Master's Thesis - Information Studies

Assessing User Awareness of the Environmental Impact of GenAI

: A Mixed-Methods Approach to Reinforce Sustainable Practices

Authors: Bruna Tomaic(20220817), Sehee Lee(20220365)

Study Programme: Information Studies, Aalborg University – Copenhagen

Semester: Master's Thesis

Supervisor: Florian Maximilian Meier

Submission Date: 31 May, 2024

Character Count: 186 976 (77.9 Pages)

AALBORG UNIVERSITY

Abstract

In the era of rapid technological growth, Generative Artificial Intelligence (GenAI) has gained global attention for its ability to produce diverse types of content, ranging from text, and images to music. Unlike traditional AI, which focuses on data analysis and prediction, GenAI creates novel data similar to its training datasets. The launch of tools like OpenAI's ChatGPT, which quickly accumulated millions of users, shows the potential of GenAI applications. However, this advancement comes with significant environmental concerns, particularly the digital carbon footprint associated with the energy-intensive processes of training and inference.

This master thesis explores the environmental impact and user awareness related to ChatGPT's digital carbon footprint. We aim to look into the CO2 emissions generated by GenAI, with a focus on ChatGPT, and develop potential strategies for sustainable practices regarding user behaviour and ChatGPT prompt creation. Our research includes a thorough literature review, a questionnaire and contextual inquiry, as well as methods for analysis to understand the environmental impact, user patterns, and user experiences with ChatGPT. Additionally, we investigate user awareness of the environmental impact, propose changes in interface design, and prompt sustainable practices.

Keywords: Generative AI, ChatGPT, carbon footprint, user behaviour, sustainable practices

Table of Content

Abstract **2**
1. Introduction **6**
1.1. Problem Statement 7
1.1.1. Research Questions 7
1.2. Significance of the Project 8
1.3. Research Scope and Limitations 9
1.4. Overview of the Report 9
2. Literature Review **10**
2.1. Literature Search Process 11
2.2. Related Works 11
2.2.1. Digital Carbon Footprint 12
2.2.2. Innovation of Generative AI 13
2.2.3. User Behaviour with Generative AI 13
2.2.4. Environmental Impact of Generative AI 14
2.2.5. Prompt Engineering 16
3. Research Design and Methods **17**
3.1. Research Design 17
3.2. Data Collection Methods 18
3.2.1. Questionnaires 19
3.2.1.1. General Description of Questionnaires 19
3.2.1.2. Implementation of Questionnaires 21
3.2.2. Contextual Inquiry 25
3.2.2.1. General Description of Contextual Inquiry 25
3.2.2.2. Implementation of Contextual Inquiry 27
3.3. Data Analysis Methods 29
3.3.1. Quantitative Analysis 30
3.3.1.1. General Description of Quantitative Analysis 30
3.3.1.2. Implementation of Quantitative Analysis 30
3.3.2. Thematic Analysis 31
3.3.2.1. General Description of Thematic Analysis 31
3.3.2.2. Implementation of Thematic Analysis 33
3.3.3. Content Analysis 33
3.3.3.1. General Description of Content Analysis 33
3.3.3.2. Text Mining 35
3.3.3.3. Sentiment Analysis 37
3.3.3.3.1. Cohen's Kappa 40
3.4. Validity and Reliability of Methods 41
3.5. Ethical Considerations and Limitations of Methods 43
3.5.1. Ethical Considerations 43
3.5.2. Limitations of Methods 45
4. Findings **45**
4.1. Findings from the Questionnaire 46
4.1.1. Quantitative Analysis 46
4.1.2. Text Mining Analysis 53
4.1.3. Sentiment Analysis 56
4.1.4. Thematic Analysis 59
4.2. Findings from Contextual Inquiry 64
5. Discussion and Conclusion **67**
5.1. Reflection on the Research Questions 68
RQ1: What are the environmental impacts of Generative Artificial Intelligence (GenAI) technologies,

specifically regarding their CO2 emissions footprint? 68
RQ2: What is the most commonly utilised Generative AI tool, and in which specific contexts is it used? 68
RQ3: What are the common usage patterns and preferences of ChatGPT users? 69
RQ4: What factors contribute to positive and negative experiences among ChatGPT users? 70
RQ5: How do users utilise prompts and optimise their use of prompts to obtain desired outcomes, interacting with ChatGPT? 71
RQ6: To what extent are individuals aware of the environmental impact of ChatGPT on the digital carbon footprint? 72
RQ7: How can sustainable practices, in terms of user behaviour, be enhanced to prevent CO2 emissions? 72
5.2. Conclusion 74
Bibliography **75**

List of Figures

Figure 1. Frequency of Using ChatGPT by Age Group
Figure 2. ChatGPT Usage Preference
Figure 3. Distribution of Average Rephrase Prompts
Figure 4. Distribution of Creation Prompts in the Latest Chat
Figure 5. Awareness of Climate Change
Figure 6. Awareness of ChatGPT's Environmental Impact
Figure 7. Ranked Energy Type
Figure 8. Context of Using GenAI
Figure 9. Most Frequently Occurring Words Regarding Responsibility
Figure 10. Sentiment Analysis Per Coder
Figure 11. Most Mentioned Words in Sentiment Analysis
Figure 12. Initial Themes of Positive Factors
Figure 13. Initial Themes of Negative Factors

List of Tables

Table 1. Result of Cohen's Kappa
Table 2. Defined Themes of Positive Factors
Table 3. Defined Themes of Negative Factors

1. Introduction

In the era of rapid technological revolution, Generative Artificial Intelligence, referred to as Generative AI or GenAI, has been gathering widespread attention. GenAI refers to a category of artificial intelligence technology that allows users to produce various types of content (Cao et al., 2023). "*Generative AI models use neural networks to identify the patterns and structures within existing data to generate new and original content*" (NVIDIA, n.d., How Does Generative AI Work? section).

The key distinction between traditional AI and Generative AI resides in their capabilities and applications. While traditional AI is mainly used for data analysis and prediction tasks, Generative AI goes beyond this capability by creating entirely novel data similar to its training data (Marr, 2023).

In November 2022, OpenAI launched ChatGPT, a large language model (LLM) accessible to the public, with the ability to produce detailed responses based on instructions provided in prompts. It has become the most popular product, setting a record for the highest number of users, reaching 1 million users within 5 days; ChatGPT had 180.5 million users and 1.6 billion visits in January 2024 (Duarte, 2024). Moreover, following OpenAI's partnership with Microsoft, numerous global tech companies, including Google and Meta, have unveiled GenAI tools and it is widely expected that these GenAI applications will become common in daily activities and business (Chien et al., 2023).

GenAI applications like ChatGPT allow broad audiences to effortlessly engage in generating human-like creations, including texts, realistic images and even music (García-Peñalvo & Vázquez-Ingelmo, 2023). For example, DALL-E 2, produced by OpenAI, enables users to create new images guided by natural language prompts, and GitHub Copilot assists in code completion, producing new code based on natural language prompts (Brady, 2023).

With unprecedented advancements, concerns regarding environmental issues arose due to the energy-intensive process of Machine Learning models. Two crucial concepts for the issues are carbon emissions (CO2 emissions) and carbon footprint. Carbon footprint is recognised as the life cycle of carbon-equivalent emissions and their impacts associated with products or services; when considering carbon emissions, commonly related terms include traffic, factories, airlines, and other transportation (Malmodin & Dag 2018). Electricity consumption has a carbon footprint unless all electricity is generated by renewable energy sources, yet renewable energy sources currently struggle to meet the substantial demand for GenAI (Das and Modak, 2023).

The carbon emissions generated by training a single GenAI model for corporate use amount to 626,000 pounds of CO2, which is five times greater than the emissions produced by an average car in the US throughout the entire period of its use from manufacturing to disposal (Hao, 2019). The emission caused by GenAI extends beyond training. During the inference phase, where predictions are made and AI responds to queries, it is estimated that 60 per cent of AI energy usage is allocated to inference and it surpasses that of training; ChatGPT's daily carbon footprint is estimated to equal 50 pounds of CO2 or 8.4 tons annually based on GPT-3 (Cho, 2023).

Given this context, it is crucial to address the carbon footprint resulting from the inference of GenAI. As highlighted by Luccioni et al. (2023), while large AI models undergo training only once, they are employed billions of times.

This report demonstrates our approach to addressing user awareness of carbon footprint resulting from the use of Generative AI, particularly ChatGPT, which is the most widely used GenAI application. It also outlines how to contribute to reducing carbon emissions through modifications in the interface design of ChatGPT and enhancing sustainable practices.

1.1. Problem Statement

“What is the extent of the environmental impact and users’ awareness related to ChatGPT, and how can strategies be developed to foster sustainable practices of users and minimise its digital carbon footprint?”

1.1.1. Research Questions

Based on the above problem statement and literature review we formulated seven research questions that we will be looking into and answering throughout this report.

RQ1: What are the environmental impacts of Generative Artificial Intelligence (GenAI) technologies, specifically regarding their CO2 emissions footprint?

RQ2: What is the most commonly utilised Generative AI tool, and in which specific contexts is it used?

RQ3: What are the common usage patterns and preferences of ChatGPT users?

RQ4: What factors contribute to positive and negative experiences among ChatGPT users?

RQ5: How do users utilise prompts and optimise their use of prompts to obtain desired outcomes, interacting with ChatGPT?

RQ6: To what extent are individuals aware of the environmental impact of ChatGPT on the digital carbon footprint?

RQ7: How can sustainable practices, in terms of user behaviour, be enhanced to prevent CO2 emissions?

First, we decided to explore to what extent Generative AI impacts the CO2 footprint. Furthermore, we will examine which GenAI tool is used most and in which context it is used. With these research questions, we will gather essential information to solidify our starting point in our research. In order to answer these questions we will be looking through the topic's relevant literature. Moreover, we will further acquire answers by conducting a questionnaire and doing content analysis that will be used on the questionnaire results.

From the third research question, one of the applications of GenAI, which is ChatGPT will be mainly focused since ChatGPT is currently the most used GenAI technology, according to our research. We will dive into the user behaviour with ChatGPT, experiences, prompt usage and awareness of the environmental impact it causes. We will acquire this knowledge by conducting a questionnaire and contextual inquiry. Furthermore, to analyse the data we gained we will be implementing qualitative analysis, which is a thematic analysis of the questionnaire results as well as quantitative analysis.

Lastly, we will examine how to reinforce sustainable practices to prevent CO_2 emissions by investigating how users optimise prompts and interact with ChatGPT. This research question will be answered by conducting a sentiment analysis of the questionnaire results and the findings from the contextual inquiry.

The research questions were formulated based on the issues identified during our research and literature review. As there is a lack of studies on this topic we aim to provide significant insights into user behaviour and awareness level. Moreover, our study seeks to suggest sustainable practices to reduce CO_2 emissions by answering our research questions.

1.2. Significance of the Project

The increase of Generative AI applications and their advancements has been remarkable, with OpenAI's ChatGPT rising in popularity and technological growth. However, the environmental impact of these technologies, particularly in terms of CO_2 emissions associated with their energy consumption, has become evident as a significant concern.

The significance of this project lies in its response to a critical challenge: the essential digital CO_2 footprint of GenAI and ChatGPT. As the utilisation of GenAI models like ChatGPT becomes increasingly common in daily life and work life, their environmental impact should not be overlooked. This project is based on the understanding that while GenAI offers unparalleled opportunities for innovation and efficiency, its sustainability must be prioritised to ensure a balance between technological advancement and environmental impact.

The primary objective of this project is to get a better understanding of the user's awareness regarding Generative AI and the environmental impact it causes. It will also look into users' behaviour concerning ChatGPT and possibly develop potential solutions to minimise the CO_2 emission it produces. Ultimately this research is crucial for several reasons: user awareness, user behaviour with ChatGPT and sustainable practice.

1.3. Research Scope and Limitations

While this study is significant in detecting the environmental impact of ChatGPT and user awareness regarding these emerging issues, there are limitations due to the relatively recent unveiling of ChatGPT in 2022. OpenAI shares its carbon footprint and sustainability-related information less transparently than other companies like Google, which are more transparent about their decarbonisation efforts. This lack of transparency prevents us from providing consistent and accurate measurements of ChatGPT's environmental impact, such as concrete numbers for energy consumption and CO_2 emissions.

A major focus of our study is user behaviour with ChatGPT and their awareness of its environmental impact. We conducted user research using questionnaires and contextual inquiries. During this process, we did not segment users of ChatGPT. Instead, we primarily considered users of ChatGPT-3, who use the service for free. This choice was made due to the higher availability of data from free users, but it may limit the generalisability of our findings to users of a paid version.

Comparing our findings with existing studies on user behaviour with ChatGPT, particularly interactions with prompts, is challenging due to the lack of research in this area. Additionally, due to the rapid evolution of AI technologies, our findings could quickly become outdated as newer versions of ChatGPT are released and user behaviours evolve.

1.4. Overview of the Report

This master thesis report presents an investigation into the awareness, user behaviour and environmental impact associated with Generative AI and ChatGPT. The structure of the report is carefully thought out and organised to provide a detailed exploration of the research questions, literature review, methods, analysis, findings and discussion.

Our project begins with an introduction (Chapter 1) that presents the problem statement, research questions, the project's significance and research scope and limitations. In Chapter 2, the literature review follows, exploring topics like digital carbon footprint, innovations in Generative AI, user behaviour and environmental impact of GenAI, and prompt engineering.

Following that, the research design and methods, Chapter 3, shows a detailed research design, data collection methods such as questionnaires and contextual inquiry, and data analysis methods including quantitative analysis, text mining, thematic analysis, content analysis and sentiment analysis. It also addresses the validity, reliability, ethical considerations, and limitations of methods. Chapter 4 presents findings from the implementation of the methods. Lastly in Chapter 5, the discussion and conclusion, we answer our research questions and connect them to all the findings and literature.

This structure provided us with a comprehensive presentation of our research, contributing to the exploration of the awareness and environmental implications of GenAI, particularly ChatGPT.

