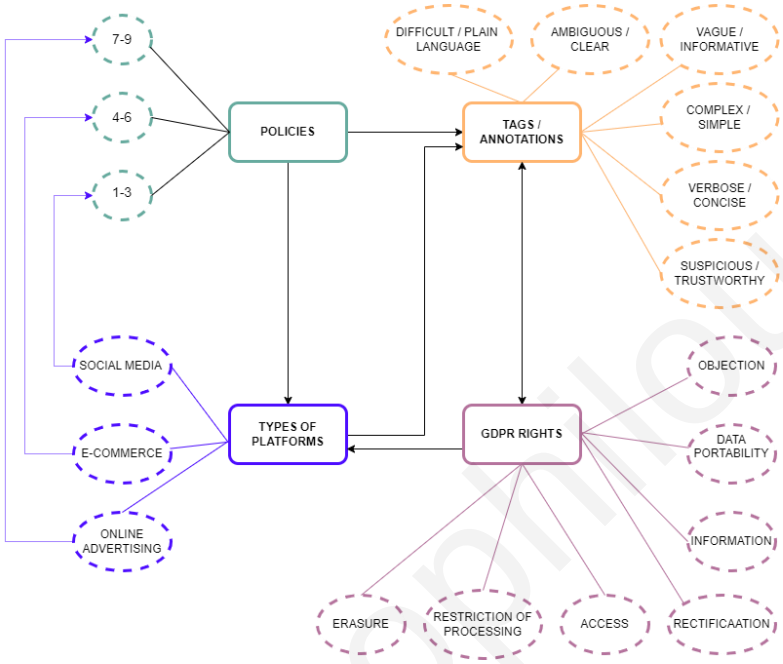
More specifically, we were trying to detect how many negative annotations were used in the assessments and whether these annotations were coming from people with no legal background. We were also trying to find how many negative ratings each annotation had in order to propose ways to solve the specific negative characteristics. Neutral responses are also important because even though they are not negative, they indicate that there is room for improvement in terms of understandability and readability.

The survey was running for a month and many results were gathered. The analysis of these results as long as the methodology used for the analysis are discussed below.

4.3 Methodology

The first step in trying to create a methodology was to identify the structure and how each component of said structure is related to the other or even finding if there was any correlation at all. The structure is also described in more technical terms in the Back-end Architecture of Chapter 3, although for the purposes of finding correlations in the structure, the following figure was very helpful to do so.

Figure 29 Comprehensibility Analysis - Methodology



The figure shows the four main components of the platform. These are : The policies, the Tags /Annotations , the GDPR rights , and the Types of platforms. The arrows show the correlation found between the components. Some are multidirectional. The correlations in the figure are the outcome of the questions shown in the table below:

Table 8 Comprehensibility Analysis - Methodology Questions

Questions			
How many times each annotation was used	*Percentage of negatives*	*What type of platform was more verbose*	*How many negative annotations were there for vagueness in general*
How many tags were used for each policy	*Percentage of neutrals*	*What type of platform was more suspicious*	*How many negative annotations were there for verbosity in general*
How many annotations were there annotated at 2 and below for each tag	*Percentage of positives*	*What type of platform was more ambiguous*	*How many negative annotations were there for complexity in general*
How many annotations were there annotated as neutral (at 3)	*What type of platform was more complex*	*What type of platform uses the most difficult language*	*How many negative annotations were there for suspiciousness in general*
How many of each tag were annotated as 4 and above	*What type of platform*	*How many negative annotations were*	*How many negative annotations were*

	was more vague	*there for ambiguity in general*	*there for language difficulty in general*

To relate the questions with the figure above, the questions can be characterized as the arrows in the figure connecting two components. For example: the question 'What type of platform was more verbose?' correlates the Type of platforms component with the Annotations/ Tags.

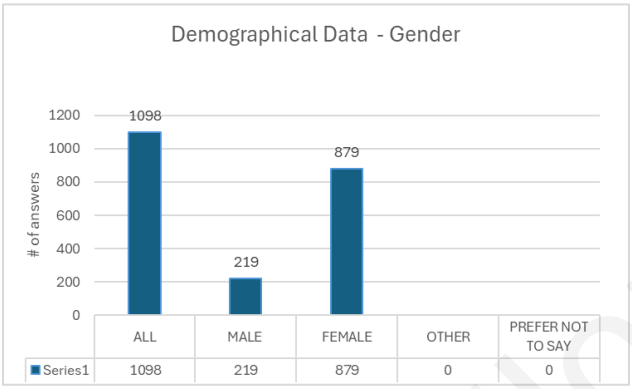
While overviewing the questions, it was realized that there is a need to define which annotations will be marked as positive, negative or neutral. **Positive** annotations are the ones rated ***four*** and ***five*** in the Likert scale, **neutral** annotations are the ones rated as ***three*** and **negative** annotations are the ones rated with ***two*** and ***one***. In order to answer the questions above, an extensive analysis of the results was conducted. This analysis along with the corresponding graphs can be seen in ANNEX I . In the analysis section below, only the most important and relevant outcomes will be highlighted and discussed.

4.4 Analysis

4.4.1 Demographical Data

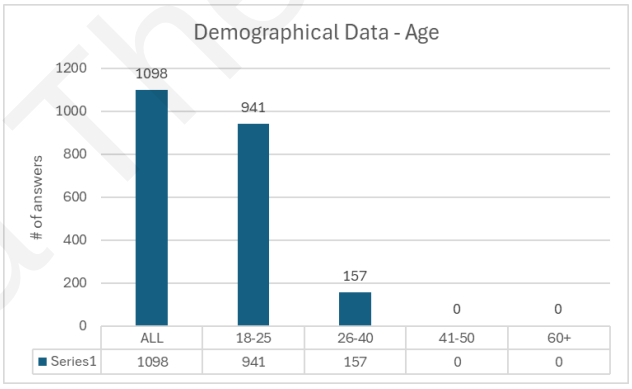
As explained in chapter 3, the survey was anonymous, although in order to understand the background of the participants, some demographics were collected. Firstly, they had to select their age group, gender and whether they have a legal background. Knowing their background will show if Privacy Policy Comprehension is better when someone studied law. Although not all law degrees study the GDPR, some might assume that it would be easier for someone with this background to understand the legal terminology in it.

Figure 30 Comprehensibility Analysis - Demographics / Gender



As shown in the figure above, there were 1098 total annotations from which 897 were female participants and 219 male. The most important result of this graph is the number of total annotations.

Figure 31 Comprehensibility Analysis - Demographics / Age



From the 1098 participants, almost 86% are in the ‘ 18-15’ age group and 14% are in the ’26-40’ age group. We can assume then that the participants have some experience with online platforms since these age groups are considered to be more into technology. Stating again that all privacy policies in the platforms are from companies that are well known, so participants may have used some or even all the platforms of which the Privacy policies are included in the tool. Although due to the anonymization procedure explained in chapter 3, they were not able to identify any of them.

Figure 32 Comprehensibility Analysis - Demographics / Legal Background

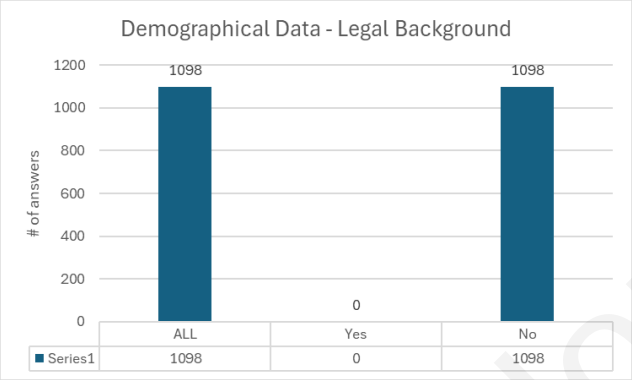


Figure 32 shows that all annotations were done by participants with no legal background. This will be taken into account in the conclusions since comprehension might be affected.

4.4.2 General Information

Figure 33 Comprehensibility Analysis - General Info / Number of Annotations on each type.

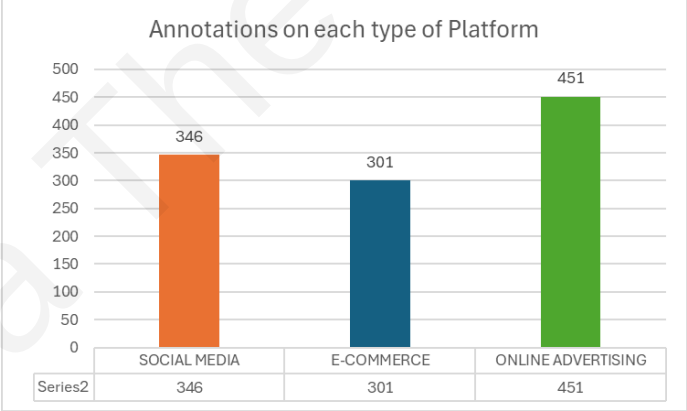
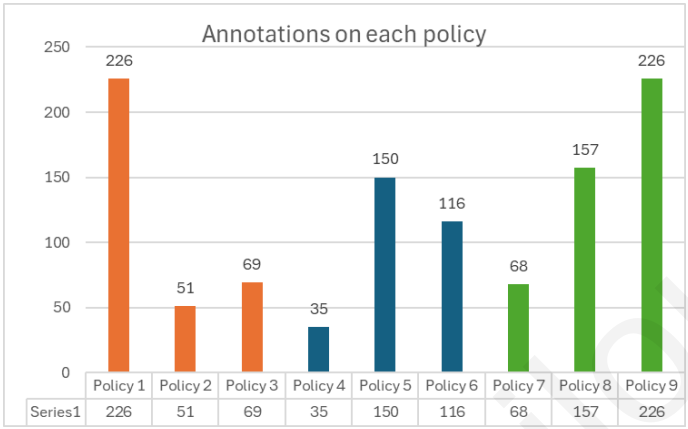


Figure 33 shows that most annotations were done in online advertising platforms, and this might be due them being lengthier. The Figure below, shows how many annotations were there in general, for each policy. The policies below are color-coded to the corresponding platform type shown above. Meaning that orange policies are social media policies , blue is e-commerce and so forth.

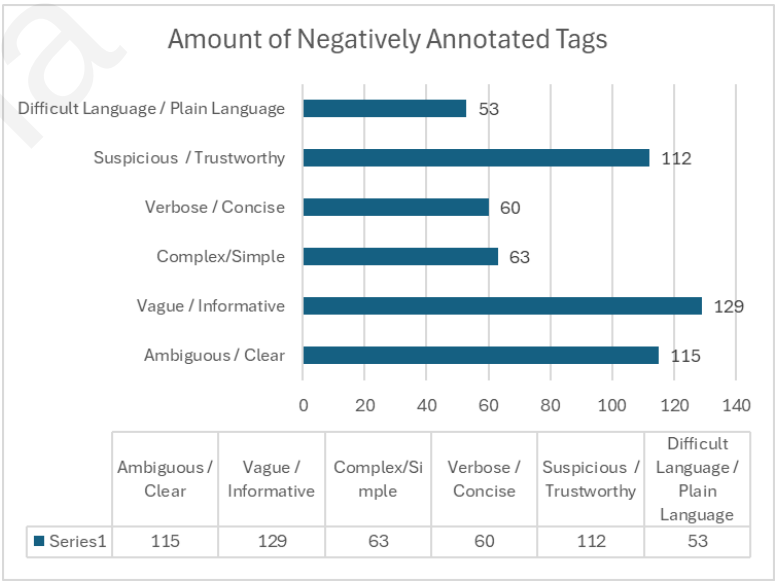
Figure 34 Comprehensibility Analysis - General Info / Number of Annotations on each policy



4.4.3 Negative Assessments per Annotation

As mentioned in chapter 4.3, ***negatively*** annotated tags are the tags rated by users as '2' or '1' in the Likert scale. The figure below shows how many negative annotations were there for each annotation. For example, 'Vague / Informative = 129' means that the survey had 129 texts tagged as 'Vague'.

Figure 35 Comprehensibility Analysis - Negative Assessments per Annotation / Amount



It can be seen that the tag with most negative annotations is the 'Vague / Informative.' Vagueness then seems to be the biggest problem. After that, the second problem seems to be 'Ambiguity' and then 'Suspiciousness'. Fourth comes the 'Complexity' , fifth the 'Verbosity' and last the 'Language Difficulty'.

Figure 36 Comprehensibility Analysis - Negative Assessments per Annotation / Percentage.

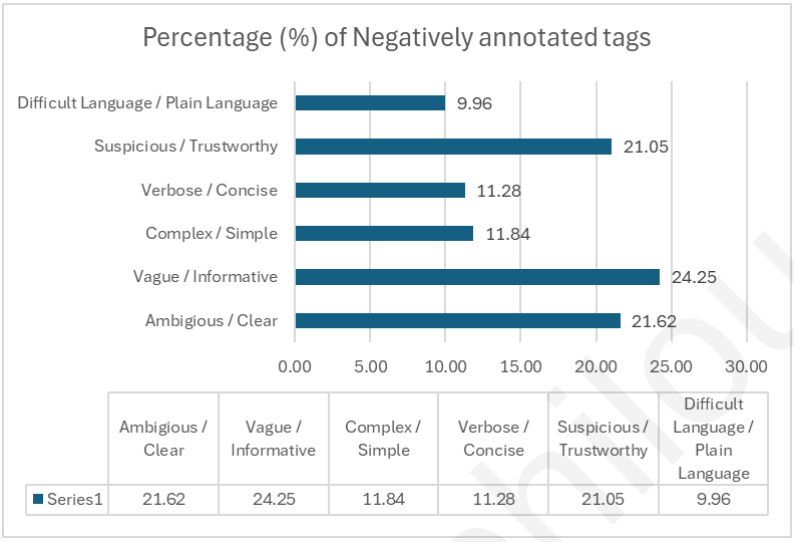
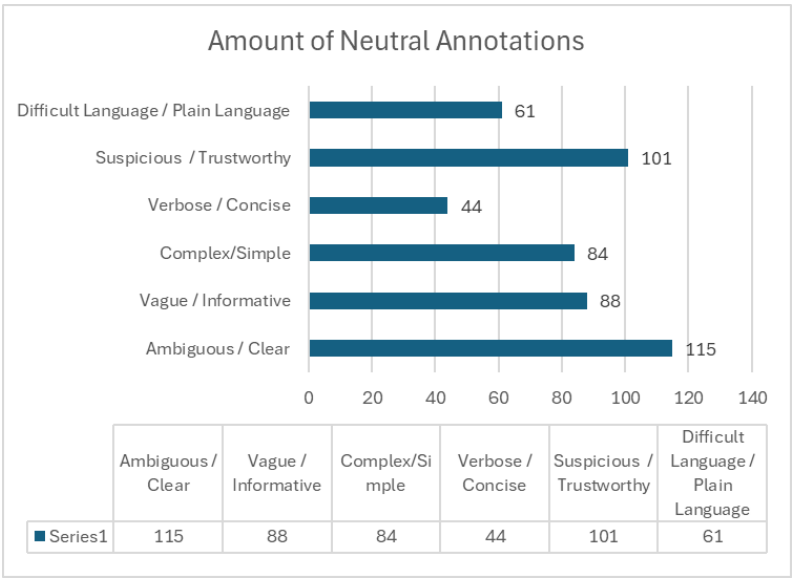


Figure 36 describes the percentages of Negatively Annotated tags. Vagueness being 24.25% means that of all negative annotations, the 24.25% were about Vagueness.

4.4.4 Neutral Assessments per Annotation

Neutral annotations show uncertainty in comprehension. So, they are as important as negative annotations, since the general purpose of privacy policies are to inform and be transparent to the end user.

Figure 37 Comprehensibility Analysis - Neutral Assessments per Annotation / Amount



Many neutral assessments appear for Ambiguity, Suspiciousness, Vagueness , Complexity , Language difficulty and Verbosity, listed in descending order of neutrality.

Figure 38 Comprehensibility Analysis - Neutral Assessments per Annotation / Percentage

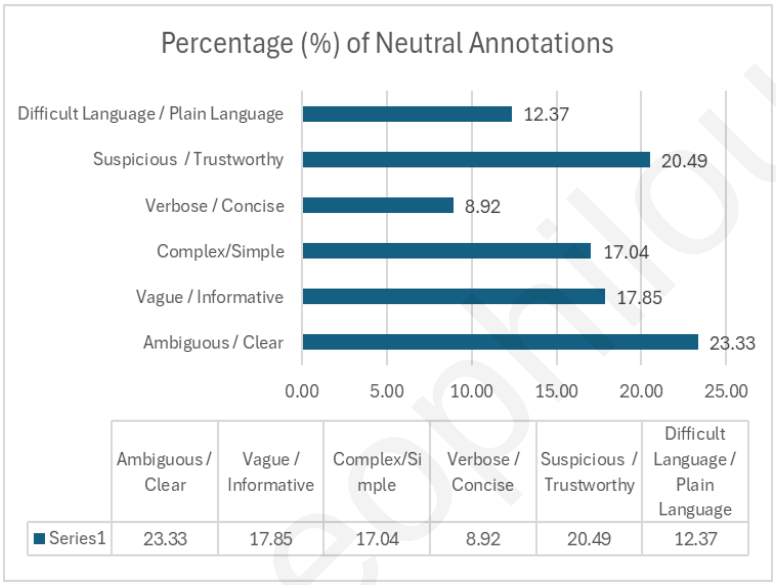


Figure 38 describes the percentages of Neutrally Annotated tags. Ambiguity being 23.33% means that of all neutral annotations, the 23.33% were about ambiguity.

4.5 Conclusions

In conclusion, the examination of Privacy Policy annotation, in terms of comprehensibility, reveals several key findings. Firstly, the inclusion of a diverse group of participants across various age demographics ensures a comprehensive assessment. Secondly, with a significant number of annotations (1098), the data pool is extensive, offering rich insights. Thirdly, it seems to be a relatively even distribution of annotations across different platform types, indicating a balanced representation. However, the presence of negative annotations highlights the urgent need for effective solutions to address comprehension challenges. Furthermore, neutral annotations show the significance of clarity, as comprehension ideally should not hinge on uncertainty but rather on clear, binary understanding.

Some other results were that the most ambiguous, vague, and suspicious platform type are social media platforms by 46% , 46.7% and 40.20% respectively. Also, the most complex and verbose

platform type appears to be the Online advertising by 50.9% and 50% respectively. Lastly the platform type with the most difficult language was proven to be e-commerce by 44.64%. Since these results are supplementary, their graphs and tables are presented in ANNEX I.

It can be observed that it is crucial to find the solution in order in creating comprehensive Privacy policies. From the literature review we can adopt the idea of different visually pleasing approaches in doing this. Since usability is a huge aspect of user experience, combining usability and visualization would be the most prefered solution. The next chapter discusses the solution to this problem along with its evaluation by its end users.

Chapter 5 : AlRíght – An easy-to-use tool for generating Comprehensible Privacy Policies

AlRíght – An easy-to-use tool for generating Comprehensible Privacy Policies

5.1 Introduction

After conducting a thorough survey on comprehension analysis in the previous chapter, it was concluded that the need for a comprehensible Privacy Policies is crucial. In this chapter a new tool is introduced as the proposed solution in creating comprehensible Privacy Policies. The purpose, the related work and requirements engineering procedures are shown in the chapters below, along with an extensive User Manual showcasing the tool.

5.2 Purpose / Need

In this chapter, the AlRíght privacy policies generator is introduced. This tool is an easy-to-use tool that covers two target groups. The indirect target group are Policy Makers whose purpose is to incorporate the comprehensible privacy policy in their actual platforms. The Policy Makers that will find the tool useful are the ones who are responsible for creating policies for Software Engineering Platforms. The direct target group are the end-users of said platforms, since the purpose is to incorporate the comprehensible privacy policy in order to assure their understanding of it.

In the OhKéy Privacy Policies assessment tool, participants were asked to tag texts from Policies as pairs rated in a Likert scale. These pairs were words selected to assess comprehension. Each pair included a positive and a negative characterization. In the Comprehensibility Analysis it was shown that many negative annotations exist, and many neutrals show uncertainty although it was stated that comprehension should be binary. In the

AlRíght policy generation tool, we decided to address all negative characteristics in a visually usable way through the platform.

It is crucial to address all negative characteristics since they are all important. The characteristics addressed are : ‘Vagueness’, ‘Language Difficulty’, ‘Verbosity’, ‘Complexity’, ‘Ambiguity’ and ‘Suspiciousness’. To address all these, we added functionalities and visual aids in the Privacy Policies.

To overcome Verbosity the Policy Maker is obliged to insert a summary of the text regarding a specific GDPR right. To address Language difficulty, there are two panels in the platform, that present useful terminology of technical words and legal words along with search engines for each. To address Ambiguity and Vagueness, Policy Makers are asked to answer some questions. The answers to the questions are shown to the end user through visual aids as well. An option to redirect the user to the corresponding GDPR article is also available for this purpose. Suspiciousness though is not very easy to address, because each website has their own rules and regulations. To address Complexity , the user has the option to see an information video. To overcome suspiciousness, we present the user with a warning that suggests they do not consent if they read something they believe is suspicious. In the next chapters these functionalities will be further discussed.

5.3 Related Work

Before suggesting the idea presented above, some related work was overviewed in order to identify gaps and make sure the proposed solution is innovative. First of all, lets review which GDPR rights are related to Software and Web Platforms. The first GDPR right regarding Software is the ‘Right to Access’. The Right to Access says that individuals have the right to know if their personal data are being processed, and if they are, they have the right to access these data and verify the lawfulness the of the processing [41]. The second right regarding software is the Right to Rectification. The Right to Rectification says that the user has the right

to request correction on their data held by a data controller [42]. The third one is the Right to Erasure, or the Right to be Forgotten, where users can delete their account without due delay [43]. The Right to Restrict Processing says that users can request to stop the processing of their data for a duration of time until they lift this request [44]. Another GDPR right regarding software is the Data Portability, that describes the ability of the user to download their data in a machine-readable format in order to transmit them to another controller [45]. After comes the right to Object, where user can opt out from any processing procedures done by the controller [46]. Lastly, the Right to Information, is the right that ensures transparency between the platform and its end users [47]. We decided to include these GDPR rights in the Policy Maker's panel, since they are related to Software and Web platforms.

According to [48] the goal of Privacy Policies is to inform the user and enhance transparency although they are purposely long and complex. The authors here had created a tool that is a web browser extension and presents users with visual explanations to privacy-related information. In [49] it is mentioned that users sometimes avoid online services due to privacy concerns. To enhance user's trust, transparency of privacy mechanisms are needed. They found many Transparency Enhancing tools and they overviewed them. They categorized them in terms of transparency. In the Transparency in Intended Data Collection category, they reviewed : Mozilla Privacy Icons, Privacy Bird and Privacyscore. In terms of Transparency Collected/ Stored Data, they reviewed : PrimeLife Privacy Dashboard and Google Dashboard. Regarding Third Party Tracking, they reviewed: Collusion and Netograph. They have also reviewed Web of Trust (WOT) and in terms of Awareness Promotion they reviewed the following tools : Me & My Shadow, Priveazy, Firesheep, Panopticlick , and Creepy. Enhancing transparency though does not enhance comprehension. Lastly, in [50] the objective was to extract information from Privacy Policies as 'sequence-labeling problem'. The approach was to create a large dataset of sentences from 30 different real-world policies. They created PI-Extract, an automated system that uses NLP to extract privacy practices. The results were very positive since it was shown

that they improved readability by 26%, which highlighted the significance in enhancing users' comprehension.

In conclusion, different approaches exist in trying to enhance comprehensibility, although none was found to exist with the same target groups and goals as the proposed solution. Hence the solution is considered to be innovative.

5.4 Requirements Engineering

5.4.1 Tool Specification

The tool specifications are shown in the table below:

Table 9 AlRíght - Tool Specifications

Specification	Purpose	Details
Home page	The homepage is necessary to present the purpose of the tool, and to briefly describe the contents of it.	- Description of tool and its purpose - Consent form. - "START" button that redirects the user to the first page of the tool
Policy Maker's Panel or the 'GDPR' page	The purpose of this page is to create a panel where Policy Makers insert the text of their Privacy Policies regarding the seven GDPR right about Software and Web Services.	- The panel must have a warning button that informs the user that in order to be fully compliant with the GDPR, they must include all necessary fields for each GDPR right
		- 'Accordion'-like structure - Clickable buttons for each GDPR right
		- For each GDPR right, the policy maker must include : ✓ The actual text ✓ The summary of the text (max 100 words) ✓ Important Questions answering - Each of the above will be a step in the completion procedure, meaning that in order to go to the next step, the previous must be completed.
		- The questions that will be presented to the policy maker will include, Yes or No questions, open questions, or multiple choice. - They will help users in understanding the privacy policy better.
		- 'GENERATE' Button that redirects the user to the new representation of privacy policies.
Enhanced Policy Page	The purpose of this policy is to provide the policy maker with a new	- Translate button. - The Google Translate API will be used in order to provide the option for automatic translation of the whole page.

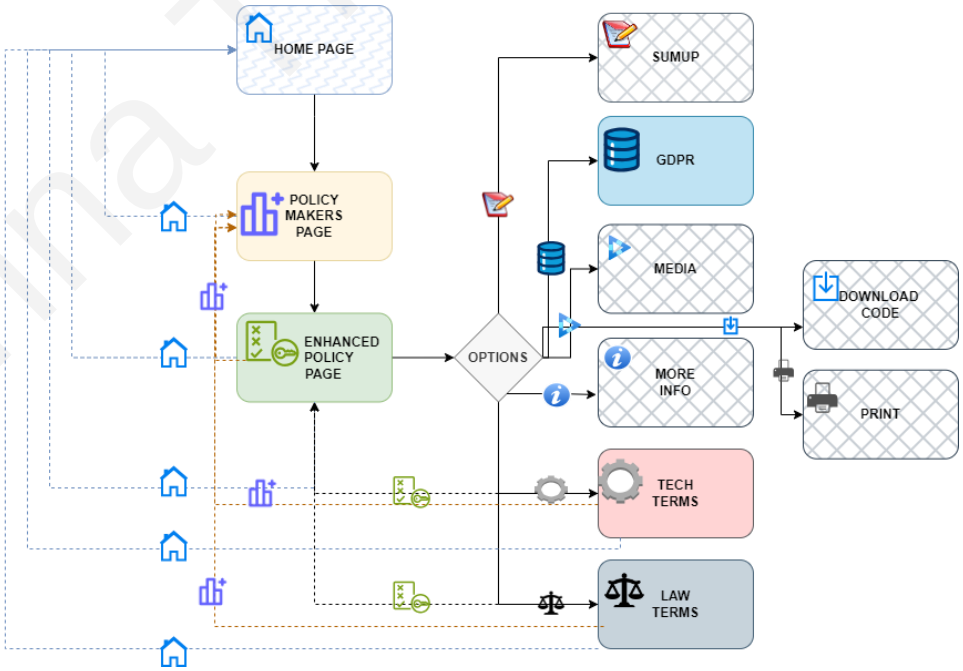
	representation of their policy, in order use it in their actual website. The most important purpose of this page though is for the platform's end users to use this new representation in order to understand it better.	
		- 'TECH TERMS' Button - This button will help in addressing 'Language Difficulty' because non tech users may read the policies and therefore, they may not understand some terminology
		- 'LAW TERMS' Button - This button will help in addressing 'Language Difficulty' because users with no legal background may read the policies and therefore, they may not understand some terminology.
		- 'WARNING' button - Encouraging the user to reject the policy or change their setting, or even contact the person responsible if they find something suspicious in the policy. - This will help in addressing 'Suspiciousness'.
	GDPR Rights	- For each GDPR right, the actual text will be presented along with four extra options
		- Option 1 : 'SUMUP' Button - This button will help in addressing 'Verbosity' - This page will present the summary that the policy maker inserted in the previous panel
		- Option 2 : 'GDPR' Button - This button will redirect the user to the corresponding GDPR article. - With this button we will address 'Vagueness' and 'Ambiguity'
		- Option 3: 'MEDIA' Button - This button will popup an explanatory video for the corresponding GDPR right. - By including this, we will overcome 'Complexity'.
		- Option 4: 'MORE INFO' button - This is where the answered questions from the Policy Maker will appear. - The answered questions may be used by the end-users in finding out important information like where the account panel of the platform is. - With this option we also address Vagueness and Ambiguity.
	Downloading	- 'DOWNLOAD CODE' button - This button will let the policy maker download the code of the GDPR page and incorporate it to their actual platform.
	Printing	- The Policy Maker will let users print the page or save it as a PDF file. - This will be done through the 'PRINT' button
Technical Terminology Page	This page, helps in addressing Language Difficulty as described above.	- In the OhKéy platform we provided users with 9 Privacy policies. For this page, we will collect technical words that appear in these policies and present them in a table-like structure. - The table will be scrollable, and it will provide a small description for each terminology.

		- Search engine. - There will be a search engine above the table in order to make the searching procedure easier.
Legal Terminology Page	This page, also helps in addressing Language Difficulty as described above.	- In the OhKéy platform we provided users with 9 Privacy policies. For this page, we will collect legal words that appear in these policies and present them in a table-like structure. - The table will be scrollable, and it will provide a small description for each terminology.
		- Search engine. - There will be a search engine above the table in order to make the searching procedure easier.

5.4.2 Architecture

The general architecture of this platform is the same as the one used in OhKéy described in Chapter 3. There is no Back-end architecture here and the Front-end architecture is different since they are two different platforms. Therefore, in this section the Front-end architecture will be described.

