

# THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of  
Master of Business Administration

Presented to the

Swiss School of Management



## AN ANALYSIS OF THE READABILITY OF PRIVACY NOTICES

by

Kashfun Nazir

Singapore

22IND-M06001

Reviewer: Amit Kumar Padhy

April, 2024

# **Abstract**

The proliferation of data-driven services has placed privacy notices at the forefront of consumer awareness and regulatory scrutiny. This thesis examines the readability and efficacy of privacy policies from the top 100 organisations listed in Forbes' Global 500 companies in 2023. This study addresses the growing complexity of privacy notices and the challenges users face in understanding the terms to which they are consenting. Through a quantitative analysis, this research evaluates the length, reading level, and time required to read privacy notices, while exploring variances across different industries and regions. The findings suggest a significant readability challenge, emphasising a need for clearer communication and standardisation to facilitate informed consent. The implications of this thesis extend to legal and regulatory frameworks, as well as organisational practices around transparency and user education. This critical analysis aims to pave the way for more user-friendly privacy policies that foster trust and empower users in the digital economy.

Keywords: privacy notices, readability, hemingway, forbes global 500

# Acknowledgements

The entire process of writing this thesis was an experience like never before. I have learned a lot and gained plenty of knowledge and experience from this experience. Completing this thesis gave me a huge sense of accomplishment that I would never forget.

I would firstly like to thank my wife for being very supportive and understanding throughout my journey of attaining my Master's. It was not easy juggling time while having young kids in the house, always in constant need of attention. My wife has sacrificed a lot for me so that I could complete my thesis, especially towards the last lap. I would also like to thank my parents and siblings for believing in me and providing me with moral support, especially during the many times I felt like giving up.

Next, I would like to thank my thesis advisor, Amit Kumar Padhy for his unwavering support and patience in me. Having changed my topic a few times, I can only imagine how difficult of a student I must have been to him. Despite that, he did not cast me aside and continued to give me the support I required.

Lastly, I would like to give thanks to my course mates and colleagues for their support and ideas. They have made my journey more fun and memorable.

**Kashfun Nazir**

# Content

Abstract.....	i
Acknowledgements.....	ii
1. Introduction .....	8
1.1. Research Motivation .....	8
1.1.1. Privacy Notices.....	9
1.1.2. Legal Requirements .....	9
1.1.3. Consumer Trust .....	10
1.1.4. Technological Advancements .....	10
1.1.5. Readability .....	10
1.2. Research Question .....	11
1.3. Research Challenges .....	11
1.4. Research Scope.....	11
1.5. Research Hypotheses.....	12
1.5.1. Hypothesis #1 .....	12
1.5.2. Hypothesis #2.....	12
1.6. Research Goal .....	12
1.7. Research Relevance .....	12
1.7.1. Practical Implications .....	12
1.7.2. Theoretical Implications .....	13
1.8. Research Design.....	13
1.9. Research Summary .....	13
2. Background.....	15
2.1. Privacy Notices.....	15
2.1.1. Complexity of Privacy Notices .....	15
2.2. Readability .....	16

2.2.2. Chosen Tool and Method.....	18
2.2.3. Attention span.....	20
2.2.4. Retention .....	20
2.2.5. Reading time .....	21
2.2.6. Word count .....	22
2.2.7. Sentence count.....	22
2.3. Forbes Global 500.....	23
3. Method .....	26
3.1. Sample List .....	26
3.2. Privacy Notices Collection.....	36
3.3. Metrics Gathering.....	39
3.4. Data Cleanup .....	39
3.5. Data Analysis .....	39
4. Results and Analysis.....	41
4.1. Industry .....	41
4.1.1. Reading Time by Industry .....	43
4.1.2. Word Count by Industry .....	43
4.1.3. Sentence Count by Industry .....	43
4.2. Country.....	43
4.2.1. Reading Time by Country .....	45
4.2.2. Word Count by Country .....	45
4.2.3. Sentence Count by Country .....	45
4.3. Grade Level of Readability.....	45
4.3.1. Reading Time by Grade Level Of Readability .....	47
4.3.2. Word Count by Grade Level Of Readability.....	47
4.3.3. Sentence Count by Grade level of readability .....	47

5. Discussion.....	48
5.1. Industry .....	48
5.2. Countries .....	49
5.3. Grade Level of Readability .....	50
5.4. Hypotheses .....	50
5.4.1. Hypothesis #1 .....	50
5.4.2. Hypothesis #2 .....	51
5.5. Recommendations .....	51
5.6. Limitations .....	51
6. Conclusion .....	53
References .....	54

## List of Tables

Table 1: Final sample list .....	36
Table 2: Results by Industry .....	43
Table 3: Results by Country .....	45
Table 4: Results by Grade Level of Readability .....	46

# 1. Introduction

Privacy notices, also commonly known as privacy policies, privacy statements or data protection statements, are crucial documents that inform individuals about how their personal information is collected, used, and protected by organisations (Tikkinen-Piri et al., 2018). In theory, these documents should be read and understood by every user of the service, product, website, etc. Over the years, there has been growing concern that privacy notices are lengthy and use complex language with legal jargons that make it unnecessarily difficult for laymen to understand.

However, many individuals resort to skimming through or ignoring the privacy notice in its entirety because of its length and complexity. As a result, this potentially puts the user's personal information at risk, as they may unknowingly agree to terms they do not fully comprehend (Alsulaiman & Alrodhan, 2014).

With more privacy laws enacted worldwide, such as the General Data Protection Regulation (GDPR) in the European Union (EU), individuals must have a clear understanding of their rights. While there are no standards or requirements to ensure that privacy notices are readable by a layman, organisations are encouraged to use plain and simple language so as to improve the readability of privacy notices and help facilitate informed decision-making by their users regarding their personal information.

## 1.1. Research Motivation

With the ever-evolving technology, the collection and sharing of data have been made a lot easier. Users have also become more aware of the potential dangers and risks of sharing their personal information online. The need for transparency and informed consent in data handling practices has become paramount.

In this thesis, we will analyse the readability of the privacy notices made available to the general public of the top 100 organisations from Forbes' list of Global 500 companies in 2023.

According to (Goldstein et al., 2011), there is an increasing demand for better content accessibility. This demand is partly driven by societal awareness of the needs of people with disabilities, as well as by increasing legal requirements worldwide to make digital content accessible. In the context of privacy notices, accessibility refers not only



to providing clear and concise information but also to ensuring that users with different levels of understanding are able to comprehend and make informed decisions about their personal information.

Crossler, 2011, found that concerns have been raised that privacy notices are often "excessively lengthy, dense, and packed with details," making them challenging for users to understand. The content of these notices frequently becomes "overly theoretical, technical, or steeped in legal jargon," which can prevent users from fully understanding the implications of their privacy choices. In addition, a lack of standardisation of data protection regulations across jurisdictions globally adds to the complexity.

Efforts are being made to standardise content accessibility regulations on a global scale. For example, the uptake of the Web Content Accessibility Guidelines 2.1 in Europe, seeks to improve the accessibility of web content for all users (Abou-Zahra, 2018). There are similar efforts globally, such as that by the World Wide Web Consortium (W3C) Web Accessibility Initiative which develops web accessibility standards to address and integrate accessibility into web technologies (Essential Components of Web Accessibility, 2018).

As privacy awareness increases with time, the demand for more accessible and readable privacy notices will also increase.

### **1.1.1. Privacy Notices**

Privacy notices became more common with the rise of the internet and online services, particularly during the late 1990s and early 2000s. They came about in response to growing concerns about personal data protection as more private information began to be collected, stored, and processed digitally. The reasons for the rise in need of privacy notices include:

### **1.1.2. Legal Requirements**

The implementation of data protection laws, such as the European Union's Data Protection Directive in 1995 (and later the General Data Protection Regulation in 2016), mandated that organisations inform users about the collection and use of their personal data. Similar laws in other jurisdictions, such as the Children's Online Privacy

Protection Act in 1998 in the United States, also placed emphasis on the importance of privacy notices.

### **1.1.3. Consumer Trust**

Organisations recognised the need to communicate their privacy practices to users to build trust and credibility. As users became more aware of their digital footprint, transparency about data practices became an important factor in establishing and maintaining customer relationships.

### **1.1.4. Technological Advancements**

With the advancement of online technologies, organisations equip themselves with tools to collect vast amounts of data. This led to greater scrutiny regarding what data was collected and how it was used, fostering the need for clear privacy notices.