2. Literature Review

The first step in any report or research paper on which the review is based is to conduct a proper literature review or literature search. A well-known quote made by Rau (2004) is "*Literature search is a systematic and well-organised search from the already published data to identify a breadth of good quality references on a specific topic*" (p. 1244). To put it in other words what Rau was referring to was that a literature review is an important section of any report be it either as part of a course of study or as a key step in the research process. It can contain a simple summary of a source paper or article, but in most cases, it is structured, has organisational patterns and combines both summary and synthesis (Ramdhani et al., 2014).

Although it might appear to be a simple task a literature search demands a range of skills, such as defining the correct topics for exploration, acquiring skills in the literature search process and retrieval, developing the ability to analyse as well as rephrase the data, often needed to be completed within a limited time (Ramdhani et al., 2014).

The main purpose of a literature review is to display one's knowledge and understanding of a major research topic (The University of New Castle Australia, n.d.). By doing so it helps readers to understand better the project as well as reassures readers that the research paper is credible and reliable (Cantero, 2019). Furthermore, the justification of the research paper can be summarised in four main objectives: review published literature, critique the literature, identify gaps in the literature and inform proposed research (Young, 2017).

Other benefits that can arise from a well-conducted literature search are additional ideas that might come to light after reviewing and summarising others' work on a similar topic. Bolderston (2008) put it in great words: "*New theories can be built from the evidence discussed, and new directions for future research can be suggested*" (p. 86). Moreover, from a well-made literature review new methods, techniques of analysis and methods of measuring concepts can be identified or become apparent (Grewal et al., 2016).

Lastly, the approaches of literature search. In the last decade, searching libraries for journals, books, articles, etc. was the normal procedure. Physical literature exploration is still an essential practice for a systematic review search process, but it is not the only way (Paré et al., 2017). With the advantages of technology, the internet has become a maze of vast literature.

There are two main literature search approaches. The first one is a physical approach that includes: protocol-driven (hand search journals), snowballing (reference chasing, tracking citations), personal knowledge (existing theories and basis, particular contacts and academic system), etc. The second approach is a web-based approach which consists of search engines (Google, Google Scholar, etc.) and electronic sources of databases (SCOPUS, ProQuest, ERIC, etc.) (Grewal et al., 2016).

When it comes to our project, we used the literature web-based approach. The main reasoning behind that was the white-spread information the internet provides. It is well known that the Internet is one of the most popular sources of information regarding articles and scholarly research (Kibirige et al., 2000).

2.1. Literature Search Process

In order to conduct the literature search we first had to choose which type of literature review we wanted to work with. In general, there are four basic types: traditional, narrative, systematic, and meta-analytic reviews (Rhoades, 2011). We chose to work with the traditional literature review which needs to summarise the relevant articles and group them by topics.

To start our literature review we carried out a brainstorm about relevant keywords and possible topics. We categorised the keywords related to our main research subjects. For example, we had a label called fundamental keywords: digital carbon footprint, Generative AI, and most popular Generative AI. Another label was called behaviour-related themes regarding GenAI which had: behaviour and awareness of GenAI, prompt usage and the environmental impact of GenAI. After the brainstorming session, we divided the parts and started to search for various relevant research papers, literature and articles. Along the process, some new keywords and topics came to light such as digital decarbonisation, GenAI user characteristics, etc.

Moreover, we used Cooper's taxonomy of literature review (Cooper, 1988) to further determine the focus of our literature review (Sipe et al., 1996). Cooper’s taxonomy can be classified into five characteristics: focus, goal, perspective, coverage, organisation and audience (Randolph, 2019). For focus, we looked into research outcomes and findings, mostly concerning the user behavioural topic. The goal was used in the digital carbon footprint and digital decarbonisation topic to generalise and identify central issues. Perspective can be identified in our findings from a neutral perspective and we predominantly used primary sources. Next, the coverage of the literature we selected was representative, as we included numerous articles from various years and authors. When it comes to organisational characteristics, we grouped our literature by topics, meaning the literature review is conceptual. Lastly, the primary audience is the supervisor and reviewers of the project. The scholars within the field that this topic relates to are the secondary audience (Randolph, 2019).

Regarding sources, we used Google Scholar, Scopus, Eric and primarily Aalborg University’s library. We focused on the keyword search strategy, for narrowing down the searches. Initially, we began with broad queries and steadily refunded them to discover more relevant literature. Additionally, we came across extra literature by examining documents that were cited in the articles that we had originally found. That method of searching is called the "*Pearl growing*" method (Ramer, 2005).

2.2. Related Works

Within this chapter, we will be summarising our literature review of numerous research papers related to our topic. First, we will begin with a general understanding of Digital Carbon Footprint followed by the Innovation of Generative AI. In the Innovation of Generative AI, we will go through a general knowledge of Generative AI and establish the current most popular application in the GenAI industry. Moreover, we will look at the User Behaviour with GenAI mostly focusing on ChatGPT and how users interact with it. Lastly, we will touch on the Environmental Impact of Generative AI and the damage unknowingly caused by its massive usage.

2.2.1. Digital Carbon Footprint

Digital technologies (DT) are advancing in the fundamental mode of global economic leading and development in production and organisations (Shen et al., 2023). However, with growing digitalisation, the digital carbon footprint increases with every online action (Sharma & Dash, 2022). In other words, using technology results in significant greenhouse gas (GHG) emissions and carbon footprints (Narula, n.d.).

Specifically, AI technologies have recently become a main topic concerning climate change. For instance, Generative AI applications such as ChatGPT (for text generation), DALL-E (for images generation), etc. are estimated "*Use of 11 million requests/hour produces emissions of 12.8k metric ton CO2/year, 25 times the emissions for training GPT-3. Inference is critical to environmental and power cost*" (Chien et al., 2023, pp. 1-2). Moreover, in accordance with Strubell et al. (2019), It is estimated that training a single large language model is equal to around 300,000 kg of carbon dioxide emissions. To put it into a better perspective it can be compared to 125 round-trip flights between New York and Beijing (Dhar, 2020).

On the other hand, according to Dhar (2020), AI seems to play a dual role. It can lessen the effects of climate crises, for instance, smart grid design, modelling climate change predictions and developing low-emission infrastructure. Additionally, digital decarbonisation efforts and AI development go alongside one another because of their similar goals: new approaches to technology innovation and sustainability, best use and enabling widespread information across data, and Improving processes to drive Industry 4.0 (Advice, n.d.).

However, a large emission of carbon is emitted by AI itself (Dhar, 2020). Research carried out by Transforma Insights (n.d.) shows the global deployment of devices is expected to increase from 2.7 billion to 7.8 billion over the next decade. Moreover, the appearance of transformative technologies like AI-generated content has caused a significant increase in data volume (Wen et al., 2024). In that regard, the rapid growth in power consumption and the consumption of energy resources present significant energy challenges to digital technology, leading to new environmental impacts.

Decarbonisation

Decarbonisation is a process "*by which countries or other entities aim to achieve a low-carbon economy, or by which individuals aim to reduce their consumption of carbon*" (IPCC, 2014, p. 121). The Intergovernmental Panel on Climate Change (IPCC) provides a broad definition of decarbonisation but the message and aim are clear; the term aims to achieve an economy that provides as little carbon dioxide as possible.

Although it was seen that researchers placed digitalisation as a key driver for decarbonisation it is still being determined if these researchers may have overestimated the increased demand for resource efficiency and the negative environmental impact that digital infrastructures cause (Martín & Ortega, 2021; Li X. et al., 2021; Li Y. et al., 2021). As Martin and Ortega (2021) mention, digitalisation "*has become an indispensable tool for achieving* *the objectives of a green economy without pollution. Yet technology itself is also responsible for a significant amount of pollution*" (p. 3).

2.2.2. Innovation of Generative AI

Artificial intelligence (AI) is one of the main branches of computer science. It requires developing computer programs in order to complete tasks that would otherwise involve human intelligence. Furthermore, AI algorithms can implement learning, problem-solving, perception, language understanding and logical reasoning (Saleh, 2019).

Throughout our history, it was commonly assumed that artistic, creative tasks such as writing, coding software, designing fashion, and composing songs could only be achieved by humans. This assumption has seen a significant change in recent years with the growth of artificial intelligence (AI) that is capable of generating new content in ways that closely resemble human craftsmanship (Feuerriegel et al., 2024).

Generative artificial intelligence (GenAI) is the latest trend in the AI industry, which has the key ability to generate novel, open-ended content (Albrecht & Aliaga, 2023). The term "*Generative AI*" comes from computational techniques that have the ability to generate what seems like new, meaningful content such as text, images and audio from training data based on user prompts. Some of the most known training data models of this technology are Dall-E 2, GPT-4, and Copilot which are currently revolutionising the way we work and communicate with each other (Feuerriegel et al., 2024).

The most popular of these applications are developed in the form of chatbots, like ChatGPT, a conversational web app based on large language models (LLMs) (Singh, n.d.). Combining words based on patterns found in large sets of text data, such as on the internet (Albrecht & Aliaga, 2023). It took OpenAI's ChatGPT about five days to reach one million users compared to the popular social media application Instagram which took 2.5 months (Singh, n.d; Duarte, 2024). Additionally, we can compare ChatGPT's record-shattering 100 million monthly users in just three months to the time it took TikTok nine months to reach the same milestone (Albrecht & Aliaga, 2023; Duarte, 2024).

Furthermore, advances in "*machine learning (ML), massive datasets, and substantial increases in computing power have propelled such tools to human-level performance on academic and professional benchmarks, comparable to the bar exam.*" (Singh, n.d., p. 1)

2.2.3. User Behaviour with Generative AI

Regarding user behaviour with GenAI, there are several studies investigating how users perceive and use ChatGPT, which is the first product to become generally accessible and was adopted by many users early on.

Skjuve et al. (2023) conducted a qualitative questionnaire to explore characteristics of the user experience with ChatGPT, analysing responses from 194 participants who are early users of ChatGPT. First, it was observed that the majority of participants use ChatGPT quite frequently, once a week or more. In terms of user experience both good and poor experiences were disclosed: Most participants mentioned good experiences with ChatGPT in the context of creative activities and knowledge in both work and daily life. For instance, several respondents described that ChatGPT assists them in executing their work faster, enhancing the quality of their work related to work or schools and understanding new topics. It was also observed that they use ChatGPT to facilitate inspiration for ideas.

However, Skjuve et al. (2023) also observed poor experiences with ChatGPT. Most poor experiences occurred when ChatGPT provided irrelevant and useless outcomes to queries. Moreover, some participants reported that they felt frustrated when ChatGPT did not understand their queries in the first place and they needed to formulate several prompts to get relevant outcomes. Besides, the respondents illustrated negative experiences with technical issues, which were caused by high demands and server overload, making them unable to access the service later on.

In conclusion, Skjuve et al. (2023) highlighted that research in prompt writing support and optimising the process of prompt refinement is crucial to prevent user frustration. For example, detecting inefficient prompts and suggesting alternatives based on inferred user intent.

On the other hand, when it comes to user practices with ChaptGPT, Eiden (2023) pointed out that people have a tendency to use the largest LLMs to get a response for trivial stuff which can be handled by a smaller model, such as a search engine and just play with it without clear purpose.

According to the study by Eugenie and Gelles-Watnick (2023), user behaviour with ChatGPT was divided based on generation. The study with early users of ChatGPT in the U.S. showed that younger adults under 30 had different tendencies regarding the usage of ChatGPT. They were more likely to use it for education or amusement. Moreover, 20 per cent of U.S. adults used it for entertainment; 19 per cent of respondents answered that they used it for learning something, followed by 16 per cent of respondents who used it to complete tasks at work.

2.2.4. Environmental Impact of Generative AI

Recent advancements in machine learning are leading to remarkable performance across various tasks that have been energy-intensive since the development and deployment of ML models require access to computational resources such as Graphical Processing Units (GPUs) (Luccioni & Hernandez-Garcia, 2023). For instance, during training of a large language model like ChatGPT, which takes weeks or months, the model's performance is enhanced through parameter tuning and the parameters significantly increase as the generation of these models evolves (Cho, 2023).

In this context, GPUs are commonly used due to their capacity to execute numerous calculations or processes simultaneously, which require 10 to 15 times more energy to power machine-learning models, compared to CPUs (Central Processing Units) which are traditional computational resources (Kumar & Davenport, 2023). According to Kumar & Davenport (2023), most well-known GenAI models are generated using thousands of servers powered by GPU chips, resulting in a substantial carbon footprint.

As models become more sophisticated, the energy consumption of data centres for training and utilising these models can become unsustainable, particularly with major technology companies developing their own models, which leads to a substantial training load on data centres (McQuate, 2023). Koot & Wijnhoven (2021) estimated that the global data centre electricity consumption could rise from approximately 1.15% of global electricity demand in 2016 to 1.86% by 2030. In this context, the concept of Green AI was coined by Schwartz et al. (2019) to advocate for *"AI research that is more environmentally friendly and inclusive"*. To achieve Green AI it is essential to understand the AI system's life cycle as AI systems are significantly complex (Verdecchia et al., 2023).

Carbon Footprint from Training and Inference

According to Kumar & Davenport (2023), the carbon footprint life cycle of machine learning models is divided into three categories: 1) the carbon footprint from training the model, 2) the carbon footprint from running inference and 3) the carbon footprint required to produce all of the needed computing hardware and cloud data centre capabilities. While training models represent the most energy-intensive phase, inference and using the models for user prompts require less energy. However, as inference eventually involves enormous sessions used by millions of users, the carbon footprint becomes significant.

In a similar vein, Chien et al. (2023) emphasise that "*much previous work has focused on the carbon impact of model training, and we believe inference (operation) can also be problematic, particularly with rapid user growth and integration into everyday applications*" (p. 1). When considering hundreds of millions of queries on ChatGPT costs around 1 GWh each day, which is equivalent to the energy consumption of 33,000 U.S. households (McQuate, 2023).