Figure 39 AlRíght - Front End Architecture



The pages in the architecture are presented in solid colors and the multiple functionalities that do not redirect to a new page are presented with a grey striped outline. The user first starts with the Home page and can navigate to the Policy Maker's page. Then when they press Generate, they can then go to the Enhanced Policy Page where they can see all eight options. The SUMUP

option is a popup, hence its in a striped grey outline. If the user presses the GDPR button a new window will appear with the corresponding article. Since this redirects to the EU's website and its not a part of the AlRíght platform, it can be seen that there is no backtracking from this page. The MEDIA option is another popup that shows informational video for the corresponding GDPR right. The MORE INFO popup shows important questions answered by the policy maker. This option helps in securing comprehension and understanding. Last but not least, the TECH TERMS and LAW TERMS pages are the ones that present users with lists of technical and legal terminologies respectively along with search engines for easier navigation. Finally the options DOWNLOAD and PRINT are the most important part of the architecture, since they allow policy makers in incorporating the Enhanced Policy in their platforms.

In the figure above, it can be seen that all pages can backtrack to the HOME page and the Policy Maker's page. This is because these two pages are always available through the navigation tab. Backtracking to the Home Page is represented in blue dashed lines. Backtracking to the Enhanced policy Page is represented in black dashed lines and finally, backtracking to the Policy Maker's page is presented in orange dashed lines.

To conclude, it is important to mention that the Technologies used for this implementation are the same as the ones used in the OhKéy platform.

5.5 Content Creation Methodology

As mentioned in the previous chapters, the platform asks the Policy Maker to answer questions regarding their platform strategies in some situations that relate to specific GDPR rights. The rights included in the platform are explained in the section above. Each GDPR right has different questions regarding the corresponding right. To create the questions regarding each right, many forums (like QUORA and Reddit) were studied in order to find frequently asked questions. The questions are not only from forums but common questions that occurred while

reading the actual articles of the regulations. All questions used in the platform for each type are presented in the tables below.

Table 10 AlRíght - Content Creation / Right to Access.

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can have access to their personal data?	Yes / No	ROOT
1.1	Is there a panel where the users can visit?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the user have access to their data? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can the users contact to obtain their personal data?	Short Text	Follow up – if NO in 1
2	What does the copy of the user's personal data include?	Checkboxes	ROOT
3	In which format can the data be downloaded? (eg. PDF, XML,ZIP etc)	Short Text	ROOT
4	What other supplementary information is provided in the copy?	Long Text	ROOT
5	Who can the users contact for any inconvenience in the copy they download?	Short Text	ROOT

It can be seen that there are four types of questions. Long Text, Short Text , Yes or No questions and Checkboxes.

Table 11 AlRíght - Content Creation / Right to Rectification.

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can request to modify their personal data?	Yes / No	ROOT
1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users request modifications in their data? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can the users contact to modify their personal data ?	Short Text	Follow up – if NO in 1
2	How long will the modification process take ?	Long Text	ROOT
3	Is everything modifiable ?	Short Text	ROOT
3.1	Which data is not and why ?	Long Text	Follow up – if NO in 3
4	Who can the users contact for any inconvenience ?	Long Text	ROOT

For each question, it is indicated whether they are the Root question of they are a Follow up question, meaning that they appear depending on the user's previous answer.

Table 12 AlRíght - Content Creation / Right to Erasure.

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can request to delete their account ?	Yes / No	ROOT
1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users request deletion of their data? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can the users contact to request a full data deletion ?	Short Text	Follow up – if NO in 1
2	How long is the data stored after deletion and why ? Please explain.	Long Text	ROOT
3	Are the data actually being deleted or are there any backups ?	They are permanently deleted / There are backups	ROOT
3.1	Is the user informed in the original text and the summary text about the backups?	Yes / No	Follow up – if backups in 3
3.2	Is there an option to delete the backups ?	Yes / No	Follow up – if backups in 3
3.2.1	For how long are the backups still stored?	Long Text	Follow up – if NO in 3.2

The methodologies described above are true for all tables that present the questions to the GDPR rights.

Table 13 AlRíght - Content Creation / Right to Restrict Processing.

Number	Question	Type	Follow up or root ?
1	Does your platform have a way were logged in users can limit the processing of their data ?	Yes / No	ROOT
1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users change the setting in their accounts in order to limit the processing of their data? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can the users contact to restrict the processing of their personal data?	Short Text	Follow up – if NO in 1

2	Should the users specify the reason of requesting the restriction of data?	Yes / No	ROOT
2.1	How can they do that ?	Short Text	Follow up – if YES in 2

Table 14 AlRíght - Content Creation / Right to Data Portability.

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can download their data in a format compatible in most online platforms ?	Yes / No	ROOT
1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users download that ? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Where can the users find this option?	Short Text	Follow up – if NO in 1
2	In which structure/format can the users download their data ?	Long Text	ROOT
3	Which personal data are included in the downloaded document ?	Long Text	ROOT
4	After how long will the user receive a copy of their data?	Long Text	ROOT

Table 15 AlRíght - Content Creation / Right to Object

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can object to their data being processed ?	Yes / No	ROOT
1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users object to that if there is not a panel? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can they contact in order to find this option?	Short Text	Follow up – if NO in 1
2	When does the right to object apply immediately? What are the circumstances?	Long Text	ROOT

Table 16 AlRíght - Content Creation / Right to Information

Number	Question	Type	Follow up or root ?
1	Does your platform have a way where logged in users can learn which of their personal data are used in the platform and in which way?	Yes / No	ROOT

1.1	Is there a panel where the users can visit ?	Yes/ No	Follow up – if YES in 1
1.1.1	Can you please provide the link?	Yes / No	Follow up – if YES in 1.1
1.1.2	How can the users learn about that? Please explain.	Long Text	Follow up – if NO in 1.1
1.2	Who can they contact in order to find out which of their personal data are used in the platform?	Short Text	Follow up – if NO in 1
2	Which of the information below will be displayed in the informational message that the user will receive if needed ?	Checkboxes	ROOT
3	How is the controller going to inform the user about any changes on the usage of their personal data ? (eg. via email , via the platform , written letter etc.)	Long Text	ROOT
4	Who can the users contact for any inconvenience ?	Long Text	ROOT

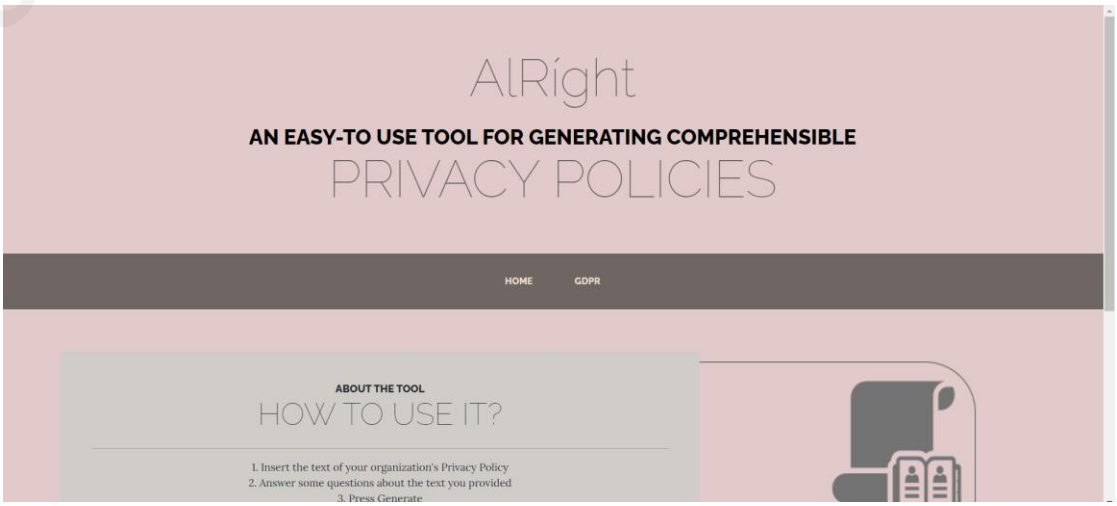
In this section, the content methodology was presented. For each GDPR right there was a corresponding table displaying the questions that the Policy Maker will be asked to answer in their panel. In the chapter below, the platform will be showcased in detail.

5.6 User Manual

5.6.1 Home Page

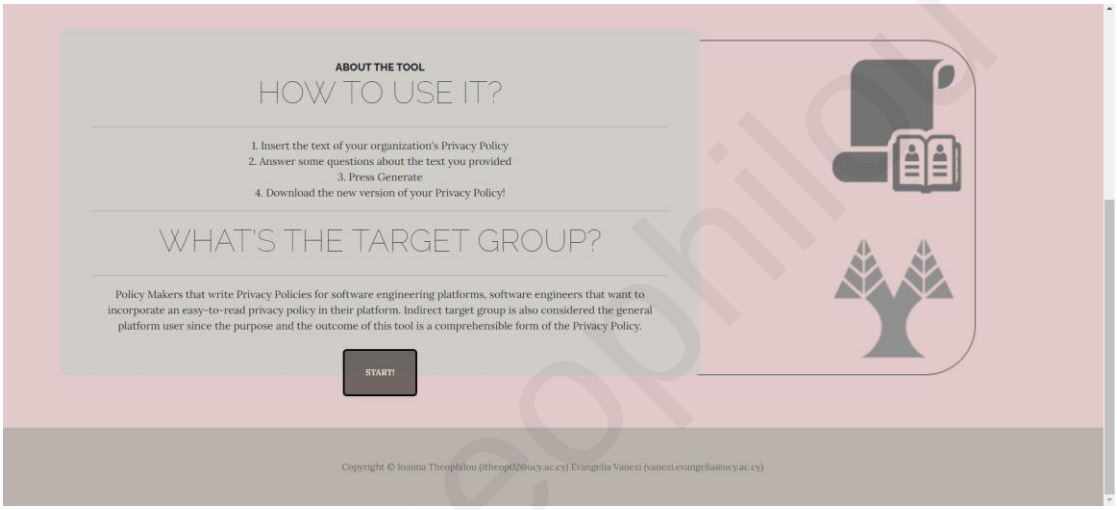
In this section of the dissertation, the AlRíght platform will be showcased. The platform is online and its available here.

Figure 40 AlRíght - Home Page part 1/2.



The tool specification presented in the previous section was followed. In the Home Page, there is an explanation of the purposes of the platform and its target groups.

Figure 41 AlRíght - Home Page part 2/2.

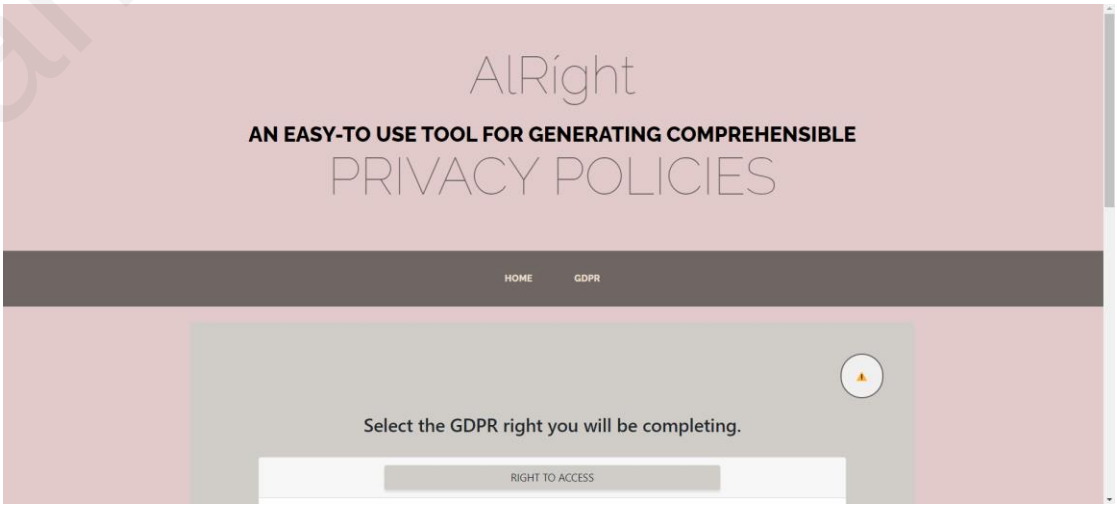


To navigate to the next page, the user must press START or use the GDPR button on the navigation menu.

5.6.2 Policy Make's Page

5.6.2.1 General User Interface

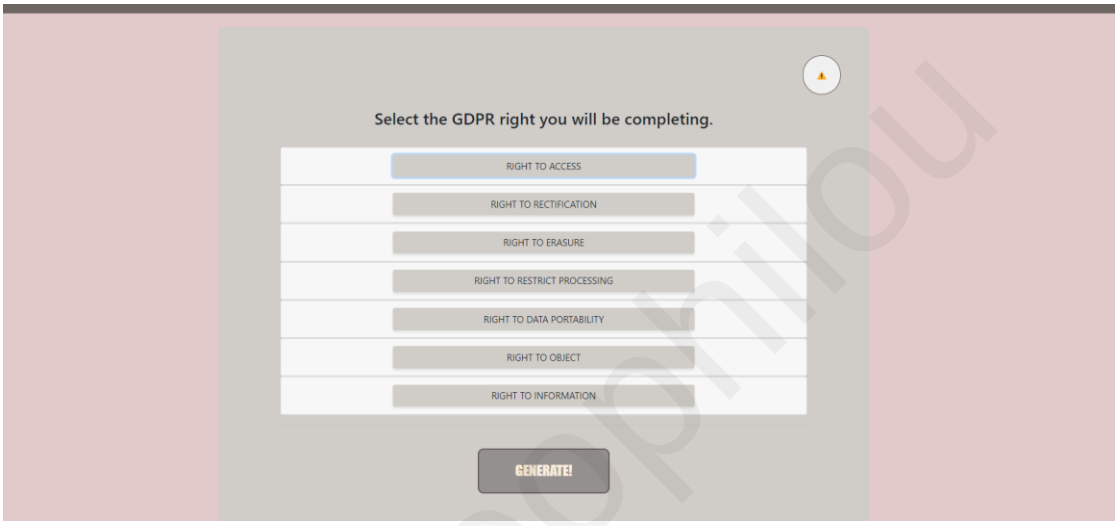
Figure 42 AlRíght - General UI part 1/3.



The first page that the user sees is this 'Accordion' like structure. The structure presents all GDPR rights related to Software and Web Applications in the form of clickable boxes. When

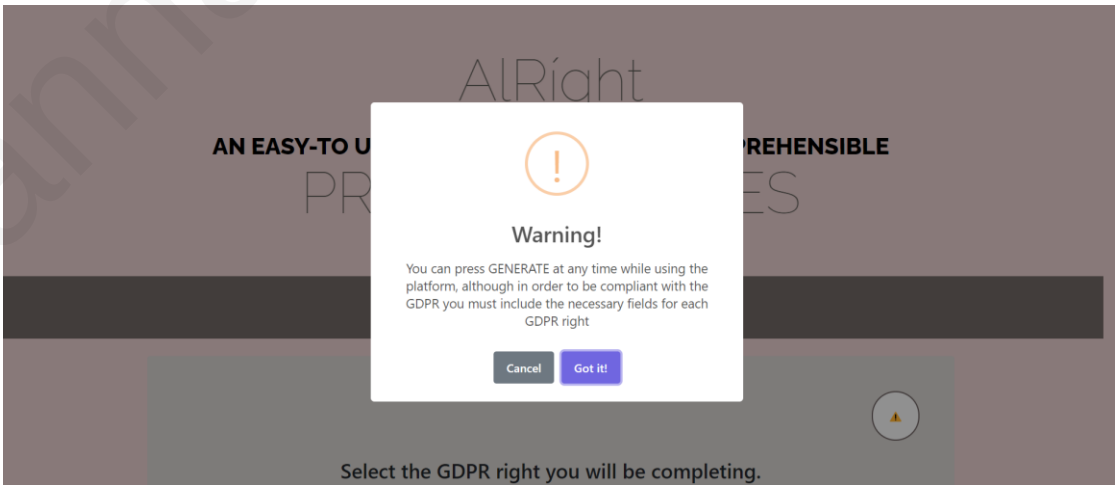
clicked, each GDPR requires the actual policy text regarding the specific right, and its summary. After that, a series of questions are presented for the Policy maker to answer.

Figure 43 AlRíght - General UI part 2/3.



In the figure above, the 'Accordion' like structure is presented, along with the GENERATE button that indicates that the policy maker has completed all fields and is ready to preview the Enhanced Policy generated.

Figure 44 AlRíght - General UI part 3/3.

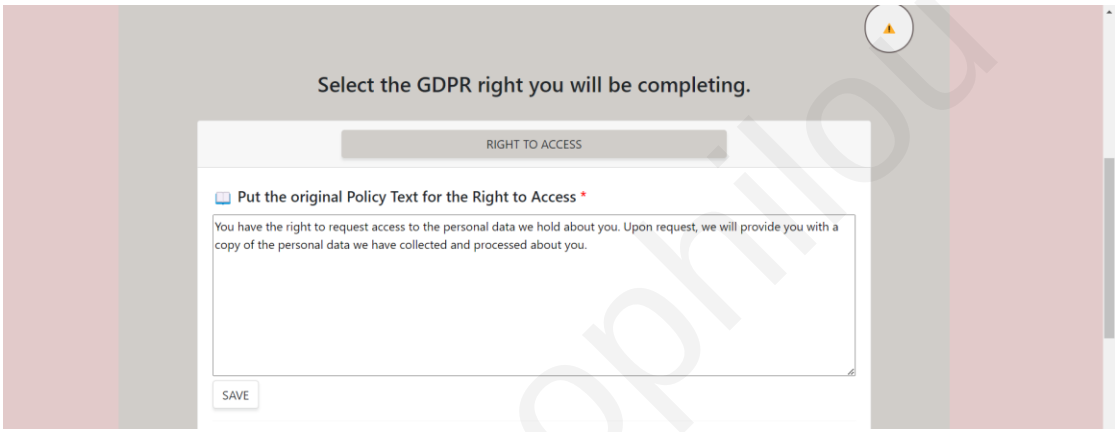


There are cases where the user may have already used the tool and completed all GDPR rights, but in the case of changes they may want to press generate with the part they would like to modify. For this reason, the warning button informs the Policy Maker that they can use the generate button even if not all fields are completed, although in order to be GPDR compliant,

they are all necessary. In the subsections below, we will analyze the questions and interfaces for each GDPR right.

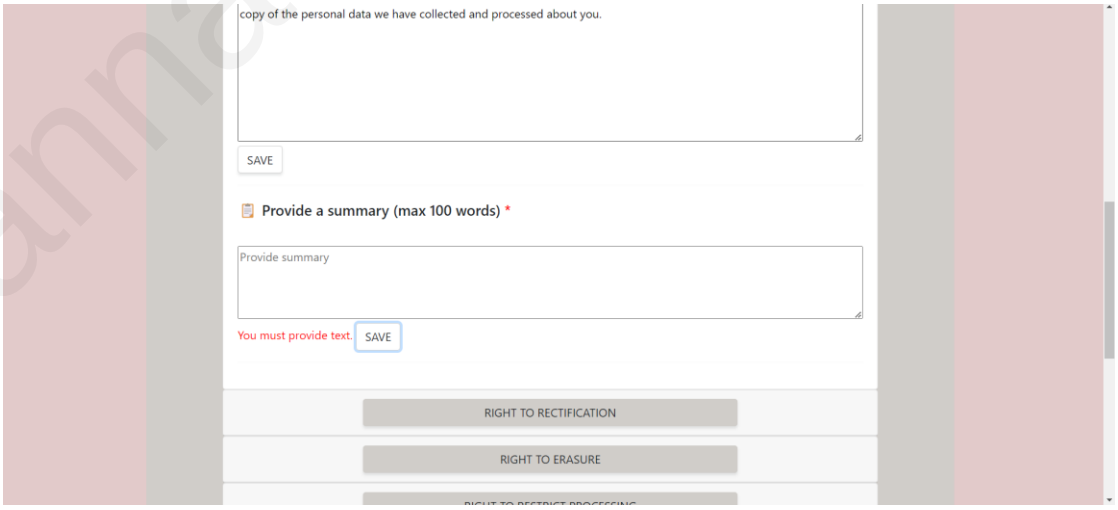
5.6.2.2 GDPR : Right to Access

Figure 45 AlRíght - Right to Access part 1/6.



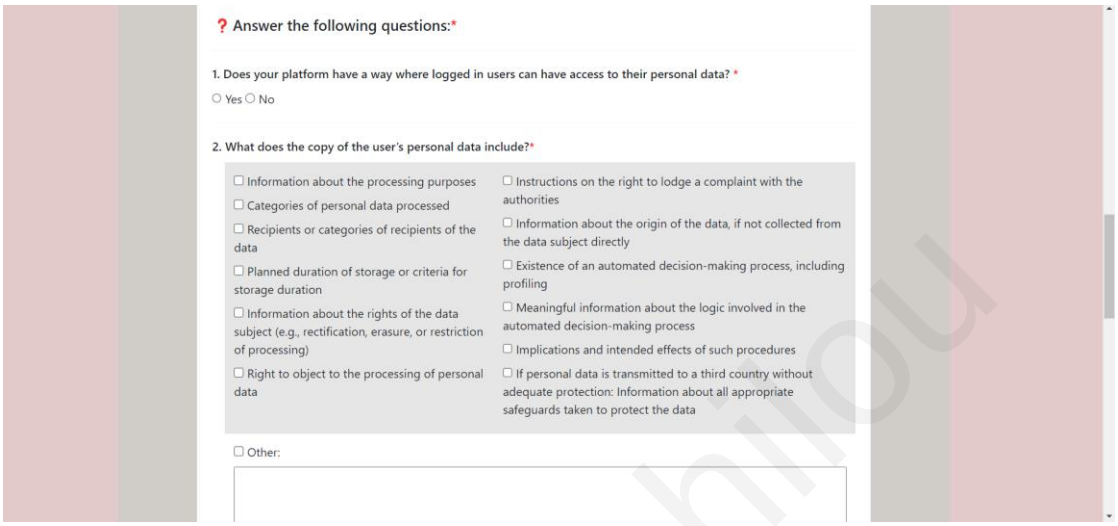
In this figure, it can be seen that the first part in completing the Policy Maker's panel is to insert the actual text of the corresponding GPDR Right. When pressing SAVE the next step will popup asking the user to provide the summary of said text.

Figure 46 AlRíght - Right to Access part 2/6.



The user must provide the summary of the actual text of the privacy policy. It is obligatory to include the text, and as written in the title, the summary should not be more than 100 words long. When the user presses SAVE, a series of questions will pop up.

Figure 47 AlRíght - Right to Access part 3/6.



This is the empty form of the questions panel. The options presented in the second question, were gathered from the corresponding article [41]. By presenting policy makers with all options we guide them to answer the questionnaire.

Figure 48 AlRíght - Right to Access part 4/6.

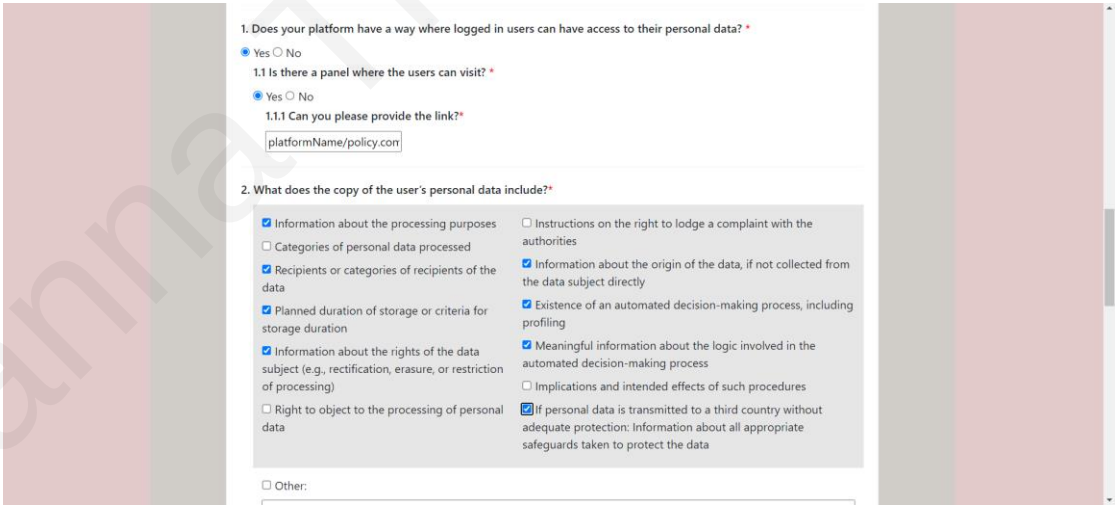
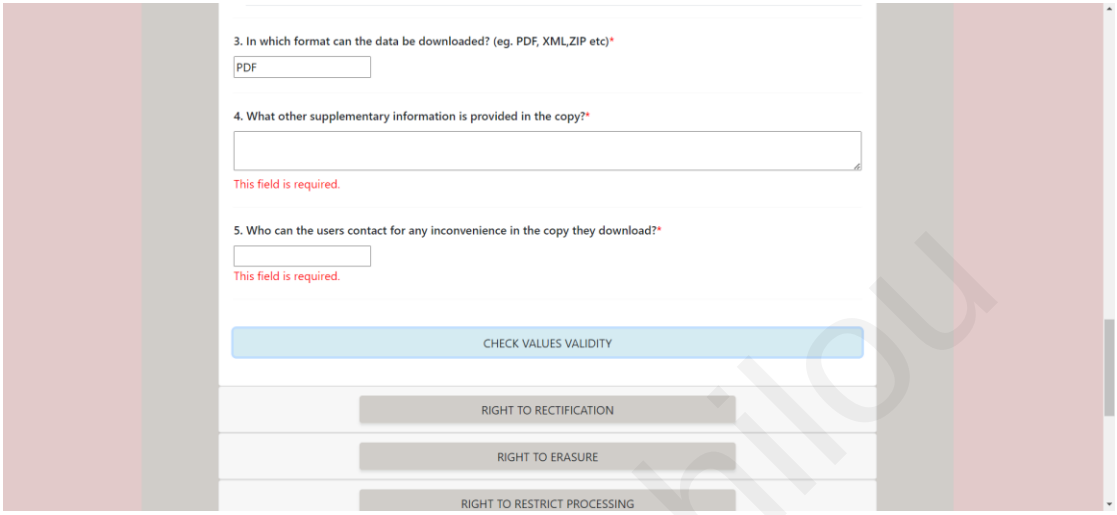


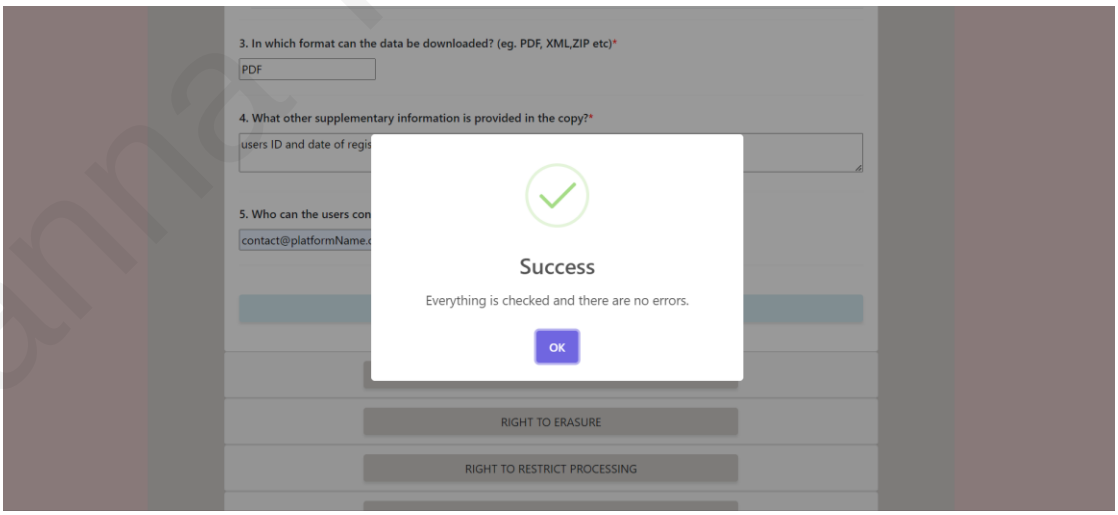
Figure 48 shows the completed form by the user.

Figure 49 AlRíght - Right to Access part 5/6.



The 'Check Values Validity' button can be seen. This button is available for all GDPR rights. It can be used by the user in order to check the validity before pressing GENERATE. It can be seen that the user in figure 49, forgot to answer questions 4 and 5. This button reminds them to do so.

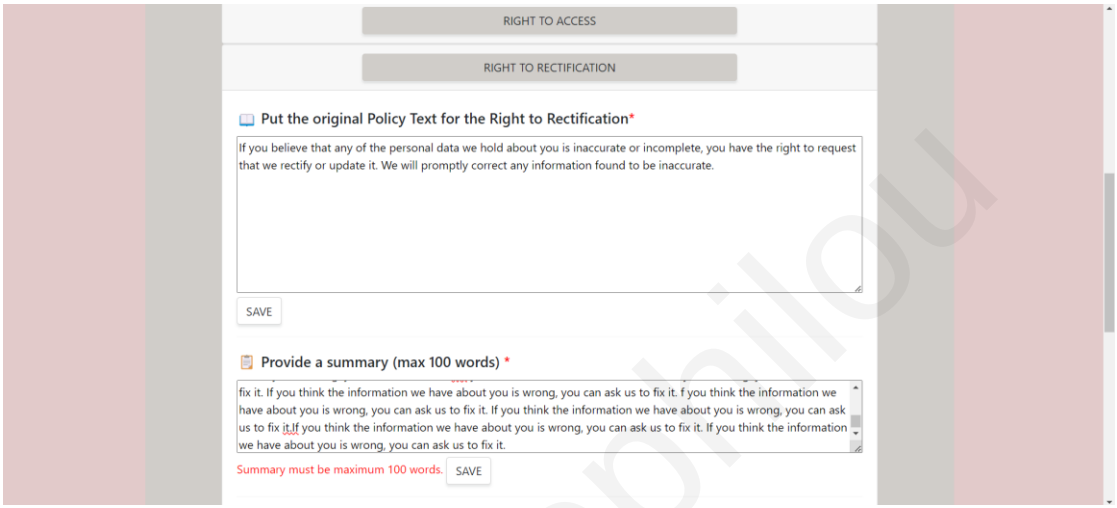
Figure 50 AlRíght - Right to Access part 6/6.