### **1.1.5. Readability**

Readability refers to how easy it is for a reader to understand written text. It is an objective measure of the reading skills an individual must possess to comprehend material comfortably. Readability is often expressed in terms of "grade levels"; for example, a text with a ninth-grade readability level should be understandable to someone with a ninth-grade reading level or higher. Factors that influence readability include sentence length, word complexity, grammar, vocabulary, and the structure of the text.

Poor readability can pose barriers to understanding, particularly in areas such as healthcare, where it is crucial that patients grasp medical information to make informed decisions (Badarudeen & Sabharwal, 2010). Other areas include information security and privacy, where it is of significant importance that the intended users are aware of the dangers and risks of sharing their data with any party.

It is important for writers, content creators, policymakers and alike, especially of educational or technical content, to consider readability so their messages are accessible to a broad audience (Fanning et al., 2023).

There are tools and formulas that measure text readability to help ensure materials can be understood by the intended audience. An example of such a tool is the

Hemingway Editor, which was used to assess the readability of privacy notices in this thesis.

## **1.2. Research Question**

To what extent are privacy notices readable and understandable for the general public?

To investigate the readability and understandability of privacy notices, the following sub-questions will be addressed:

- How lengthy in word and sentence count are the privacy notices?
- What reading level or grade are the privacy notices written at?
- How much time does it take to read the content of the privacy notices?

## **1.3. Research Challenges**

One of the main challenges that researchers face when studying the readability of privacy notices is that there is no universally agreed-upon standard for writing or presenting a privacy notice. This makes it difficult to directly compare one notice with another.

Another challenge is the varied language in which the privacy notice is written due to the geographic location of the organisation and its target audience. These variations in language can affect the readability of the notice for different individuals.

## **1.4. Research Scope**

This research was carried out acknowledging that there are certain limitations and was scoped as follows:

- The sample for this research was limited to the list of the top 100 organisations of the Forbes' Global 500 Companies list in 2023 which was published in early 2024.
- The chosen research sample may not be representative of privacy notices across all industries or regions.
- The samples were also limited to only those written in English language, no translation or analysis of privacy notices in other languages was conducted.

This research also relied on the use of Hemingway, a text readability tool, to assess the readability of the sampled privacy notices.

## **1.5. Research Hypotheses**

This thesis follows an inductive research approach. The research hypotheses were derived from the data collected.

### **1.5.1. Hypothesis #1**

The readability of privacy notices for the top 100 organisations from Forbes' list of Global 500 companies in 2023 is significantly low.

### **1.5.2. Hypothesis #2**

There is significant variance in the readability of privacy notices across different industries and geographic regions.

## **1.6. Research Goal**

The goal of this thesis is to analyse the readability of privacy notices for the top 100 organisations from Forbes' list of Global 500 companies in 2023. This will be achieved by assessing the length, reading level or grade, and time required to read the content of the privacy notices. The research also aims to explore the variance in the readability of privacy notices across different industries and geographic regions in English language.

Ultimately, the goal is to provide insights into the current state of privacy notices' readability for the general public and to identify potential areas for improvement in making privacy notices more accessible and comprehensible.

## **1.7. Research Relevance**

This research hopes to be of valuable use to organisations in pursuit of building trust and transparency with their target audience through more user-friendly written privacy notices.

### **1.7.1. Practical Implications**

The findings of this research can be used by organisations to understand the current state of privacy notices' readability and the factors that influence it. This can help them make informed decisions in terms of improving the readability of their privacy notices, such as simplifying language, organising information more effectively and developing user-friendly formats.

### **1.7.2. Theoretical Implications**

The findings of this research extend beyond private organisations, as it may also inform policymakers in developing regulations and guidelines for privacy notices.

## **1.8. Research Design**

As the purpose of this thesis is to understand the current state of privacy notices' readability for the general public through the analysis of the privacy notices of the top 100 organisations on Forbes' Global 500 Companies list in 2023, a quantitative research design was adopted.

## **1.9. Research Summary**

The goal of this thesis was to analyse the length, reading level, and time required to read the content of the privacy notices, along with exploring variance in readability across different industries and regions.

This thesis aimed to address the challenges related to the lack of standardisation in privacy notice writing and the variations in language due to geographical locations. The scope of this thesis was limited to the top 100 organisations on Forbes' Global 500 Companies list in 2023, focusing exclusively on privacy notices written in English. The research hypotheses suggested that the readability of privacy notices for these organisations would be significantly low and that there would be significant variance in readability across different industries and geographic regions.

The relevance of the research was highlighted in its potential value to organisations in building trust and transparency with their target audience through more user-friendly privacy notices. The findings, it was suggested, could have practical implications for organisations in improving the readability of their privacy notices. Furthermore, the

theoretical implications were mentioned, indicating how the findings could inform policymakers in developing regulations and guidelines for privacy notices.

To achieve these goals, a quantitative research design was adopted to analyse the readability of privacy notices. The findings of this thesis were expected to shed light on the current state of privacy notices' readability, aiming to identify potential areas for improvement in making privacy notices more accessible and comprehensible.

## **2. Background**

### **2.1. Privacy Notices**

One of the commonly publicly available documents by almost every organisation today is the privacy notice. Privacy notices are not only important for the users but for the organisation as well. There are many reasons why privacy notices have almost become a mandatory document.

A study found that privacy notices offer transparency to users by informing the individuals about the types of personal information that are being collected and the purposes for which this information is used. This helps to build and foster trust between individuals and organizations by disclosing the organisation's privacy practices. Users may feel more comfortable sharing personal information if they are informed about what will happen to their data (Anwar et al., 2020).

Another study found that many jurisdictions require organizations to provide privacy notices as part of legal compliance with data protection laws, such as the General Data Protection Regulation in the European Union. Privacy notices also serve as internal guidelines for organizations, detailing the proper handling and protection of personal data (Milne et al., 2006).

While it may be more commonly mandated by regulations around the globe, there are still some jurisdictions that do not mandate privacy notices, leaving it up to the discretion of organisations to provide them.

#### **2.1.1. Complexity of Privacy Notices**

Unfortunately, the complexity of privacy notices has become a growing concern. Some factors that contribute to its complexity include the use of legal jargon, the need to cover multiple jurisdictions' requirements, comprehensive details and technical language that may be difficult for ordinary users to understand. To add on, privacy notices are often written for a broad audience, to be applicable to a wide range of scenarios.

Milne et al., 2006, evaluated the readability of online privacy notices found that they had become more complex over time, suggesting that efforts to make these

documents more accessible have not been entirely successful. Another perspective suggests that while it's necessary to be transparent about data practices, presenting everything at once may not be effective; hence, a multilayered approach is recommended to balance comprehensiveness and understandability.

## **2.2. Readability**

Readability refers to the ease with which a reader can understand a written text. The concept of readability encompasses not just the legibility of letters and words, but also the structure of sentences, the complexity of vocabulary, and the overall style of writing that contributes to how quickly and thoroughly a reader can comprehend the material. Readability can be measured using specific formulas designed to estimate the level of education needed to understand the text, such as the Flesch Reading Ease Score, Flesch-Kincaid Grade Level, Gunning Fog Index, Coleman-Liau Index, and others.

For example, in the study "Critical Analysis of the Quality, Readability, and Technical Aspects of Online Information Provided for Neck-Lifts" by Hani Rayess, Giancarlo Zuliani, Amar Gupta, et al., readability assessments were used to analyze the quality, readability, and technical aspects of online material. This study employed several readability tools, including the Flesch Reading Ease Score and the Flesch-Kincaid Grade Level, to determine the complexity of the information provided about neck-lift procedures (Rayess et al., 2017).

Similarly, "Content and Readability of Online Recommendations for Breast Implant Size Selection" by James E. Fanning, Lauren A. Okamoto, Emma C. Levine, et al., discussed the importance of readability in medical information shared online. The paper indicated that while online medical information is significant for pre-operative planning, it should be accessible and readable to a broad patient audience (Fanning et al., 2023).

In the education sector, "Readability, content, and mechanical feature analysis of selected commercial science textbooks intended for third grade Filipino learners" by Apler J. Bansiong focused on assessing the readability of science textbooks for third graders (Bansiong, 2019).



These examples illustrate how readability is a key factor in analyzing and designing communications to be understood by the intended audience.

### Tools and Methods for Assessing Readability

Several tools and methods have been designed to assess the readability of texts, each with its own set of metrics for quantifying the ease with which a text can be read and understood by the intended audience. Below are descriptions of some common tools and methods

#### **2.2.1.1. Automated Readability Index**

The ARI is a readability test that calculates an approximate representation of the US grade level needed to comprehend a text. It uses character counts, word counts, and sentence lengths to produce a score that corresponds to a grade level .