Despite the efforts of data centre operators to use more renewable energy sources, the substantial energy demand of GenAI makes it inevitable to rely on traditional electricity (Cohan, 2023). Research by Luccioni and Hernandez-Garcia (2023), investigating the carbon emissions of 95 ML models over time and across various tasks in natural language processing and computer vision, shows that the majority of ML models from their sample relied on high-carbon energy sources like coal and natural gas as their primary energy source.

Approaches to Reduce the Environmental Impact of GenAI

Electricity consumption has a carbon footprint unless all electricity is generated by renewable energy sources (Das & Modak, 2023). In an attempt to reduce carbon footprint and energy consumption, Google introduced the 4Ms framework: Model, Machine, Mechanization and Map Optimization (Patterson et al., 2022). Patterson et al. (2022) elaborate on the framework by arguing that the best practices begin with the selection of efficient ML model architectures. Moreover, they emphasise the utilisation of processors and systems that are optimised for ML training and computing; it is recommended to conduct these operations in the cloud rather than on-premise.

Additionally, from the perspective of specific engineering practices, it is required to use existing models and carefully consider the appropriate size and complexity of AI models, rather than always using LLMs (Su, 2023). AI practitioners and engineers also contribute to advancing towards greater algorithmic efficiency by using techniques such as transfer learning, fine-tuning, and mixed-expert models (Bashir et al., 2024).

Gaur et al. (2023) argue that AI can be deployed to lower carbon emissions and decrease the effects of climate change, emphasising AI is a double-edged sword. Nonetheless, they stress that it is essential to ensure the use of AI in a sustainable way.

2.2.5. Prompt Engineering

In Generative AI models, a prompt refers to the text user's input to direct the model's output. Their inputs can vary from simple questions to detailed descriptions or specified tasks. In image generation models like DALLE-3, prompts typically describe, whereas, in large language models such as GPT-4 or Gemini, they can cover anything from simple questions to intricate problem scenarios (Amatriain, 2024).

Prompts can be classified into various types according to their structure, purpose, and the degree of guidance they offer to the AI model. There are three types of most common prompts explicit, implicit, and creative (Data Science Horizon, n.d.);

- **Explicit prompts:** give clear and straight instructions to the AI model, more accurately they provide the exact format or information required in the generated output.
- **Implicit prompts:** require less guidance in their instructions, allowing the AI model the flexibility to interpret the desired outcome.
- **Creative prompts:** are formulated to inspire AI models to produce innovative, imaginative, or unconventional outputs (Data Science Horizon, n.d.)

Prompt engineering, however, involves crafting text prompts to assist large language models in generating outputs that are more accurate, consistent, and creative. With careful selection of words and phrases in a prompt, prompt engineers can impact in which way the LLM understands a task and the results that it provides (TutorialsPoint, n.d.). It is a challenging but crucial step in optimising the performance of LLM on personalised tasks. The challenging parts are also the requirement needed for complex reasoning to be able to examine the prompt errors, hypothesise what is missing or misleading and still communicate the task with clarity (Ye et al., 2024).

Prompt engineering entails analysing data and task requirements, adjusting the language model accordingly and crafting and improving prompts. Modifications to prompt parameters like length, complexity, format, and structure are implemented to enhance the model's performance for the given task (Takyar, n.d.).

Furthermore, prompt engineering mirrors traditional software engineering practices like version control (the practice of tracking and managing changes to software code) and regression testing (testing if updates or changes had caused new defects in the existing functions), involving an iterative and exploratory approach. The rapid growth of this field indicates its capacity to transform specific aspects of machine learning, surpassing conventional methods like feature or architecture engineering, particularly within the realm of large neural networks. On the other hand, traditional engineering methods such as version control and regression testing require adaptation to align with this new model, similar to how they were modified for other machine learning approaches (Amatriain, 2024).

Exemplary instances of prompt engineering can be seen in OpenAI's GPT-3 model utilised for translation and creative writing, Google's Smart Reply feature for automatic message responses, and DeepMind's AlphaGo for excelling in the game of Go. In all these scenarios, meticulously designed prompts were employed to train the models and steer their outputs towards accomplishing specific results (Takyar, n.d.).

Lastly, in ChatGPT's Prompt Engineering for Developers course, Isa Fulford and Andrew Ng defined two primary facets of prompt engineering:

- **Formulation of clear and specific instructions:** this concept defines the significance of being conscious and specific when constructing prompts. The clearer the prompts are the greater the quality of output results are provided (Takyar, n.d.).
- **Allowing the model time to "*Think*":** this concept highlights the importance of allowing the model sufficient time to process the provided information. Including "*thinking time*" within the prompts can enhance the model's ability to process and comprehend the input, resulting in better outputs (Takyar, n.d.).

3. Research Design and Methods

The research design and methods section provides a detailed description and implementation of the methods used for both the data collection as well as data analysis, which is a crucial part of our project. This section outlines the framework for collecting, analysing, and interpreting data, ensuring the research is accurate, consistent, and aligned with our project's objectives. By detailing the research design, data collection methods and data analysis methods, this chapter establishes the validity and reliability of the findings. Moreover, at the end of this section, we go more in-depth about the validity and reliability of the methods as well as into the ethical considerations and methods limitations.

3.1. Research Design

As a framework for data collection and analysis of data, it is crucial to build a research design thoroughly for researchers to gain high-quality data and analyse adequately (Clark et al., 2021). In this project to make use of its advantages a mixed-method research design was adopted, which is "*a study that employs both quantitative and qualitative methods as part of a single research strategy*" (Clark et al., 2021, p. 557).

Research design can be divided into three categories in a high-level: quantitative, qualitative and mixed-methods research design (Asenahabi, 2019). In general, quantitative and qualitative research designs are employed independently (Caruth, 2013). Each approach offers distinct advantages: qualitative research provides deeper insight into a study, while quantitative research is beneficial to gain better objectivity and generalisability (Lund, 2012 as cited in Caruth, 2013).

Combining both quantitative and qualitative methods offers strengthened insights into the research problems and questions than deploying one of the methods for the research (Creswell, 2012; Frels & Onwuegbuzie, 2013; Hong & Espelage, 2011, as cited in Caruth, 2013). In terms of a combination of methods, several combinations of methods can be utilised in the same study as mixed methods. According to Clark et al. (2021), for instance, both a questionnaire and focus groups are conducted within a single research design.

When diving deeper into mixed-method research design, there are three distinct approaches: equal status, qualitative dominant and quantitative dominant mixed-method research design (Johnson et al., 2007);

- **Equal status mixed-methods research:** Researchers who use this approach perceive the value of qualitative and quantitative data equally and believe that they can robust insights for research questions by combining them.
- **Qualitative dominant mixed-methods research:** Qualitative methods are emphasised in the study, including quantitative data and approaches to enrich their findings in qualitative research projects.
- **Quantitative dominant mixed-methods research:** In this way, researchers focus on quantitative methods, incorporating qualitative data to provide additional context and depth to their quantitative projects.

Quantitative dominant mixed-methods research is used in this project for several reasons. To investigate user behaviour with ChatGPT and their awareness of climate change and ChatGPT's CO2 footprint, the questionnaire is primarily used to collect data in this study. Simultaneously, it is fundamental to gather qualitative data to understand users deeply in terms of their thoughts, behaviour and experiences regarding GenAI, ChatGPT and CO2 footprint. For this reason, one of the five sections consists of entirely open-ended questions in the questionnaire. Furthermore, contextual inquiry is used to observe how users interact with ChatGPT and to get insights into their ideas on sustainable practices. In conclusion, our study is focused on quantitative data, incorporating qualitative data providing an in-depth understanding of users.

3.2. Data Collection Methods

Section 3.2 describes the data collection methods used in this master thesis, focusing on the general description and the implementation of the questionnaire and contextual inquiry. These methods were chosen to capture both quantitative and qualitative data. The questionnaire method focuses on the collection of numerical and qualitative data to determine the statistical analysis of trends and patterns as well as some further insight into participants' thoughts. On the other hand, contextual inquiries offered in-depth insights into participants' behaviours and experiences within ChatGPT, enhancing the questionnaire findings with qualitative depth.

3.2.1. Questionnaires

3.2.1.1. General Description of Questionnaires

A questionnaire refers to a list of questions completed by respondents to provide their opinion and a survey as a process enables people to illustrate their opinions, interests, and preferences in a structured way (Roopa & Rani, 2012; Goodman et al., 2012). The success of surveys lies in the design of questionnaires, and through surveys both quantitative and qualitative information from a target audience can be collected (Grimshaw, 2014; Roopa & Rani, 2012). By conducting surveys researchers can reveal not only broad features and patterns about users but also a higher level of certainty about the overall user population (Goodman et al., 2012).

Clark et al. (2021) elaborate on self-completion questionnaires by comparing them with structured interviews in their book. According to them, self-completion questionnaires have more close-ended questions to make respondents easier to answer and are designed in an easy-to-follow way to minimise risk and fatigue leading the respondents to end up not completing the questionnaire.

When designing questionnaires, it is essential to consider the cognitive process as an iterative complex information processing (Lietz, 2010). According to Lietz (2010), When answering a questionnaire, the process starts with understanding the question and retrieving related information from memory to generate matched answers to the questions. Therefore, the following elements need to be considered to avoid cognitive overload: *question length, grammar, specificity and simplicity, social desirability, double-barrelled questions, negatively worded questions, adverbs of frequency and question order.*

It is recommended to formulate questions in an active way rather than the passive voice and repeat nouns instead of pronouns. Additionally, researchers need to bear in mind that "*respondents might choose to select a certain position that is thought to be one that is favoured by society*" (Lietz, 2010, p. 253).

In terms of format, the most common format of a survey is a digital format which is online due to its convenience (Clark et al., 2021). For instance, responses are available for analysis as soon as respondents complete the survey; additionally, researchers have an unlimited scope in determining sampling due to the large scale of web users (Umbach, 2004).

Questionnaire Design

When designing questionnaires there are important principles to follow. As an initial step, researchers should consider what needs to be accomplished through the survey. In that regard, there are two different types of goals. First, the descriptive goal aims to profile users such as their personal characteristics, what they want and how they behave. On the other hand, the explanatory goal aims to explain people's beliefs and behaviours by detecting relationships between their answers (Goodman et al., 2012).

Once the goal has been set, questions need to be formulated through brainstorming. According to Goodman et al. (2012), the survey can be divided into several parts:

- **Introduction**: Presenting the purpose of the survey, instruction, duration and contact information of the researcher.
- **Beginning and middle:** Drawing respondent's interests and questions grouped thematically.
- **End:** Concluding with demographic questions and open-ended fields.

The type of question is applied differently according to the purpose of the survey. The most representative types are close-ended and open-ended questions. Close-ended questions where respondents' answers are limited to a set of possible responses include *Yes/No questions, multiple choice and scaled questions* (Roopa & Rani, 2012). On the other hand, when answering open-ended questions where they are asked in a variety of forms such as completely unstructured format, word association and sentence completion, respondents are not constrained (Roopa & Rani, 2012).

Moreover, a close-ended format is easier for not only respondents to answer but also for researchers to compare the answers; On the other hand, open-ended questions allow researchers to evaluate the levels of knowledge and understanding of the topic (Clark et al., 2021). Throughout creating questions, making unambiguous questions is the most crucial (Goodman et al., 2012).

Pilot Testing

The purpose of pilot testing is to detect any potential flaw in the questionnaire prior to launching the main survey (Srinivasan et al., 2017). It is necessary to conduct pilot testing once the initial draft of the survey has been finalised (Ruel et al., 2015). Ruel et al. (2015) elaborate that pilot testing is mainly used to assess the clarity and comprehensibility of both questions and response options from respondents' points of view. Additionally, they emphasise that pilot testing plays a crucial role in ensuring a shared understanding between researchers and respondents by observing that respondents interpret the survey in the same way as researchers.

Pros and Cons of a Questionnaire

When it comes to advantages of a survey, it is more cost and time-effective than interviews since they do not involve interviewers when conducting the survey; researchers do not affect the answers respondents give, which means respondents feel easy with less social desirability bias leading to behave according to researcher's expectations (Clark et al., 2021).

However, there are limitations of a survey since researchers can not interrupt and help when respondents have difficulty answering the question owing to ambiguity; even though the respondents miss so many questions researchers can not prompt immediately and it can affect the quality of data collection (Clark et al., 2021). Furthermore, if a survey is not designed carefully, it can easily go wrong by asking the wrong questions to the wrong people, generating inaccurate results (Goodman et al., 2012).

3.2.1.2. Implementation of Questionnaires

Designing a Questionnaire

To begin with, the aim of the questionnaire was discussed. The goal of the survey was to investigate the user behaviour with GenAI & ChatGPT and the user's awareness of climate change and the CO2 footprint of ChatGPT. To achieve both descriptive and explanatory goals, the questions were created in a variety of forms and placed in different sections appropriately. Moreover, as this research follows the mixed method approach the questions were combined with close-ended questions and open-ended questions for both quantitative and qualitative analysis.

The questionnaire consists of five sections: *1) Demographic, 2) Utilisation of GenAI, 3) Utilisation of ChatGPT, 4) the latest experience of ChatGPT and 5) Awareness of the environmental impact of ChatGPT*. Each section has its instructions to gain desired responses and prevent participants from misunderstandings. Furthermore, a description of the survey and consent are introduced in the initial section. The survey was implemented using Microsoft Forms. The questionnaire used for the survey can be found in this link and Appendix A.

Section1: Demographic

The participants were asked to answer their age, gender identity, and employment status. Commonly, demographic questions are used to detect "*whether identity is causing an individual to do a specific thing or if something is causing an individual to adopt a certain identity*" (Hughes et al., 2016, p. 138).

In this case, the demographic questions were created with seven questions to see if there are remarkable differences in user behaviour and awareness based on identity. On top of that, the awareness of climate change was asked to measure how well people are aware of climate change and how much they consider climate change in their daily activities in a Likert scale format from 0 to 10. The two questions related to the awareness of climate change were placed in the first section to remind respondents of one of the main topics of the questionnaire, which is the awareness of the environmental impact of ChatGPT.