When all questions are answered, the 'Check values validity' button, shows a success message to let the user know.

5.6.2.3 GDPR : Right to Rectification

Figure 51 AlRíght - Right to Rectification part 1/3.



It can be seen in Figure 51, that when the user provides a summary bigger than 100 words, they cannot proceed to the next step. This is implemented in all summary fields for all GDPR rights.

Figure 52 AlRíght - Right to Rectification part 2/3.

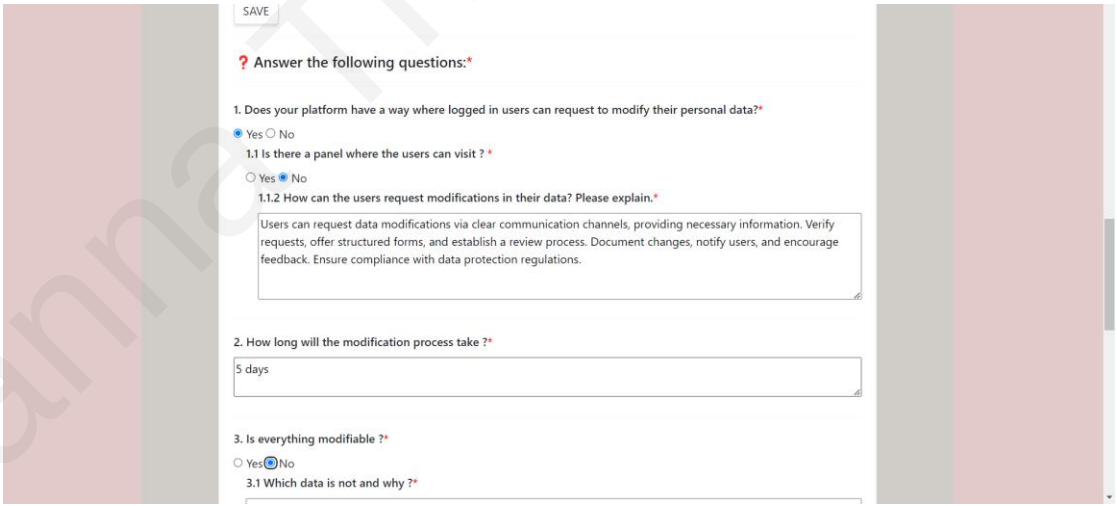
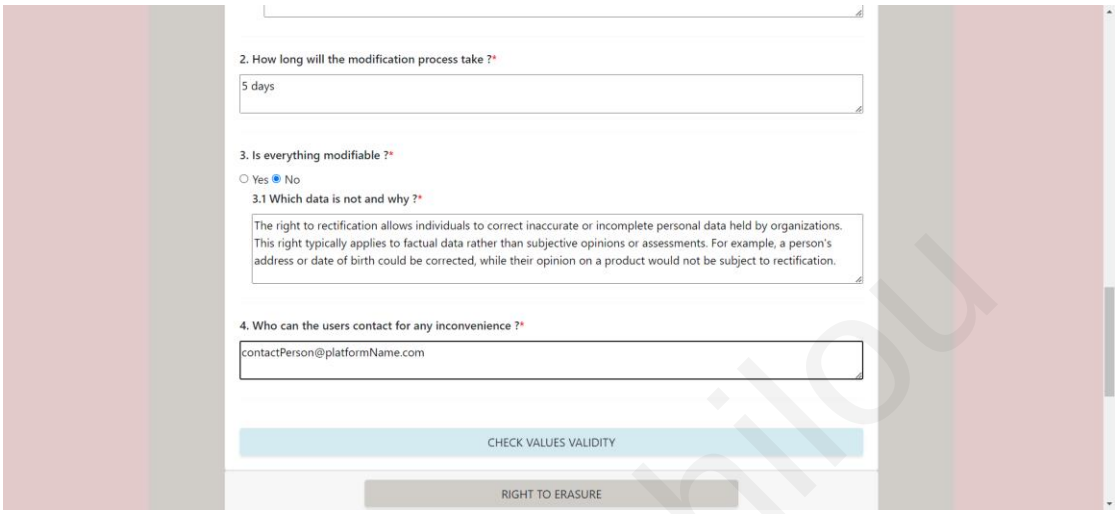


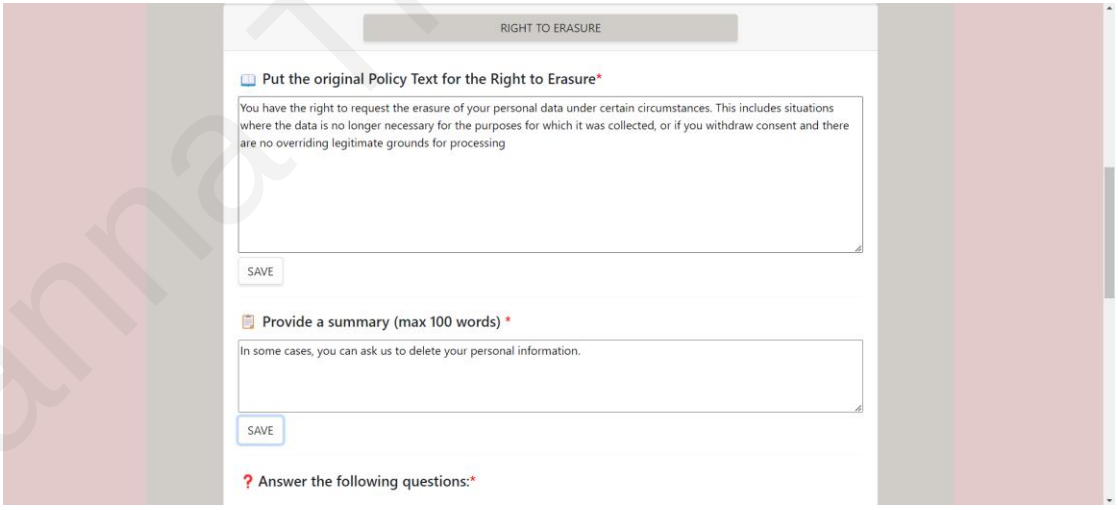
Figure 53 AlRíght - Right to Rectification part 3/3.



In Figures 52 and 53, it is seen that the user answers the questions presented to them regarding the Right to Rectification.

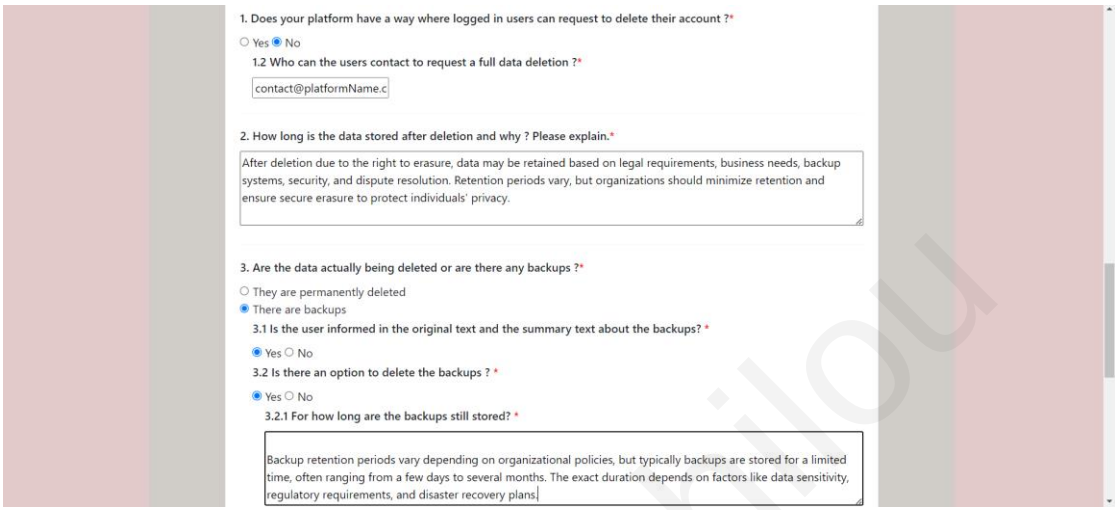
5.6.2.4 GDPR : Right to Erasure

Figure 54 AlRíght - Right to Erasure part 1/2.



Here as well, the user inputs the information in the fields asked. It can be seen in the following figure as well.

Figure 55 AlRíght - Right to Erasure part 2/2.



The user responds to the questions and proceeds to the next rights.

5.6.2.5 GDPR : Right to Restriction of Processing

Reminding that the structure is 'Accordion' like and therefore, it can be seen that when a new GDPR right will be answered, the previous collapses. The answers though are saved and the user can see them if they press to the previously visited rights.

Figure 56 AlRíght - Right to Restriction of Processing part 1/2.

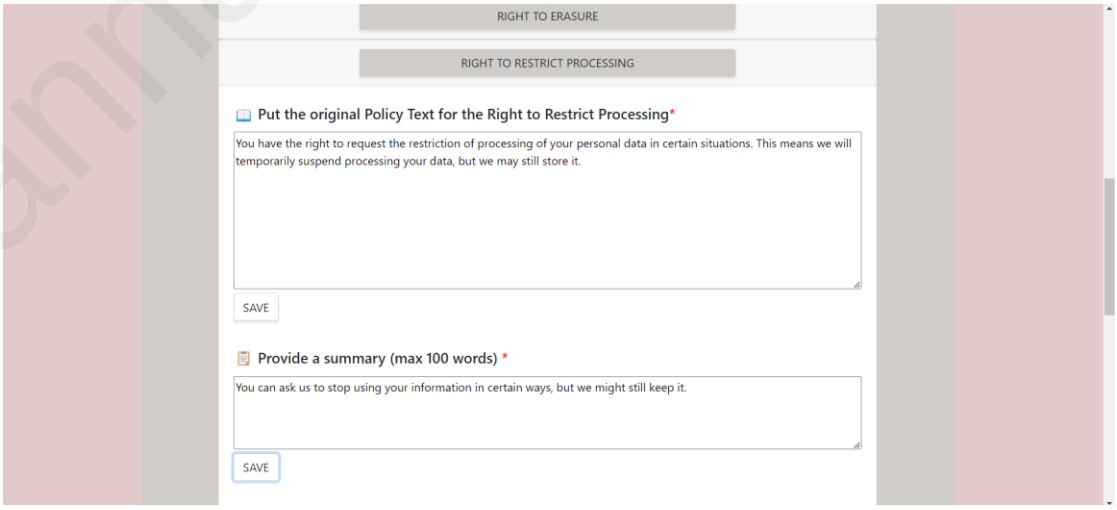
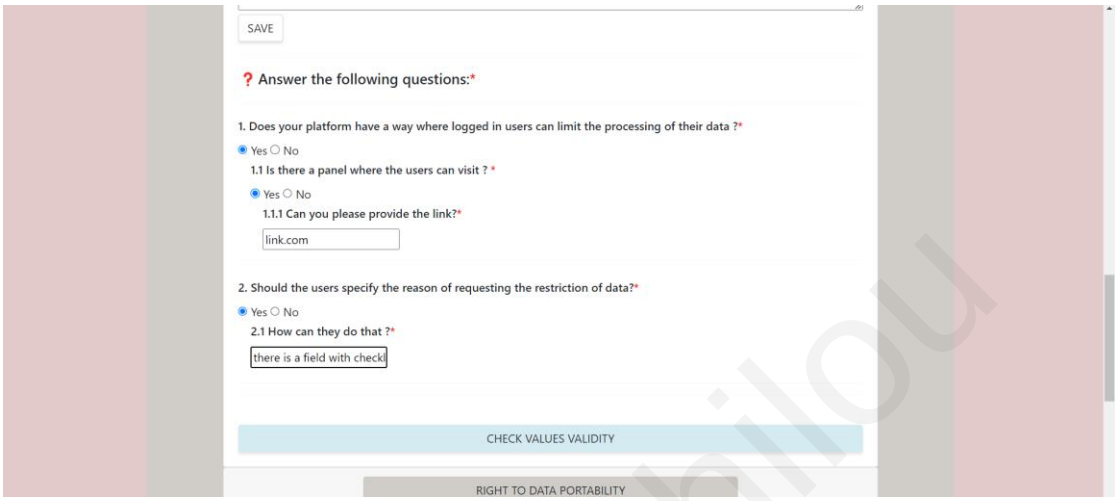


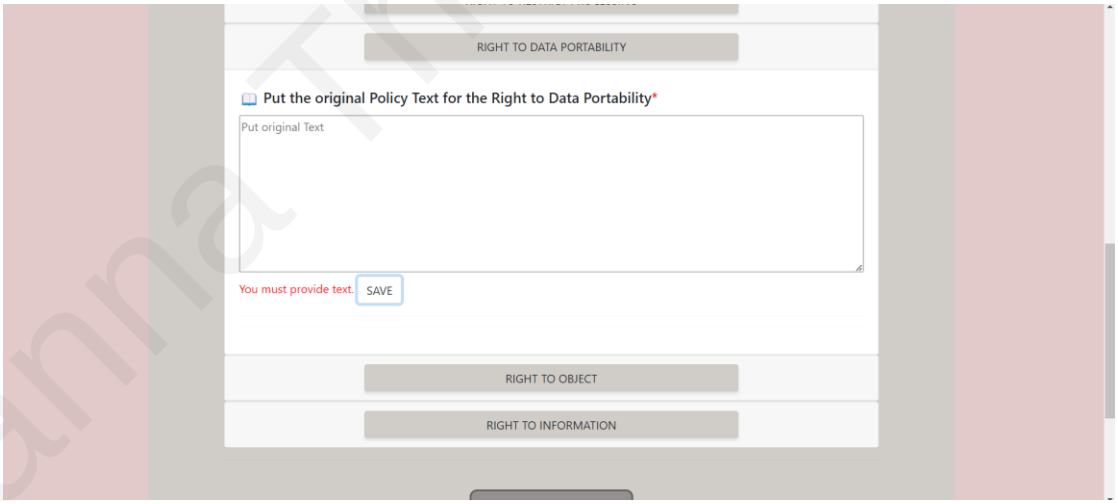
Figure 57 AlRíght - Right to Restriction of Processing part 2/2.



The user then continues to the next GDPR right.

5.6.2.6 GDPR : Right to Data Portability

Figure 58 AlRíght - Right to Data Portability part 1/2.



It can be seen in Figure 58 that when the original text of the Privacy Policy is not inserted, the user cannot move forward to the next step.

Figure 59 AlRíght - Right to Data Portability part 2/2.

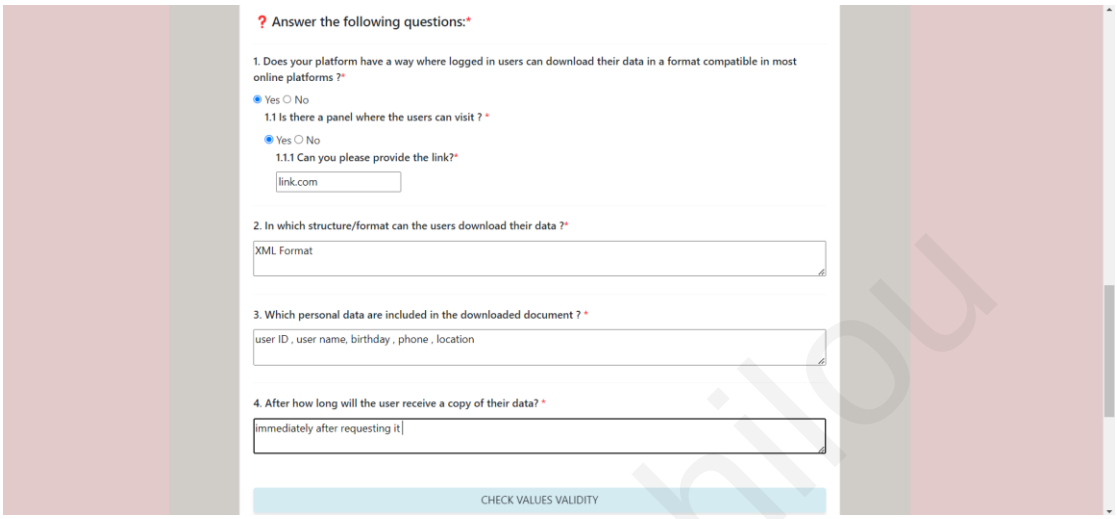


Figure 59 shows the questions that the Policy Maker had to answer for the Right to Data Portability.

5.6.2.7 GDPR : Right to Objection

Figure 60 AlRíght - Right to Object.

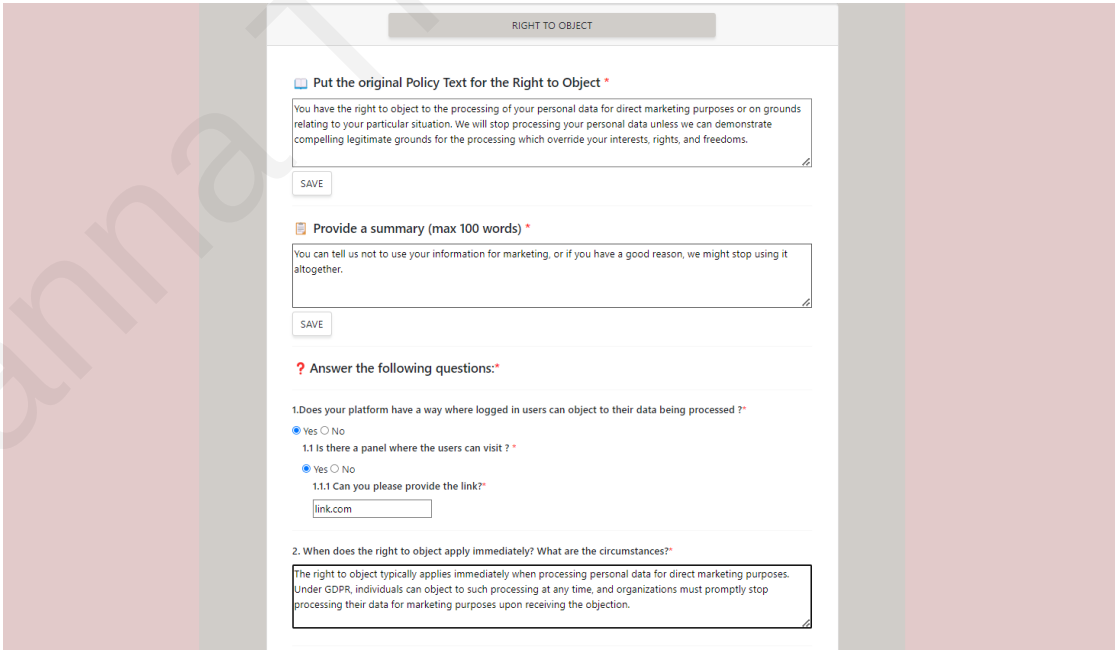


Figure 60 shows the whole panel of the Right to Objection. Here the user had to answer only two questions.

5.6.2.8 GDPR : Right to Information

Last but not least, the following figures show the interface for the right to information.

Figure 61 AlRíght - Right to Information part 1/3.

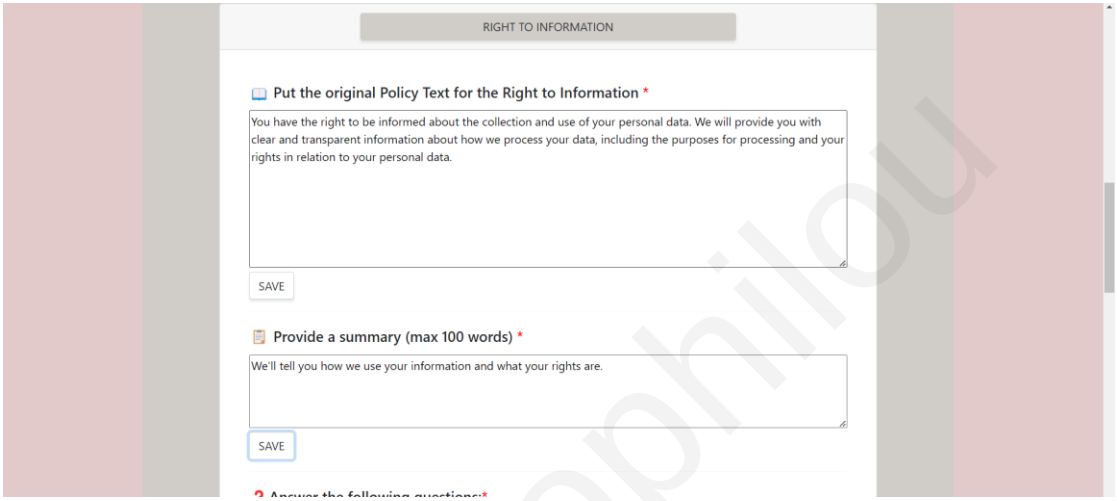
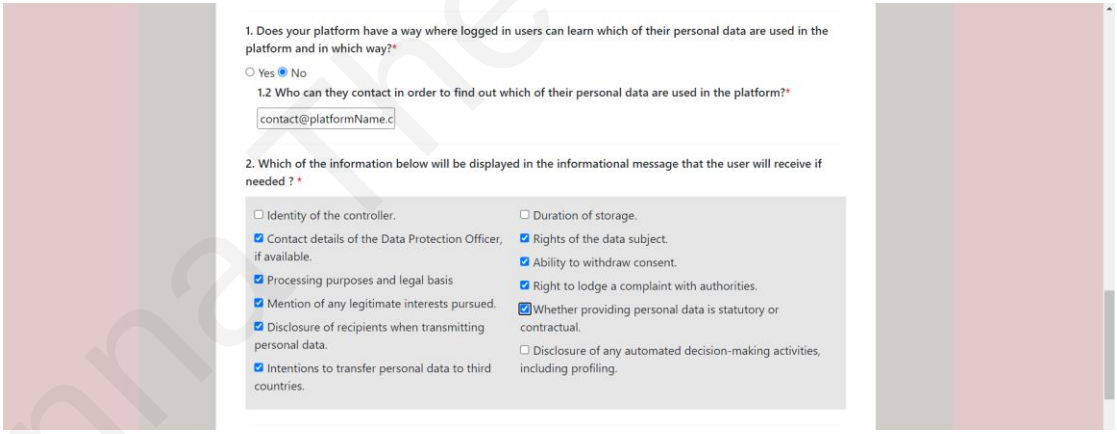
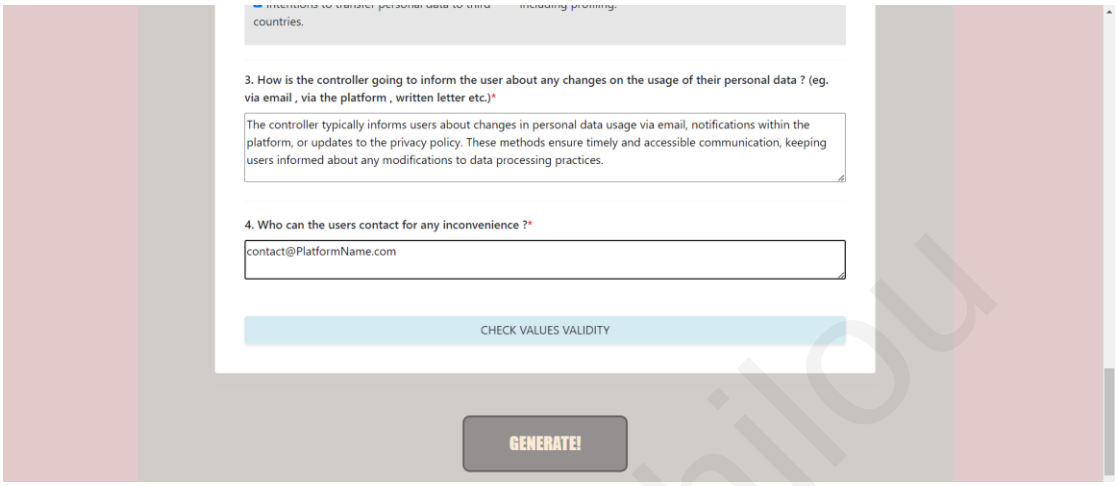


Figure 62 AlRíght - Right to Information part 2/3.



The options for the checkboxes here are also driven by the article of the corresponding right in order to guide the user [47].

Figure 63 AlRíght - Right to Information part 3/3.



The user then presses 'GENERATE' in order to navigate to the Enhanced Policy page.

5.6.3 Enhanced Privacy Policy

5.6.3.1 General User Interface

In the figure below, the Enhanced Policy interface will be shown. The goal for the Enhanced policy is to provide options to the end users that overcome comprehensibility issues. Each functionality addresses a different characteristic as described in the sections above.

Figure 64 AlRíght - Enhanced Policy / General UI part 1/7.

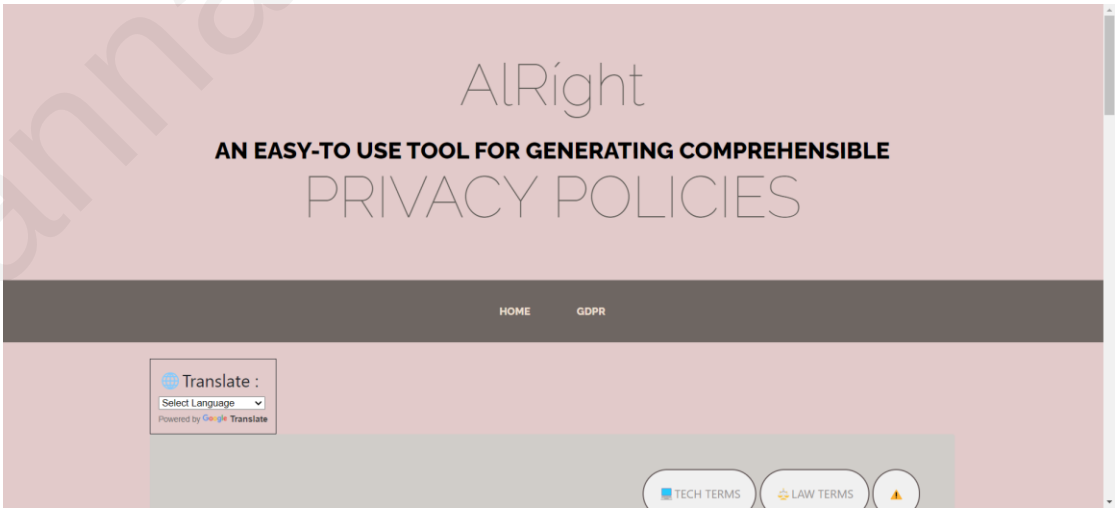
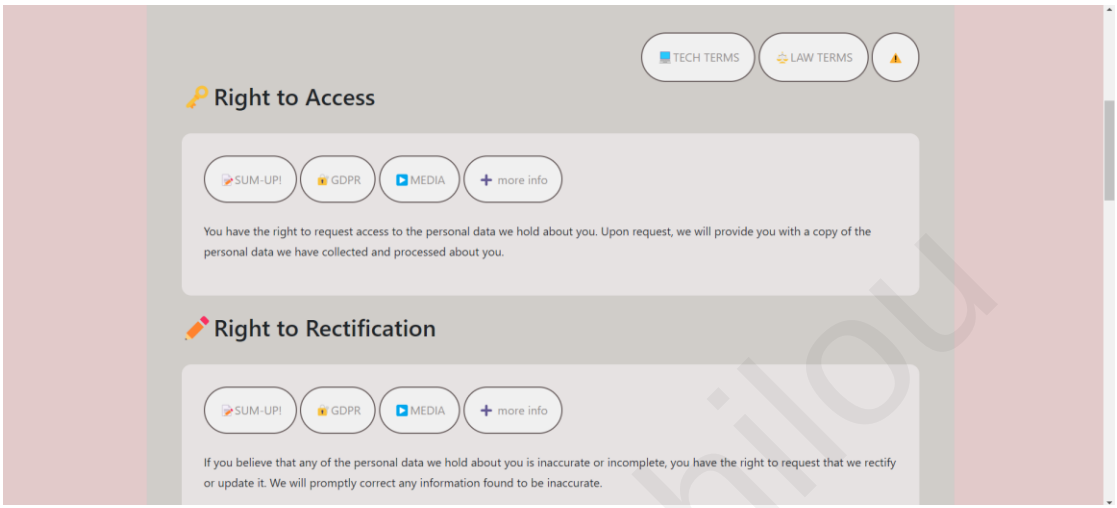


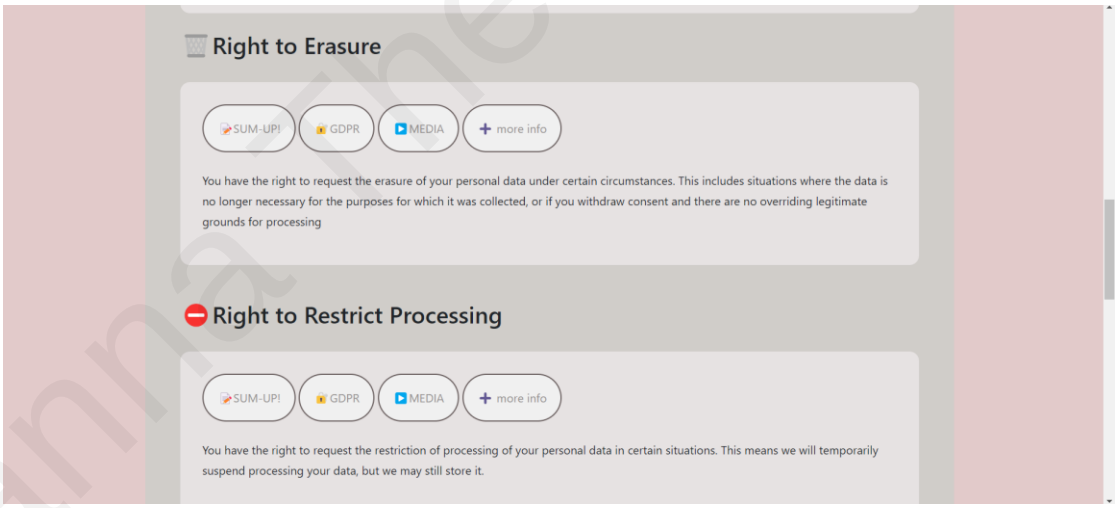
Figure 64 describes the first interaction with the page. It can be seen that the Home and GDPR pages are always available through the navigation menu.