#### **2.2.1.2. Flesch Reading Ease/Flesch-Kincaid Grade Level**

Developed by Rudolph Flesch, these readability tests are widely used. The Flesch Reading Ease score rates texts on a 100-point scale; the higher the score, the easier it is to understand the document. Conversely, the Flesch-Kincaid Grade Level formula translates the score into a US grade level .

#### **2.2.1.3. Gunning Fog Index**

This tool measures the complexity of English writing and gives a grade level that a person must have reached to understand the text upon first reading..

#### **2.2.1.4. Hemingway Editor**

A digital writing tool that highlights complex sentences, use of passive voice, and overuse of adverbs. It also provides a readability grade for the overall text .

#### **2.2.1.5. Coleman-Liau Index**

Similar to the ARI, the Coleman-Liau Index predicts the grade level required to understand a piece of text based on characters instead of syllables .

#### **2.2.1.6. SMOG Index**

The SMOG formula estimates the years of education needed to understand a piece of writing. It focuses on polysyllabic words, those with three or more syllables .

### **2.2.1.7. Linsear Write Formula**

This method is specifically designed for English texts and calculates a text's US grade level readability based on sentence length and the number of easy or difficult words .

Each of these methods provides different insights into the readability of texts and can be used in various combinations to achieve a comprehensive analysis of textual material. To find citations of specific studies or articles that use these tools, you could search in academic databases or journals related to linguistics, education, or communication.

## **2.2.2. Chosen Tool and Method**

For this thesis, the chose tool to assess readability of privacy notices is Hemingway Editor and the method that Hemingway Editor uses is the Automated Readability Index.

### **2.2.2.1. Hemingway Editor**

The Hemingway Editor is a free online web application that is available for public use. It is an editor that helps improve readability through analysing the number of adverbs, instances of passive voice, words or phrases, sentences that are hard to read and sentences that are very hard to read.

The Hemingway Editor provides recommendations to help users make their writing more direct and understandable. Here are some of its key recommendations:

- **Use Simple Language:** It suggests replacing complex words and phrases with simpler alternatives.
- **Avoid Adverbs:** The editor recommends minimizing the use of adverbs, especially those that can weaken the prose, like "very" or "really."
- **Use Active Voice:** Hemingway Editor encourages the use of active voice over passive voice to make sentences clearer and more dynamic.
- **Keep Sentences Short:** It highlights long, complex sentences and recommends breaking them into shorter, more digestible ones.
- **Limit Use of Qualifiers:** Just like with adverbs, the tool advises against using qualifiers that can dilute the message, such as "just" or "almost."
- **Grade Level:** The editor provides a readability grade level that indicates the complexity of the text based on these factors. The goal is to keep the grade

level relatively low to ensure a wider audience can easily understand the content. It adopts the method, Automated Readability Index, although, it has additional grades from 1 through 16 and anything above that would be classified as Post-graduate grade.

#### **2.2.2.2. Automated Readability Index**

The Automated Readability Index (ARI) is a readability formula designed to gauge the understandability of a text. The formula estimates the US grade level needed to comprehend the text. According to the ARI, the recommended grade levels for text are as follows:

- A score of 1-3 corresponds to elementary school levels (ages 6-9).
- A score of 4-6 corresponds to middle school levels (ages 9-12).
- A score of 7-8 corresponds to early high school levels (ages 12-14).
- A score of 9-10 represents standard high school level (ages 14-16).
- A score of 11-12 corresponds to a junior or senior high school level (ages 16-18).
- A score above 13 is considered college level.

The ARI is intended to ensure materials are written in a style that is appropriate for the intended audience. For general audiences and public materials, it is often recommended to aim for a readability level around grade 8 or below to ensure they are accessible and understandable to a wide audience .

The recommendation for materials to be written at a 6th to 8th grade reading level for general audiences, which may align with some uses of the Automated Readability Index, comes from health communication research rather than exclusively from the ARI itself. For instance, the American Medical Association and the National Institutes of Health recommend that patient information materials should not exceed a 6th-grade reading level, as the average reading skill of US adults is at the 8th grade level . This recommendation is based on the recognition that materials written at lower reading levels are more likely to be understood by a broader audience, including those with varying levels of literacy and education.

While this research and recommendations focus primarily on health literature and patient information, the principle behind the recommendation—to enhance

comprehension and accessibility—is widely applicable across different types of texts intended for public consumption (Eltorai et al., 2015).

### **2.2.3. Attention span**

Attention span is a factor that greatly affects the readability of privacy notices. It is generally known that the average attention span of an average person reduces over time, making it more challenging for individuals to focus and comprehend lengthy and complex texts.

In the study "Effect of online social networking on student academic performance" by Jomon Aliyas Paul, Hope Baker, and Justin Cochran, it is noted that the average attention span seems to be lessening, and students often exhibit frustration and seek help before thoroughly reading the supporting text and working through problems themselves (Paul et al., 2012). This may indicate a trend in decreased attention spans in academic settings related to reading and comprehension, especially among students who are balancing multiple demands on their time.

It's important to note that the study doesn't provide a specific average time for attention span during reading, and it does not claim these findings are representative of the general population. Furthermore, attention spans can vary widely based on the type of reading material and the individual's interest and expertise in the subject.

### **2.2.4. Retention**

Retention is another important factor that affects the readability of privacy notices. Several factors can influence retention in reading and comprehension.

Xu and Durgunoglu found, in 2019, that cognitive abilities such as working memory, attention, and prior knowledge contribute significantly to how well someone can retain information from reading.

Another study found that factors such as motivation, text difficulty, reading strategies, decoding skills, vocabulary, medical and physical conditions are contributing factors to retention (Gilakjani & Sabouri, 2016).

**Motivation:** A reader's interest and motivation are highly important in developing reading comprehension skills. If a reader is interested and motivated, they are more likely to engage deeply with the material and remember it.

**Text Difficulty:** The complexity of the text can affect retention. Texts that are too difficult may hinder understanding and recall, while texts that are too easy may not be engaging enough for proper learning to take place.

**Reading Strategies:** Effective reading strategies such as summarizing, questioning, clarifying, and predicting can help improve comprehension and retention. These strategies actively involve the reader with the text, thus fostering better recall.

**Decoding Skills and Vocabulary:** Proficiency in recognizing words quickly and understanding their meanings can lead to better comprehension and retention, as less cognitive effort is spent on decoding words and more on understanding and integrating content.

**Medical and Physical Conditions:** Medical conditions such as speech and hearing difficulties can affect reading comprehension and subsequent retention. Learners with these issues may be less likely to engage actively with reading material, impacting their ability to retain information.

### **2.2.5. Reading time**

Reading time is an estimate of how long it will take a reader to read through a piece of written content. It is often calculated based on the average reading speed of an adult (typically around 200-250 words per minute) but can vary depending on the complexity of the text and the familiarity of the reader with the content.

The relationship between reading time and readability revolves around how quickly and easily a text can be understood and processed by a reader. If a text is complex, uses a lot of technical jargon, or has long sentences, it will generally take longer to read because readers may need to reread the content to fully comprehend it. Conversely, if a text is clear, well-structured, and uses simpler language, it will be easier to understand, thus reducing the time needed to read it.

Many readability formulas incorporate sentence length and word difficulty, which indirectly influences the estimated reading time. For instance, texts with a lower

Flesch-Kincaid Grade Level, which implies higher readability, will on average take less time to read because they are easier to understand.

Readability is a pivotal factor in determining how engaging a text is for a reader, and influencing whether they will actually spend the time to read through it completely. As with the information provided in the "Content and Readability of Online Recommendations for Breast Implant Size Selection" study, the average time to read each webpage was found to be 2 minutes and 57 seconds, showing that readability assessments take reading time into account, and shorter reading times can be a sign of more accessible content (Fanning et al., 2023).

### **2.2.6. Word count**

Word count, in relation to readability, impacts how digestible the text is for the reader. Shorter word counts generally imply conciseness, which can enhance readability by making the content less daunting and helping readers grasp concepts quickly. The readability of a text increases when it's concise and clear, without unnecessary verbosity.

Moreover, research has shown that there is a positive correlation between the number of citations in a paper and the word count of the paper (Knight-Davis & Sung, 2008). This implies that longer documents may provide more detailed information and be more informative. However, longer word counts can also reduce readability if they lead to overly complex and dense information that is hard for the reader to process.