Section 2: Utilisation of GenAI

The second section consists of 5 questions asking about their general behaviour with GenAI. The purpose of GenAI and their experience and behaviour with GenAI tools were asked.

For instance, what is the purpose of using GenAI between text generation and image generation, what type of GenAI tools they have used and which GenAI tools they used most frequently were asked. Unlike other questions in this section, question 8: *"What types of Generative AI tools have you used? (Select all that apply)"* is created in a multiple choice format with representative GenAI tools to discover which tool is most used. The options of 4 representative tools were chosen based on the list of the Most Visited Websites In The World (See Howarth, 2024, for more detail).

The open-ended question, question 10: *“In what contexts do you use Generative AI? Please provide further details on the usage of Generative AI”* was also created, enabling us to gain qualitative data in terms of in-depth understanding. Lastly, the question asking whether they have used ChatGPT is placed as the last question since the next section is linked to the utilisation of ChatGPT.

Section 3: Utilisation of ChatGPT

As the vast majority of the research questions are directly connected to ChatGPT the questions from this third section are deeply focused on ChatGPT. The “*Utilisation of ChatGPT*” part, consisting of five questions, was built to detect the user’s behaviour with ChatGPT and the use of prompts.

First, participants were requested to answer the close-ended questions: *“How often do you use ChatGPT?”* and *“For what purpose do you use ChatGPT?”*. Afterwards, they were asked to rate their usage of ChatGPT on seven different statements, indicating a scale from 0 (no usage) to 5 (frequent usage). This scale question was created to investigate RQ3, closely related to users’ ChatGPT usage patterns and preferences with representative examples. The seven examples are inspired by findings from the previous studies, User Behaviour with GenAI in Chapter 2. The statements are as follows: *Ask simple questions, Get help with writing (e.g., email, invitation), Summarise content, Get assistance with coding, Creation (e.g., write a song, novel), Translate text and Generate images.*

Furthermore, question 15: *“On average, how many attempts do you rephrase your prompts until you achieve the desired outcome when using ChatGPT?”* was created with five options from 1 to over 20. Participants who said that they rephrased their prompts over eleven times on average were supposed to answer the reason in an open-ended question, question 16: *“In which situations or contexts do you often find yourself rephrasing prompts?*”. Through these close-ended and open-ended questions, we assume that we would figure out the average number of prompts and challenges or reasons to cause many prompts.

Section 4: ChatGPT's latest experience

For an in-depth examination of the usage of ChatGPT, consisting of five open-ended questions, was created. To reduce the risk of relying on participants’ vague memories, we created this section requiring participants to access their ChatGPT account and review their chat history.

The questions in this section were built to achieve an explanatory goal where researchers can detect the relationship between their answers and draw their behaviours (Goodman et al., 2012). For example, the section starts with the questions: *“What is your latest chat about?”* and *“How many times did you create prompts in the latest chat?”* Following the initial questions respondents were asked about positive and poor experiences during their most recent chat and any thoughts on how ChatCPT can be enhanced to meet their needs. Investigating the relationship between the chat topics, the number of prompts, and participants' satisfaction can be significant in understanding their interactions.

Section 5: Awareness of the environmental impact of ChatGPT
In this last section, we aimed to comprehend how well users are aware of the environmental impact of ChatGPT. Different types of questions were formed: Likert scale, ranking question, Yes/No question, multiple choice question and open-ended question. For the question, question 22: *“How well do you believe you understand the CO2 footprint of ChatGPT?”* participants were asked to rate to what extent they are aware of the CO2 footprint of ChatGPT between 0 (not at all) to 10 (extremely well). As the concept of awareness is subjective, the Likert scale was used to measure their awareness in a consistent way.

The ranking question (Q23) was formed to observe to what level participants perceive the environmental impact of ChatGPT in real life. Furthermore, it can be detected whether their thoughts correspond to their knowledge; in other words, what they actually knew. To measure it, five different digital activities created based on several research (See Eiden, 2023; Walkley, 2022; Hölzle, 2009 for more detail) were given to be ranked. Table A1 of Appendix A shows digital activities and CO2 amount used as the options in Q23.

On the other hand, we aimed to discover to what extent they are likely to contribute to reducing CO2 emission in terms of behavioural change. With the Yes/No question, we also formed an open-ended question, question 25: *“Please specify the reasons for your response”* to examine if users are willing to change their behaviour to decrease the CO2 impact from ChatGPT and their attitudes and thoughts regarding the issue. Additionally, question 26: *“In your opinion, who is primarily in charge of limiting the environmental impact of ChatGPT?”* was asked with the options: Government, Companies, Individuals and others.

Conducting Pilot Testing

Before releasing the survey, pilot testing was conducted. Through the pilot testing, it was observed how respondents interacted with the initial questionnaire form and some misunderstandings caused by improperly formatted questions in terms of bias, grammar specificity and simplicity.

Participant 1
The participant was provided with the link ahead of time to go through the questionnaire and complete it. After the participant finished the questionnaire, they informed us and we had a meeting to hear and discuss the feedback. The main response that we acquired from this session is that the structure of the questionnaire was well made and all the questions were understandable. The negative critique was on the ranking question as they were not sure on what to base the statistic on and were confused as to why we had option 1 and option 2 at the time. Overall the participant had no issues with going through the questionnaire, appreciated the small introductions on sections and understood all the questions.

Participant 2
The pilot testing for participant 2 was conducted in real-time. The participant was provided with the link to the Microsoft Form and asked to fill out the survey. The researcher took a note while observing how the participant interacted with the survey. Afterwards, an interview with the participant was implemented. Overall, the participant did not have challenges to fill out the survey and it took eleven minutes to finish the form, corresponding to the estimated time. However, it was seen that the participant was struggling with comprehending one question, which asked people to rank the amount of CO2 emissions. The participant spent

some time figuring out how to adjust the options in MS Forms. Moreover, during the interview, it was mentioned that the options in the ranking question seemed not relevant to the purpose of the survey. For this reason, the participant was not fully motivated to answer that question.

In summary, the pilot testing gave us great insights into developing clear options for the ranking question (Q23), previously described in Section 5. Originally, the options were three, including the amount of CO2 emissions from three different sectors such as transportation and aviation. However, based on the results from the pilot testing, the options were reformulated with either digital activities or daily activities that can be easily imagined by people.

Prolific

To conduct our surveys we used a website called Prolific. Prolific is a platform that links researchers with participants for online studies and surveys. In Prolifics (2024) words it was "*created by researchers for researchers, we built Prolific to provide a better way to get high-quality human data and input for cutting-edge research.*". It was founded in 2014 at Oxford University. Some interesting facts about it are that over 35 thousand researchers are using Prolific and over 700 studies are launched every day (Prolific, 2024).

The application is popular because researchers can easily create projects and recruit participants, while participants can earn money by completing studies in various fields like psychology, sociology, economics, etc. It's known for its user-friendly interface and efficient matching system (Prolific, 2024).

Implementation

To use Prolific we needed to create accounts and set up a project or study on the website. The set-up process was divided into four steps: 1) Study Details, 2) Data Collection, 3) Recruit Participants, and 4) Study Cost.

In the first section, Study Details, we needed to provide our project with a title, internal name and a short description of what the survey is researching and what the participants are expected to be carrying out. There were also optional selections such as labelling the study, adding a content warning, and choosing which devices the participants could take the survey on.

The next section was Data Collection. Within this segment, we provided Prolific with the external link to our Microsoft Form survey. Additionally, for easier analysis, we added a question to our survey asking for Prolific ID. With the unique ID, we could track individual participants' submissions. Lastly, after the participants submitted the survey a link showing them the unique completion code was provided. The completion code served as a way for the participants to receive payment.

In the third step, Recruit Participants, we were required to establish the participants segment in terms of sampling. We decided on the number of participants we were looking to recruit, the participant's location, as well as study distribution. Within this section, we added a condition that only users with experience with ChatGPT can participate for more valid and reliable results. After setting all the conditions that we wanted we ended up with about 50

000 eligible participants who have been active in Prolific. The population for questionnaire data collection were users from all over the world. Since the study is about Generative AI and ChatGPT we decided to keep a broad perspective on the users we gather data from. We listed a couple of characteristics of the population:

- Location: Worldwide
- The main focus was on online users who have used ChatGPT at least once previously, so we can collect more accurate information.

Lastly, in the Study Cost section, the project needed to be provided with an approximate time to complete. In our case, the approximate time was set to 13 minutes as that was the time it took our pilot test participants to finish the survey. With time the cost was automatically calculated and it ended up being about 2 pounds per participant.

3.2.2. Contextual Inquiry

3.2.2.1. General Description of Contextual Inquiry

Contextual Inquiry is also known for being a part of Contextual Design used as a field data-gathering technique. In many studies, it is used to reveal unconscious and implied aspects of life, such as going out and gathering field information, talking with people about their work and observing them. By doing so, information can be gathered while observing the users engaging in their daily activities (Kip et al., 2018; Kim et al., 2012).

Gathering information for a product or research paper is not just about conducting interviews or simply asking users what they need. You can not directly ask users about their needs for a technology they do not fully understand or know the capabilities of. Moreover, everyday tasks become habitual and unconscious things a person does. For that reason, users often have difficulty expressing their practices (Holtzblatt & Beyer, 2015).

The main principle of contextual inquiry (CI) is to incorporate concentrated observation and if required engage in conversation while the participant is performing a task. The task can vary and as such can be: taking field notes, gathering information, etc. Conversations with the user should be conducted in a seemingly informal manner (Privitera, 2015).

Although this method seems easy to complete at first glance by just talking and observing the participants, it does have its specific requirements. To start the data collection process it required specific approaches and certain interview skills to be able to capture all the relevant information in a CI study (Privitera, 2015).

Furthermore, CI observation needs to respect the participant's rights, needs and desires. Consent for observing the users and using the data gathered is also a requirement. There are five approaches of observation. When choosing which one to use there is no right or wrong answer only better or worse suited for the selected study. The approaches are: *overt observation approach, think-aloud approach, use of simulation with reflection, role-playing approach and note taking* (Privitera, 2015).

Along with observing approaches, there are also interviews to be taken into consideration. The CI interviewers need to have a general knowledge about the study they are researching so, the choice of interview approaches is narrowed down (Beyer & Holtzblatt, 1997). The interviewers need to be able to keep the flow of the conversation going and not strictly follow the questions. For that reason, interview approaches can be defined as three primary ones: *semi-structured, unstructured and narrative interviews* (Privitera, 2015).

Framework of Contextual Inquiry

To conduct a good contextual inquiry it can be said that a "*master/apprentice relationship*" with the participant is desired. That way the user can teach the interviewer, the apprentice, how to complete a certain task or function. The interviewer's job throughout the session is to continuously ask questions to acquire a better understanding of why certain activities are performed in a certain way (Kronberg, 2020). CI interviews are one-on-one interviews which can last up to 1.5 to 2 hours (Duda et al., 2020). Moreover, the roles of a CI can be divided into that of an observer and an interviewer. Thus the best possible insights can be discovered, including the participant's unconscious actions (Bednar, 2010).

Additionally, the interviewer needs to be able to empathise with the user, be attentive to details and be modest. According to Holtzblatt & Beyer (2015), "*People can respond with reliable information when they are doing their own activities in their own life context*" (p. 72). Interviewers should if possible try emerging themselves into the world of the user to discover what is important from their point of view (Duda et al., 2020).

The base guidelines for any CI are "*Be concrete, focus on describing a specific task or event, and avoid all generalisations and summaries: don't allow the user to summarise, abstract, or report*" (Duda et al., 2020, p. 3). Trying to focus on a specific task may help the participant to better explain as well as help them remember past events (Duda et al., 2020).

Although the observer will be taking notes as the CI session is being carried out it is highly advised to have the interviewer take notes themself as well. The reason is that a recording or the second person, the observer, might not be able to catch everything. The interviewer is in close proximity to the participant and can therefore capture emotional responses and tiny reactions better than the observer or a recording of a session (Holtzblatt & Beyer, 2015). According to Holtzblatt & Beyer (2015), there is a structure and four principles od CI.

The four principles of contextual inquiry are**:**

- **Context:** Creating a familiar atmosphere in which the user can recall a specific action or past task regarding the study.
- **Partnership:** Let the user lead the interview and have the interviewer learn and observe the tasks being completed. The "*master/apprentice relationship*" in which the interviewer, the apprentice, learns from the user, the master, and asks questions in regards to the tasks.
- **Interpretation:** Having both the user and interviewer share the understanding of the tasks. After the tasks are completed the interviewer shares what they interpreted and the user either confirms or corrects them.
- **Focus:** The focus point of the study needs to be established to steer the conversation of the interview to fall within the project scope.

The structure of a contextual inquiry is as follows:

- **Introduction:** Welcoming and explaining, getting to know the participant and assessing the overall knowledge of the subject of study.
- **Transition:** Explaining how contextual inquiry works.
- **Contextual Interview:** Observing the participant's actions and body language, asking relevant questions and interpreting. Looking for habits or emotional reactions.
- **The wrap-up:** The interviewer may share a summary of the session and what they have learned for validation.

Pros and Cons of Contextual Inquiry

The benefit of conducting a CI is mainly in the information we can receive from it. Doing interviews, and asking participants how they accomplish a task or why they use a specific function is not always as reliable as seeing it happening in real time. Observing participants in context can provide more reliable and valid information about what they are doing, how they are doing it, and why they are doing it (Duda et al., 2020). Moreover, since CI is conducted in the participant's comfort zones, the quality and accuracy of the information provided are higher (Holtzblatt & Beyer, 2015).

Despite all the benefits, CI can be a challenging method to use. For instance, in many cases, it takes more effort to recruit participants, who are willing to let their activities and actions be observed. Sometimes, some legal issues may occur as well, especially in a work environment (for confidentiality reasons, or because it would be too disruptive) (Duda et al., 2020). Another problem that may occur is that the participant feels pressure from being observed. Lastly, the method in itself is relatively time-consuming since one session can last up to 2 hours (Holtzblatt & Beyer, 2015).