Figure 65 AlRíght - Enhanced Policy / General UI part 2/7



With a first scroll, the user can see the text that they inserted in the previous page, under the corresponding GDPR right.

Figure 66 AlRíght - Enhanced Policy / General UI part 3/7.



The figure above shows the Right to Erasure and Restriction of Processing's original texts.

Figure 67 AlRíght - Enhanced Policy / General UI part 4/7.

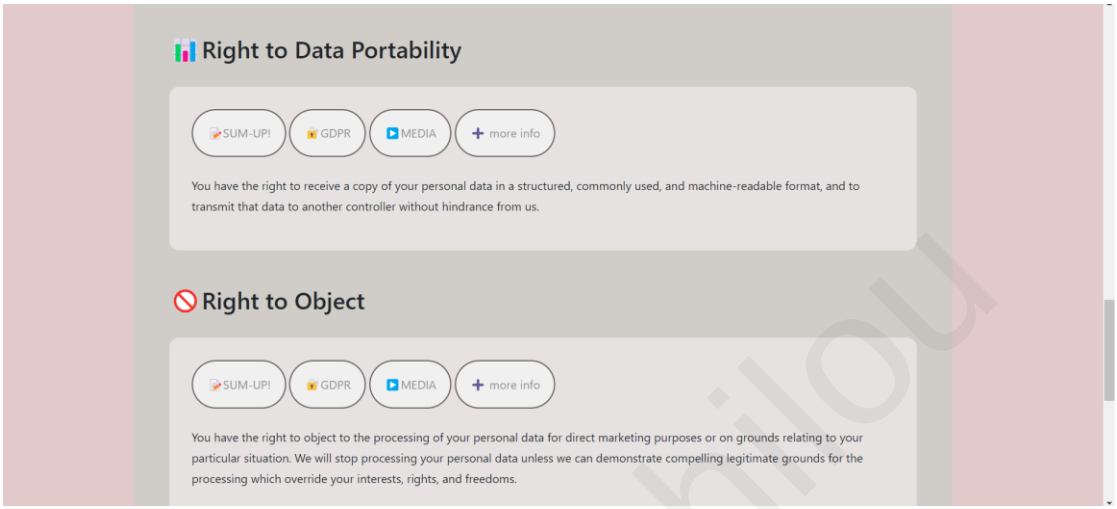
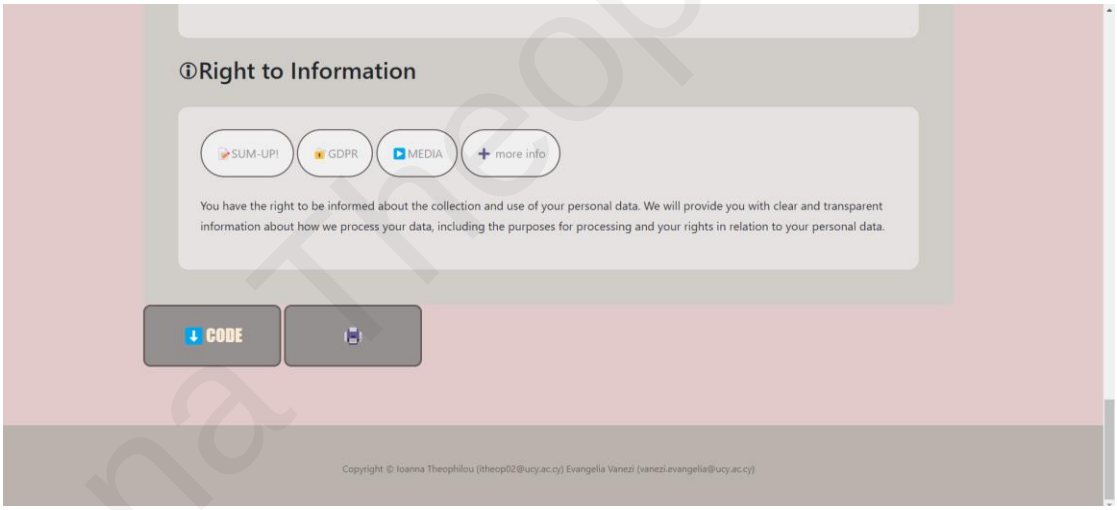
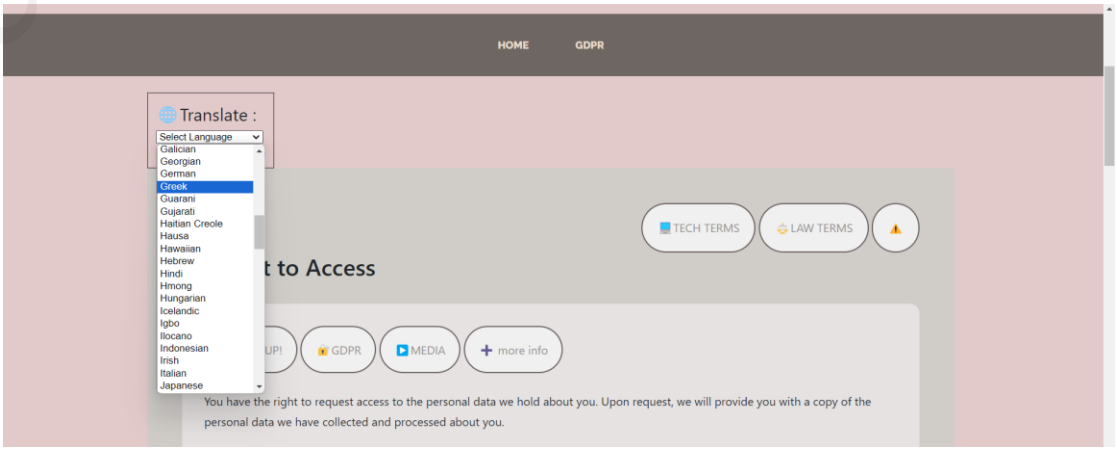


Figure 68 AlRíght - Enhanced Policy / General UI part 5/7.



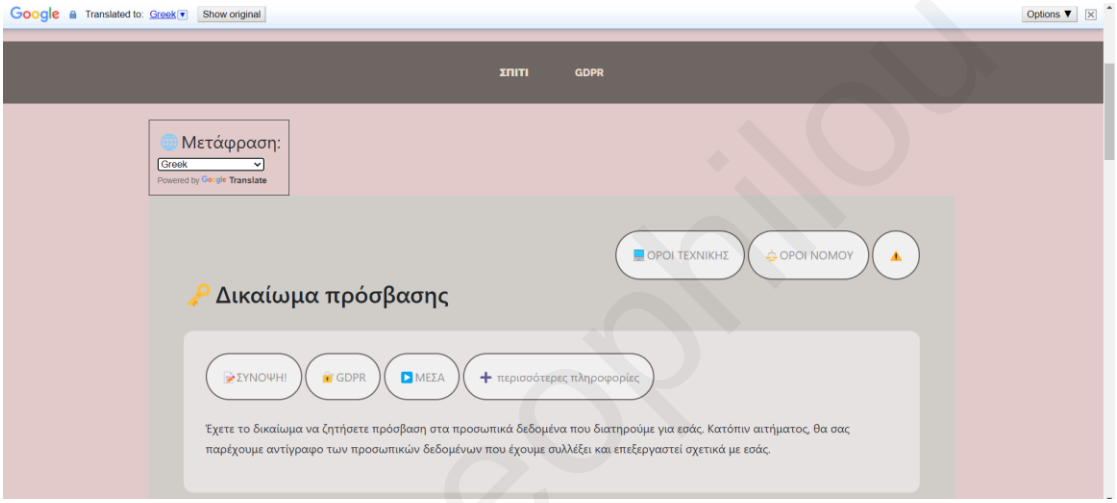
When the user scrolls to the bottom of the page, they can see the options download and print.

Figure 69 AlRíght - Enhanced Policy / General UI part 6/7.



On the left side of the page in the corner, there is a Translate button, where the user can choose any language and the automatic translation will appear. In technical terms, this was done with the usage of the google Translator's API.

Figure 70 AlRíght - Enhanced Policy / General UI part 7/7.

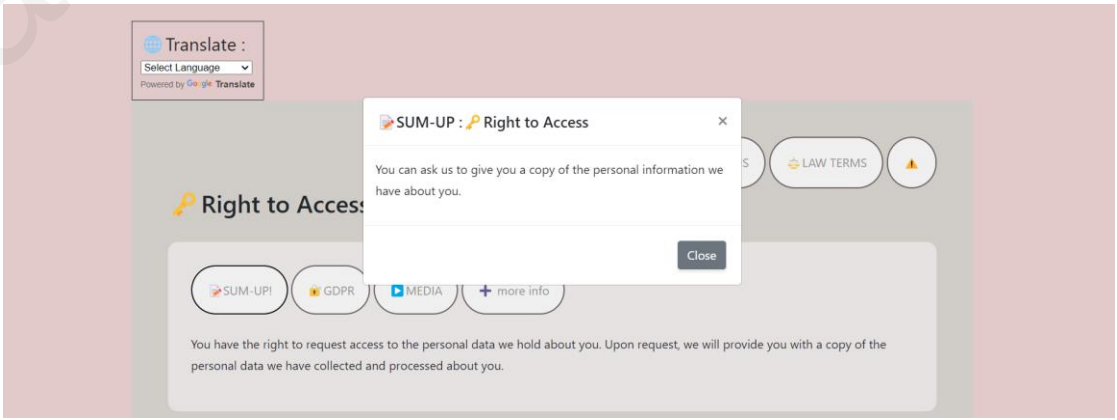


In figure 70 we can see the translated page, when the user selects the language 'Greek'.

5.6.3.2 SUMUP Functionality

The SUMUP functionality addresses 'Verbosity' in terms of comprehensibility issues. With this option, the end user can see the summary that the policy maker has provided and avoid reading the whole document. For demonstration purposes, the user in all the tool's figures, uses a fake privacy policy. Original Policy texts are lengthier and more complex as proved in the literature.

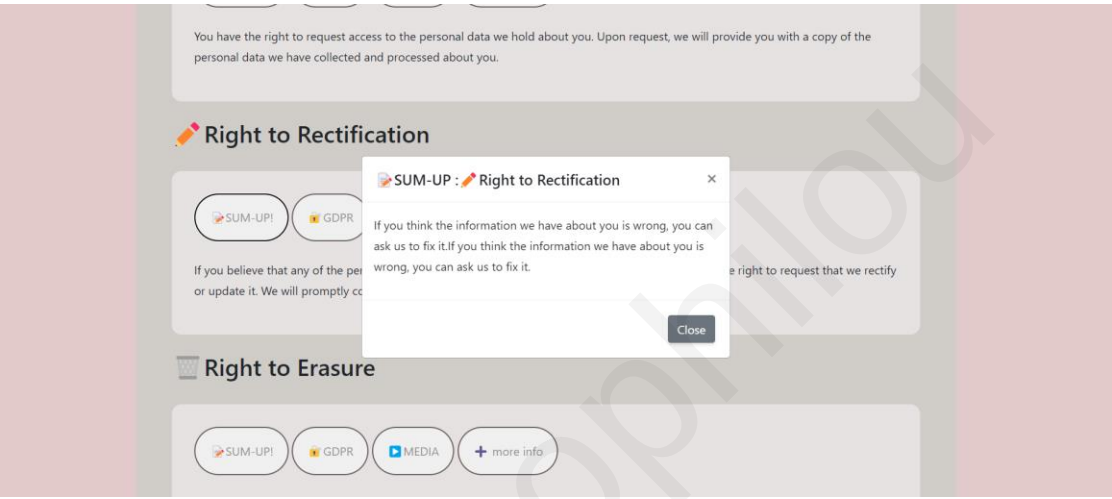
Figure 71 AlRíght - Enhanced Policy / Sumup 1/2.



In the figure above, we can see the pop up when the user pressed to see the summary of the right to access GDPR right. They can press the 'X' button on the right corner of the pop up to

discard the popup or the ‘Close’ button on the bottom. They can even press anywhere else on their screen and the pop up will disappear.

Figure 72 AlRíght - Enhanced Policy / Sumup 2/2.



In this figure, we can see the summary for the Right to Rectification.

5.6.3.3 GDPR Functionality

Above each policy text for each GPDR right, we can see in the figures above that there is a GDPR button. With the GDPR button, the user will be redirected to the actual GDPR’s page. Since the GDPR’s page is not part of the AlRíght platform, there is no backtracking. The user though will not loose their Enhanced Policy since the GDPR articles will be shown on a new tab on their browser.

Having this option, we overcome Ambiguity and Vagueness since, the option to read the actual article may help in understanding their rights better.

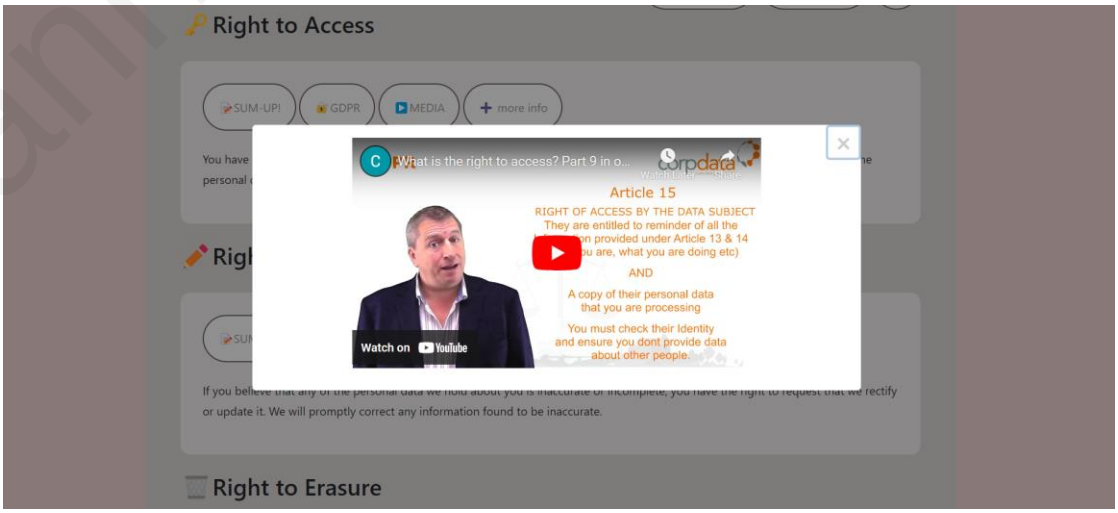
Figure 73 AlRíght - Enhanced Policy / GDPR Redirection.

For example, in figure 73 we can see the page that the user will be redirected to when they press the GDPR button above the Right to Access.

5.6.3.4 MEDIA Functionality

Having the option to see visual aids helps the user in understanding complex terminology as visual aids are proven to be the most effective method of learning. Therefore, by including Media, we overcome Complexity. It can be seen in the figures below that for each right the corresponding media will show up.

Figure 74 AlRíght - Enhanced Policy / Media 1/2.



It can be seen that the user in figure 74 pressed the MEDIA Button above the Right to Access. The popup showed up with an educational and explanatory video about the specific GDPR right.

Figure 75 AlRíght - Enhanced Policy / Media 2/2.

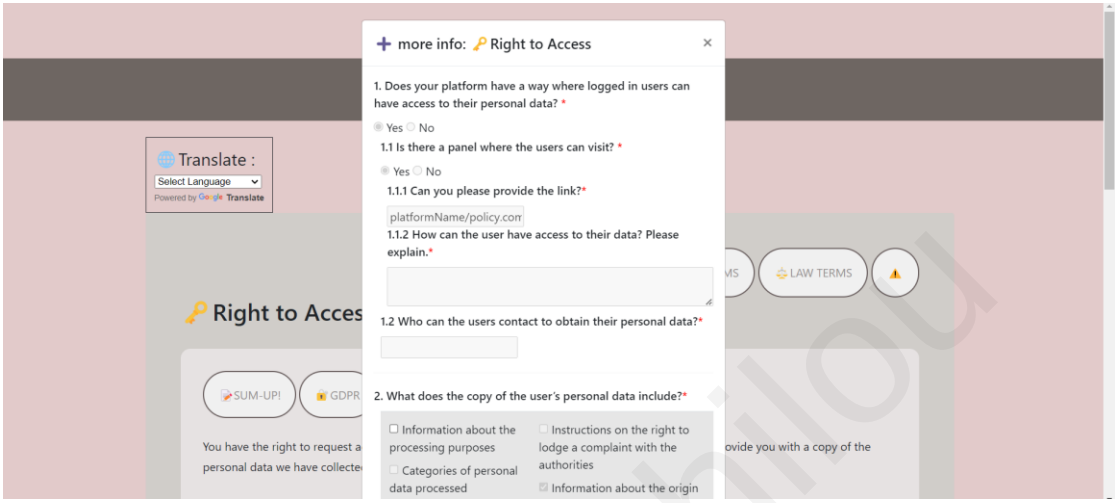
Although, in this figure it can be seen that the user has pressed the MEDIA button above the right to Data Portability.

5.6.3.5 MORE INFO Functionality

Another way in addressing Ambiguity and Vagueness is to provide users with the option to read important information that the policy maker provides in a form of a Question – Answer forum. This option is available when pressing the MORE INFO button above the corresponding GDPR right.

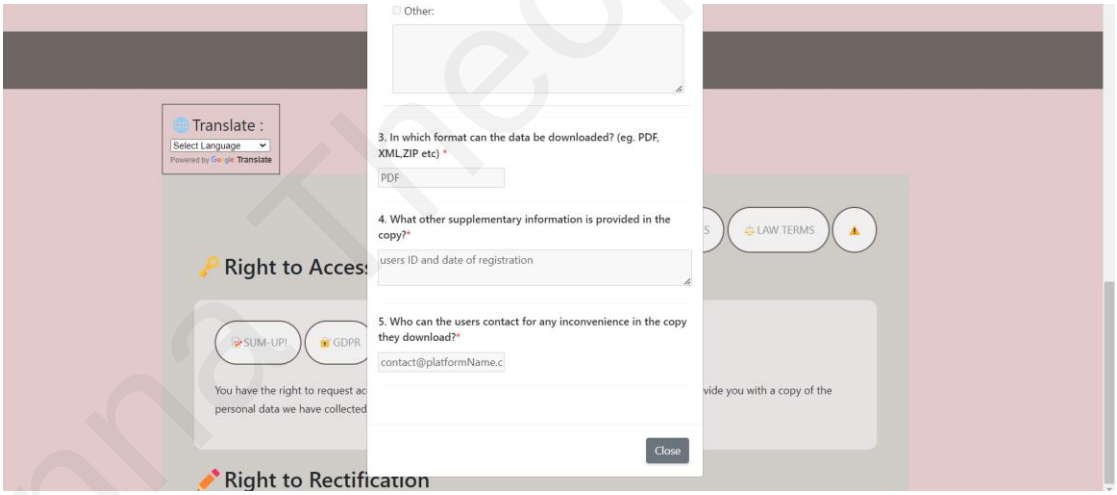
As it can be seen in the figure below, this option provides users with the answers of the questions displayed in all subchapters of 5.6.2. The answers are presented in disabled text fields and all questions are displayed. Although, filled are only the questions that the Policy Maker chose to answer.

Figure 76 AlRíght - Enhanced Policy / More info 1/4.



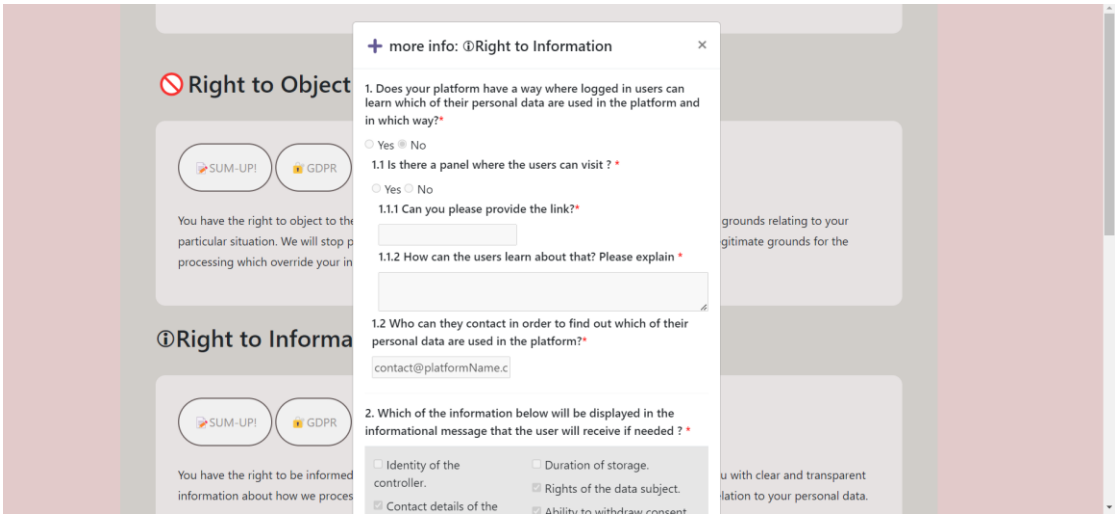
In figure 76 we can see part of the completed forum of the Right to Access.

Figure 77 AlRíght - Enhanced Policy / More info 2/4.



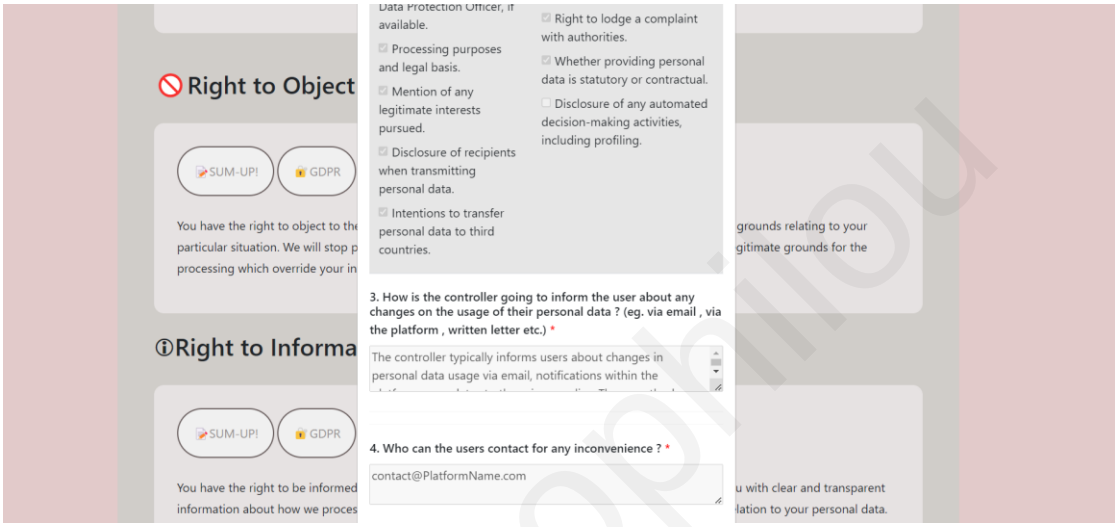
This is the second part of the Right to Access more info panel.

Figure 78 AlRíght - Enhanced Policy / More info 3/4.



In figures 78 and 79 we can see the more info panel about the right to information.

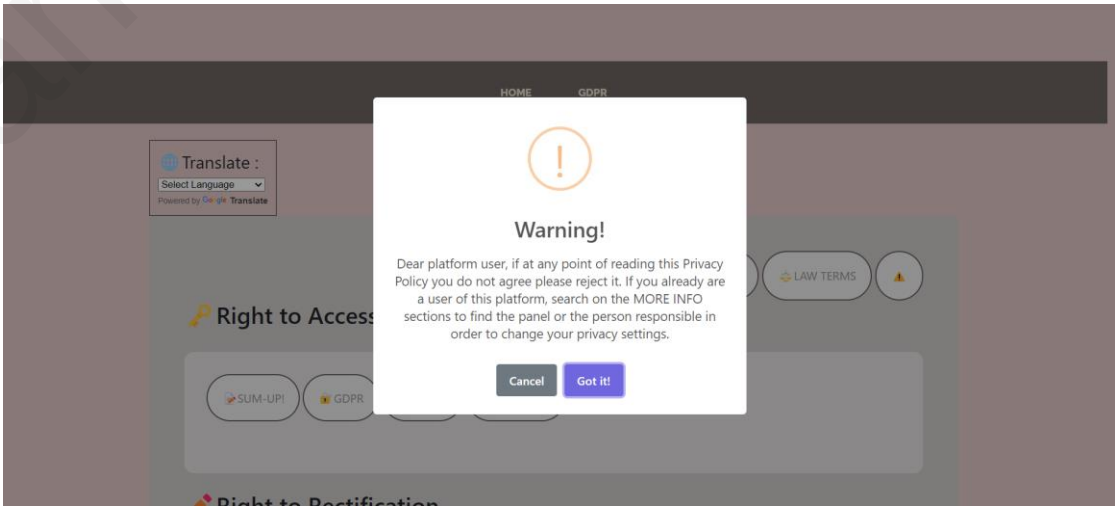
Figure 79 AlRíght - Enhanced Policy / More info 4/4.



5.6.3.6 WARNING Functionality

The warning functionality was made in order to overcome Suspiciousness, since each platform has their own rules and regulations. The goal here is to push the user into rejecting the policy in case they find something that does not suit them. The warning also suggests in searching on the MORE INFO sections to find possible privacy panels, or the contact person responsible.

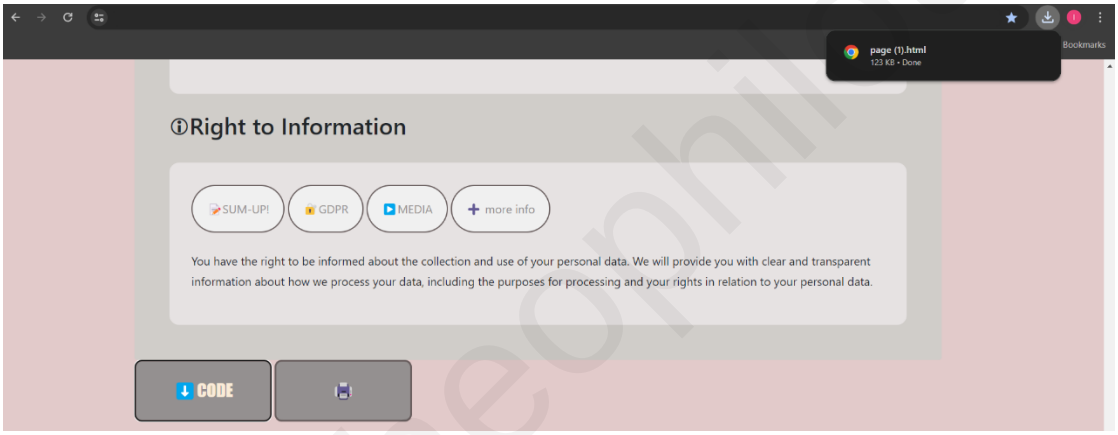
Figure 80 AlRíght - Enhanced Policy / Warning



5.6.3.7 DOWNLOAD CODE Functionality

The download button shown below is meant to be used by the Policy Maker in order to incorporate the presented Policy in their platform. This button downloads the actual code, so the Policy Maker can send it to the responsible department and the incorporation is very easy.