Rayess et al., 2017, mentioned that it is important to find a balance; text should be long enough to be informative and thorough, but also concise enough to maintain readability and reader engagement. Readability formulas like the Flesch-Kincaid Grade Level can help assess how accessible a text will be for a specific audience based on sentence length and word difficulty.

### **2.2.7. Sentence count**

Sentence count contributes to readability by affecting the density of information and the cognitive load on the reader. Shorter sentences can make text easier to understand because they tend to contain simpler structures and clearer ideas. The

Gunning Frequency of Gobbledygook (or Gunning Fog Index) is one readability measure that assesses text difficulty based on sentence length (average number of words per sentence) along with the number of complex words. Texts with lower sentence counts and less complicated structures typically receive a better (lower) Gunning Fog Index score, indicating they are easier to read (Ehrenberg, 1982).

Additionally, the Flesch Reading Ease Score is another commonly used measure of readability that takes into account sentence length, among other factors, to determine how easy a text will be for readers to understand. Shorter sentences tend to improve Flesch Reading Ease scores, indicating better readability.

In the context of understanding how text structure impacts readability, sentence count is just one facet. It should be balanced with content needs and textual coherence to ensure that while the text is readable, it also comprehensively conveys the intended message.

## **2.3. Forbes Global 500**

Forbes Global 500, more accurately called the Forbes Global 2000, is an annual ranking compiled by Forbes magazine of the top 2,000 public companies in the world. The list is based on a mix of four metrics: sales, profit, assets, and market value. The ranking is meant to reflect the largest and most powerful enterprises which are considered as the leading players in their industries and are ranked among the most profitable and successful in the world.

Forbes releases this list to provide a comprehensive picture of the corporate landscape globally, showcasing the most significant companies that have the greatest impact on the economy and the business world. It's an important reference for investors, policymakers, and business leaders to understand the dynamics of global commerce.

In a separate list, Forbes also ranks "The World's Best Employers," which evaluates companies based on factors like working conditions and diversity. LG, for example, was ranked 10th on Forbes magazine's list of the world's 500 Best Employers. This is distinct from this list and focuses specifically on employers rather than all aspects of the companies' operations.

As a ranking of the world's largest and most influential companies, this list can influence the standard for privacy notices and policies in several ways:

**Setting Precedents:** Large global companies often set precedents for best practices in various domains, including privacy. When such a company updates its privacy policy to better protect consumer data, it may encourage or pressure others to follow suit.

**Benchmarking:** Being on the list suggests that a company has significant market power and resources, which can be leveraged to establish comprehensive privacy practices that set benchmarks for other companies.

**Public Expectations:** Consumers may have higher expectations for how their data is managed by companies on this list, given their prominence and resources. This can drive these companies to invest more in privacy measures.

**Regulatory Compliance:** Given their size and scope, these companies are often subject to various international privacy laws and regulations. By complying with stringent regulations like the EU's General Data Protection Regulation, they can set a standard for the industry.

**Innovation in Privacy:** These companies have the capacity to invest in new and innovative privacy technologies, such as advanced encryption or data anonymization techniques. Smaller companies might look to these leaders for the latest in privacy protection trends.

**Reputation Management:** Companies on this list are highly visible and their brand reputations are critical. Privacy breaches can cause significant damage, so these companies often lead the way in establishing rigorous privacy notices and policies to protect their reputations.

**Industry Leadership:** These companies often have the influence to lead industry groups and consortiums that develop privacy standards and best practices.

The influence of these companies extends to shaping privacy norms and standardizing consumer expectations across industries on a global scale. Their policies can essentially become a de facto standard, particularly if they are recognised as being comprehensive and user-centric.





### 3. Method

In this section, we look into the sample list which comprises the top 100 organisations on Forbes' Global 500 Companies list in 2023, how their privacy notices were collected, how the notices were analysed and the variables collected. We then take a look into comparing the variables among each notice for their readability grade, length and estimated reading time.

#### 3.1. Sample List

The sample list for this thesis consisted of the top 100 organisations on Forbes' Global 500 Companies list in 2023. This list was gathered from publicly available sources such as the Forbes website. The list was then populated with each organisation's official and main website URL. Where available, its international, global or English website is selected. The list did not include any of the organisation's region or country-specified websites.

The final sample list was put together in Table 1, which includes the organisation's rank within the Global 500, name, industry, country and website URL.

Rank	Company	Industry	State / Country	Country	Website
1	Walmart	General Merchandisers	Arkansas	U.S.	<a href="http://www.walmart.com">www.walmart.com</a>
2	<a href="http://Amazon.com">Amazon.com</a>	Internet Services and Retailing	Washington	U.S.	<a href="http://www.amazon.com">www.amazon.com</a>
3	State Grid	Utilities	China	China	<a href="http://www.sgcc.com.cn">www.sgcc.com.cn</a>
4	China National Petroleum	Petroleum Refining	China	China	<a href="http://www.cnpc.com.cn">www.cnpc.com.cn</a>

5	Sinopec Group	Petroleum Refining	China	China	<a href="http://www.sinopec.com">www.sinopec.com</a>
6	Saudi Aramco	Mining, Crude-Oil Production	Saudi Arabia	Saudi Arabia	<a href="http://www.aramco.com">www.aramco.com</a>
7	Apple	Computers, Office Equipment	California	U.S.	<a href="http://www.apple.com">www.apple.com</a>
8	Volkswagen	Motor Vehicles & Parts	Germany	Germany	<a href="http://www.volkswagenag.com">www.volkswagenag.com</a>
9	China State Construction Engineering	Engineering & Construction	China	China	<a href="http://www.cscec.com">www.cscec.com</a>
10	CVS Health	Health Care: Pharmacy and Other Services	Rhode Island	U.S.	<a href="http://www.cvshealth.com">www.cvshealth.com</a>
11	UnitedHealth Group	Health Care: Insurance and Managed Care	Minnesota	U.S.	<a href="http://www.unitedhealthgroup.com">www.unitedhealthgroup.com</a>
12	Exxon Mobil	Petroleum Refining	Texas	U.S.	<a href="http://www.exxonmobil.com">www.exxonmobil.com</a>
13	Toyota Motor	Motor Vehicles & Parts	Japan	Japan	<a href="http://www.global.toyota/en/">www.global.toyota/en/</a>
14	Berkshire Hathaway	Insurance: Property and	Nebraska	U.S.	<a href="http://www.berkshirehathaway.com">www.berkshirehathaway.com</a>

		Casualty (Stock)			
15	Shell	Petroleum Refining	Britain	Britain	<a href="http://www.shell.com">www.shell.com</a>
16	McKesson	Wholesalers: Health Care	Texas	U.S.	<a href="http://www.mckesson.com">www.mckesson.com</a>
17	Alphabet	Internet Services and Retailing	California	U.S.	<a href="http://www.abc.xyz">www.abc.xyz</a>
18	Samsung Electronics	Electronics, Electrical Equip.	South Korea	South Korea	<a href="http://www.samsung.com">www.samsung.com</a>
19	Trafigura Group	Trading	Singapor e	Singapor e	<a href="http://www.trafigura.com">www.trafigura.com</a>
20	Hon Hai Precision Industry	Electronics, Electrical Equip.	Taiwan	Taiwan	<a href="http://www.foxconn.com">www.foxconn.com</a>
21	AmerisourceB ergen	Wholesalers: Health Care	Pennsylv ania	U.S.	<a href="http://www.amerisourcebergen.com">www.amerisourcebergen.com</a>
22	Industrial & Commercial Bank of China	Banks: Commercial and Savings	China	China	<a href="http://www.icbc-ltd.com">www.icbc-ltd.com</a>
23	Glencore	Mining, Crude- Oil Production	Switzerla nd	Switzerla nd	<a href="http://www.glencore.com">www.glencore.com</a>

24	China Construction Bank	Banks: Commercial and Savings	China	China	<a href="http://www.ccb.com">www.ccb.com</a>
25	Ping An Insurance	Insurance: Life, Health (Stock)	China	China	<a href="http://www.group.pingan.com">www.group.pingan.com</a>
26	Costco Wholesale	General Merchandisers	Washington	U.S.	<a href="http://www.costco.com">www.costco.com</a>
27	TotalEnergies	Petroleum Refining	France	France	<a href="http://www.totalenergies.com">www.totalenergies.com</a>
28	Agricultural Bank of China	Banks: Commercial and Savings	China	China	<a href="http://www.abchina.com">www.abchina.com</a>
29	Stellantis	Motor Vehicles & Parts	Netherlands	Netherlands	<a href="http://www.stellantis.com">www.stellantis.com</a>
30	Cigna	Health Care: Pharmacy and Other Services	Connecticut	U.S.	<a href="http://www.cigna.com">www.cigna.com</a>
31	Sinochem Holdings	Chemicals	China	China	<a href="http://www.sinochem.com">www.sinochem.com</a>
32	AT&T	Telecommunications	Texas	U.S.	<a href="http://www.att.com">www.att.com</a>
33	Microsoft	Computer Software	Washington	U.S.	<a href="http://www.microsoft.com">www.microsoft.com</a>