3.2.2.2. Implementation of Contextual Inquiry

Set up of Contextual Inquiry

The aim of contextual inquiry was to dive deeper into user behaviour and their habits. To be more specific we wanted to gather in-person data on how a user uses ChatGPT and what their thought process is while going through the application. To attain that information we thought of two scenarios to play out with our participants as well as had a semi-structured interview session afterwards.

Before arranging the CI sessions we had to first think of scenarios on how we wanted to test the participants. Simply thinking of the tasks was not as easy as we had imagined as such we first had a brainstorm about the aim of the method itself. Some of our notes included ideas of giving situations instead of instruction, implementing a think-out-loud method and doing a deep dive into user behaviour. These notes served as a reminder during the setup process to not lose focus on what our aim of the method was.

In the end, we came up with two scenarios, which are:

1. You might wonder whether Labor Day is a public holiday in Denmark (or other country). How would you use the ChatGPT to get an answer?
2. You are applying for a job by email. How would you get assistance with ChatGPT? Please write an email based on your role and situation.

We carefully chose those scenarios because we wanted to keep the tasks relevant for all participants for reliability and validity purposes. Moreover, these scenarios alone provided us with vital information that we wanted to gather. With the first task, the participants can show how they use ChatGPT for simple questions and explain why (if applicable) they would not simply google that fact. While with the second task, we can see the participants use ChatGPT more in-depth to get a satisfactory response. Additionally, we will ask the participants to carry out the thinking-out-loud method while conducting the scenarios.

Afterwards, we began thinking of questions for the semi-structured interview we were planning to have with our participants at the end of the sessions. We wanted the questions to be relevant to the previous scenarios as well as connect them a bit to the questionnaire results.

As such we ended up with five questions and two sub-questions:

1. For the first question we asked, would you ask that type of question on Google instead of ChatGPT?
2. In case you need to search again the same topic you already searched before, would you rather create a new chat or revisit the history and continue to ask?
    a. Do you usually create a new chat or revisit the history and continue to ask in case you need to search again for the same topic you already searched previously?
3. Have you ever rephrased prompts because ChatGPT provided you with false information?
4. In what context do you usually rephrase prompts?
5. According to our research, the more prompts you create the more CO2 is emitted. How would you cut down the number of prompts?
    a. Any suggestions - how can the interface of ChatGPT be enhanced to reduce the number of prompts? e.g., reduce misspellings, and guidelines embedded in the interface.

We arranged the order of the questions to make sense with the order of the tasks completed previously. Additionally, since we chose the semi-structured format for the interview we did not want to create too many required questions but rather have an open discussion at the end.

Lastly, when it came to recruiting the participants we wanted to have two examples. To do that we looked at our close associates to have one participant be a heavy user and the other a normal user of ChatGPT. By heavy user what we meant was that the participant used the application daily and had a paid version while the normal user used it often and had an unpaid version. With those participants, we could see two different ways of thinking and completing the same tasks in ChatGPT.

Testing

As Contextual Inquiry requires researchers to observe how participants interact with software it was inevitable to record the meetings. For this reason, consent for recording the meeting and sharing their screen was asked, as well as the background of the project and the aim of the CI session in the process of recruiting participants. Once they agreed with those requirements beforehand they were invited to the meeting and it was confirmed before starting the CI session once again.

To conduct the contextual inquiry method, virtual meeting platforms were used such as Zoom meetings and Microsoft Teams which have functions to record and transcribe the meetings, as well as good accessibility for users as a free version of software.

The participants joined the online meetings through the researcher's personal meeting room with a password. To begin with, it was consented that the recording and transcript will not be used outside of the project, the data will be deleted right away once the project is finished and the names of participants will not be revealed on the transcript which is required to be submitted for school once again. It is significantly considered to protect participant's personal information during the implementation of CI.

The sessions follow this structure on a high level: Asking for consent, Providing task1, Providing task2 and semi-structured interviews. Each participant was invited to the online meeting separately and both researchers joined each meeting, dividing the roles into leading the session and taking notes from observation.

During the think-aloud method, participants received some questions and they were asked to implement additional actions from researchers. The vast majority of questions were related to their behaviour, for instance, asking whether the participants had a specific reason to act that way and how they could do it differently. Furthermore, additional actions were asked in the situation where they had done tasks too quickly. In this case, follow-up scenarios were given to observe how they interact with ChatGPT as thoroughly as possible, including how they create the initial prompt and how they react to the chat when creating further prompts.

3.3. Data Analysis Methods

This section outlines the data analysis methods used in this master thesis, including analysis such as quantitative analysis, thematic analysis, content analysis, sentiment analysis, and text mining. These diverse analysis techniques were selected to carefully interpret the previously collected data, providing both numerical and significant insights. Quantitative analysis provides the statistical examination of numerical data, revealing patterns and correlations. Thematic and content analyses were applied to qualitative data, identifying recurring themes and underlying meanings. Additionally, sentiment analysis was conducted with content analysis to evaluate the emotional tones within the data, while text mining provided the extraction of relevant information from the text. Together, these methods ensure a thorough and all-round analysis, enhancing the depth of our research findings.

3.3.1. Quantitative Analysis

3.3.1.1. General Description of Quantitative Analysis

The quantitative method entails gathering and analysing numerical data to address scientific research inquiries. It can be used to summarise, find patterns, make predictions and test associations in numerical results (Rana et al., 2021).

Moreover, when it comes to analysing the data there are two main approaches: descriptive and inferential statistics. Descriptive statistics offer a numerical overview of data and summarise information. Data description typically involves frequency, measures of central tendency, and measures of dispersion (Patel, 2009). As for inferential statistics they extend beyond mere description by striving to achieve the intensity and direction of relationships among variables. Some of the most frequently used inferential statistics are hypothesis testing, simple regression, multiple regressions, chi-square, etc. (Rana et al., 2021).

Pros and Cons of Quantitative Analysis

The advantage of this analysis is that the findings are likely to be very consistent, precise and reliable. Besides that, the data analysis is less time-consuming as it is achieved by coding and not manually. Lastly, it benefits from being a cost-efficient analysis (Rahman, 2016).

Despite providing the advantages of the analysis above, the analysis does have its limitations. One of its main downfalls is the inability to gather a deeper understanding of meanings and explanations. Another issue is the reliability of the data gathered which highly depends on the quality of the answers provided (Almeida et al., 2017). Additionally, the quantitative analysis overlooks the participant's experiences and perspectives because it lacks direct connections between the researchers and participants when gathering data (Rahman, 2016).

3.3.1.2. Implementation of Quantitative Analysis

In order to conduct a quantitative analysis we first had to decide on which data we wanted to work with. Since the questionnaire was conducted with close-ended questions it was the ideal data set to work with.

We began by conducting a quick brainstorm of which answers fit together. Afterwards, we looked through our already defined research questions to determine which questionnaire questions/answers fit best with which research question.

At the end of the brainstorming, we had a list of research questions with the questionnaire questions that answered them. For example, we had placed together the RQ3: "*What are the common usage patterns and preferences of ChatGPT users?"* and question 14 of the questionnaire: *"Please rate your usage of ChatGPT based on the following statements, on a scale from 0 to 5. "0" indicates no usage, while "5" indicates frequent usage"*. With question 14 we can determine ChatGPT usage patterns and their preference among various statements.

Moreover, we also immediately thought of which graph should go with which questionnaire question. For example, for the provided above question 14 we decided to visualise the results in a stacked bar chart.

Lastly, the programming language we used to conduct the analysis is R. R is a programming language that is strictly case and character-sensitive. This implies that the imputed instructions follow the specific syntactic rules of the language in a console or command-line interface. The software then interprets and executes the code and returns the results (Tilman, 2016). According to Tilman (2016), *"R is what's known as a high-level programming language. Level refers to the level of abstraction away from the fundamental details of computer execution"* (p. 4).

3.3.2. Thematic Analysis

3.3.2.1. General Description of Thematic Analysis

Thematic analysis is a representative method for analysing qualitative data, involving identification, analysis and interpretation of patterns of meaning, which are themes (Clarke & Braun, 2017). Thematic analysis is utilised to determine patterns within and across data collected in connection with *"participants' lived experience, views and perspectives, and behaviour and practices; 'experiential' research which seeks to understand what participants' think, feel, and do"* (Clarke & Braun, 2017, p. 297).

It is commonly compared to content analysis which allows researchers to analyse data qualitatively and also quantify the data (Vaismoradi et al., 2013). *"Content analysis uses a descriptive approach in both coding of the data and its interpretation of quantitative counts of the codes"* (Downe-Wamboldt, 1992; Morgan, 1993, as cited in Vaismoradi et al., 2013), whereas thematic analysis produces a solely qualitative, detailed, and nuanced interpretation of data (Braun & Clarke, 2006, as cited in Vaismoradi et al., 2013). A variety of qualitative data exist such as *"recorded observations, focus groups, texts, documents, multimedia, public domain sources, policy manuals, and photographs"* (Nowell et al., 2017, p. 4).

According to Clarke and Braun (2017), thematic analysis is not a method to simply summarise the data content. Instead, it plays a key role in identifying and interpreting essential characteristics of the data in accordance with the research question. Additionally, they stress that the research question has the possibility to evolve as coding and themes are developed.

The essential concepts of thematic analysis are codes and themes. Codes are the smallest unit of analysis and they capture interesting properties of the data related to the research question, playing an essential role in building blocks for themes and patterns of meaning (Clarke & Braun, 2017). On the other hand, a theme is defined as a classification of interest detected by the analyst, building on codes (Clark et al., 2021). Clarke & Braun (2017) describe *"Themes provide a framework for organising and reporting the researcher's analytic observations"* (p. 297).

The remarkable property of thematic analysis lies in its flexibility. For instance, it is not limited by sample size, data collection method, or meaning generation including research questions (Clarke & Braun, 2017). Moreover, specific theoretical and technological knowledge is not required to conduct thematic knowledge (Braun & Clarke, 2006, as cited in Nowell et al, 2017). For those reasons, thematic analysis is known as a common and reachable method for early researchers.

Framework of Thematic Analysis

When it comes to the implementation of thematic analysis, Braun and Clarke suggest a six-phase process to follow for trustworthy thematic analysis: *familiarisation, initial coding, identifying themes, reviewing themes, defining themes and evidencing themes* (Clark et al., 2021, p. 538).

- **Familiarisation**: It requires transcribing interviews, writing and reading fieldnotes or looking into documents and other materials.
- **Initial coding**: Researchers can do coding to capture emerging characteristics of the data and proceed to do more theoretical coding of concepts. In the process of initial coding, the research questions needed to be considered so that interesting chunks of data, which are relevant to the research questions, are coded. Researchers can conduct initial coding manually with pens and highlighters or employ software such as Nvivo (Maguire & Delahunt, 2017).
- **Identifying themes:** Once the initial coding has been shaped the researchers need to identify themes. As mentioned above, *"a theme is a pattern that captures something significant or interesting about the data and/or research question"* (Maguire & Delahunt, 2017, p. 6). Therefore, researchers need to form a theme by its significance (Maguire & Delahunt, 2017). Furthermore, each category needs to be concrete and as short as possible as well as in a neutral way and common language (Popping, 2015).
- **Reviewing themes:** In this phase, researchers can merge them into high-order constructs and seek sub-themes by investigating identified themes.
- **Defining themes:** The narrative describing the characteristics of themes and sub-themes can be improved in this step. For instance, researchers can consider the relationship between each theme and sub-theme to figure out how they interact and relate to each other by asking what the theme is saying and what the theme is about.
- **Evidencing themes:** As a final step, the researcher is required to use evidence from the codes to link the themes to the wider literature and write-up.

Pros and Cons of Thematic Analysis

Thematic analysis has relatively simple procedures, so researchers are able to comprehend and learn the method quickly (Nowell et al., 2017). Owing to its simplicity, thematic analysis is useful for those unfamiliar with qualitative methods. *"Thematic analysis is also useful for summarising key features of a large data set, as it forces the researcher to take a well-structured approach to handling data, helping to produce a clear and organised final report"* (King, 2004, as cited in Nowell et al., 2017, p. 2).

Compared to other research methods such as grounded theory, thematic analysis lacks solid literature (Nowell et al., 2017). Nowell et al. (2018) point out that it can cause researchers to feel uncertain about how to implement a rigorous thematic analysis. Additionally, they argue that its flexibility affects the consistency and coherence of themes derived from research data.

3.3.2.2. Implementation of Thematic Analysis

The goal of the thematic analysis was to identify patterns in users' actual experiences and practices, based on data collected from two open-ended questions included in the survey, which are question 19: "*What positive experiences have you had while interacting with ChatGPT in your most recent chats?"* and question 20: "*What poor experiences have you had while interacting with ChatGPT in your most recent chats?"*

As we had two perspectives about users' experience, it involved a repetitive process to detect positive factors and negative factors. To prepare thematic analysis, one of the researchers created a framework with some guidelines in a Miro board in advance. Given the manageable size of the data set, advanced analytic software tools were not used. Instead, basic functions of Microsoft Excel and Miro board were used for effective analysis between two researchers.

As a first step of implementation, we tried to familiarise ourselves with the text from responses by going over the responses in the Microsoft Excel file and examining whether there was invalid information. Entering the coding phase, we documented the codes, which are an interesting data set related to RQ4: "*What factors contribute to positive and negative experiences among ChatGPT users?"* With an active discussion between two researchers, the data that will be analysed in the thematic analysis were coded in a new Microsoft Excel file.

Once the data was coded, it was copied and pasted into the Miro board to be classified by themes. First, codes were grouped on a high level and the chunk of groups were split into firm clusters with specific themes. Each cluster was examined to see if codes in the cluster were relevant to each other and fit the theme. In this process, some groups were merged into one cluster with a concrete theme. Finally, the themes were reviewed to see whether they represented the cluster precisely. This process was repeated two times to detect the themes for positive factors and negative factors, respectively.