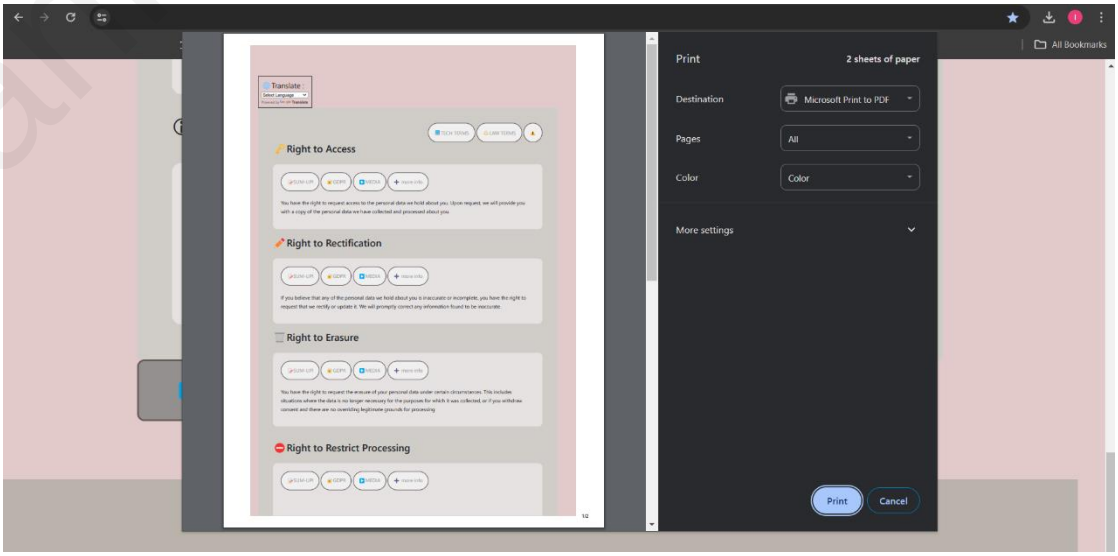
Figure 81 AlRíght - Enhanced Policy / Download



5.6.3.8 PRINT Functionality

This functionality is for printing purposes or saving the Enhanced policy as PDF. This can also be used by the Policy Maker.

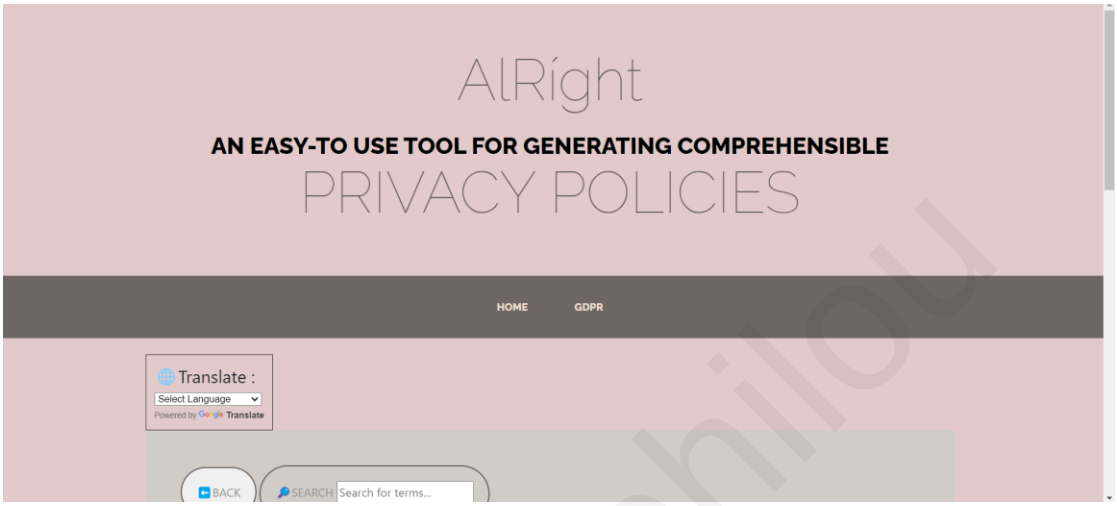
Figure 82 AlRíght - Enhanced Policy / Print



5.6.3.9 TECH TERMS Page

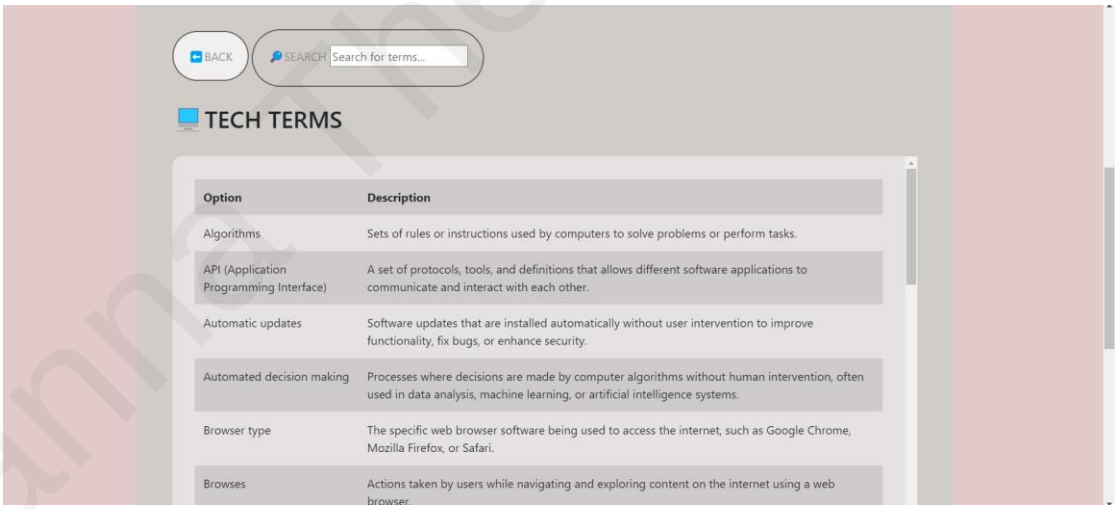
The Technical terminologies page was created to overcome Language Difficulty.

Figure 83 AlRíght - Enhanced Policy / Tech Terminology part 1/4.



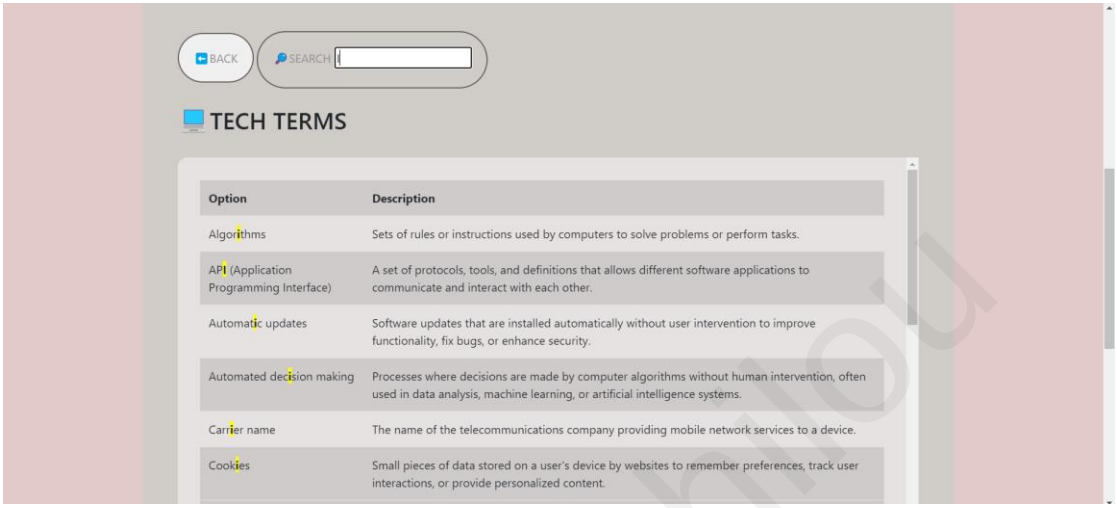
This figure shows the first interaction with the Technical Terminologies Page. The user can user the Translate option on the left side of the page. There is also a search engine.

Figure 84 AlRíght - Enhanced Policy / Tech Terminology part 2/4.



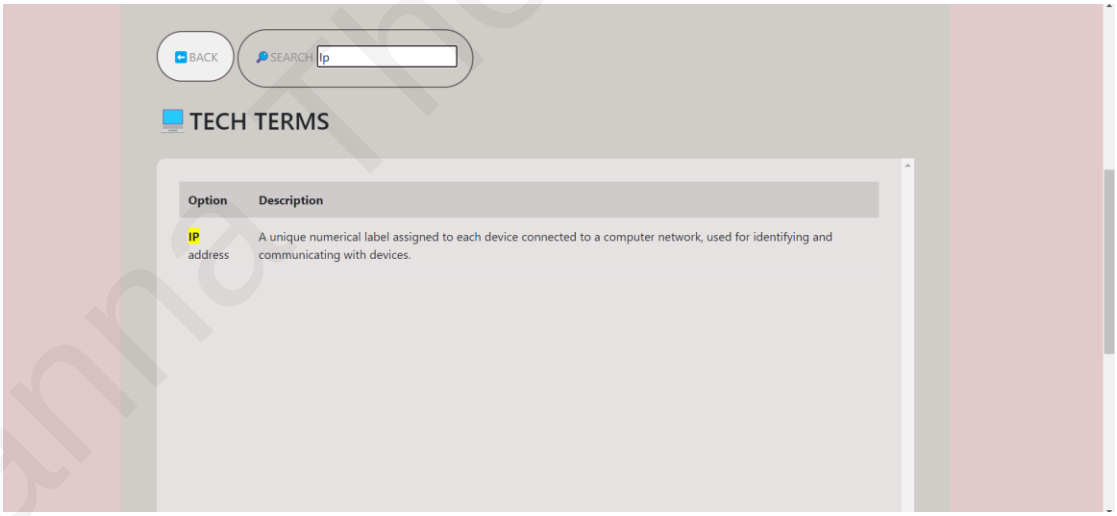
The user is presented with 40 technical terminologies and their description. These terminologies were extracted from all nine policies that were presented in the OhKéy platform.

Figure 85 AlRíght - Enhanced Policy / Tech Terminology part 3/4.



It can be seen that the search is being done while the user types. In the figure above, the user typed the character ‘I’ and the search engine started highlighting the words with this character.

Figure 86 AlRíght - Enhanced Policy / Tech Terminology part 4/4.

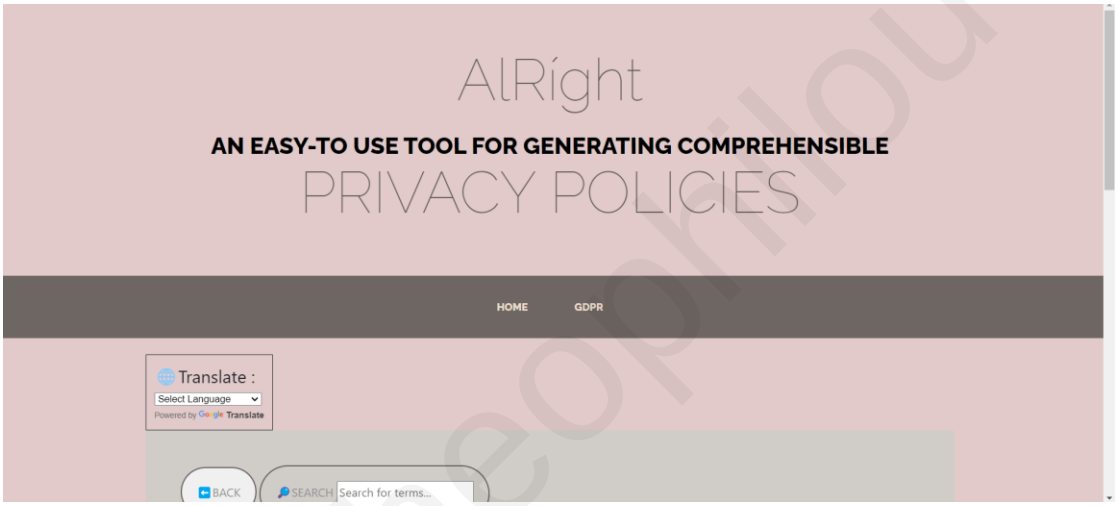


It can be seen that the user finally searched the word IP, and the corresponding terminology appeared with its description.

5.6.3.10 LAW TERMS Page

The Legal terminologies page was also created to overcome Language Difficulty. The legal terminology in this page was also found from the nine policies of the OhKéy platform.

Figure 87 AlRíght - Enhanced Policy / Legal Terminology part 1/3.



This is also the first interaction with the legal terminologies page and it is the same as the Technical Terminologies presented above. Here there are 40 terminologies.

Figure 88 AlRíght – Enhanced Policy / Legal Terminology part 2/3.

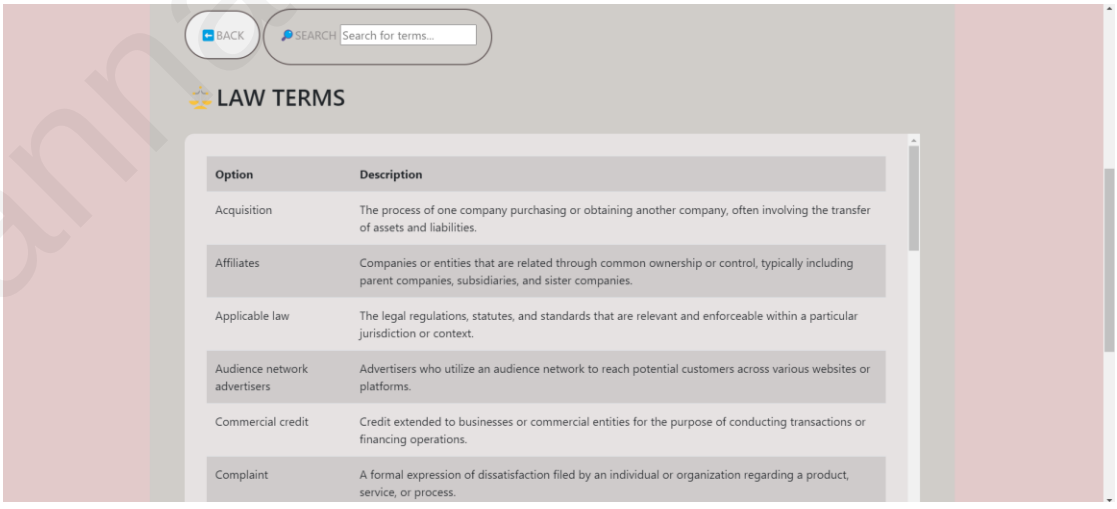
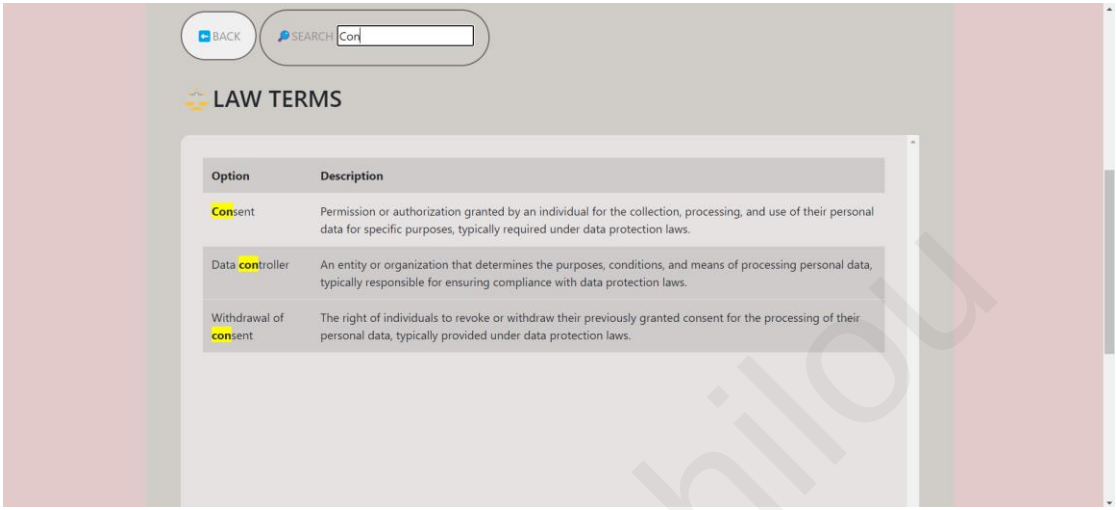


Figure 89 AlRíght - Enhanced Policy / Legal Terminology part 3/3.



The user here searched the word ‘con’ and the suggestions below are ‘Consent’ ‘ Data controler’ and ‘Withdrawal of consent’. This summarizes the demonstration of the AlRight platform. The next chapter will show its evaluation with two different surveys.

Chapter 6 : Tool Evaluation

Tool Evaluation

6.1 Introduction

In this chapter the evaluation of the AlRíght platform will be presented. The evaluation is split into two evaluation forms. Each form examines different aspects of it. The first methodology of evaluation is through the User Experience Questionnaire (UEQ) [51]. The second methodology is through simple questions to non technical users. In this chapter the two methodologies are presented and explained thoroughly. Conclusions are also presented at the end.

6.2 AlRíght Platform : UEQ Evaluation

6.2.1 Introduction: What is the UEQ?

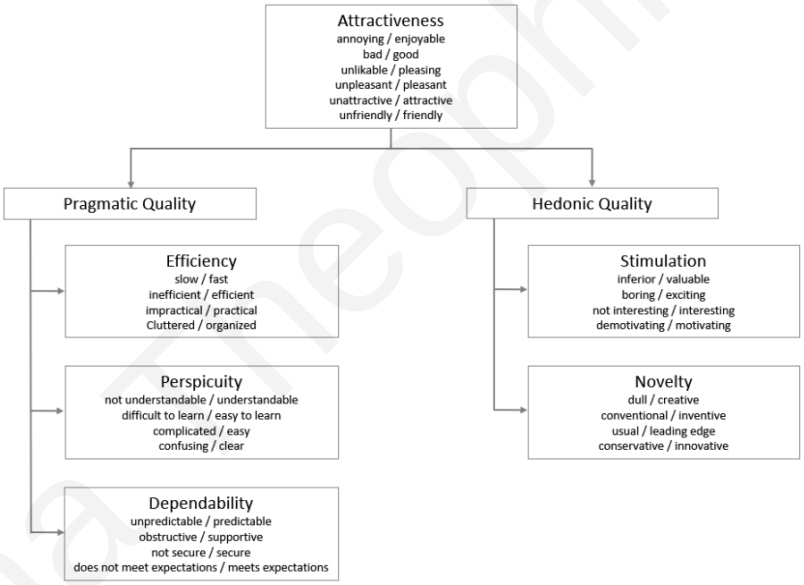
In general, the UEQ is a User Experience Questionnaire that measures the 'Attractiveness', 'Perspicuity', 'Efficiency', 'Dependability', 'Stimulation', and 'Novelty' of an interactive product. These are the three scales that are measured. Each scale has 26 items. The first scale measures Attractiveness, the overall impression of the product. Perspicuity measures whether the product was easy for users to learn. Efficiency measures how much effort is needed for users to use this tool. Dependability measures whether the product gives the feeling of control to the users. Stimulation measures excitement and motivation. And finally, Novelty measures the creativity of the product [52]. The items presented are in the form of 'semantic differential' meaning that they are represented by two terms with opposite meaning. They are presented in a seven-stage scale. An example described in [52] is the following:

attractive o o o o o o o unattractive

The items are rated on a scale ranging from -3 to +3. Therefore, a rating of -3 indicates the most negative response, 0 represents a neutral response, and +3 signifies the most positive response. The scales are then categorized in three aspects: Valence Dimension Aspect , Pragmatic Quality Aspect, Hedonic Quality Aspect [52].

The figure below, shows the three aspects , which scale corresponds to which aspect and which item corresponds to which scale.

Figure 90 Assumed Scale Structure of UEQ [51]



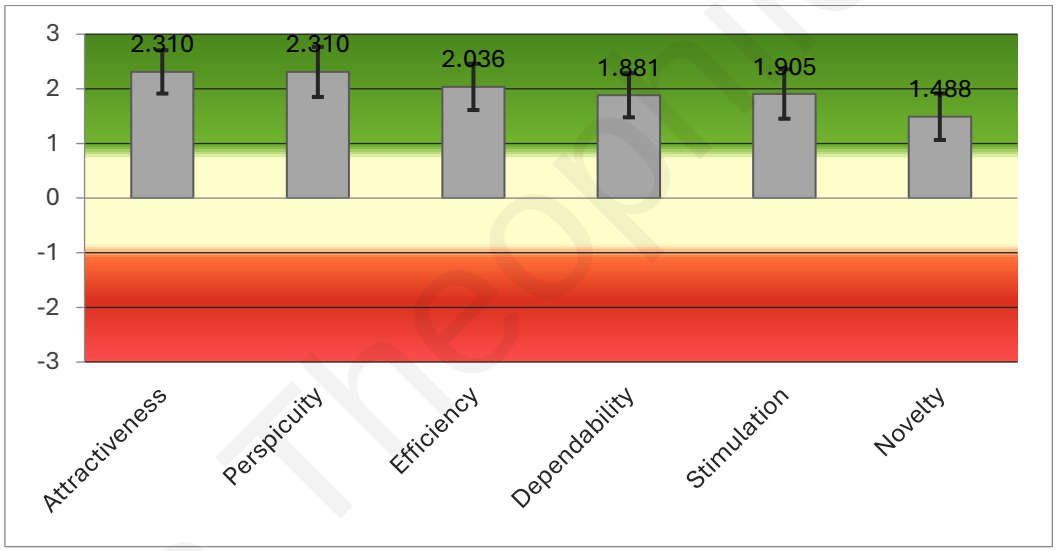
The presented figure was extracted from [51]'s handbook and shows the assumed scale structure of UEQ. In the following sub-chapters we will analyze sufficiency and benchmark.

At this point, it is important to mention that in total for this evaluation 20 participants answered the UEQ questionnaire. The goal was to include as many people as possible, so there were two options in which they could evaluate the platform. Either by using it (they were provided with a fake privacy policy for testing purposes), or by watching a video showcasing the platform. The An extensive presentation of the results exists in ANNEX II.

6.2.2 Testing Sufficiency

According to [53] in order to evaluate whether a product fulfils the general expectations concerning user experience we have to evaluate if it is sufficient. Recall on chapter 6.2.1 where we described the scaling. The items were re-scaled / rated from -3 to +3. [53] says that values above 2 are rare and show excellent results.

Figure 91 AlRíght Platform : UEQ Evaluation / Sufficiency.



It can be seen that attractiveness and perspicuity are on the same level on the scale. This means that the product was easy to learn and made a good impression to its users. After comes efficiency with a score of 2.04 indicating that the user made less effort to learn how to use the tool. With a score of 1.9 follows stimulation meaning that users showed excitement and motivation. Dependability shows whether the user felt like they had control and in this product (the AlRíght platform) a score of 1.88 is a very good outcome. Last but not least comes novelty which measures creativity with a score of 1.49.

Confidence intervals are the intervals that show how precise our calculations are. According to [53] the error bars represent the changes in the results in case of many repetitions of the experiment under the same conditions. This difference may be random influences that will

affect the scale mean. The error bars in the figure above do not differ much from the actual values which shows a good precision rate.

6.2.3 Benchmark

[53] also describes the notion of Benchmark. This measure technically classifies the tool into five categories since it compares it with other products in order to create these classifiers. It is worth mentioning that the classifiers now involve 468 studies from 21175 people. The classifiers as mentioned in [51]'s handbook are :

- ✓ *'Excellent*: In the range of the 10% best results.'
- ✓ *'Good*: 10% of the results in the benchmark data set are better and 75% of the results are worse'.
- ✓ *'Above average*: 25% of the results in the benchmark are better than the result for the evaluated product, 50% of the results are worse'.
- ✓ *'Below average*: 50% of the results in the benchmark are better than the result for the evaluated product, 25% of the results are worse'.
- ✓ *'Bad*: In the range of the 25% worst results.'

Figure 92 AlRíght Platform : UEQ Evaluation / Benchmark 1/2.

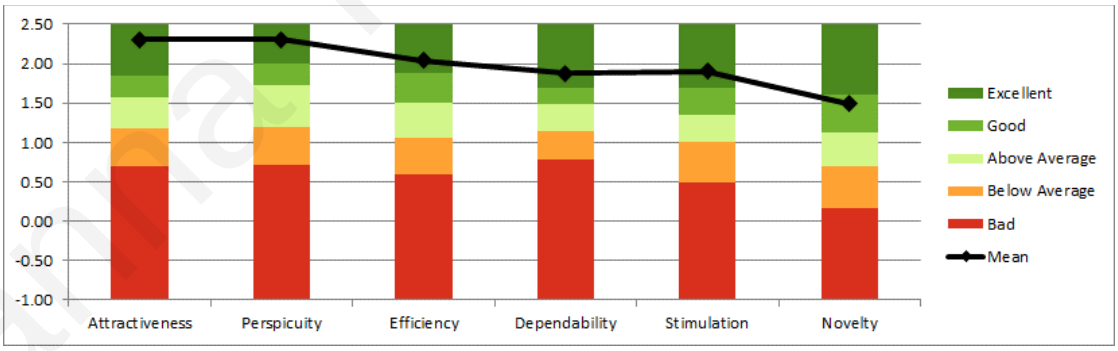


Figure 92 shows the mean scores per scale proving the excellence in the first five scales and 'good' score in novelty.

Figure 93 AlRíght Platform : UEQ Evaluation / Benchmark 2/2.

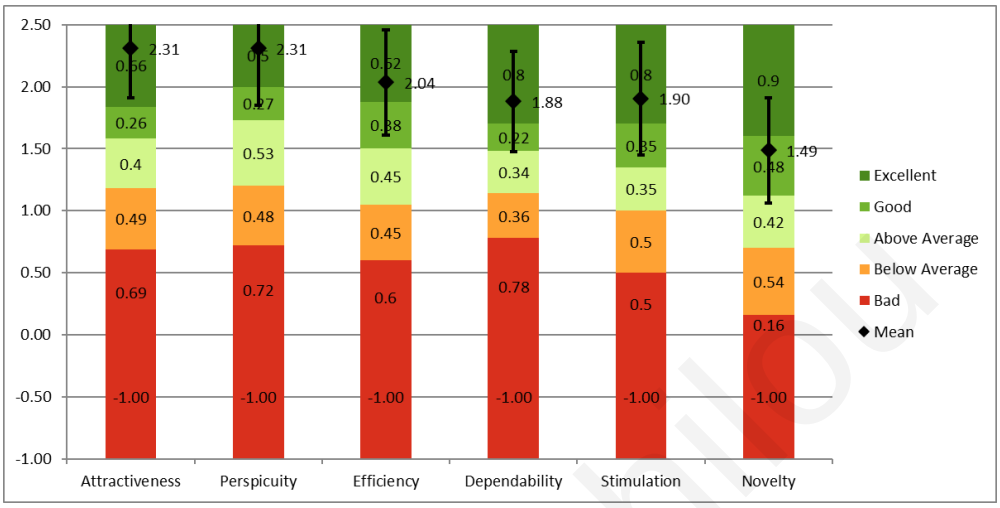


Figure 93 shows the confidence intervals of the scale scores. It can be seen from the error bars that even with the possibility of other repetitions of the experiment, the results would not be below average. Most results would be between 'good' and 'excellent'. The form given to users is presented in ANNEX III.

6.3 AlRíght Platform : Google Forms Evaluation

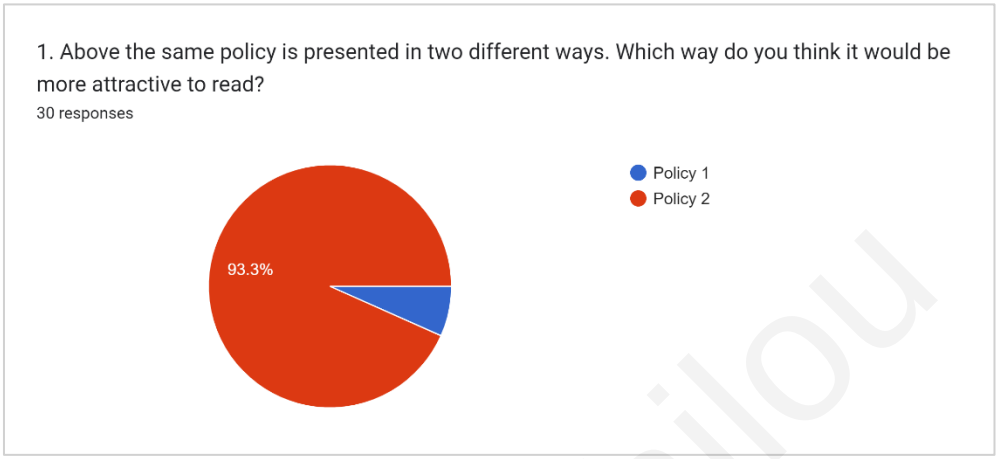
6.3.1 Introduction : What was the purpose?

The purpose of this evaluation was to present a simpler way to non-technical audience to evaluate the tool without needing to actually use it. This survey's purpose was to evaluate the general need for the platform and its functionalities by end-users. The form given to users is displayed in ANNEX IV.

6.3.2 Results

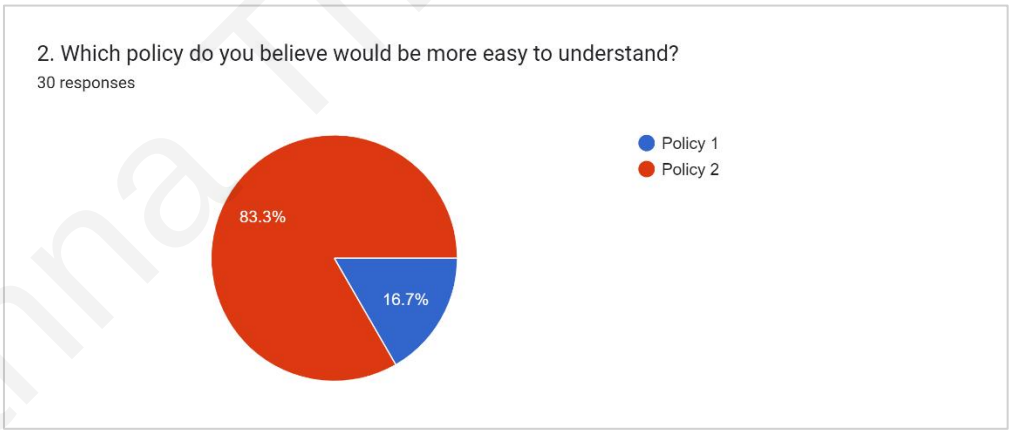
The survey had 30 participants from technical and non-technical users. Each user could only answer the survey once.