34	China Railway Engineering Group	Engineering & Construction	China	China	<a href="http://www.crecg.com">www.crecg.com</a>
35	BP	Petroleum Refining	Britain	Britain	<a href="http://www.bp.com">www.bp.com</a>
36	Cardinal Health	Wholesalers: Health Care	Ohio	U.S.	<a href="http://www.cardinalhealth.com">www.cardinalhealth.com</a>
37	Chevron	Petroleum Refining	California	U.S.	<a href="http://www.chevron.com">www.chevron.com</a>
38	Mercedes-Benz Group	Motor Vehicles & Parts	Germany	Germany	<a href="http://group.mercedes-benz.com">group.mercedes-benz.com</a>
39	China Railway Construction	Engineering & Construction	China	China	<a href="http://www.crcc.cn">www.crcc.cn</a>
40	China Life Insurance	Insurance: Life, Health (Stock)	China	China	<a href="http://www.chinalife.com.cn">www.chinalife.com.cn</a>
41	Mitsubishi	Trading	Japan	Japan	<a href="http://www.mitsubishicorp.com">www.mitsubishicorp.com</a>
42	Bank of China	Banks: Commercial and Savings	China	China	<a href="http://www.boc.cn">www.boc.cn</a>
43	Home Depot	Specialty Retailers	Georgia	U.S.	<a href="http://www.homedepot.com">www.homedepot.com</a>
44	China Baowu Steel Group	Metals	China	China	<a href="http://www.baowugroup.com">www.baowugroup.com</a>

45	Walgreens Boots Alliance	Food & Drug Stores	Illinois	U.S.	<a href="http://www.walgreensbootsalliance.com">www.walgreensbootsalliance.com</a>
46	<a href="http://JD.com">JD.com</a>	Internet Services and Retailing	China	China	<a href="http://www.jd.com">www.jd.com</a>
47	Allianz	Insurance: Life, Health (Stock)	Germany	Germany	<a href="http://www.allianz.com">www.allianz.com</a>
48	AXA	Insurance: Life, Health (Stock)	France	France	<a href="http://www.axa.com">www.axa.com</a>
49	Marathon Petroleum	Petroleum Refining	Ohio	U.S.	<a href="http://www.marathonpetroleum.com">www.marathonpetroleum.com</a>
50	Elevance Health	Health Care: Insurance and Managed Care	Indiana	U.S.	<a href="http://www.elevancehealth.com">www.elevancehealth.com</a>
51	Kroger	Food & Drug Stores	Ohio	U.S.	<a href="http://www.thekrogerco.com">www.thekrogerco.com</a>
52	Gazprom	Energy	Russia	Russia	<a href="http://www.gazprom.com">www.gazprom.com</a>
53	Ford Motor	Motor Vehicles & Parts	Michigan	U.S.	<a href="http://www.ford.com">www.ford.com</a>
54	Verizon Communications	Telecommunications	New York	U.S.	<a href="http://www.verizon.com">www.verizon.com</a>

55	Alibaba Group Holding	Internet Services and Retailing	China	China	<a href="http://www.alibabagroup.com">www.alibabagroup.com</a>
56	Fortum	Energy	Finland	Finland	<a href="http://www.fortum.com">www.fortum.com</a>
57	China Mobile Communications	Telecommunications	China	China	<a href="http://www.10086.cn">www.10086.cn</a>
58	China Minmetals	Metals	China	China	<a href="http://www.minmetals.com">www.minmetals.com</a>
59	BMW Group	Motor Vehicles & Parts	Germany	Germany	<a href="http://www.bmwgroup.com">www.bmwgroup.com</a>
60	China Communications Construction	Engineering & Construction	China	China	<a href="http://www.ccccltd.cn">www.ccccltd.cn</a>
61	Honda Motor	Motor Vehicles & Parts	Japan	Japan	<a href="https://global.honda/en/">https://global.honda/en/</a>
62	Deutsche Telekom	Telecommunications	Germany	Germany	<a href="http://www.telekom.com">www.telekom.com</a>
63	JPMorgan Chase	Banks: Commercial and Savings	New York	U.S.	<a href="http://www.jpmorganchase.com">www.jpmorganchase.com</a>
64	General Motors	Motor Vehicles & Parts	Michigan	U.S.	<a href="http://www.gm.com">www.gm.com</a>



65	China National Offshore Oil	Mining, Crude-Oil Production	China	China	<a href="http://www.cnooc.com.cn">www.cnooc.com.cn</a>
66	Centene	Health Care: Insurance and Managed Care	Missouri	U.S.	<a href="http://www.centene.com">www.centene.com</a>
67	Lukoil	Petroleum Refining	Russia	Russia	<a href="http://www.lukoil.com">www.lukoil.com</a>
68	SAIC Motor	Motor Vehicles & Parts	China	China	<a href="http://www.saicmotor.com">www.saicmotor.com</a>
69	Shandong Energy Group	Mining, Crude-Oil Production	China	China	<a href="http://www.shandong-energy.com">www.shandong-energy.com</a>
70	China Resources	Pharmaceuticals	China	China	<a href="http://www.crc.com.hk">www.crc.com.hk</a>
71	Meta Platforms	Internet Services and Retailing	California	U.S.	<a href="http://www.fb.com">www.fb.com</a>
72	Assicurazioni Generali	Insurance: Life, Health (Stock)	Italy	Italy	<a href="http://www.generali.com">www.generali.com</a>
73	Comcast	Telecommunications	Pennsylvania	U.S.	<a href="http://www.comcastcorporation.com">www.comcastcorporation.com</a>
74	Phillips 66	Petroleum Refining	Texas	U.S.	<a href="http://www.phillips66.com">www.phillips66.com</a>
75	Hengli Group	Textiles	China	China	<a href="http://www.hengli.com">www.hengli.com</a>

76	Amer International Group	Metals	China	China	<a href="http://www.amer.com.cn">www.amer.com.cn</a>
77	Xiamen C&D	Trading	China	China	<a href="http://www.chinacdc.com">www.chinacdc.com</a>
78	Itochu	Trading	Japan	Japan	<a href="http://www.itochu.co.jp">www.itochu.co.jp</a>
79	China FAW Group	Motor Vehicles & Parts	China	China	<a href="http://www.faw.com.cn">www.faw.com.cn</a>
80	Sinopharm	Pharmaceuticals	China	China	<a href="http://www.sinopharm.com">www.sinopharm.com</a>
81	China Post Group	Mail, Package and Freight Delivery	China	China	<a href="http://www.chinapost.com.cn">www.chinapost.com.cn</a>
82	Valero Energy	Petroleum Refining	Texas	U.S.	<a href="http://www.valero.com">www.valero.com</a>
83	Nippon Telegraph and Telephone	Telecommunications	Japan	Japan	<a href="http://www.group.ntt">www.group.ntt</a>
84	Crédit Agricole	Banks: Commercial and Savings	France	France	<a href="http://www.credit-agricole.com">www.credit-agricole.com</a>
85	China Energy Investment	Mining, Crude-Oil Production	China	China	<a href="http://www.ceic.com">www.ceic.com</a>

86	Dell Technologies	Computers, Office Equipment	Texas	U.S.	<a href="http://www.delltechnologies.com">www.delltechnologies.com</a>
87	Target	General Merchandisers	Minnesota	U.S.	<a href="http://www.target.com">www.target.com</a>
88	Mitsui	Trading	Japan	Japan	<a href="http://www.mitsui.com">www.mitsui.com</a>
89	China Southern Power Grid	Utilities	China	China	<a href="http://www.csg.cn">www.csg.cn</a>
90	Enel	Utilities	Italy	Italy	<a href="http://www.enel.com">www.enel.com</a>
91	COFCO	Trading	China	China	<a href="http://www.cofco.com">www.cofco.com</a>
92	Hyundai Motor	Motor Vehicles & Parts	South Korea	South Korea	<a href="http://worldwide.hyundai.com">worldwide.hyundai.com</a>
93	Fannie Mae	Diversified Financials	District Of Columbia	U.S.	<a href="http://www.fanniemae.com">www.fanniemae.com</a>
94	Japan Post Holdings	Insurance: Life, Health (Stock)	Japan	Japan	<a href="http://www.japanpost.jp">www.japanpost.jp</a>
95	Electricité de France	Utilities	France	France	<a href="http://www.edf.fr">www.edf.fr</a>
96	Huawei Investment & Holding	Network & Other Communications Equipment	China	China	<a href="http://www.huawei.com">www.huawei.com</a>