3.3.3. Content Analysis

3.3.3.1. General Description of Content Analysis

Content analysis involves analysing documents and texts, whether they're printed or visual, with the benefit of being systematic, replicable, and highly adaptable (Clark et al., 2021). With high flexibility, this research method has been utilised in library and information science (LIS) studies with diverse research goals and objectives. Furthermore, this method is applied across qualitative, quantitative, and occasionally mixed research frameworks, employing a broad spectrum of analytical techniques to produce findings and contextualise them (White & Marsh, 2006).

Content analyses are broadly used to illustrate phenomena, examine their relationship and predict the nature of these relationships in social science (Riffe et al., 2019). Qualitative content analysis stands among various qualitative methods applied for analysing data and explaining its significance (Elo et al., 2014). It does not simply count words or extract objective content from texts to explore meanings, themes, and patterns that may be evident or concealed within a specific text. This approach enables researchers to comprehend social reality subjectively yet scientifically (Zhang & Wildermuth, 2005).

On the other hand, quantitative content analysis is mostly used in mass communication to quantify explicit textual elements. It follows a deductive approach, aiming to examine hypotheses or tackle inquiries derived from theories or previous studies. One of the key requirements is the selection of data through random sampling or other probabilistic methods to ensure the validity and reliability of the method (Zhang & Wildermuth, 2005).

Pros and Cons Content Analysis

The biggest strength of content analysis is its flexibility as it can be applied to a vast variety of text sources. With the help of computer software programs, it can also handle large amounts of data. Moreover, content analysis can be used as an unobtrusive research approach in a way to analyse naturally occurring data (Rose et al., 2014).

Lastly, the weakness of content analysis can be seen in its process of sampling and coding. The sampling process and document availability can present biases. Additionally, extracting the content from its context can cause issues. Taking a word or phrase out of its text structure can result in it losing its meaning. Similarly, it risks excluding particular text that might hold significance as that of the included (Rose et al., 2014).

Implementation of Content Analysis

In our mixed-method research, we applied content analysis both quantitatively and qualitatively, using a hybrid approach that incorporates sentiment analysis and text mining. First, to gather a better understanding of human views, sentiment-content analysis was used. Sentiment analysis facilitates the systematic exploration of affective states through natural language processing techniques, whereas content analysis provides a quantitative description of a given communication form. Therefore, this fusion of sentiment and content analysis enables a comprehensive examination of both emotional states and the structural aspects of communication (Niklander S. & Niklander G., 2017). In that regard, for more accurate and better quality results we decided to use the manual technique approach. It helped us better interpret the context in which the text was placed (Niklander S. & Niklander G., 2017).

Additionally, we implemented another hybrid approach that combines text mining with content analysis, focusing on revealing the facts behind the text. According to Fahmi et al. (2018), combining text mining with qualitative analysis can reduce both technical and ethical shortcomings. While text mining helps with the process of extracting information from unstructured text, content analysis provides in-depth qualitative insights. By combining these methods we made use of the fast, quantitative benefits of text mining and the detailed, qualitative strengths of content analysis.

To get an in-depth understanding of the main methods used in the hybrid approaches, the following sub-chapters explain the general description, framework, pros and cons and implementation of sentiment analysis and text mining.

3.3.3.2. Text Mining

General Description of Text Mining

Text mining is a process of extracting valuable information and insights from collections of documents by identifying and investigating patterns (Feldman & Sanger, 2007, as cited in Yu et al., 2011). It primarily handles unstructured data referred to as natural language text such as emails, social media, and the web (Hassani et al., 2020).

The principal objective of text mining is to capture and examine all possible meanings embedded in the text and it can be implemented both qualitatively and quantitatively (Benchimol et al., 2022). Tan (1999) illustrates that *"Text mining is a multidisciplinary field, involving information retrieval, text analysis, information extraction, clustering, categorisation, visualisation, database technology, machine learning, and data mining"* (p. 65).

A variety of mining techniques exist based on how to analyse text patterns and how to conduct mining processes (Talib et al., 2016) and there are *"rule-based, knowledge-based, statistical and machine-learning-based approaches"* (Jusoh & Alfawareh, 2021, p. 432). The most well-known methods are natural language processing (NLP) and information extraction (IE) techniques (Jusoh & Alfawareh, 2021).

Natural Language Processing aims for a better understanding of natural language using computers (Hotho et al., 2005). It includes natural language generation, natural language understanding, and linguistic analysis techniques and they are used to ensure grammatical correctness and fluency in the generated text (Hotho et al., 2005; Jusoh & Alfawareh, 2021).

On the other hand, the Information Extraction (IE) technique is to extract valuable information and insights from enormous amounts of text and the extracted data is stored in a database for further use (Talib et al., 2016). In other words, the primary objective of IE is extracting specific information from text documents (Hotho et al., 2005).

Framework of Text Mining

Text mining involves several sequential steps. Due to the unstructured nature of text data, a consistent and replicable method is required to assign meaningful quantitative measures to this data type (Benchimol et al., 2022). Benchimol et al. (2022) introduce the general steps of text mining: Data selection, Data cleaning, Information extraction and Analysis of extracted information.

As an initial step of text mining, the data from different sources such as websites, emails, social media, blogs and others mentioned in the general description are collected. Once the documents are retrieved, they are transformed into a suitable form to be applied by text mining techniques (Choudhary et al., 2009). This data-cleaning phase is called pre-processing, which is the main part of text mining and involves a lot of time and effort

(Bhattarai, 2022). In particular, the tokenisation process is required to obtain all words from the given text and during this process, *“text document is split into a stream of words by removing all punctuation marks and by replacing tabs and other non-text characters by single white spaces"* (Hotho et al., 2005, p. 25).

After the preprocessing phase, a variety of techniques are used (Bhattarai, 2022). For instance, information extraction, categorisation, clustering, summarisation and visualisation are used as common methods; Finally, the results are evaluated to get valuable and meaningful patterns (Bhattarai, 2022).

Pros and Cons of Text Mining

Text mining enables us to discover knowledge from previously unknown information by efficiently extracting the names of distinct entities and the relationships between them from a corpus of documents (Sukanya & Biruntha, 2012). Additionally, text mining is beneficial for addressing the challenge of managing vast amounts of unstructured information, as it facilitates the easy extraction of patterns (Sukanya & Biruntha, 2012).

However, while text mining excels at analysing complex unstructured data, it also faces significant challenges due to the complexity of natural language. The ambiguity of natural language means that a single word might have multiple meanings, which can introduce noise into the analysis and result in differing interpretations among researchers (Gaikwad et al., 2014).

Implementation of Text Mining

Text mining was used to capture the meanings of three specific questions asking respondents’ experiences and thoughts in the questionnaire. Apart from it, we used text mining to extract significant keywords from sentiment analysis.

The three questions are:

- Question 10: *“In what contexts do you use Generative AI? Please provide further details on the usage of Generative AI”*
- Question 25: “*Please specify the reasons for your response*”, which is the follow-up question to question 24: “*Research indicates that ChatGPT has 180.5 million users and 1.6 billion visits (January 2024) and hundreds of millions of prompts on ChatGPT costs around 1 GWh each day, which is equivalent to the energy consumption of 33,000 U.S. households. Would this influence your behaviour using ChatGPT?”*
- Question 27: *“Please specify the reasons for your response*”, which is a follow-up question to question 26: “*In your opinion, who is primarily in charge of limiting the environmental impact of ChatGPT?”*

Text mining analysis was implemented using R. Within the R environment, we used the *“tm”* package, which serves as a comprehensive toolkit designed for text-mining tasks (Feinerer, 2013). This package provides essential functionalities for text mining like preprocessing, tokenisation, text transformation, and information extraction (Feinerer, 2013). The most fundamental concept of the “*tm”* package is “*corpus*” referring to a collection of textual data (Feinerer, 2013). In our study, a corpus was represented by the responses from these three questions and the data set used in sentiment analysis.

To apply the *“tm”* package to the data set, relevant responses for these questions were selected after importing an Excel file to R. With the selected columns the preprocessing phase was conducted, involving converting data to lowercase and removing punctuation, numbers and common English stop words like “*the*”, “*and*” and “*is*”. Once the preprocessing phase was done, a corpus was transformed into a document-term matrix (DTM) in which rows represent words with *TermDocumentMatrix*. This allowed us to analyse the frequency of each word from the responses to each selected question. The procedure, counting the frequent words of responses, was conducted for each question and the sentiment analysis data set separately.

Lastly, visualisation was implemented to illustrate the results from the analysis clearly and intuitively. The results are shown in Chapter 4.1.2. Text Mining Analysis.

3.3.3.3. Sentiment Analysis

General Description of Sentiment Analysis

In the present day, most people express their opinions, feelings, and experiences through the internet and social media. Companies have realised that analysing this data is useful for understanding whether the user's opinions are positive, negative, or neutral. As a result, sentiment analysis (SA) emerged due to the vast information exchange on the Internet (Aqlan et al., 2019).

Sentiment analysis, also known as opinion mining (OM), “*is the field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organisations, individuals, issues, events, topics, and their attributes*” (Liu, 2012, p. 7). It can represent a large spect of problems. It used to have many names with slightly different tasks, for example, opinion extraction, sentiment mining, subjective analysis, affect analysis, review mining, etc. However, nowadays, it all falls under the term sentiment analysis or opinion mining (Liu, 2012).

The main objective of sentiment analysis is to properly identify whether the given text is subjective or objective or whether it is positive, negative or neutral. Often, opinion mining and sentiment analysis are used mutually together. OM would deal with extracting and analysing the opinions of participants while SA would deal with the specific sentiments expressed in the text and analyse them (Bhonde & Prasad, 2015).

Although the research for sentiment analysis mainly began in the early 2000s, there have been papers that share similarities such as metaphors, sentiment adjectives, subjectivity and viewpoint to SA (Hatzivassiloglou and McKeown, 1997; Hearst, 1992; Wiebe et al., 1999). These books serve as an introduction to the analysis as a whole (Liu, 2012).

Framework of Sentiment Analysis

With the advancements in technology, there has been a rise in user-generated opinions, such as reviews, images, comments, etc. Alongside came a need for researchers, companies, politicians, service providers, etc., to analyse them in order to implement better decision choices (D'Andrea et al., 2015). For that to be accomplished the best approach is sentiment analysis.

The sentiment analysis approach is a complex process that requires five steps to complete the analysis. These steps are: (D'Andrea et al., 2015)

- **Data Collection:** Data collection is the first step in sentiment analysis in which the collected data is gathered from user-generated content contained in social networks, blogs, and forms. This gathered data is disorganised because it has a variety of expressions as well as a variety of different vocabularies and contexts of writing.
- **Text Preparation:** Before initiating the analysis the extracted data needs to be cleaned. That can be achieved by identifying and eliminating all the non-textual content that is irrelevant to the analysis.
- **Sentiment Detection:** After cleaning up the data the extracted sentences of different opinions and or reviews are examined. The selected sentences can then further be assigned as either subjective expressions (opinions, beliefs and views) or objective communication (facts, factual information). The objective communication sentences are then discarded.
- **Sentiment Classification:** The remaining subjective expressions are then classified as positive, neutral, or negative.
- **presentation of output:** Lastly, the presentation of output provides the main objective of the whole analysis which is to transfer the unstructured text into meaningful information. Once the analysis is completed the text results are typically displayed on graphs, such as bar charts, line charts or pie charts.

Moreover, sentimental classification can be divided by techniques. One of the techniques is machine learning: algorithms or techniques best used for brief classification and sentiment predictions (Umar et al., 2022). Lexicon-based technique is a group of words or phrases that convey feelings (Han et al., 2018). These sentiment techniques can be divided into positive, neutral and negative categories (Umar et al., 2022; Han et al., 2018). Lastly, hybrid-based techniques are a combination of lexicon-based and machine-learning approaches (Ahmad et al., 2017). The sentimental classification we choose to work with is a Lexicon-based technique and sentence-level approach. It classifies the sentiment expressed in each sentence, determining whether it is subjective and categorising it into positive, negative or neutral opinions (Umar et al., 2022).

Pros and Cons of Sentiment Analysis

The advantages SA provides are for the most part to determine the general attitude towards brands, products or services. Companies can also use the analysis to monitor reputation metrics like sentiment scores and customer satisfaction scores, etc. Moreover, it can assist in understanding the most relevant and impactful feedback from participants (Determ, 2023).

The main downfall the analysis method faces is that, unlike factual information, opinions and sentiments have an important characteristic, namely, they are subjective. For that reason, it is emphasised that opinions need to be gathered from many individuals rather than have an opinion from a singular person (Liu, 2012). Some other problems with the method are biased opinions, fake reviews and miscommunication (Dilmegani, 2024).

Implementation of Sentiment Analysis

In order to get a deeper understanding of our questionnaire results (questions 24 and 25) we conducted the sentiment analysis. The analysis was conducted regarding whether the presented fact about ChatGPT energy usage would influence the participant's behaviour. Our goal with using sentiment analysis was to acquire a better understanding of the user's emotions.

We chose to work with the lexicon-based technique which allowed us to extract full sentences from our open-ended questions from the questionnaire. The reason was that conducting a normal lexicon-based technique by extracting just words from the context would not be sufficient enough for our project and aimed results. Extracting full sentences enabled us to see the full context of the sentences and helped us with sentiment classification. However, we also decided to conduct sentiment classification in a different way than what was considered normal, meaning we had our own criteria. As such we chose to conduct it manually. Some of the examples of why we worked on it manually are that we flagged sentences that were "*too useful*" or "*good*" as negatives and "*will change*" or "*was not aware*" as positive.

Furthermore, looking back at the questionnaire although the participants could have already selected the option Yes and No in the previous question 24 we wanted to see if some participants might have been conflicted in their decision. With this information, we could get a better grasp of how the participants feel on an emotional level regarding the question. As such we decided to conduct the sentiment analysis to see the positive, neutral and negative emotions the participant presented.

The first step we did was to copy all the results from question 25 and paste them into an excel sheet. Afterwards, we needed to clean up the data as much as possible for a proper sentiment analysis to be conducted. To start the clean up we removed all the non-textual content as well as all objective comments such as facts or factual information.