Figure 94 AlRíght Platform : Google Forms Evaluation - Q1/13.



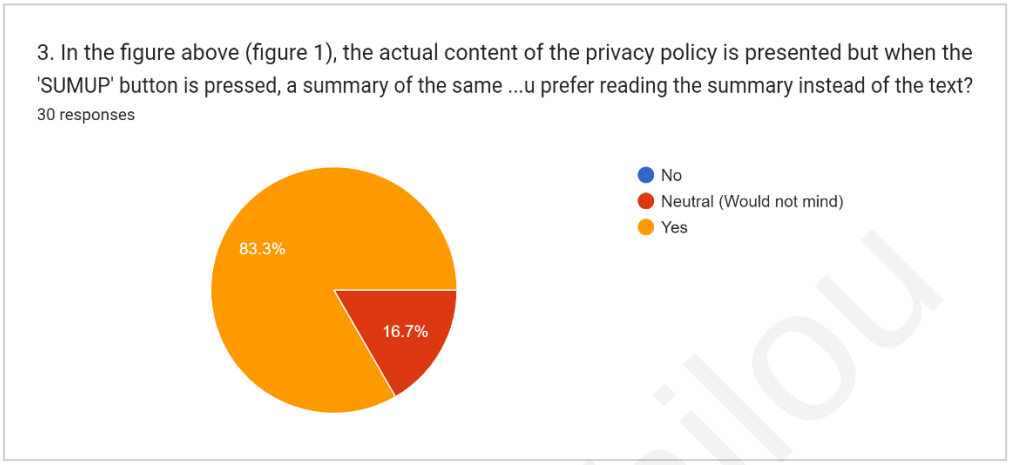
The user was presented with two policies. The first policy (Policy 1) was the one that was lengthier and dull presented in a simple document. The second policy (Policy 2) was the policy extracted by the AlRíght tool, as described above (the Enhanced Policy page). It looks like Policy 2 was preferred by the audience with the score of 93.3%.

Figure 95 AlRíght Platform : Google Forms Evaluation - Q2/13.



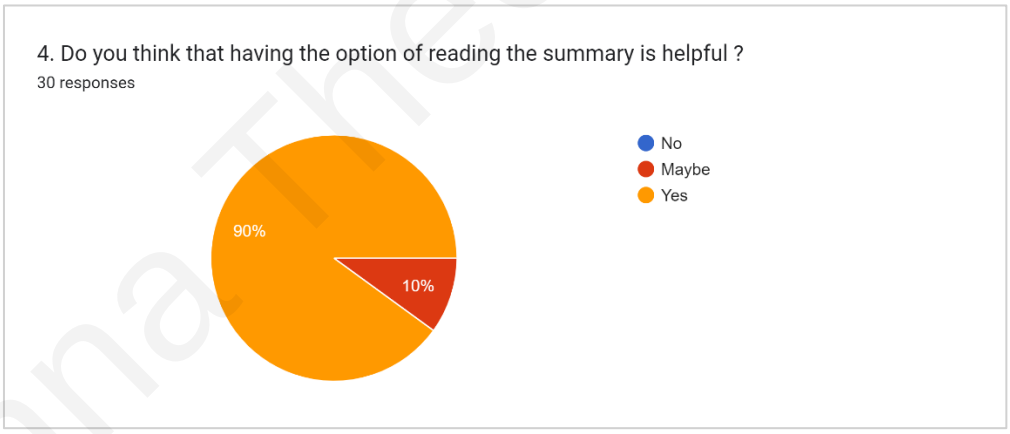
Question 2 rephrases question 1 in terms of comprehension. 83.3% respondents preferred policy 2 (the Enhanced Policy).

Figure 96 AlRíght Platform : Google Forms Evaluation - Q3/13.



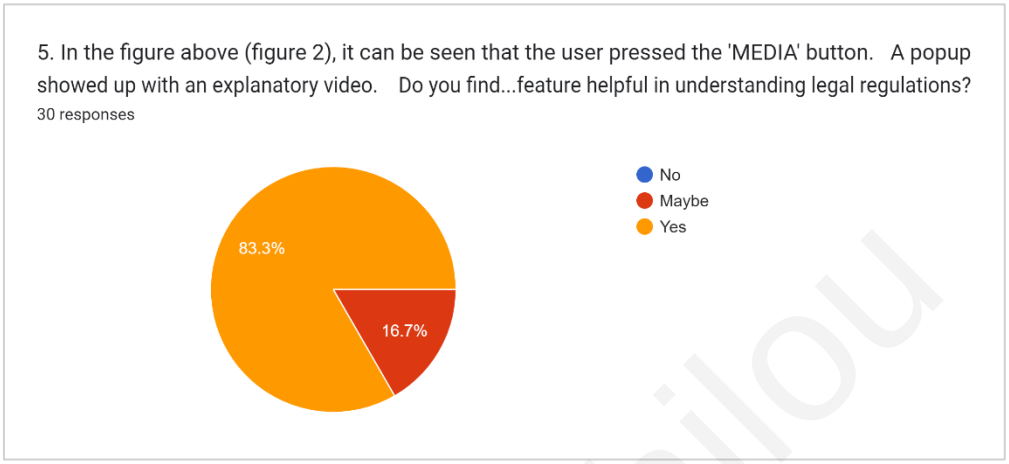
Question 3 assesses the SUMUP functionality. 83.3% of participants stated that they would prefer reading the summary rather than the actual text of the policy. The rest answered 'maybe.' No negative reactions.

Figure 97 AlRíght Platform : Google Forms Evaluation - Q4/13.



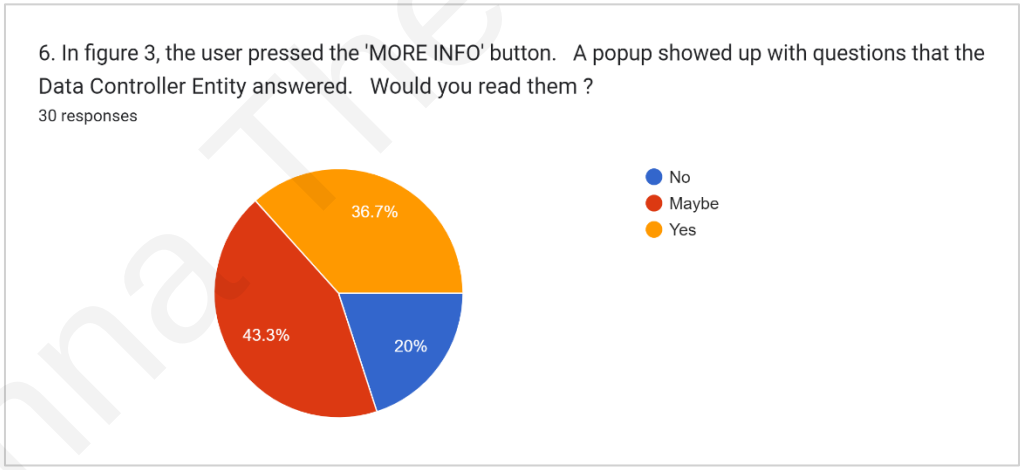
In terms of comprehension there were more positive responses for the SUMUP Functionality with 90% positive and 10% uncertain but again no negatives.

Figure 98 AlRíght Platform : Google Forms Evaluation - Q5/13.



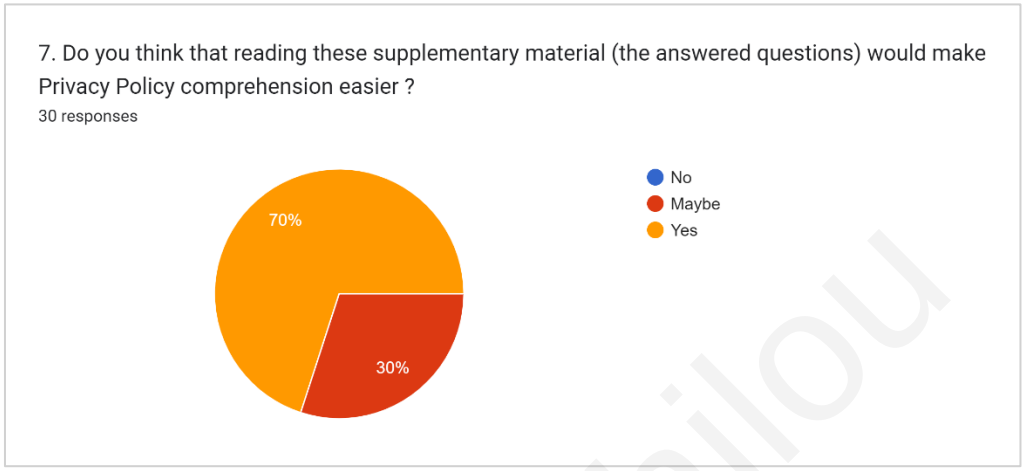
Question 5 assesses the MEDIA functionality. It also had 83.3% positive assessments and 16.7% uncertain in terms of comprehensibility. No negatives.

Figure 99 AlRíght Platform : Google Forms Evaluation - Q6/13.



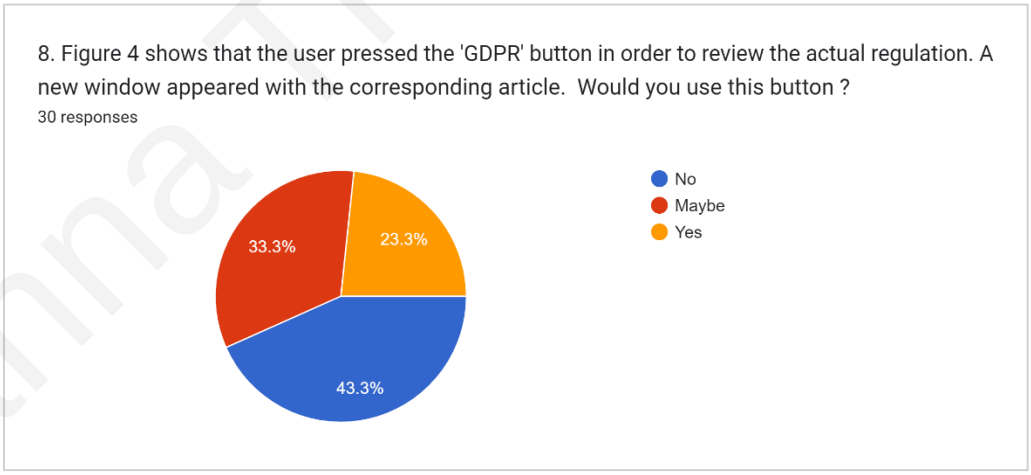
The question presented in figure 99 assesses the MORE INFO button where the user would see the answered questions from the Policy Maker. 43.3% said they would maybe use it and 36.7% said they would use it for sure. The other 20% responded negatively.

Figure 100 AlRíght Platform : Google Forms Evaluation - Q7/13.



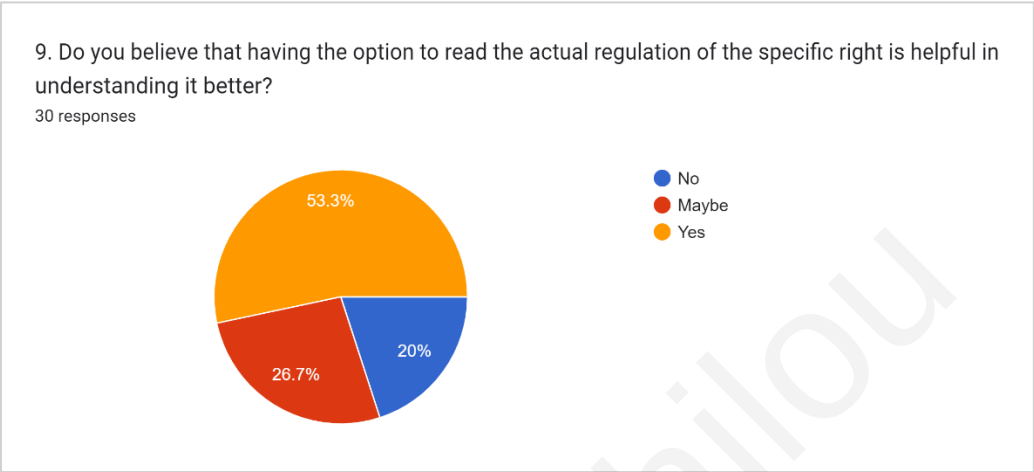
Although 20% said they would not use it, it seems that 70% believed it definitely helps in understanding Privacy Policies more and 30% are not sure. No negative assessments for the MORE INFO functionality in terms of its contribution to comprehensibility.

Figure 101 AlRíght Platform : Google Forms Evaluation - Q8/13.



Many negative assessments though are presented for the GDPR button. 43.3% said that they would not use it and 23.3% said they would. Others stated uncertainty.

Figure 102 AlRíght Platform : Google Forms Evaluation - Q9/13.



However, not being willing to use it does not beat its purpose which is enhancing comprehension. This is proven by the answers to question 9 where 53.3% said that even though they would not use it they do believe it enhances comprehension of Privacy Policies. An explanation of the biggest negative percentage presented in figure 102 above may be that users are just not willing to read the actual GDPR Articles.

Figure 103 AlRíght Platform : Google Forms Evaluation - Q10/13.

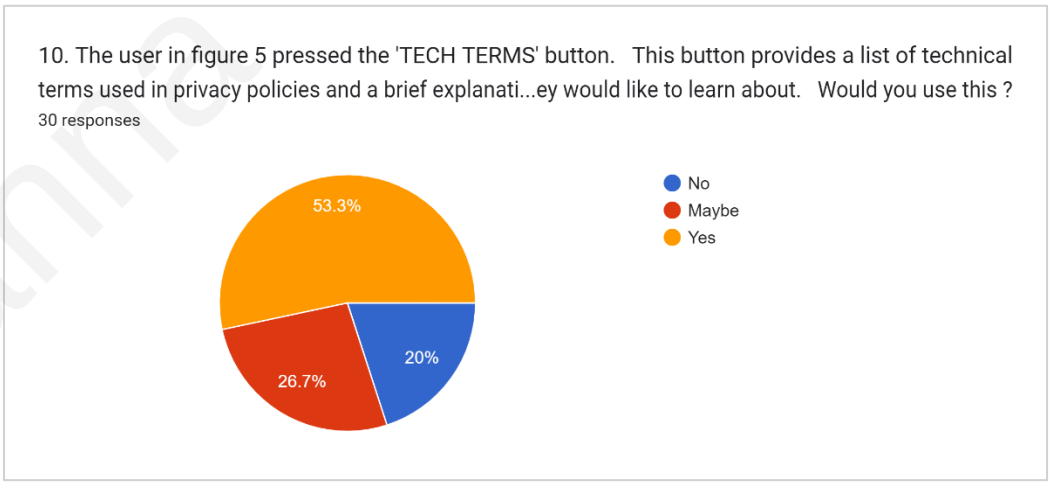
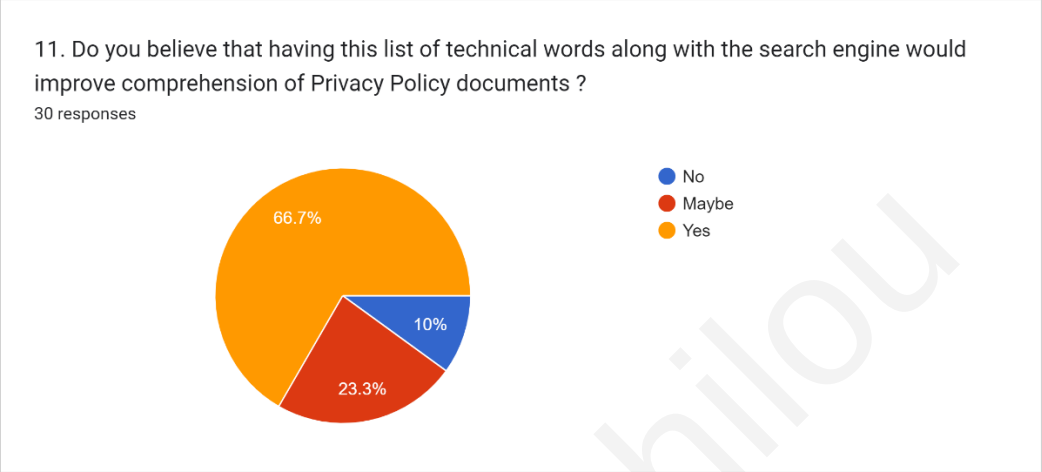


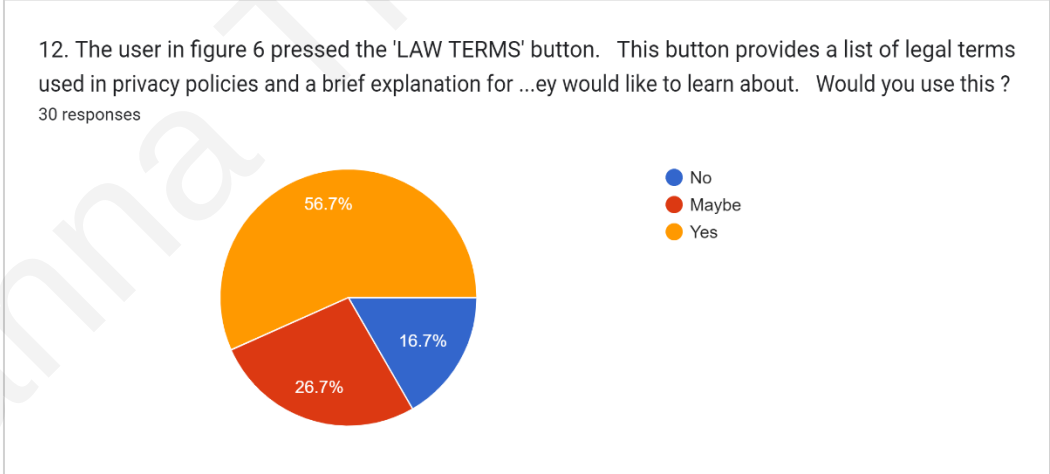
Figure 103 shows the results in assessing the TECH TERMS page of the platform. It can be seen that 53.3% would use the page in finding relevant terminology they didn't understand in the Policy. 26.7% said that might use it and 20% said they would not. Negative responses may be from participants with technical background. Either computer science degrees, software engineering jobs and others.

Figure 104 AlRíght Platform : Google Forms Evaluation - Q11/13.



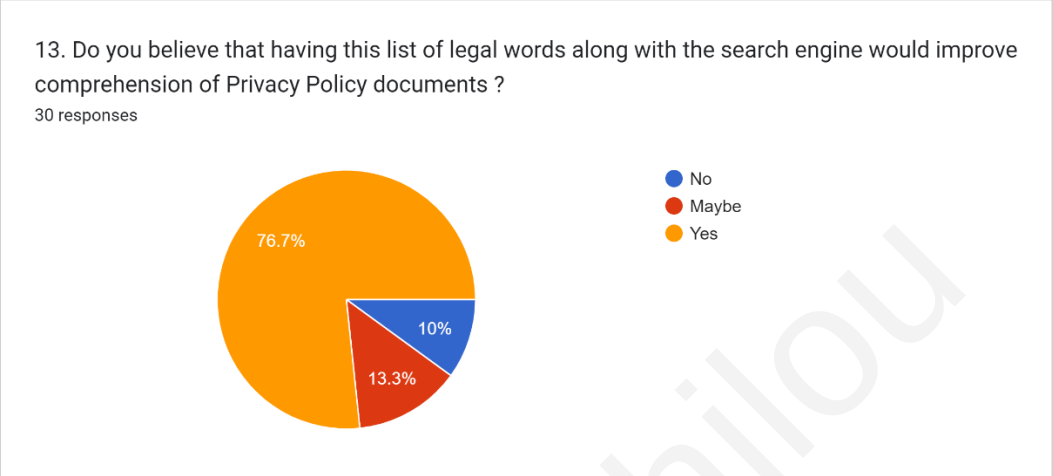
In terms of comprehension, many positive responses were gathered decreasing the negative responses by 10%. This shows that even though some would not use it they still think it's a useful functionality in terms of understanding Privacy Policies better.

Figure 105 AlRíght Platform : Google Forms Evaluation - Q12/13.



Similar responses can be observed with the LAW TERMS page. More than half stated they would use it and 26.7% said they might.

Figure 106 AlRíght Platform : Google Forms Evaluation - Q13/13.



As described, in terms of comprehension the negative assessments of this page were decreased by more than 6% leaving positive responses to 76.7%.

6.4 Conclusions

In this chapter, the AlRíght platform was assessed in two different methods. The first method used was the UEQ - User Experience Questionnaire and the second method was by simple questions to end users using Google Forms. In the first method it was discussed that six scales of usability measures were used. Scientifically proven formulas found the tool to be attractive, and efficient as well as perspicuous , that it promotes dependent user interaction and that it is creative and motivating. The tool was also compared to other platforms since the benchmark measure was calculated with very good results. Of the five categories, five out of six scales were marked as ‘excellent’ and one as ‘good’ which are very good results since in the introduction it was mentioned that results above 2 are considered to be rare. It was also mentioned that in some cases the different intervals of the experiments may alter the results. Although our error bars were still on the Excellent, Good and Above Average categories which shows the consistency and persistence of our results.

Recall of the second survey’s purpose that was to evaluate the general need for the platform and its functionalities by end users. The survey had 30 participants of different backgrounds. Generally in terms of comprehension, 83.3% respondents preferred policy 2 which is the new

version of Policies proposed by the AlRíght platform. In terms of functionalities' scores, SUMUP got the positive score of 83.3% and 90% in terms of comprehensibility. MEDIA functionality also had a positive score of 83.3%. MORE INFO had 43.3% marked as 'maybe' but it seems that 70% believed it definitely helps in understanding Privacy Policies. The GDPR functionality was negatively assessed, and this just shows that most people may not read the actual GDPR articles but, in terms of comprehension many positive results were gathered. Lastly, the TECH and LAW TERMS pages were positively assessed by more than half of the participants and in terms of comprehension they seemed to decrease negative answers by almost 10% each.

In conclusion, the evaluation from both surveys was positive. The UEQ survey showed only positive results and in the second survey a pattern was found. The pattern is that the results in terms of actual usage of functionalities by users and in terms of comprehension, are not the same. In many cases something they would not use personally use they still believe helps in understanding Privacy policies which is the actual purpose of the tool. These assessments were for sure biased by their backgrounds (like their jobs/knowledge) , and their everyday usage of online platforms.

Chapter 7: Conclusions

Conclusions

7.1 Privacy and Software Engineering

From the literature review on Privacy and Software Engineering it can be derived that Privacy and Security are often misunderstood. Privacy concerns the confidentiality of personal information, while Security focuses on protecting systems from unauthorized access. Due to the complexity of privacy regulations, many software engineers lack the expertise to ensure compliance. Thus, integrating Privacy by Design (PbD) approaches into software development is essential. PbD involves incorporating privacy considerations early in the design process and throughout development, through strategies like Privacy by Architecture and Policy.

While the General Data Protection Regulation (GDPR) provides a comprehensive framework for privacy protection, compliance should extend beyond written policies to practical implementation within software systems. This requires measures such as privacy-preserving data collection, data minimization, and secure storage and transmission. Despite the availability of tools like Privacy Impact Assessments and privacy-enhancing technologies, their adoption remains limited. Therefore, greater efforts were proven to be needed to promote awareness and utilization of these resources in software engineering practices.

7.2 Privacy and Usability

In terms of Privacy and Usability, the review highlighted the need of privacy policy visualization approaches, although stating concerns about their practicality. Research revealed widespread ignorance about privacy issues and available actions, leading to the question: How can individuals adopt solutions they're unaware of?

The primary reason for this ignorance is the lack of user-friendly ways to learn about privacy concerns. This dissertation proposed using visualization techniques to make Privacy Policies easier to understand, aiming to increase awareness. This empowers users to make more informed decisions about their privacy.

7.3 Comprehension Analysis of Privacy Policies

The analysis of Privacy Policy annotations revealed key insights into comprehensibility. Annotations are evenly distributed across platform types with negative annotations indicating the need for effective solutions. Neutral annotations highlight the importance of clarity in achieving clear comprehension. Specific findings show social media platforms as the most ambiguous, vague, and suspicious, while online advertising platforms are the most complex and verbose, and e-commerce platforms feature the most difficult language. Supplementary results are detailed in Annex I. Given these challenges, creating comprehensive Privacy Policies is crucial. Drawing from literature, visually pleasing approaches were proposed, with usability and visualization integration as the preferred solution.

7.4 Proposed Solution Evaluation

An enhanced privacy policy representation was extracted from the AlRíght platform. Its evaluation was conducted through two distinct methods: the User Experience Questionnaire (UEQ) and a Google Forms survey targeting end users. The UEQ analysis revealed positive assessments across six usability measures, demonstrating the platform's attractiveness, efficiency, perspicuity, user interactivity, creativity, and motivation. Comparative benchmarking also yielded favorable results, with the majority of scales rated 'excellent' and one rated as 'good'. The second survey reviewed the end users' perceptions of the platform's necessity and functionalities. Among participants, Policy 2 (the enhanced) highest preference for comprehension. Functionality-wise, the platform received positive scores in general, although it revealed a distinction between actual usage and comprehensibility enhancement.

Despite potential biases stemming from participants' backgrounds and online habits, the assessments affirm the platform's efficacy in enhancing comprehension of Privacy Policies.

7.5 General Conclusions

In general, the solution proposed was proven to be effective since many positive evaluations were gathered. At this point of the dissertation is important to answer to the three Research Questions declared in Chapter 1. ‘Do users understand the contents of privacy policies’? From the OhKéy platform we extracted that policies use difficult language by 9.96%, suspicious by 21.05%, verbose by 11.28% , complex by 11.84% , vague by 24.25% and ambiguous by 21.62%. Therefore, users did not feel confident enough to make informed decisions. The second question was ‘Can we assist Policy Makers & Software Engineers in creating comprehensible Privacy Policies ?’. It seems that the AlRíght platform helps in achieving that. It managed to gain 2.3/3 in terms of Attractiveness and Perspicuity, 2/3 in terms of Efficiency, 1.9/3 in terms of Stimulation, 1.8/3 in terms of Dependability and 1.5 in terms of Novelty. The final question to be answered is : ‘Can we enhance users’ comprehension ?’. With the enhanced Policy representation, comprehension is proven to be increased by 83.3%.

7.6 Future Work

It can be assumed that the contents of the proposed platform (AlRíght) can be extended continuously. The MORE INFO functionality could be extended with more questions to be answered by policy makers and the MEDIA button could be extended in providing more material related to the specific GDPR rights. Another expansion could be generating summaries with the usage of ready-made APIs through Artificial Intelligence tools.

Bibliography

[1] E. Vanezi, G. Kapitsaki, D. Kouzapas and A. Philippou, "A Formal Modeling Scheme for Analyzing a Software System Design against the GDPR," in *14th International Conference on Evaluation of Novel Approaches to Software Engineering*, Heraklion, Crete, Greece, 2019.

[2] E. Vanezi, "GDPR Compliance in the Design of the INFORM e-Learning Platform: a Case Study," in *13th International Conference on Research Challenges in Information Science (RCIS)*, 2019.

[3] N. B. Ruparelia, "Software development lifecycle models," *SIGSOFT Softw. Eng. Notes*, vol. 35, no. 3, pp. 8-13, 2010.

[4] P. b. design, "Privacy by design," General Data Protection Regulation (GDPR), 22 10 2021. [Online]. Available: https://gdpr-info.eu/issues/privacy-by-design/.

[5] J. A. Obar and A. Oeldorf-Hirsch, "The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services," *Information, Communication & Society*, vol. 23, pp. 128-147, 2018.

[6] N. Bevan, J. Kirakowski and J. Maissel, "What is Usability?," 1991.

[7] M. E. Morales-Trujillo, G. A. García-Mireles, E. O. Matla-Cruz and M. Piattini, "A systematic mapping study on privacy by Design in Software Engineering," *CLEI Electronic Journal*, vol. vol. 22, no. no. 1, 2019.

[8] B. Kostova, S. Gürses and C. Troncoso, "Privacy engineering meets software engineering. on the challenges of engineering privacy bydesign," *arXiv.org*, 16-Jul-2020.

[9] M. T. Baldassarre, V. S. Barletta, D. Caivano and M. Scalera, "Integrating security and privacy in software development - software quality journal," *SpringerLink*, 28-Feb-2020.

[10] M. T. Baldassarre, V. S. Barletta, D. Caivano and A. Piccinno, "A visual tool for supporting decision-making in Privacy Oriented Software Development: Proceedings of the International Conference on Advanced Visual interfaces," in *ACM Other conferences*, 01-Sep-2020.

[11] N. A. Arachchilage and M. A. Hameed, "Designing a serious game," in *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering Workshops*, 2020.

[12] S. Lamari, N. Benblidia, C. Tibermacine, C. Urtado and S. Vauttier, "A Process for Assisting Privacy-by-Design," *ICSR*, 2022.

[13] Y. Benbenisty, I. Hadar, G. Luria and P. Spoletini, "Privacy as first-class requirements in software development: A socio-technical approach," *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2021.

[14] C. Ke, F. Xiao, Z. Huang and F. Xiao, "A user requirements-oriented privacy policy self-adaption scheme in cloud computing,," *Frontiers of Computer Science*, vol. 17, no. 2, 2022.