97	United Parcel Service	Mail, Package and Freight Delivery	Georgia	U.S.	<a href="http://www.ups.com">www.ups.com</a>
98	Life Insurance Corp. of India	Insurance: Life, Health (Stock)	India	India	<a href="http://www.licindia.in">www.licindia.in</a>
99	Deutsche Post DHL Group	Mail, Package and Freight Delivery	Germany	German y	<a href="http://www.dpdhl.com">www.dpdhl.com</a>
100	PowerChina	Engineering & Construction	China	China	<a href="http://www.powerchina.cn">www.powerchina.cn</a>

Table 1: Final sample list

### 3.2. Privacy Notices Collection

The privacy notices of these organisations were then collected from their official websites. Only privacy notices that were publicly accessible and written in English were included in the sample. None of the sampled privacy notices were translated, edited or altered in any way for the purpose of this thesis. All content retrieved from the privacy notices was used as it was presented on the websites.

Where multiple privacy notices were made available to the general public, the most generic version of the privacy notice was selected for analysis. For example, if a company had different privacy notices for its various products or services, the privacy notice that provided the most comprehensive information applicable to most users was chosen.

Where no dedicated privacy notice was found on an organisation's website, the terms of service, terms of use or data protection policy were collected for analysis instead. The reasoning behind this choice was that these documents often contain privacy-related information and serve a similar purpose to a privacy notice. The alternate

documents were inspected to ensure that they contained relevant information regarding data collection, use, and sharing practices.

If no publicly available privacy notice or related document could be found on an organisation's website, that organisation was excluded from the analysis. If the organisation's website was not accessible or encountered any technical issues preventing the collection of the privacy notice, it was also excluded from the analysis. The collection process was repeated on different devices to ensure accuracy and to account for any potential device-specific display variations. When needed, a Virtual Private Network (VPN) service was used to ensure access to regionally restricted content and overcome any geographical limitations.

To ensure that the privacy notices collected were up-to-date, the data collection process was conducted within a period of 2 days.

The collection process for most of the organisations was straightforward and is depicted in Figure 1.

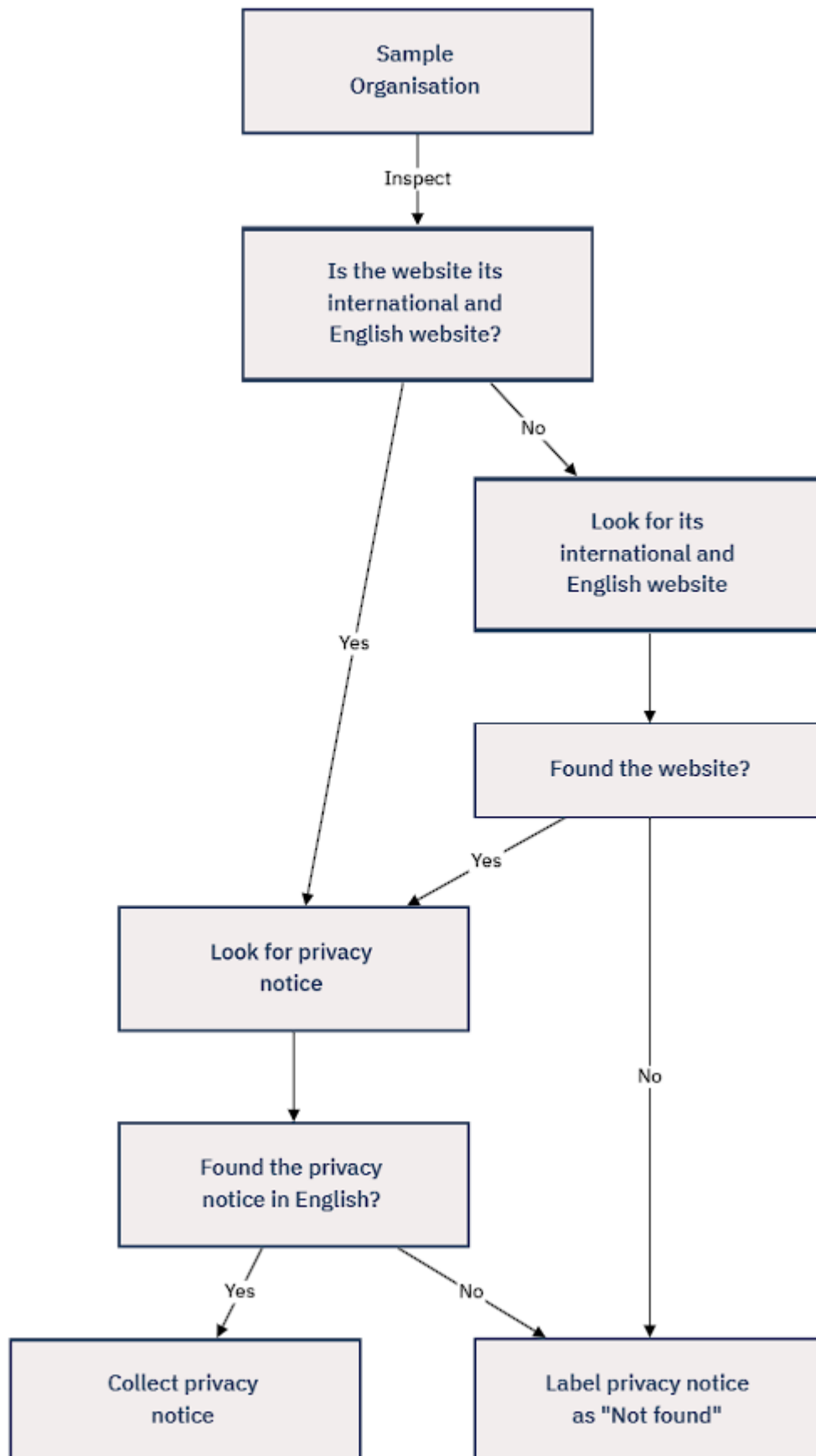


Figure 1: Privacy notices collection flowchart

### **3.3. Metrics Gathering**

Upon completion of the collection of the privacy notices, they were then analysed for their readability using the Hemingway Editor. Hemingway Editor is a web application that provides readability scores and text metrics to assess the complexity of written content.

The Hemingway Editor was used as it was when one visits its website. No account login, configuration changes or customisations were made to the Hemingway Editor during the analysis process.

Each privacy notice was simply copied and pasted into the Hemingway Editor and the tool automatically analysed the text for readability.

While various metrics were made available on the Hemingway Editor, the metrics gathered for the analysis were: readability grade, total word count, total sentence count and reading time. The metrics were then copied into the sample list as its respective columns.

### **3.4. Data Cleanup**

Upon completion of data collection, the gathered data was cleaned and organised for further analysis. All organisations for which no privacy notices were found on their websites were removed from the dataset. All data were formatted in a consistent manner and any inconsistencies or errors were corrected for proper analysis.

As the Hemingway Editor classifies the reading grades up to grade 16 and anything above it is classified as "Post-graduate", the grade "Post-graduate" was reclassified as grade "17" for consistency in the analysis.

### **3.5. Data Analysis**

After gathering the data from the metrics for each privacy notice, the next step was to analyse them. The data for each privacy notice were analysed to determine the readability of the documents.

The data collected was categorised by industry, country and grade. The categorisation was provided in the original dataset. No alteration was made to the categorisation during the analysis process.



## 4. Results and Analysis

In this section, we take a look at the results and analyse the results of the data analysis described in the Method section. The data cleanup that was performed as described in the Method section resulted in a final dataset of 87 organisations with publicly available privacy notices found, following the collection conditions and process also described in the Method section.

### 4.1. Industry

The results revealed that the organisations were spread across 26 industries. The five industries with the most number of organisations were Motor Vehicles & Parts , Petroleum Refining , Insurance: Life, Health (Stock) , Telecommunications and Trading. There were seven industries with the least number of organisations, one. They were Special Retailers, Diversified Financials, Computer Software, Network & Other Communications Equipment, Textile, Engineering & Construction and Chemicals.

The industry with the highest average grade level of readability for its privacy notices was Textiles with an average grade level of 17, Post-graduate. The industry with the lowest average grade level of readability was Mining, Crude-Oil Production with an average grade level of readability of 10. The median grade level of readability across the industries was 12.