After making sure that both of us were satisfied with the clean up we started the sentiment analysis. We did so by determining which sentences were positive, neutral and negative. As the method is subjective we made sure to conduct the analysis separately to avoid biased answers. Once the analysis was completed and the sentences were labelled, we combined the answers to discuss the results.

3.3.3.3.1. Cohen’s Kappa

General Description of Cohen’s Kappa

The Kappa coefficient, also known as Cohen's Kappa, is widely recognised and used in assessing categorical agreements between two raters or two methods (Taube, 2010).
Cohen’s kappa was first introduced by Jacob Cohen in 1690, as a measurement for agreements (Cohen, 1960). It calculates the interior agreement factor of qualitative items in a categorical form. Additionally, it is being used as a statistic interrater or interrater for reliability tests (Taube, 2010; University of York, n.d.).

When calculating Cohen’s Kappa is symbolised by the Greek letter, κ. Similarly to correlation coefficients (a numerical measure of some type of linear correlation), it is measured on a scale that varies from -1 to 1 (Kolesnyk & Khairova, 2022). With that measurement, Cohen puts 0 as a representative of the amount of agreement that can be expected from random chance, and 1 as a representative of a perfect agreement between the raters. As for below 0 values, although they are possible, Cohen himself notes that they are highly unlikely to happen in practice (McHugh, 2012).

Another thing to take note of is that Kappa is a standardised value and as such is interpreted the same across multiple studies (McHugh, 2012).

According to Cohen, the Kappa results should be interpreted as follows: “*values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81– 1.00 as almost perfect agreement*” (McHugh, 2012, p. 279).

Pros and Cons of Cohen’s Kappa

The benefit of using Kappa is that it accounts for the agreements that would otherwise occur by chance. It provides researchers with a more accurate measure of agreement between observers or raters in contrast to simple percent agreement (Li M. et al., 2023). Additionally, Cohen’s Kappa can be used in categorical variables, making it a versatile measure for a wide range of research fields (Li M. et al., 2023).

When applying Cohen Kappa to investigate agreements some problems might arise. The most notable issue that can come from this method is biased assessment and biased answers from participants (Taube, 2010). Another issue previously mentioned in the description of the method is that the values are not equally reachable. What is meant by that is that a higher value is easier to reach than a smaller value in this case values below 0 (McHugh, 2012).

Implementation of Cohen’s Kappa

In order to further solidify our sentiment analysis more specifically the sentiment classification, we decide to conduct the Cohen's kappa. By using Cohen’s kappa we can better understand the level of agreement between the two of us as raters, providing a more accurate assessment of our sentiment analysis's reliability. Furthermore, we choose to work with the programming language R.

The first step we did was to create a new project in the R studio and import the excel sheet that had the sentiment analysis. In order to import the excel file we had to run the "*library(readxl)*" which helps R read and get the data from the excel sheet into R. Next, we also ran "*library(tidyverse)*" and "*library(psych)*". "*Tidyverse*" assists with data import, tidying, manipulation, and data visualisation while "*psych*" provides various functions for psychological research and data analysis (Wickham, 2023; Revelle, 2024).

Once all the libraries were loaded we then ran a code to tidy the excel sheet and remove the extra columns we did not need. Lastly, we used "*cbind*" to combine our data frames in columns and ran the code for Cohen's kappa.

3.4. Validity and Reliability of Methods

In our project, we aimed to ensure data validity and reliability. Reliability concerns the stability of findings, while validity reflects their truthfulness. By implementing these aspects we increase transparency and reduce bias (Haradhan, 2017). Credibility depends on various research features, such as the initial question, analysis methods, conclusions, and data collection details (Roberts & Priest, 2006). In this section, we will explore our quantitative and qualitative analysis along with a questionnaire and contextual inquiry validity and reliability.

Quantitative Analysis

Although the ideas of validity and reliability are similar, they still have different qualities of the measuring instrument (Sürücü & Maslakçi, 2020). Validity refers to the capacity to assess how effectively the collected data covers the actual area of investigation. Additionally, it essentially implies measuring precisely what is intended to be measured (Taherdoost, 2016). There are many different ways for validity to be conducted such as Face Validity, Content Validity, Construct Validity, etc. (Sürücü & Maslakçi, 2020).

Additionally, reliability refers to the consistency and stability of the analysis measurements over time and its repeatability. Checking for reliability is crucial because it ensures consistency while being completed by different participants (Taherdoost, 2016). As with validity, reliability has many different ways to be conducted. The most frequently used methods are test-retest reliability, alternative forms, and internal consistency tests (Sürücü & Maslakçi, 2020).

Qualitative Analysis

The validity and reliability of qualitative research are fundamental parts of good quality research. When conducted properly it can be a key factor in distinguishing the difference between good and bad research. They assure readers that the analysis findings are trustworthy (Chetty & Thakur, 2020).

Validity in qualitative analysis is connected with the accuracy and trustworthiness of research findings. A good validity will contain a valid instrument or measure and a description of what it is supposed to measure (Brink, 1993). Countless terms have been suggested as criteria but the most used ones are truth value, credibility, trustworthiness, authenticity and goodness (Whittemore et al., 2001).

Furthermore, reliability in qualitative analysis helps us understand the purpose and situation that could otherwise be confusing (Golafshani, 2003). Some of the ways to test the reliability of analysis are to conduct techniques like the use of comprehensive data, constant testing and comparison of data and the use of tables to record data (Chetty & Thakur, 2020).

Questionnaire

To ensure our validity with the questionnaire we applied three types of validity: face validity, content validity and construct validity (Taherdoost, 2016).

- **Face validity:** The questionnaire was carefully thought out to ensure its face validity, the degree to which it appeared relevant and understandable to the participants. The introductory section explicitly stated the aim of the questionnaire, focusing on investigating user behaviour with Generative AI and ChatGPT, as well as users' awareness of the CO_2 footprint of ChatGPT. This transparent communication of the questionnaire's purpose helps to establish its face validity by aligning participants' expectations with the research objectives (Taherdoost, 2016).
- **Content validity:** Content validity was ensured by having each section of the questionnaire, carefully designed to gather only data relevant to the research goal. For example, the demographic section included questions not only about participants' age, gender identity, and employment status but also about their awareness of climate change. These questions were carefully chosen to view various user behaviour and awareness based on demographic factors. Similarly, the section focusing on the utilisation of ChatGPT included questions specifically tailored to user behaviour and experiences with the tool, ensuring the content validity of the questionnaire (Taherdoost, 2016).
- **Construct validity:** Construct validity, assesses the extent to which the questionnaire measures the underlying theoretical constructs. Questions were designed to measure constructs such as user behaviour with Generative AI and ChatGPT and awareness of CO_2 footprint and its environmental impact. By utilising both, close-ended and open-ended questions, the questionnaire aimed to capture an understanding of these constructs, thus enhancing its construct validity (Taherdoost, 2016).

Furthermore, for reliability we applied three types of reliability: consistency, clear instructions, and pilot testing.

- **Consistency in questionnaire structure:** Each section follows a similar format, with a clear introduction, followed by, a series of questions designed to gather specific information related to the research objectives. This consistency in structure reduces the likelihood of confusion among participants and ensures that the questionnaire is reliable across different sections.

- **Clear instructions:** Clear instructions and guidelines were provided at the beginning of each section to ensure participants understood how to respond to the questions accurately. These instructions helped prevent misunderstandings and inconsistencies in participant responses, thereby enhancing the reliability of the data collected.
- **Pilot testing:** Pilot testing was conducted to assess the clarity of the questionnaire before its release. Feedback from participants was carefully considered and used to refine the questionnaire further. By incorporating feedback from pilot testing, the questionnaire was refined to enhance its reliability in obtaining consistent and accurate responses.

Contextual Inquiry

Contextual inquiry is a qualitative research method used for understanding user's behaviours, preferences, and interactions within their natural environment. In this section, we assess the validity and reliability of the CI conducted to explore user behaviour and habits in utilising ChatGPT.

Validity of contextual inquiry
The scenarios used in CI were thoughtfully thought out to be relevant to the participant situations and ChatGPT usage. The tasks varied from simple to more complex to evaluate the participants' user behaviour and decision-making process. Moreover, the semi-structured interviews were designed to investigate participants' views, preferences and interactions relevant to the previously conducted tasks. By aligning the interview questions with the tasks conducted we aimed to create a clear connection between the user's behaviour and thoughts throughout the CI session. Lastly, to ensure the validity of the data gathered before starting the CI session the participants were given a clear introduction on what CI is and how it was to be conducted. As well as asked to give consent to the recording and using the data acquired.

Reliability of contextual inquiry
To ensure the reliability of CI, both sessions were conducted with both of us attending, with one having to lead the session and the other documenting the observations. This allowed us a better view of participants' behaviour as well as reduced the chances of biases in data analysis. Furthermore, for consistency purposes, we made sure to use the same tasks and semi-structured interview questions. The reason is for a more accurate and reliable result from both participants.

3.5. Ethical Considerations and Limitations of Methods

3.5.1. Ethical Considerations

Ethical considerations are essential in the development and implementation of any method within academic research. Questionnaires serve as a primary method of data collection in our research, as well as many others such as social sciences, psychology, and market research. Ensuring ethical principles are upheld throughout the process is of the utmost importance (Denzin & Lincoln, 2011).

Our second method, contextual inquiry, a qualitative research method, is used to better understand users' behaviours, needs, and challenges in their natural environment. Which on its own can present unique ethical considerations. As researchers immerse themselves in participants' lives to gather in-depth insights, it's crucial to navigate this process ethically and responsibly.

This section delves into the key ethical considerations that we considered while utilising both of the data collection methods.

- **Informed consent:** According to Denzin & Lincoln (2011) the foundation of ethical research is '*informed consent*'. The term itself consists of the most important components, that is, '*informed*' and '*consent*'. The most important principle is that the participants are fully informed about what is asked of them, how the data will be handled and lastly what (if any) consequences there could be (Fleming & Zegwaard, 2018).
- **Confidentiality and anonymity:** For a long time confidentiality and anonymity have been a part of research papers consisting of data collection. Based on that It is essential that researchers make an effort to protect the anonymity of the study participants and the privacy of data (Miles and Huberman, 1994). At the start of a data collection session, these two points must be communicated to the participants in written or verbal form. Lastly, no party other than the researchers and, if relevant, the research team members should have access to the gathered data (Mirza, 2023).
- **Voluntary participation and right to withdraw:** Participation should always be voluntary in all research, and no coercion or deception should be involved. The same can be applied to participants' rights to withdraw from the study at any time without repercussion. It is the researcher's responsibility to emphasise that participation is entirely voluntary and that participants are free to withdraw their consent or stop their involvement in the study without penalty (University of Guyana, 2022)
- **Transparency and integrity:** Maintaining transparency and integrity throughout the study is essential for building trust with the participants. Researchers should clearly communicate the purpose, scope, and limitations of the study, as well as the intended use of the data collected. Misleading practices, such as withholding information or misrepresenting the study's purpose, undermine participants' trust and compromise the validity of the research findings (Turilli & Floridi, 2009).

Conclusion

Throughout our research data collection process, we ensured that our participants' rights and well-being were being respected by applying these ethical considerations principles, such as informed consent, confidentiality and anonymity, voluntary participation and right to withdraw, and transparency and Integrity. Utilising these ethical standards has not only helped us protect participants' rights but also enhanced the credibility and trustworthiness of the research outcome.

3.5.2. Limitations of Methods

Both questionnaires and contextual inquiry are popular methods used in various research fields to gather users' insights and understand their behaviour. While both approaches provide specific advantages, they also have limitations that can affect the outcome of research. This section discusses the limitations associated with the questionnaire and contextual inquiry methods that we have encountered during our data-gathering process.

Questionnaire

There were a few limitations and concerns that we had throughout the questionnaire process. The main limitation came from the questionnaire using many close-ended questions, which made it easier to analyse but missed the in-depth information that could be gathered. Another limitation that had occurred was that the participants may have misunderstood a question. We tried our best to write questions in an easy-to-understand way as well as providing terminology in specific sections for further explanation of certain terms. Nevertheless, despite our efforts, a few misunderstandings still came to be at the end.

Contextual inquiry

The biggest limitation we have experienced with CI is that it is time-consuming. It took a lot of time from recruiting the participants to setting up all the recording equipment, and lastly, doing the method in itself. Another limitation that we experienced was biases and subjectivity that we tried our best to avoid. We did so by trying not to lead the users by providing them with the answers we wanted to hear or better said not providing our subjective thoughts, on the topic at hand. We still decided to put this into a limitation section as we may have unconsciously provided our opinions on certain topics and as such influenced the participant. The last limitation we had was recruiting participants. What we mean by this is that since the method took a long time to prepare and finalise we did not have a lot of participants. Moreover, we had our deadline for submitting the academic report as well to take into consideration.

4. Findings

This chapter presents the findings gathered from the research methods and techniques outlined in the previous chapters. On a high level, this chapter is divided into Findings from the Questionnaire and Findings from Contextual Inquiry. The descriptions of the data collected and the outcomes of the analyses performed are provided for each subchapter, offering a clear and objective view of the gathered data and findings. Each result is discussed and is connected to at least one research question. By presenting the results in a structured manner, this chapter provides the foundation for the upcoming discussion where the conclusions of these findings will be further explored and interpreted in relation to existing literature and research questions.

4.1. Findings from the Questionnaire

Our questionnaire was completed by 74 unique individuals, with the average time duration to complete being 12 minutes. The data collected from the five distinct sections in the questionnaire are: *Demographic, Utilisation of Generative AI, Utilisation of ChatGPT, ChatGPT's latest experience and Awareness of the Environmental Impact of ChatGPT* contributes to a better understanding of the interaction between Generative AI utilisation, environmental consciousness and potential ideas for the improvement of ChatGPT.

Within this section, we will be exploring the data set collected by our questionnaire and the findings of both quantitative and qualitative analysis of the data set. The collected data set and analysis are shown in an order of quantitative analysis, text mining, sentiment analysis and thematic analysis.