[15] A. Senarath and N. A. Arachchilage, "Why developers cannot embed privacy into software systems?," in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, 2018.

[16] J. Lenhard, L. Fritsch and S. Herold, "A literature study on Privacy Patterns Research,," in *2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 2017.

[17] S. Spiekermann and L. F. Cranor, "Engineering privacy," in *IEEE Transactions on Software Engineering*, 2009.

[18] M. Li, L. Yang, Q. Xia, M. Fang, G. Liang and C. Zuo, "STPChain: A crowdsourced software engineering method for software traceability and fine-grained privacy based on Blockchain,," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2022.

[19] K. N. Tongay, "Privacy Preserving Software Engineering for Data Driven Development," 2017.

[20] R. N. Zaeem, "PrivacyCheck v2: A Tool that Recaps Privacy Policies for You," in *29th ACM International Conference on Information & Knowledge Management*, Ireland, 2020.

[21] A. Gautam, "Usable, Acceptable, Appropriable: Towards Practicable Privacy," 2020.

[22] V. Bannihatti Kumar, "Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text," in *The Web Conference 2020*, Taipei, Taiwan, 2020.

[23] T. Sigmund, "Attention Paid to Privacy Policy Statements," *Information*, vol. 12, no. 4, 2021.

[24] J. Johansen and S. Fischer-Hübner, "Making GDPR Usable: A Model to Support Usability Evaluations of Privacy," Springer International Publishing, 2020, pp. 275-291.

[25] A. d. L. Salgado, P. C. K. Hung and R. P. M. Fortes, "Six usable privacy heuristics," in *XXII Brazilian Symposium on Human Factors in Computing Systems*, Maceió, Brazil, 2024.

[26] M. Tahaei, K. Vaniea and N. Saphra, "Understanding Privacy-Related Questions on Stack Overflow," in *2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI,USA, 2020.

[27] D. Reinhardt, J. Borchard and J. Hurtienne, "Visual Interactive Privacy Policy: The Better Choice?," in *2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 2021.

[28] N. Duch-Brown, The competitive landscape of online platforms, Seville: European Commission, Joint Research Centre (JRC), 2017.

[29] U. K. Alves, J. I. A. d. Albuquerque and P. D. Bondaruk., "L2 intelligibility and comprehensibility: trying out new measurements with AEPI," *Anales de Lingüística. Segunda época*, 2021.

[30] Y. Shvartzshnaider, N. Apthorpe, N. Feamster and H. Nissenbaum, "Analyzing Privacy Policies Using Contextual," *Sciendo*, 2018.

[31] A. Säuberli, S. Hansen-Schirra, F. Holzknecht, S. Gutermuth, S. Deilen, L. Schiffl and S. Ebling, "Enabling text comprehensibility assessment for people with intellectual disabilities using a mobile application," *Frontiers*, 2018.

[32] "HTML 5," [Online]. Available: https://www.w3.org/TR/2011/WD-html5-20110405/.

[33] "Introduction to CSS," [Online]. Available: https://www.w3.org/TR/2001/WD-css3-roadmap-20010523/.

[34] "Introduction Bootstrap 5," [Online]. Available: https://getbootstrap.com/docs/5.0/getting-started/introduction/.

[35] "JavaScript," [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/JavaScript.

[36] "PHP: Hypertext Processor," [Online]. Available: https://www.php.net/.

[37] "MariaDB server : the innovative opensource database," [Online]. Available: https://mariadb.org/.

[38] "The Apache HTTP Server Project," [Online]. Available: https://httpd.apache.org/.

[39] "phpMyAdmin : Bringing MySQL to the net," [Online]. Available: https://www.phpmyadmin.net/.

[40] "CSV Files," [Online]. Available: https://data.europa.eu/apps/data-visualisation-guide/csv-files.

[41] "GDPR : Right to Access," [Online]. Available: https://gdpr.eu/article-15-right-of-access/.

[42] "GDPR : Right to Rectification," [Online]. Available: https://gdpr.eu/article-16-right-to-rectification.

[43] "GDPR : Right to be Forgotten," [Online]. Available: https://gdpr.eu/article-17-right-to-be-forgotten.

[44] "GDPR : Right to Restrict Processing," [Online]. Available: https://gdpr.eu/article-18-right-to-restriction-of-processing.

[45] "GDPR : Right to Data Portability," [Online]. Available: https://gdpr.eu/article-20-right-to-data-portability/.

[46] "GDPR : Right to Object," [Online]. Available: https://gdpr.eu/article-21-right-to-object.

[47] "GDPR : Right to be Informed," [Online]. Available: https://gdpr-info.eu/issues/right-to-be-informed/.

[48] W. Brunotte, L. Chazette, L. Kohler, J. Klunder and K. Schneider, "What About My Privacy? Helping Users Understand Online Privacy Policies," *ICSSP 22*, 2022.

[49] M. Janic, J. P. Wijbenga and T. Veugen, "Transparency Enhancing Tools (TETs): An Overview," *Third Workshop on Socio-Technical Aspects in Security and Trust,* 2013.

[50] D. Bui, K. G. Shin, J.-M. Choi and J. Shin, "Automated Extraction and Presentation of Data Practices in Privacy Policies," *POPETS 21,* 2021.

[51] "User Experience Questionnaire," [Online]. Available: https://www.ueq-online.org/.

[52] B. Laugwitz, T. Held and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work,* Springer Link, 2008.

[53] M. Schrepp, A. Hinderks and J. Thomaschewski, "Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios," in *Design, User Experience , and Usability. Theories , Methods and Tools for Designing the User Experience,* SPRINGER, 2014, pp. 338-392.

ANNEX I

DEMOGRAPHICAL DATA:

Tables:

Gender:

ALL	MALE	FEMALE	OTHER	PREFER NOT TO SAY
1098	219	879	0	0

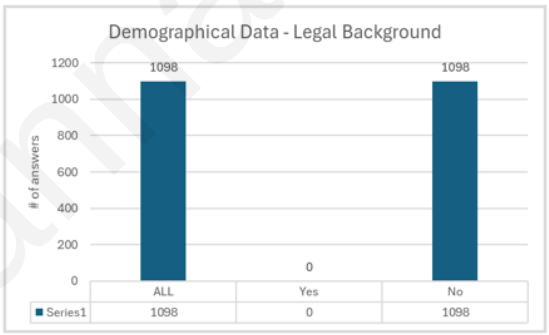
Age:

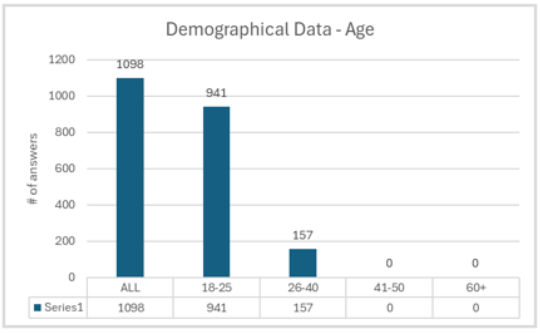
ALL	18-25	26-40	41-50	60+
1098	941	157	0	0

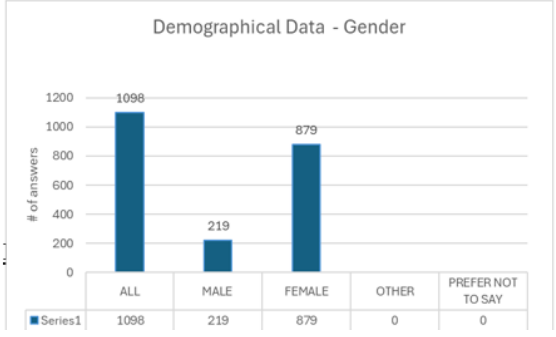
Legal Background :

ALL	Yes	No
1098	0	1098

Graphs:







Tables:

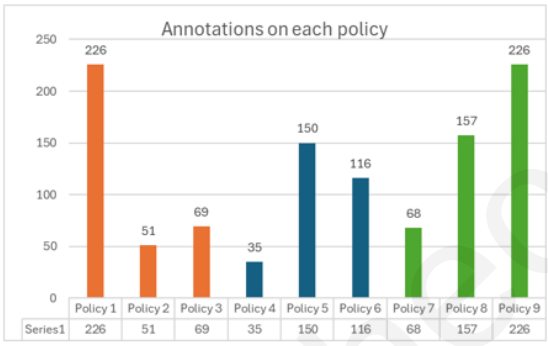
Individual Policy Annotations:

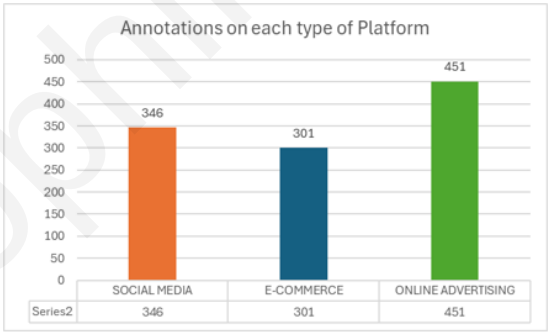
All annotations	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5	Policy 6	Policy 7	Policy 8	Policy 9
1098	226	51	69	35	150	116	68	157	226

Totals for each platform type :

SOCIAL MEDIA	E-COMMERCE	ONLINE ADVERTISING
346	301	451

Graphs:





NEGATIVE ANNOTATIONS :

Tables :

Amount of negatively annbtated tags :

Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language	ALL
115	129	63	60	112	53	532

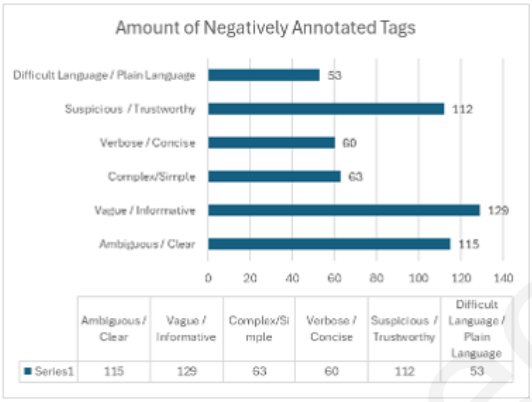
Percentage of negatively annotated tags :

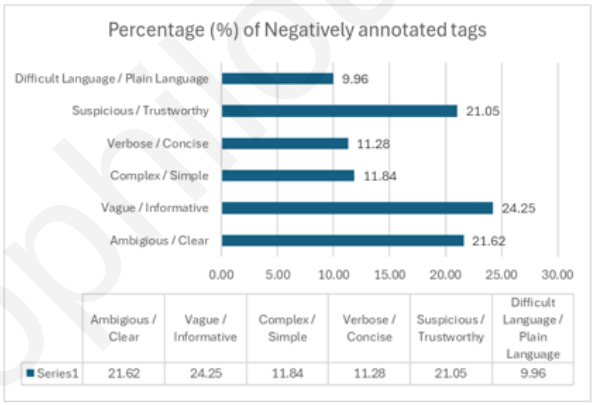
Ambiguous / Clear	Vague / Informative	Complex / Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
21.62	24.25	11.84	11.28	21.05	9.96

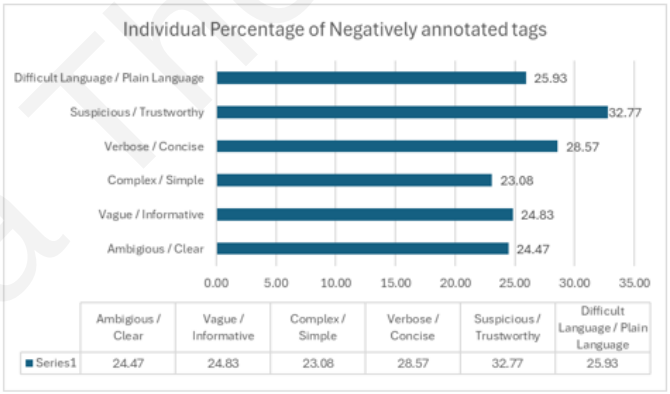
Individual Percentage of negativity - Example : Of all annotations tagged as 'Ambiguous / Clear' 24.47 % were negatively annotated.

Ambiguous / Clear	Vague / Informative	Complex / Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
24.47	24.83	23.08	28.57	32.77	25.93

Graphs :







NEUTRAL ANNOTATIONS :

Tables :

Amount of Neutrally annotated tags

Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
115	88	84	44	101	61

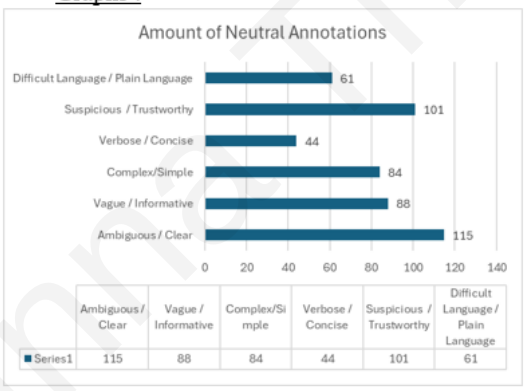
Percentage of Neutrally annotated tags

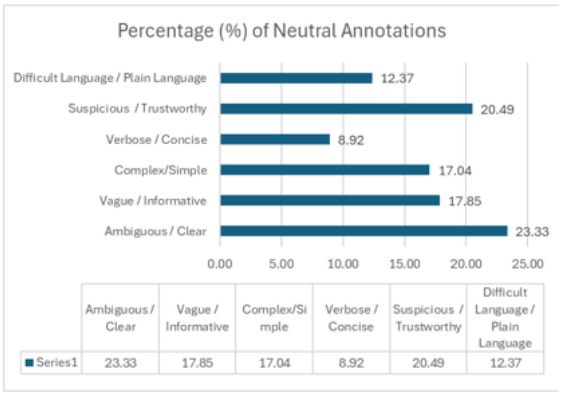
Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
23.33	17.85	17.04	8.92	20.49	12.37

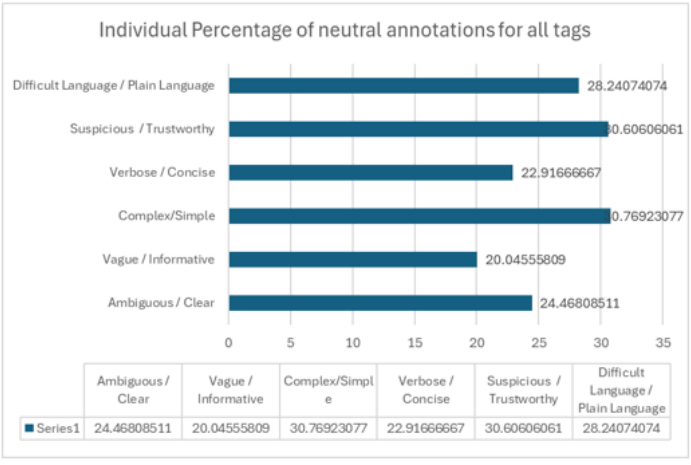
Individual Percentage of Neutrally annotated tags : *Example : Of all annotations tagged as 'Ambiguous / Clear' .47 % were negatively annotated.*

Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
51.06382979	50.56947608	46.15384615	45.83333333	35.45454545	47.22222222

Graphs :







SOCIAL MEDIA PLATFORMS :

All metrics gathered:

POLICY 1	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	133	134	73	55	86	66
POLICY 1	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	50	48	19	7	37	7
POLICY1	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	58	68	48	44	26	51
POLICY 1	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	25	18	6	4	23	8
POLICY 2	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	10	19	9	3	7	7
POLICY 2	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	0	0	0	0	0	3
POLICY2	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	6	16	9	0	7	3
POLICY 2	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	4	3	0	3	0	1
POLICY 3	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	23	14	22	10	7	11
POLICY 3	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	3	3	2	3	2	2
POLICY3	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	15	11	10	4	2	4
POLICY 3	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	5	0	12	3	3	5

E-COMMERCE PLATFORMS

All Metrics gathered :

POLICY 4	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	17	13	13	9	12	5
POLICY 4	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	4	4	3	3	7	0
POLICY 4	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	11	8	8	6	2	5
POLICY 4	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	2	1	2	0	2	0
POLICY 5	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	66	41	14	41	63	6
POLICY 5	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	11	8	2	8	20	2
POLICY 5	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	40	21	6	21	23	1
POLICY 5	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	15	12	6	12	20	3
POLICY 6	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	30	29	25	16	12	41
POLICY 6	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	6	2	5	11	4	23
POLICY 6	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	12	20	4	2	1	1
POLICY 6	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	12	7	16	3	7	17

ONLINE ADVERTISING :

All metrics gathered :

POLICY 8	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	12	7	11	5	6	4
POLICY 9	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	122	113	71	67	90	55
POLICY 9	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	13	21	12	17	18	4
POLICY 9	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	77	53	35	29	44	33
POLICY 9	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	32	39	24	21	28	18
POLICY 7	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	20	23	11	7	5	8
POLICY 7	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	5	2	3	5	2	2
POLICY 7	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	7	19	1	0	1	3
POLICY 7	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEUT	8	2	7	2	2	3
POLICY 8	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
ALL	49	53	35	16	14	17
POLICY 8	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
NEG	23	21	17	10	7	11
POLICY 8	Ambiguous / Clear	Vague / Informative	Complex/Simple	Verbose / Concise	Suspicious / Trustworthy	Difficult Language / Plain Language
POS	14	25	7	1	0	3

NEGATIVE ANNOTATIONS PER PLATFORM TYPE:

Tables:

Most Ambiguous Platform Type

	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	AMBIGUITY	TOTALS
NEG	41	21	53		115
ALL	191	113	166		470
	Online Advertising	E-commerce	Social Media		Ambiguous / Clear
%	35.65217391	18.26086957	46.08695652		24.468085

Most Vague Platform Type

	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	VAGUENESS	TOTALS
NEG	44	14	51		109
ALL	189	83	167		439
	Online Advertising	E-commerce	Social Media		Vague / Informative
%	40.36697248	12.8440367	46.78899083		24.829157

Most Complex Platform Type

	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	COMPLEXITY	TOTALS
NEG	32	10	21		63
ALL	117	52	104		273
	Online Advertising	E-commerce	Social Media		Complex / Simple
%	50.79365079	15.87301587	33.33333333		23.076923

Most Verbose Platform Type

	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	VERBOSITY	TOTALS
NEG	32	22	10		64
ALL	90	66	68		224
	Online Advertising	E-commerce	Social Media		Verbose / Concise
%	50	34.375	15.625		28.571429

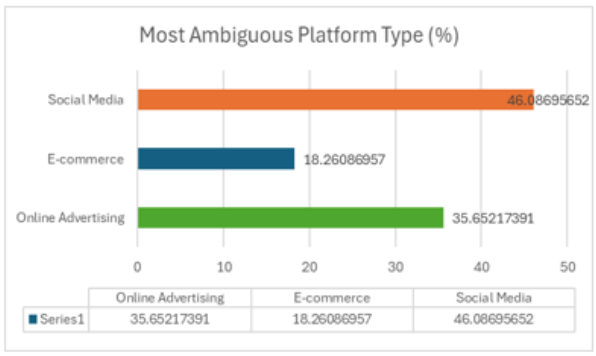
Most Suspicious Platform Type

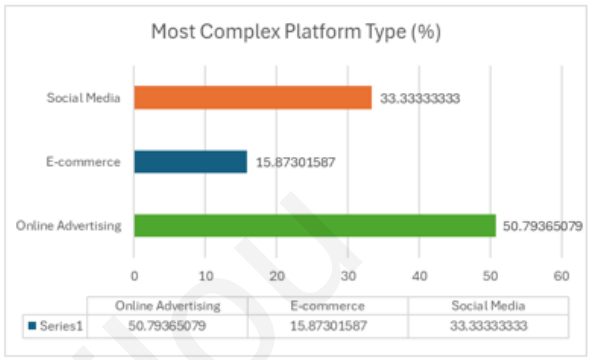
	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	SUSPICIOUSNESS	TOTALS
NEG	27	31	39		97
ALL	109	87	100		296
	Online Advertising	E-commerce	Social Media		Suspicious / Trustworthy
%	27.83505155	31.95876289	40.20618557		32.77027

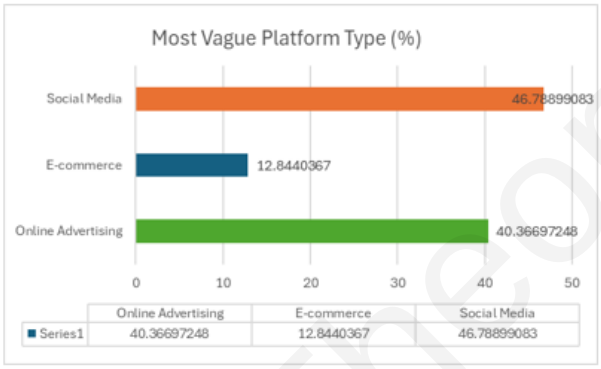
Platform type with the most Difficult Language

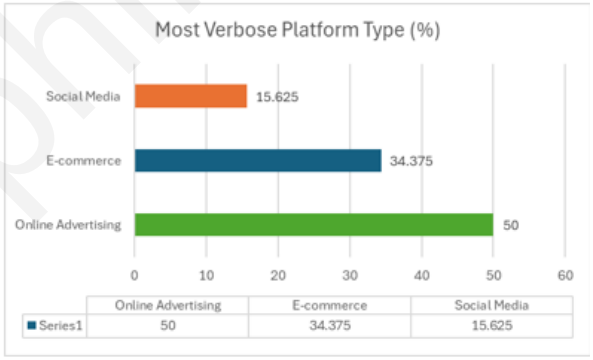
	ONLINE ADVERTISING	E-COMMERCE	SOCIAL MEDIA	LANG. DIFFICULTY	TOTALS
NEG	19	25	12		56
ALL	80	52	84		216
	Online Advertising	E-commerce	Social Media		Difficult Language / Plain Language
%	33.92857143	44.64285714	21.42857143		25.925926

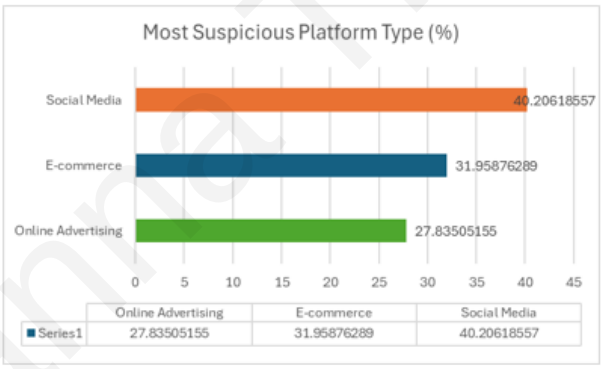
Graphs :

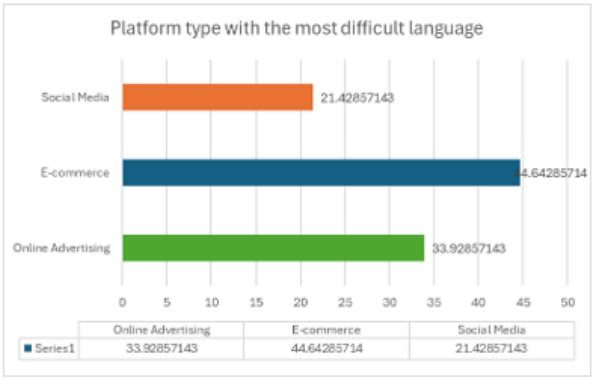












ANNEX II

Data :

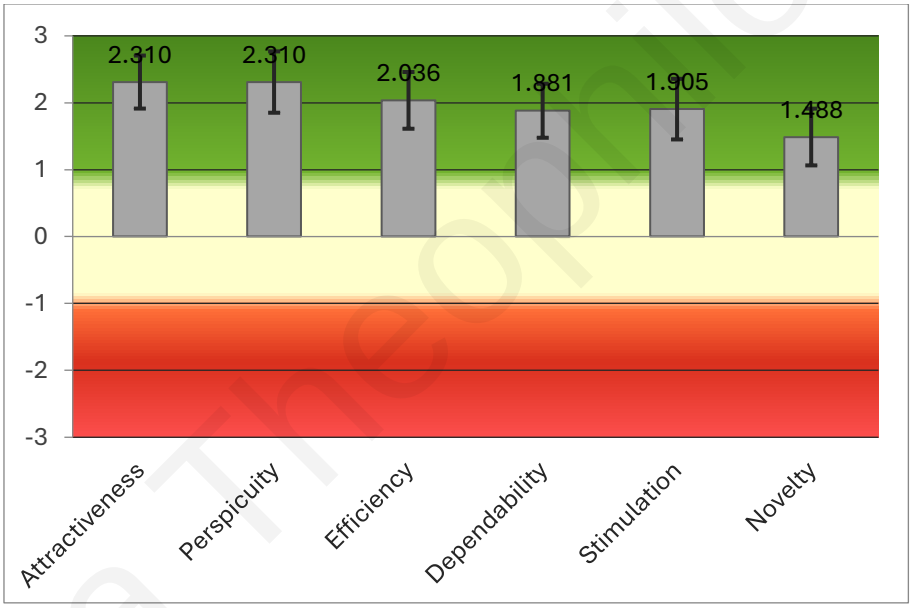
Please enter the data here!

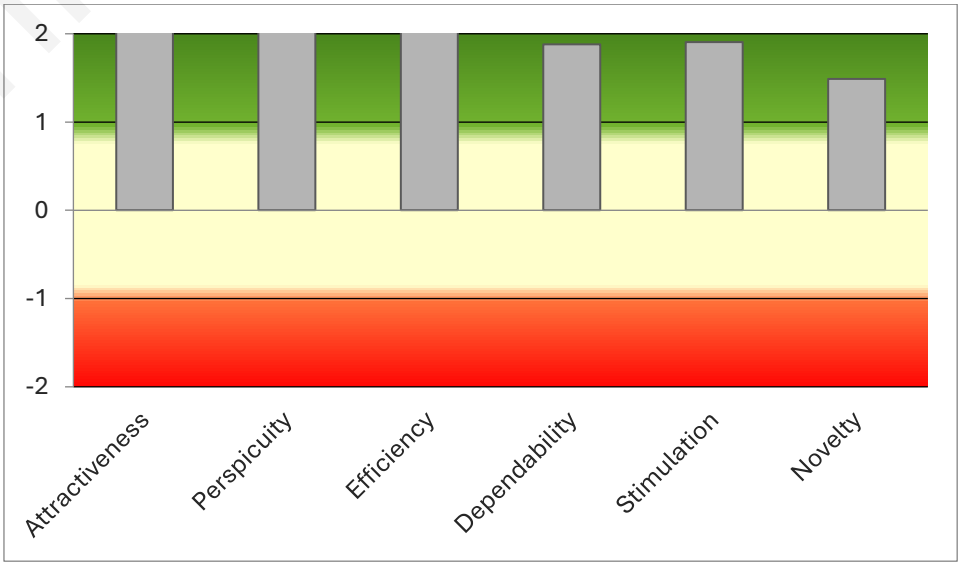
Use the item numbers in the printed questionnaire and the categories 1 (if the alternative on the extreme left is marked) to 7 (if the alternative on the extreme right is marked).
Leave the cell empty if the person has not answered the item. Please do not enter a special caracter in such cases, since this would cause errors in the calculations.
You can enter data for a **maximum of 1000** participants. If you need more, you have to adjust the fomulas in the Excel.
The data analysis tool is delivered with some example data in this work sheet. Simply delete them before you enter your data here.