Industry	Ave Grade	Ave Read	Ave Words	Ave Sent.
Banks: Commercial and Savings	15	0:20:29	5121	168
Chemicals	7	0:00:31	128	15
Computer Software	13	0:14:38	3657	187
Computers, Office Equipment	14	0:17:05	4272	194
Diversified Financials	13	0:19:57	4986	274

Electronics, Electrical Equip.	13	0:11:33	2887	138
Energy	14	0:01:27	360	19
Engineering & Construction	12	0:02:00	502	25
Food & Drug Stores	12	0:27:35	6894	486
General Merchandisers	14	0:25:49	6452	335
Health Care: Insurance and Managed Care	10	0:11:03	2761	181
Health Care: Pharmacy and Other Services	15	0:16:00	4000	160
Insurance: Life, Health (Stock)	13	0:07:23	1847	102
Internet Services and Retailing	12	0:27:04	6766	501
Mail, Package and Freight Delivery	16	0:15:43	3929	162
Mining, Crude-Oil Production	10	0:03:12	800	41
Motor Vehicles & Parts	11	0:11:47	2944	166
Network & Other Communications Equipment	11	0:10:41	2671	157
Petroleum Refining	13	0:10:54	3054	145
Pharmaceuticals	12	0:00:18	74	4
Specialty Retailers	11	0:39:32	9883	726

Telecommunications	12	0:09:24	2350	130
Textiles	17	0:15:08	3783	108
Trading	12	0:07:28	1867	105
Utilities	12	0:02:30	973	438
Wholesalers: Health Care	12	0:17:27	4363	239

Table 2: Results by Industry

#### 4.1.1. Reading Time by Industry

The industry with the longest average reading time was Specialty Retailers with an average reading time of 39 minutes and 32 seconds. The industry with the shortest average reading time was Pharmaceuticals with an average reading time of 18 seconds. The median reading time across the industries was 11 minutes and 33 seconds.

#### 4.1.2. Word Count by Industry

The industry with the highest average word count for its privacy notices was Specialty Retailers with an average word count of 9883 words. The industry with the lowest average word count was Pharmaceuticals with an average word count of 74 words. The median average word count across the industries was 2944 words.

#### 4.1.3. Sentence Count by Industry

The industry with the highest average sentence count for its privacy notices was Specialty Retailers with an average sentence count of 726 sentences. The industry with the lowest average sentence count was Pharmaceuticals with an average sentence count of 4 sentences. The median average sentence count across the industries was 160 sentences.

## 4.2. Country

The results revealed that the organisations were spread across 16 countries. The five countries with the most number of organisations were United States (U.S.), China, Japan, Germany and France. There were seven countries with the least number of organisations, one. They were Finland, India, Netherlands, Saudi Arabia, Singapore, Switzerland and Taiwan.

The country with the highest average grade level of readability for its privacy notices was Italy with an average grade level of 16. There were four countries with the lowest average grade level of readability, which were Netherlands, Saudi Arabia, Singapore and Switzerland, all with an average grade level of 10. The median grade level of readability across the countries was 12.

Country	Ave Grade	Ave Read	Ave Words	Ave Sent.
Britain	11	0:08:21	2,087	115
China	12	0:06:48	1,960	69
Finland	11	0:01:41	422	28
France	15	0:03:20	832	30
Germany	11	0:08:52	2,215	141
India	15	0:04:10	1,040	41
Italy	16	0:06:47	1,696	671
Japan	13	0:08:05	2,022	107
Netherlands	10	0:04:57	1,236	93
Russia	13	0:06:09	1,534	115
Saudi Arabia	10	0:04:49	1,206	56

Singapore	10	0:13:45	3,437	232
South Korea	14	0:19:08	4,781	234
Switzerland	10	0:01:34	393	25
Taiwan	12	0:08:44	2,182	107
U.S.	13	0:21:21	5,337	311

Table 3: Results by Country

#### 4.2.1. Reading Time by Country

The country with the longest average reading time was the U.S with an average reading time of 21 minutes and 21 seconds. The country with the shortest average reading time was Switzerland with an average reading time of 1 minute and 34 seconds. The median reading time across the countries was 6 minutes and 47 seconds.

#### 4.2.2. Word Count by Country

The country with the highest average word count for its privacy notices was the U.S. with an average word count of 5337 words. The country with the lowest average word count was Switzerland with an average word count of 393 words. The median average word count across the industries was 1696 words.

#### 4.2.3. Sentence Count by Country

The country with the highest average sentence count for its privacy notices was Italy with an average sentence count of 726 sentences. The country with the lowest average sentence count was Switzerland with an average sentence count of 671 sentences. The median average sentence count across the was 107 sentences.

### 4.3. Grade Level of Readability

The results revealed that the highest average grade level of readability for its privacy notices was 17, also classified as Post-graduate. The lowest grade level of readability was 7. There was at least one privacy notice for every grade level of readability between the lowest and highest grades.

Grade	Ave Read	Ave Words	Ave Sent.
7	0:00:31	128	15
8	0:41:34	10390	969
9	0:07:43	1929	139
10	0:05:01	1342	83
11	0:16:12	4050	274
12	0:10:49	3031	161
13	0:13:33	3385	171
14	0:13:30	3373	153
15	0:14:46	3692	268
16	0:12:07	3027	114
17	0:17:38	4408	136

Table 4: Results by Grade Level of Readability

#### **4.3.1. Reading Time by Grade Level Of Readability**

The grade level of readability with the longest average reading time was grade 8 with an average reading time of 41 minutes and 34 seconds. The grade level of readability with the shortest average reading time was grade 7 with an average reading time of 31 seconds. The median reading time across the countries was 13 minutes and 30 seconds.

#### **4.3.2. Word Count by Grade Level Of Readability**

The grade level of readability with the highest average word count for its privacy notices was grade 8 with an average word count of 10390 words. The grade level of readability with the lowest average word count was grade 7 with an average word count of 128 words. The median average word count across the industries was 3373 words.

#### **4.3.3. Sentence Count by Grade level of readability**

The grade level of readability with the highest average sentence count for its privacy notices was grade 8 with an average sentence count of 969 sentences. The grade level of readability with the lowest average sentence count was grade 7 with an average sentence count of 15 sentences. The median average sentence count across the was 153 sentences.

## 5. Discussion

In this section, we will discuss the results presented in the previous section.

The results only covered 87 out of the 100 organisations in the sample list. This is because most of the remaining organisations' websites either did not have a publicly accessible privacy notice or the privacy notice was not written in English language. All but one of those organisations originated from China. The final one was from the U.S.

For each category, I will answer the following questions:

- Is there a trend where longer reading time led to a higher grade of readability?
- Is there a trend where more word count led to a higher grade of readability?
- Is there a trend where shorter sentences help in getting a lower grade level of readability?

Sentence lengths are calculated by dividing the word count with the sentence count.

Some recommendations would also be discussed and proposed. This section ends off with discussing the limitations of this thesis.

### 5.1. Industry

In the results section, it was mentioned that the 87 organisations were spread across 26 industries. These industries were classified by Forbes. Eight of the industries only had one organisation classified under them respectively, seven industries had two organisations each and the remaining had three to eleven each. This is important to note because the results numbers shared in the results calculated the average for each category.

The industry with the highest grade of readability was Textile, with a grade of 17, which is also classified as Post-graduate grade. The industries that were right behind were Mail, Package and Freight Delivery, grade 16, Health Care: Pharmacy and Other Services, grade 15 and Banks: Commercial and Savings with grade 15. The reason for Textile to unexpectedly be the industry with the highest grade of readability could be because it only had one organisation with a publicly accessible privacy policy. These industries had average word count ranging between 3783 and 5121 words.



Is there a trend where longer reading time led to a higher grade of readability?

No, the industry with the longest read time had an average grade 11. The two highest grades had read times that were close to the median.

Is there a trend where more word count led to a higher grade of readability?

No, the industry with the highest average word counts had an average grade 11. However, it was observed that the industries with average word count of more than 2500 were more likely to have an average grade of 12 or higher.

Is there a trend where shorter sentences help in getting a lower grade level of readability?

Yes, a trend of industries with shorter sentences was observed to have lower grade of readability. The industries with the top three grades had longer average sentence lengths.

## **5.2. Countries**

In the results section, it was mentioned that the 87 organisations were spread across 16 countries. The countries were mapped to the organisations by Forbes. Seven of the countries only had one organisation classified under them respectively, four countries had two organisations each and the remaining had four to 33 each. This is important to note because the results numbers shared in the results calculated the average for each category.