Items																								
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
6	7	1	1	2	5	6	6	3	2	6	1	7	7	6	6	3	2	1	7	1	6	2	1	1	6
7	7	3	1	2	5	6	7	1	2	7	1	7	6	5	6	3	3	1	7	1	7	1	2	1	5
6	6	1	1	1	6	6	6	1	1	6	1	6	6	5	7	1	5	1	7	1	6	1	1	1	6
4	6	4	1	4	4	5	4	1	3	4	4	5	4	2	4	1	4	4	3	2	4	1	2	2	4
7	7	1	1	1	7	7	7	1	1	7	1	7	7	7	7	1	1	1	7	1	7	1	1	1	7
7	7	2	1	1	6	7	6	2	3	6	1	7	7	6	7	1	1	1	7	1	7	1	1	1	5
6	5	2	3	2	6	6	2	3	2	6	2	5	5	6	6	2	2	2	6	2	5	2	2	2	6
7	7	1	1	1	7	7	7	1	7	7	1	7	7	4	7	1	1	1	7	1	7	1	1	1	7
6	6	4	1	2	4	6	6	7	2	6	1	7	6	5	4	4	3	1	7	1	6	2	2	1	6
7	7	1	1	1	7	7	4	2	7	7	1	7	7	7	7	1	1	1	7	1	7	1	1	1	7
7	7	1	1	1	7	7	2	1	7	7	1	7	7	7	7	1	1	1	1	1	7	1	1	1	7
1	1	1	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1
2	2	1	2	1	2	2	1	1	2	1	1	1	2	2	1	1	2	1	1	1	1	1	2	1	2
7	7	1	1	1	6	6	5	2	2	6	1	6	6	6	6	1	1	1	7	2	7	1	1	1	6
6	7	1	1	1	7	7	7	1	1	7	1	7	6	7	7	2	2	1	7	1	7	1	1	1	7
6	6	2	2	2	6	6	4	3	4	7	1	7	6	6	6	1	2	2	5	3	6	1	1	1	7
5	4	3	5	6	3	5	3	2	2	2	2	6	5	2	5	2	3	2	6	6	7	7	2	2	2
7	7	1	1	1	7	7	6	1	1	7	1	7	7	7	7	1	1	1	7	1	7	1	1	1	7
7	7	1	1	1	4	4	4	3	2	6	1	7	5	4	7	2	1	1	7	1	6	1	1	1	5
7	7	1	1	1	6	7	5	1	1	7	1	7	7	5	7	1	1	1	7	1	7	7	1	1	1
6	6	2	2	2	5	7	4	1	2	7	1	7	7	5	7	1	1	1	7	1	7	1	1	1	6

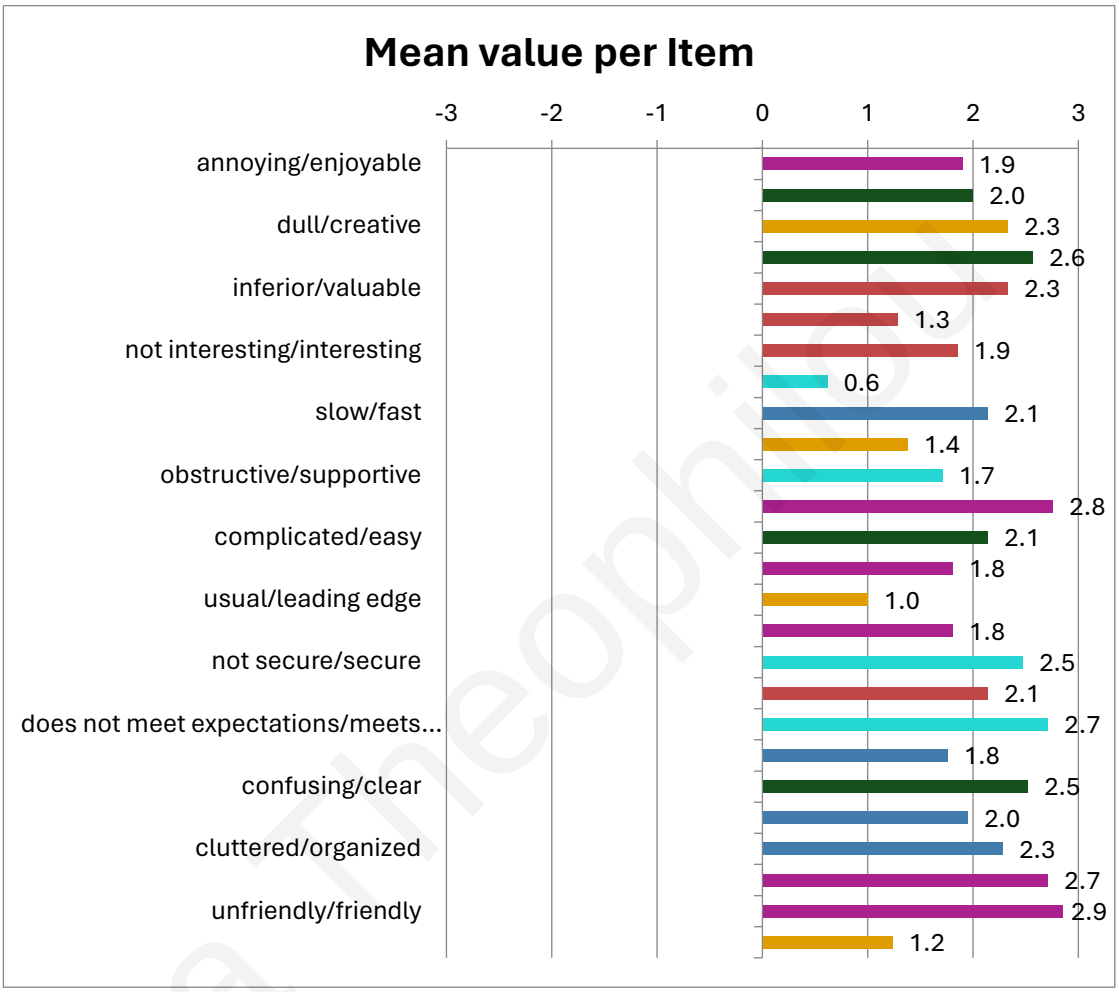
Results :

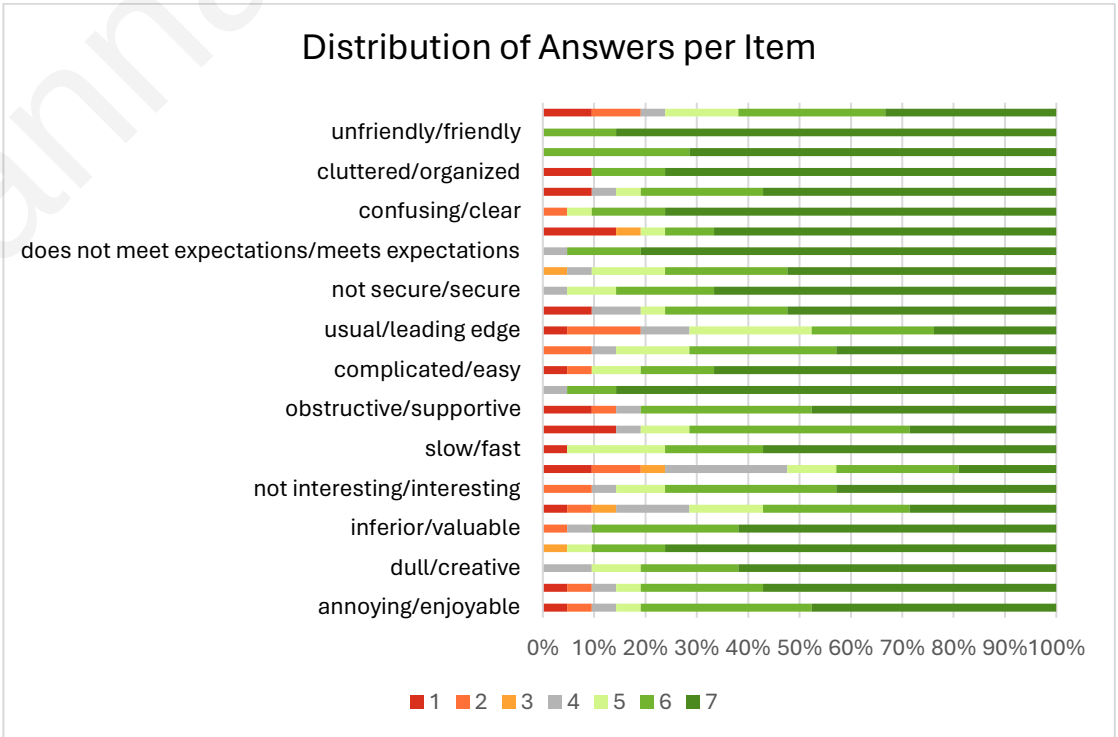
UEQ Scales (Mean and Variance)		
Attractiveness	2.310	0.86
Perspicuity	2.310	1.15
Efficiency	2.036	0.98
Dependability	1.881	0.89
Stimulation	1.905	1.12
Novelty	1.488	0.98



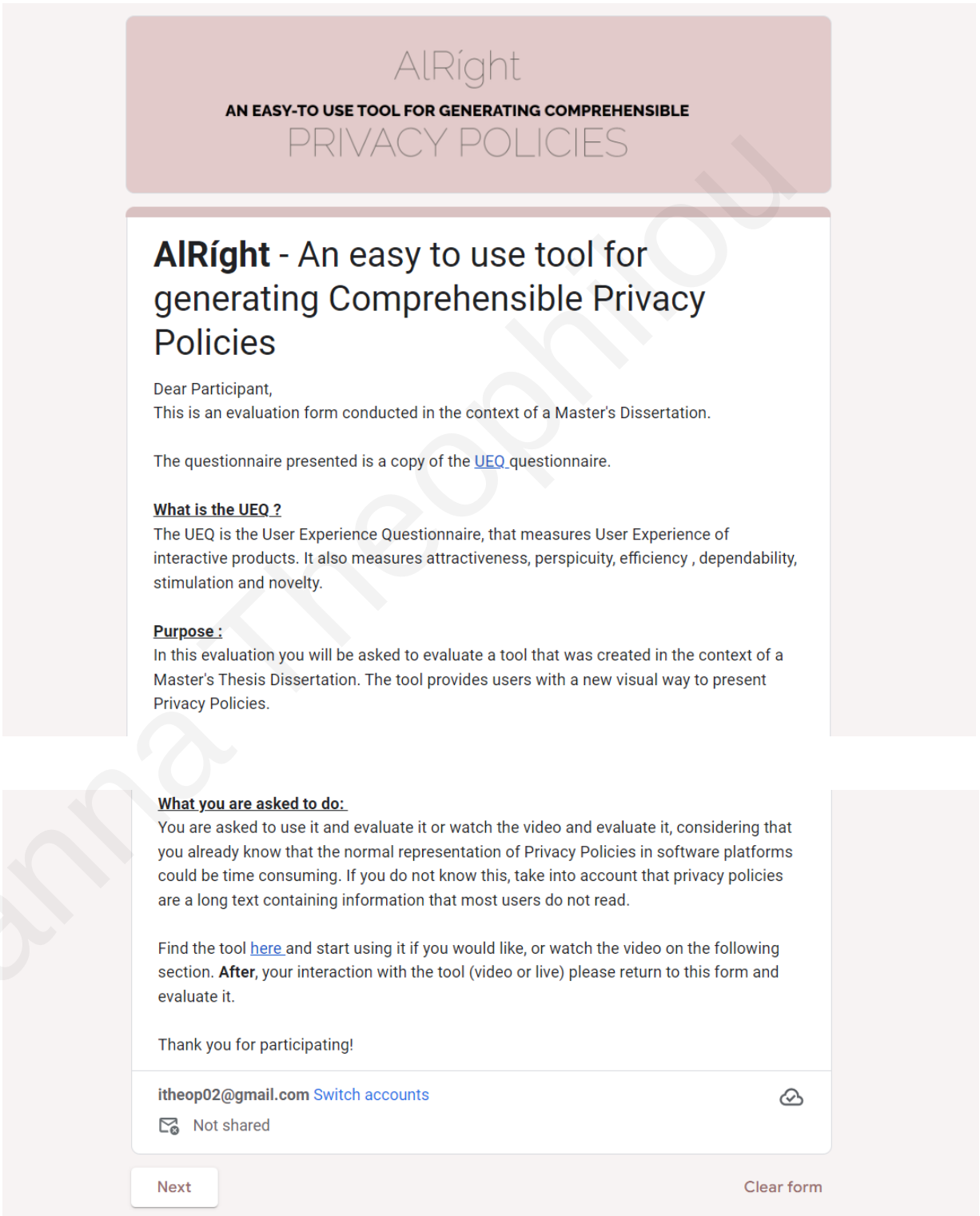


Other Metrics :





ANNEX III



AIRight

AN EASY-TO USE TOOL FOR GENERATING COMPREHENSIBLE PRIVACY POLICIES

AIRight - An easy to use tool for generating Comprehensible Privacy Policies

itheop02@gmail.com Switch accounts

Not shared

* Indicates required question

CONSENT FORM

Informed Consent Form:
This survey does not have any commercial purposes; the involved researchers do not have any monetary benefits by conducting it, and the results will be published in the form of reports and research papers based on the survey. We ask you to provide only anonymous data in this survey. All information will be safeguarded subject to any legal requirements.

This research will involve you completing a number of questions. The whole procedure is estimated to take about **10 minutes**.

By responding to this questionnaire, you confirm the following:
- You have read and understood the purpose of the survey.
- You understand that taking part is voluntary. You can withdraw from the study at any time during the survey, and do not have to give any reasons for why you no longer want to take part.
- You agree that the answers you would give will be stored in digital form. Only the involved researchers will have access to this information, and this information will not be distributed to any other person or entity.

By completing this survey, you are consenting to participate in this study voluntarily. You are indicating that you have read the research description and that you agree to the terms as described.

If you have any comments or questions, please contact the persons responsible:
- Ioanna Theophilou, University of Cyprus, Cyprus (itheop02@ucy.ac.cy)
- Evangelia Vanezi, University of Cyprus, Cyprus (vanezievangelia@gmail.com)

Consent *

- (●) Yes
- ( ) No

Back Next Clear form

AIRight

AN EASY-TO USE TOOL FOR GENERATING COMPREHENSIBLE PRIVACY POLICIES

AIRight - An easy to use tool for generating Comprehensible Privacy Policies

itheop02@gmail.com Switch accounts

Not shared

THE TOOL

You were provided with a video . Please watch it and return.

Back Next Clear form

Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression .

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product.

Nevertheless, please tick a circle in every line.
It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Back Next Clear form

Evaluation

Example

Example:

attractive	○	⊗	○	○	○	○	○	unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please assess the product now by ticking one circle *

	1	2	3	4	5	6	7
Annoying / Enjoyable	○	○	○	○	○	○	○
Not understandable / Understandable	○	○	○	○	○	○	○
Creative / Dull	○	○	○	○	○	○	○

Easy to learn / Difficult to learn	○	○	○	○	○	○	○
Valuable / Inferior	○	○	○	○	○	○	○
Boring / Exciting	○	○	○	○	○	○	○
Not interesting / Interesting	○	○	○	○	○	○	○
Unpredictable / Predictable	○	○	○	○	○	○	○
Fast / Slow	○	○	○	○	○	○	○
Inventive / Conventional	○	○	○	○	○	○	○
Obstructive / Supportive	○	○	○	○	○	○	○
Good / Bad	○	○	○	○	○	○	○
Complicated / Easy	○	○	○	○	○	○	○
Unlikable / Pleasing	○	○	○	○	○	○	○

Link to the form here.

ANNEX IV

AIRíght - An easy to use tool for generating Comprehensible Privacy Policies

Dear Participant,
This is an evaluation form conducted in the context of a Master's Dissertation. The purpose is to evaluate the generated tool regarding generating comprehensible Privacy Policies.

In this evaluation you will be presented with a new visual way to present Privacy Policies and you will be asked to evaluate it, considering that you already know that the normal representation of Privacy Policies in software platforms could be time consuming. Privacy policies are a long text containing information that most users do not read.

Please evaluate the alternative way presented in the next Sections of this survey.

Thank you for participating!

* Indicates required question

CONSENT FORM

Informed Consent Form:
This survey does not have any commercial purposes; the involved researchers do not have any monetary benefits by conducting it, and the results will be published in the form of reports and research papers based on the survey. We ask you to provide only anonymous data in this survey. All information will be safeguarded subject to any legal requirements.

This research will involve you completing a number of questions. The whole procedure is estimated to take about **10 minutes**.

By responding to this questionnaire, you confirm the following:
- You have read and understood the purpose of the survey.
- You understand that taking part is voluntary. You can withdraw from the study at any time during the survey, and do not have to give any reasons for why you no longer want to take part.
- You agree that the answers you would give will be stored in digital form. Only the involved researchers will have access to this information, and this information will not be distributed to any other person or entity.

By completing this survey, you are consenting to participate in this study voluntarily. You are indicating that you have read the research description and that you agree to the terms as described.

If you have any comments or questions, please contact the persons responsible:
- Ioanna Theophilou, University of Cyprus, Cyprus (itheop02@ucy.ac.cy)
- Evangelia Vanezi, University of Cyprus, Cyprus (vanezievangelia@gmail.com)

1. **Consent** *

   Mark only one oval.

   ◯ Yes *Skip to question 2*
   ◯ No

Tool Evaluation - General

Policy 1

If you have any questions as to how we collect and use your personal information, please contact our Customer Service. Many of our PlatformName Services also include settings that provide you with options as to how your information is being used.

As described above, you can choose not to provide certain information, but then you might not be able to take advantage of many of the PlatformName Services.

You can add or update certain information on pages such as those referenced in What Information Can I Access?. When you update information, we usually keep a copy of the prior version for our records.

If you do not want to receive email or other communications from us, please adjust your Customer Communication Preferences. If you don't want to receive in-app notifications from us, please adjust your notification settings in the app or device.

You may adjust your personalized advertising preferences by visiting Your Ads Privacy Choices.

The Help feature on most browsers and devices will tell you how to prevent your browser or device from accepting new cookies or other identifiers, how to have the browser notify you when you receive a new cookie, or how to block cookies altogether. Because cookies and identifiers allow you to take advantage of some essential features of PlatformName Services, we recommend that you leave them turned on. For instance, if you block or otherwise reject our cookies, you will not be able to add items to your Shopping Cart, proceed to Checkout, or use any Services that require you to Sign in. For more information about cookies and other identifiers, see our Cookies Notice.

If you want to browse our websites without linking the browsing history to your account, you may do so by logging out of your account here and blocking cookies on your browser.

You can manage the recommendations you receive in our store here, remove recommendations you don't want to see here by

Policy 2

2. **1. Above the same policy is presented in two different ways.** Which way do you think it would be more attractive to read? *

Mark only one oval.

- ◯ Policy 1
- ◯ Policy 2

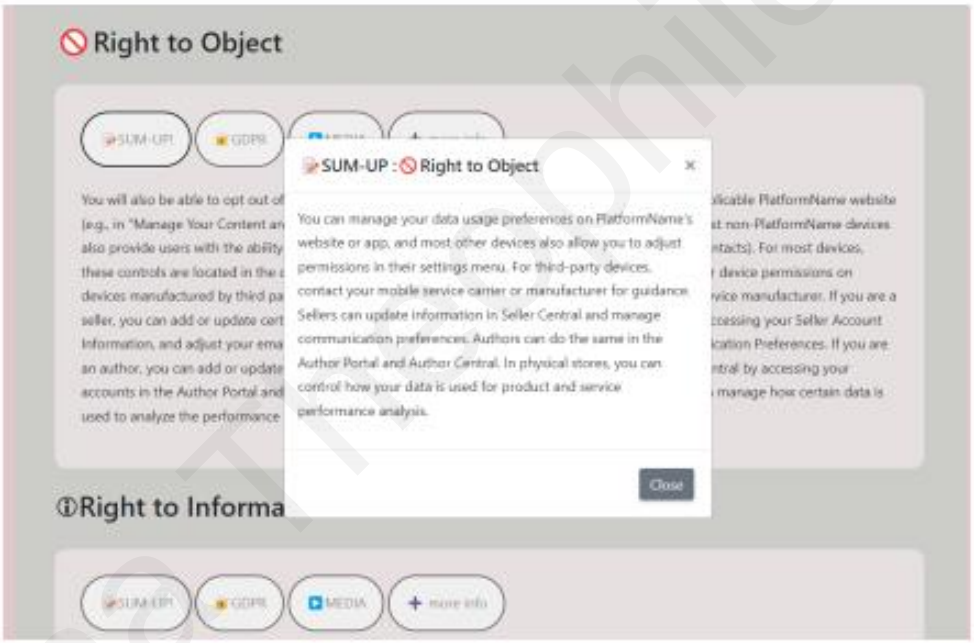
3. **2.** Which policy do you believe would be more easy to understand? *

Mark only one oval.

- ◯ Policy 1
- ◯ Policy 2

Tool Evaluation - different features

Figure 1 : Summary



4. **3.** In the figure above (figure 1), the actual content of the privacy policy is presented but when the '***SUMUP***' button is pressed, a summary of the same text appears. *

Would you prefer reading the summary instead of the text?

Mark only one oval.

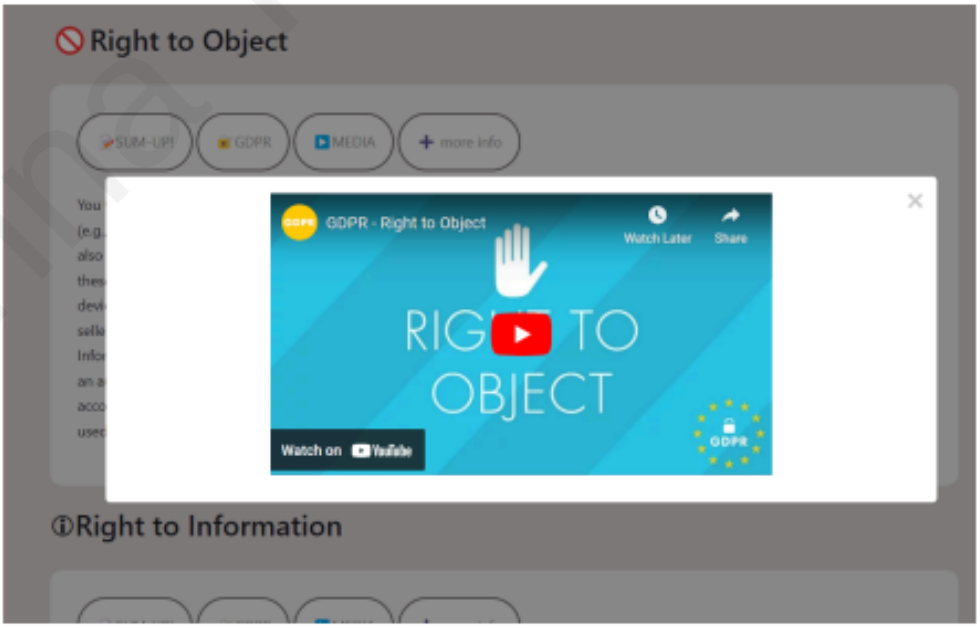
- ◯ No
- ◯ Neutral (Would not mind)
- ◯ Yes

5. **4**. Do you think that having the option of reading the summary is helpful ? *

Mark only one oval.

- ◯ No
- ◯ Maybe
- ◯ Yes

Figure 2 : Media



6. **5.** In the figure above (figure 2), it can be seen that the user pressed the '***MEDIA***' button. *

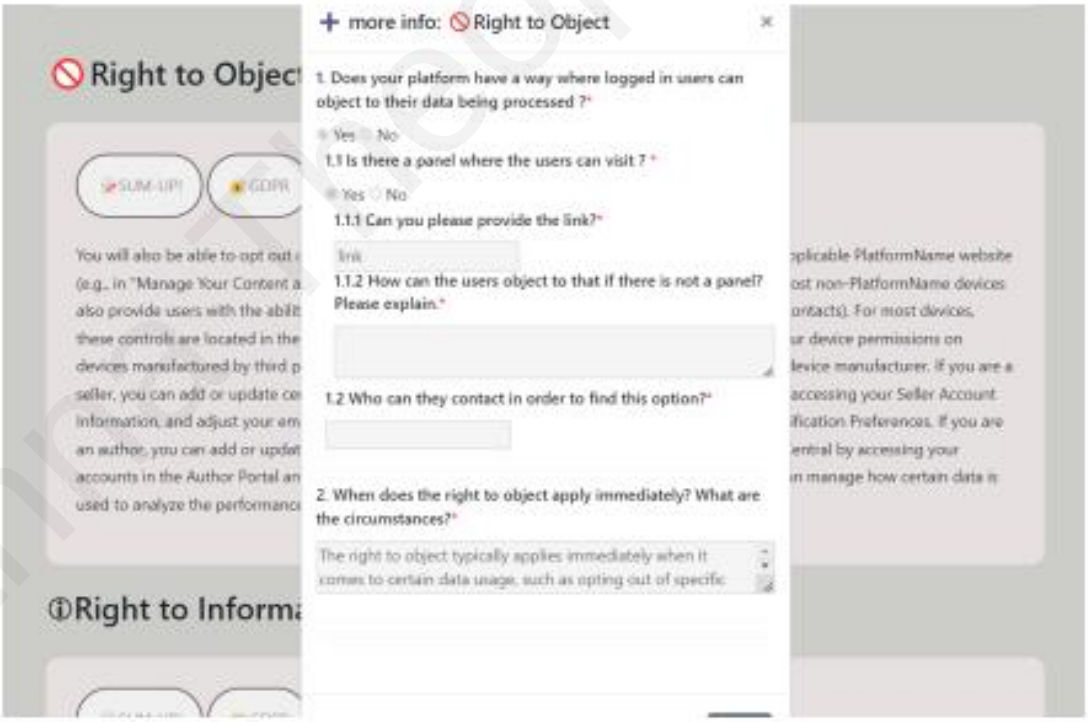
A popup showed up with an explanatory video.

Do you find this feature helpful in understanding legal regulations?

Mark only one oval.

- ◯ No
- ◯ Maybe
- ◯ Yes

Figure 3: More information



7. **6.** In figure 3, the user pressed the **'MORE INFO'** button. *

   A popup showed up with questions that the Data Controller Entity answered.

   Would you read them ?

   Mark only one oval.

   - ◯ No
   - ◯ Maybe
   - ◯ Yes

8. **7.** Do you think that reading these supplementary material (***the answered questions***) would make Privacy Policy comprehension easier ? *

   Mark only one oval.

   - ◯ No
   - ◯ Maybe
   - ◯ Yes

Figure 4: GDPR

9. **8.** Figure 4 shows that the user pressed the 'GDPR' button in order to review the actual regulation. A new window appeared with the corresponding article. *

Would you use this button ?

Mark only one oval.

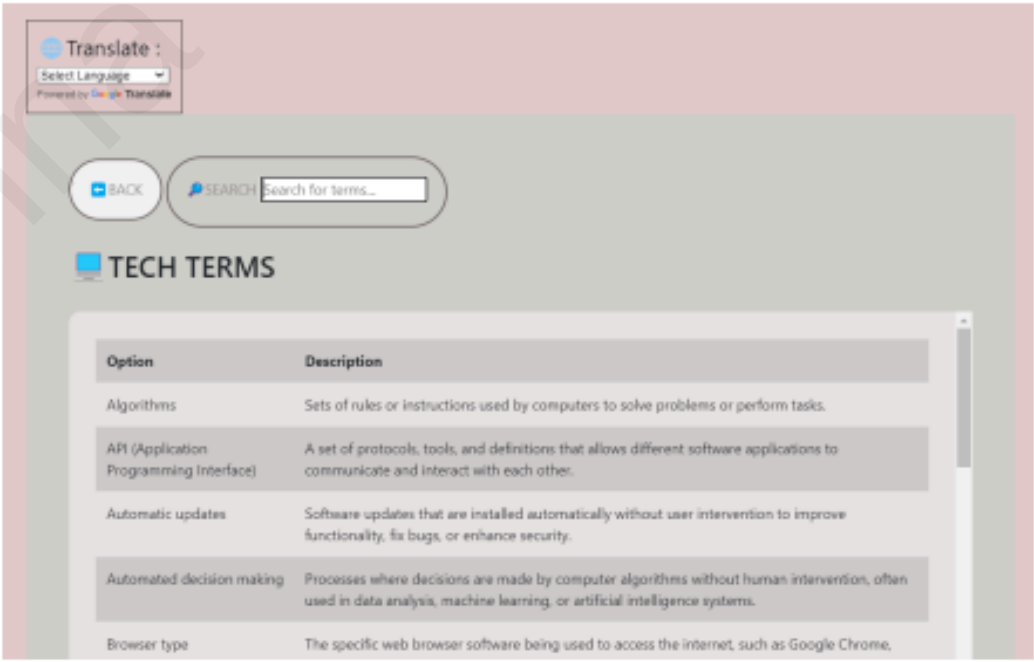
- ◯ No
- ◯ Maybe
- ◯ Yes

10. **9**. Do you believe that having the option to read the actual regulation of the specific right is helpful in understanding it better? *

Mark only one oval.

- ◯ No
- ◯ Maybe
- ◯ Yes

Figure 5: Technical Terminology



11. **10**. The user in figure 5 pressed the '**TECH TERMS**' button. *

This button provides a list of technical terms used in privacy policies and a brief explanation for each one.

It also provides a search engine, where users can type the terminology they would like to learn about.

Would you use this ?

Mark only one oval.

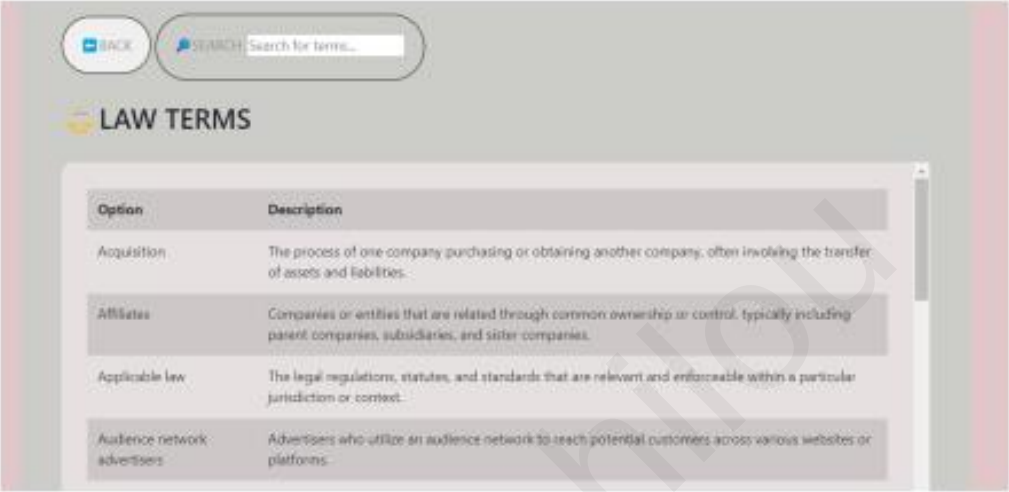
- ◯ No
- ◯ Maybe
- ◯ Yes

12. **11.** Do you believe that having this list of technical words along with the search engine would improve comprehension of Privacy Policy documents ? *

Mark only one oval.

- ◯ No
- ◯ Maybe
- ◯ Yes

Figure 6: Legal Terminology



13. **12**. The user in figure 6 pressed the '**LAW TERMS**' button. *

This button provides a list of legal terms used in privacy policies and a brief explanation for each one.

It also provides a search engine, where users can type the terminology they would like to learn about.

Would you use this ?

Mark only one oval.

- No
- Maybe
- Yes

Link to the form here.