The country with the highest grade of readability was Italy, with a grade of 16. The countries that were right behind were France, grade 15, India, grade 15 and South Korea with grade 14. Italy having the highest average grade of readability is within expectations as it is not a predominantly English speaking country. The same could be said for France, India and South Korea.

Is there a trend where longer reading time led to a higher grade of readability?

No, there are countries with significantly shorter reading time, France and India both had an average grade of 15 each. Meanwhile, the U.S., with the longest average reading time had an average grade of 13.

Is there a trend where more word count led to a higher grade of readability?

No, a similar trend was observed as that of the reading time trend.

Is there a trend where shorter sentences help in getting a lower grade level of readability?

No, the country with the longest average sentence length had a grade of 12 while Italy with the shortest average sentence length had the highest grade of 16 among all the countries.

### **5.3. Grade Level of Readability**

In the results section, it was mentioned that the 87 organisations had grade levels of readability between 7 and 17. Grade 7 had only one organisation, grades 8 and 16 had two organisations while the remaining grades had between 6 and 12 organisations each. The difference in ranges is not as wide as that of industries and countries.

Is there a trend where longer reading time led to a higher grade of readability?

No, there was a good mix. The grade with the longest average reading time was of grade 8 while the grade with the shortest average reading time was of grade 7. The grade with the second longest average reading time was of grade 17 while

Is there a trend where more word count led to a higher grade of readability?

No, a similar trend was observed as that of the reading time trend.

Is there a trend where shorter sentences help in getting a lower grade level of readability?

Yes, a trend where shorter average sentence lengths had lower grade levels of readability. The shortest average sentence length had a grade of 7 while the longest had a grade of 17.

### **5.4. Hypotheses**

Based on the results, hypothesis #1 was passed while hypothesis #2 failed.

#### **5.4.1. Hypothesis #1**

The readability of privacy notices for the top 100 organisations from Forbes' list of Global 500 companies in 2023 is significantly low.

This hypothesis passed as most of the privacy notices had a grade higher than 8, which is the average grade level of readability in the U.S.

#### **5.4.2. Hypothesis #2**

There is significant variance in the readability of privacy notices across different industries and geographic regions.

This hypothesis failed as there was no significant variance in the readability of the privacy notices within each category. The variance in grades were marginal.

### **5.5. Recommendations**

From the results and its analysis, a few recommendations can be drawn to improve readability of privacy notices.

Write with your target audience in mind. Try to attain the lowest grade level of readability of your target audience. For example, if your target audience is the average person in the U.S., you should target attaining grade 8 or below.

Write in shorter sentences. It was observed that shorter sentences helped in getting a lower grade level of readability.

While the total length of the privacy notice did not directly impact the grade level of readability, it is recommended to keep it concise.

### **5.6. Limitations**

This thesis had a few limitations. The limitations mainly falls on the fact that only a single tool and method to assess readability was adopted. In the future, more tools and methods should be adopted to attain more objective results and draw more significant trends.

Other factors contributing to readability such as measuring attention span and retention should be considered. It could provide more practical and realistic results which could lead to more practical recommendations.

More significant metrics could have been drawn to attain more diverse results. An example metric could be regulations the privacy notice is compliant to.

## 6. Conclusion

In conclusion, with evolving privacy requirements being introduced worldwide, privacy notices are becoming more lengthy and complex, potentially reducing its readability and accessibility. This thesis has underscored the pivotal role that privacy notices play in communicating about how personal information is collected, used, and protected to foster trust between organisations and its users. Our investigation of the top 100 organisations on Forbes' Global 500 list has shed light on the current state of privacy notices, highlighting a clear tension between the need for detail and the imperative of readability.

As these notices have evolved alongside legal requirements and consumer awareness, the necessity for transparency and informed consent has never been greater. The readability of such documents is crucial, as it determines whether the intended audience can readily understand and make informed decisions about their personal data.

While there is no one-size-fits-all solution, it is clear that simplification in language and structure is paramount. Therefore, organisations are encouraged to strive for brevity without sacrificing the comprehensiveness that adequately informs users about their privacy rights and the data handling practices of the company.

Going forward, privacy notices will continue to be a subject of much scrutiny and debate. Organisations must keep pace with the dynamic landscape of privacy laws and user expectations by regularly reviewing and updating their privacy notices. Ultimately, the goal should be to create a bridge between legal compliance and the provision of clear, accessible information that empowers users to make informed privacy choices.

By adhering to best practices in drafting privacy notices and being receptive to feedback, organisations can not only meet regulatory obligations but, more importantly, advance the cause of user privacy in the digital realm.

## References

- Abou-Zahra, S. (2018, February 14). Accessibility: Review of 2017 and Outlook for 2018. <https://www.w3.org/blog/2018/02/accessibility-review-of-2017-and-outlook-for-2018/>
- Alsulaiman, L A., & Alrodhan, W A. (2014, July 27). Information Privacy Status in Saudi Arabia. <https://doi.org/10.5539/cis.v7n3p102>
- Anwar, N., Shuhidan, S M., & Zaini, M K. (2020, March 1). Measuring Users Concerns For Information Privacy. <https://doi.org/10.1088/1742-6596/1496/1/012009>
- Badarudeen, S., & Sabharwal, S. (2010, October 1). Assessing Readability of Patient Education Materials: Current Role in Orthopaedics. Lippincott Williams & Wilkins, 468(10), 2572-2580. <https://doi.org/10.1007/s11999-010-1380-y>
- Bansiong, A J. (2019, January 1). Readability, content, and mechanical feature analysis of selected commercial science textbooks intended for third grade filipino learners. Taylor & Francis, 6(1), 1706395-1706395. <https://doi.org/10.1080/2331186x.2019.1706395>
- Crossler, R E. (2011, January 1). Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. <https://doi.org/10.2307/41409971>
- Ehrenberg, A S C. (1982, November 1). Writing Technical Papers or Reports. □The □American statistician, 36(4), 326-329. <https://doi.org/10.1080/00031305.1982.10483042>
- Eltorai, E M A B A., Naqvi, S A A., Ghanian, B S S., Eberson, C P., Weiss, A C., Born, C T., & Daniels, A H. (2015, December 1). Readability of Invasive Procedure Consent Forms. Wiley-Blackwell, 8(6), 830-833. <https://doi.org/10.1111/cts.12364>
- Essential Components of Web Accessibility. (2018, February 27). <https://www.w3.org/WAI/fundamentals/components/>
- Fanning, J E., Okamoto, L A., Levine, E C., McGee, S A., & Janis, J E. (2023, January 1). Content and Readability of Online Recommendations for Breast

- Implant Size Selection. Wolters Kluwer, 11(1), e4787-e4787.  
<https://doi.org/10.1097/gox.0000000000004787>
- Gilakjani, A P., & Sabouri, N B. (2016, September 23). A Study of Factors Affecting EFL Learners' Reading Comprehension Skill and the Strategies for Improvement. <https://doi.org/10.5539/ijel.v6n5p180>
- Goldstein, D G., Hill, E F., Lazar, J., Siempelkamp, A., Taylor, A., & Lepofsky, D. (2011, May 7). Increasing legal requirements for interface accessibility. <https://doi.org/10.1145/1979742.1979496>
- Knight-Davis, S., & Sung, J. (2008, September 1). Analysis of Citations in Undergraduate Papers. <https://doi.org/10.5860/crl.69.5.447>
- Milne, G R., Culnan, M J., & Greene, H. (2006, September 1). A Longitudinal Assessment of Online Privacy Notice Readability. <https://doi.org/10.1509/jppm.25.2.238>
- Paul, J A., Baker, H., & Cochran, J. (2012, November 1). Effect of online social networking on student academic performance. Elsevier BV, 28(6), 2117-2127. <https://doi.org/10.1016/j.chb.2012.06.016>
- Rayess, H., Zuliani, G., Gupta, A., Svider, P F., Folbe, A J., Eloy, J A., & Carron, M A. (2017, March 1). Critical Analysis of the Quality, Readability, and Technical Aspects of Online Information Provided for Neck-Lifts. American Medical Association, 19(2), 115-120. <https://doi.org/10.1001/jamafacial.2016.1219>
- Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018, February 1). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. <https://doi.org/10.1016/j.clsr.2017.05.015>
- Xu, H., & Durgunoğlu, A Y. (2019, March 6). Motivational factors underlying different levels of reading comprehension of English language learners. Wiley-Blackwell, 11(1). <https://doi.org/10.1002/tesj.448>