4.1.1. Quantitative Analysis

Demographic Results

We obtained meaningful demographic results. The survey revealed that a significant portion of respondents fell within the age group of 25-34, consisting of 50% of the sample population as seen in Figure C1 of Appendix C. In terms of gender, 50% of female and 47% of male respondents participated (See Figure C2 of Appendix C). Furthermore, 50% of respondents identified as full-time employees, followed by students at 26% (See Figure C3 of Appendix C).

Notably, the majority stated that they care more than average about the impact of climate change and they care about the impact of climate change in their daily activities on a moderate level. The details of these findings are seen in the last subchapter, Level of User's Awareness of the Environmental Impact of ChatGPT. These demographic insights provide valuable context for understanding our target audience and their attitude toward environmental awareness.

Usage patterns and preference of ChatGPT

Through data collected from the third section, Utilisation of ChatGPT, we found that the largest number of respondents (34%) reported that they use ChatGPT a few times a week. To look into the usage patterns deeply, we compared by age group. Figure 1 shows while ChatGPT is widely used across various age groups, the intensity of usage differs. It was observed that younger respondents (18-44) are more likely to use ChatGPT daily. For instance, while the daily usage patterns were seen in the age groups ranging from 18 to 44 they were not observed in the 45-64 age group.

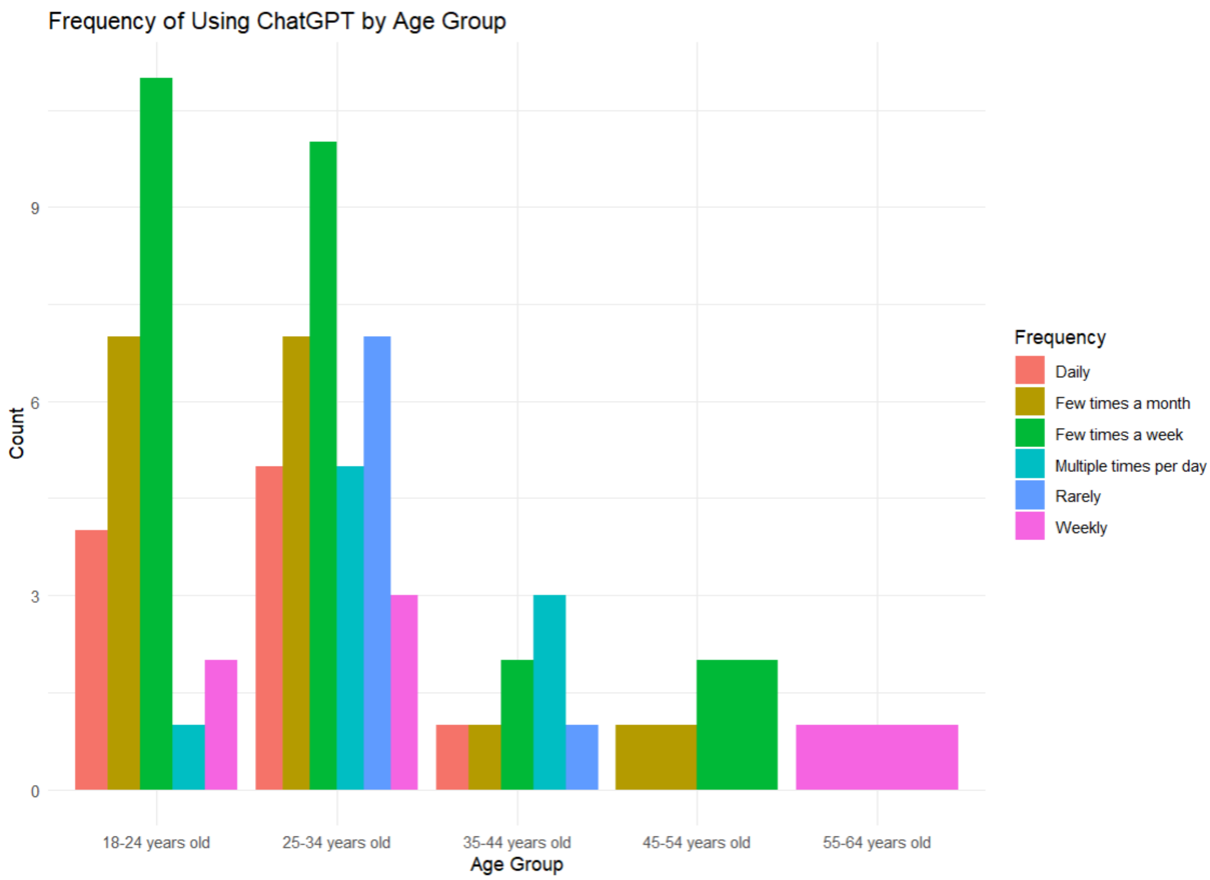

Figure 1. Frequency of Using ChatGPT by Age Group

Regarding the purpose of ChatGPT, 57% answered they use ChatGPT for both work and personal reasons. Additionally, ChatGPT's latest experience, the fourth section in the questionnaire, highlighted its frequent use in work-related activities such as email formulation, research, and content summarisation.

Furthermore, Table C1 in Appendix C shows there is no remarkable difference in the purpose of using ChatGPT across different age groups; the option, "*All of the above*" received the highest numbers across all age groups. A similar pattern is observed in occupation (See Table C2 of Appendix C). However, interestingly, among respondents who identified as "*Part-Time Employee*", the number of responses indicating "*Personal*" was slightly higher than those for "*Work-related*" by 2 points. Conversely, respondents who identified as "*Student*" reported a higher number of "*Work-related*" compared to "*Personal*" by 4 points. It suggests that many students use ChatGPT for their school-related tasks.

Overall, the vast majority of respondents answered they use ChatGPT for both personal and work-related purposes. It indicates ChatGPT's functionalities are broadly applicable regardless of the user's age and occupation.

In terms of user preference for ChatGPT, one of the significant questions is question 14: *"Please rate your usage of ChatGPT based on the following statements, on a scale from 0 to 5. "0" indicates no usage, while "5" indicates frequent usage"* formed in the 6-point Likert scale with various statements. The 6-point scale means: *0 - no usage, 1 - rarely used, 2 - infrequently used, 3 - occasionally used, 4 - often used, and 5 -frequently used.* We obtained responses indicating the level of usage for each statement: *Translate text, Summarise content, Get help with writing, Get assistance with coding, Generate images, Content creation and Ask simple questions.*

As shown in Table C3 of Appendix C, the results from descriptive statistics show that "*Get help with writing*" has the highest median score at 4, tied with "*Ask simple questions*" and "*Summarise content*". Additionally, it has the highest mean score at 3.76. In contrast, "*Generate image*" has both the lowest median score of 0 and the lowest mean score of 1.04. The statement, "*Creation*" follows closely with a mean score of 1.1.

A percentage-stacked bar chart was used to visualise the percentage distribution of each rating for the statements intuitively. As shown in Figure 2, the most common use of ChatGPT was *"Get help with writing"*, which accounted for 45% of the highest rating (frequently used), 18% of rating 4 (often used), 22% of rating 3 (occasionally used), 7% of rating 2 (infrequently used), 4% of rating 1 (rarely used) and 5% of rating 0 (no usage). This was followed by "*Asking simple questions*" and "*Summarise content*", which received 27% and 26% of frequently used, respectively. The combined percentage of the highest (frequently used) and second-highest (often used) ratings exceeds 50% for these three statements. It indicates that ChatGPT is primarily used when users want to get help with writing, ask simple questions and summarise content.

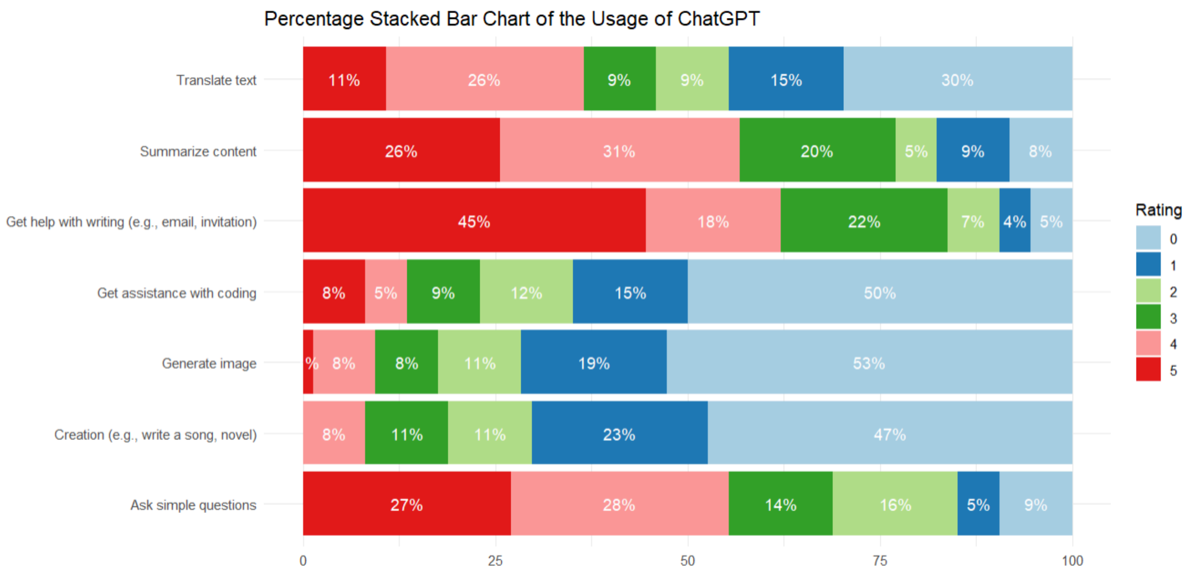

Figure 2. ChatGPT Usage Preference

In contrast, the least use of ChatGPT was "*Generate image*", which took up 53% of the lowest rating (No usage), followed by "*Get assistance with coding*" and "*Creation*", which made up 50% and 47% of the lowest ratings, respectively. The combined percentage of the lowest rating (No usage) and the second-lowest rating (Rarely used) for "*Generate image*" and "*Creation*" exceeded 70%. This means that respondents rarely use ChatGPT for the tasks relevant to generating images and creation. Regarding the statement "*Translate text*", it was observed that the ratio of ratings of 0 (No usage) and 1 (Rarely used) was only 8 percent higher than the ratio of ratings of 4 (Often used) and 5 (Frequently used)."

User Behaviour of Generating Prompts

When looking back at our questionnaire results, we found some connections to the related works that we wanted to explore further. For that reason, we decided to further investigate the questionnaire for question 15: *"On average, how many attempts do you rephrase your prompts until you achieve the desired outcome when using ChatGPT?"*; question 18: "*How many times did you create prompts in the latest chat?"*

For question 15, We decided to create a pie chart with the percentages in the middle of the correlating colour or field. This visualisation helps to easily comprehend the data distribution regarding how often participants rephrased prompts per session. As seen in Figure 3, 81.1% of participants chose that they rephrased prompts around 1-5 times per session. The second highest was 6-10 prompts per session with 14.9% of participants. The last two options have a small difference in the percentage of participants' choice, we have 2.7% of participants choosing 11-15 prompts per session and 1.4% choosing over 20 per session. On average, users create prompts approximately five times per interaction chat.

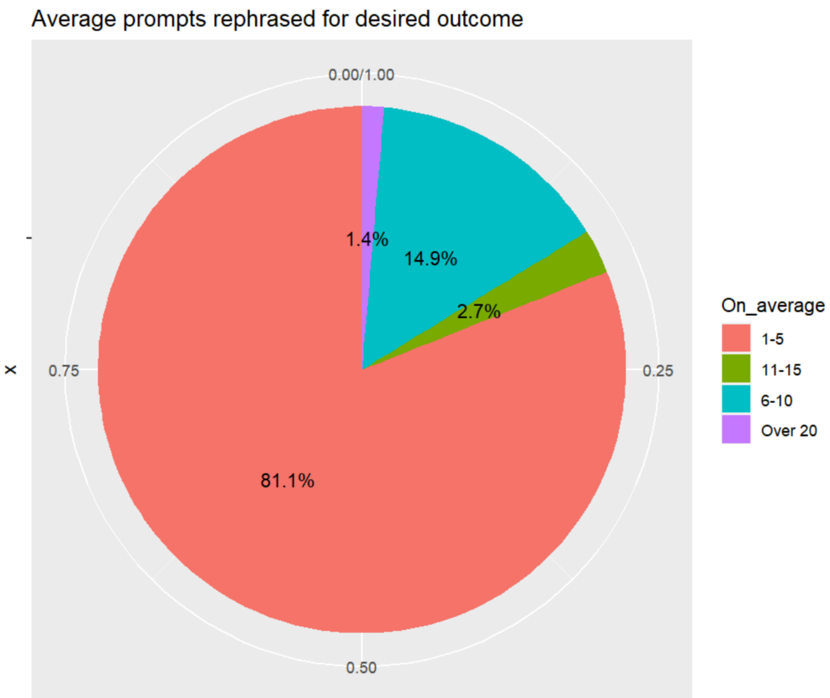

Figure 3. Distribution of Average Rephrase Prompts

What we can gather from this analysis is that most people rephrase prompts 1-5 times per session making it clear that most users do not need extensive rephrasing. Additionally, a smaller but notable group, of 14.9%, rephrases prompts 6-10 times per session. This could be seen as users who are seeking more precise answers or dealing with more complex queries. Only a small percentage of participants engage in extensive rephrasing, with 2.7% rephrasing 11-15 times and 1.4% rephrasing over 20 times per session. This suggests that prolonged prompt rephrasing for refining answers is rare.

When it comes to question 18: *“How many times did you create prompts in the latest chat?”*, we decided to visualise the results in a histogram graph. With the histogram graph, we can see the difference in participants' creation of a number of prompts from the highest to the lowest created. From Figure 4 we can read that similarly to Figure 3 the participants mostly create around 1 to 5 prompts even in their latest chats. Furthermore, there is a bit higher number in creation from 6 to 10 which once more provides similarities to the previous findings. Lastly, the highest numbers were mentioned once between 13 to 23 prompts created in the last session.

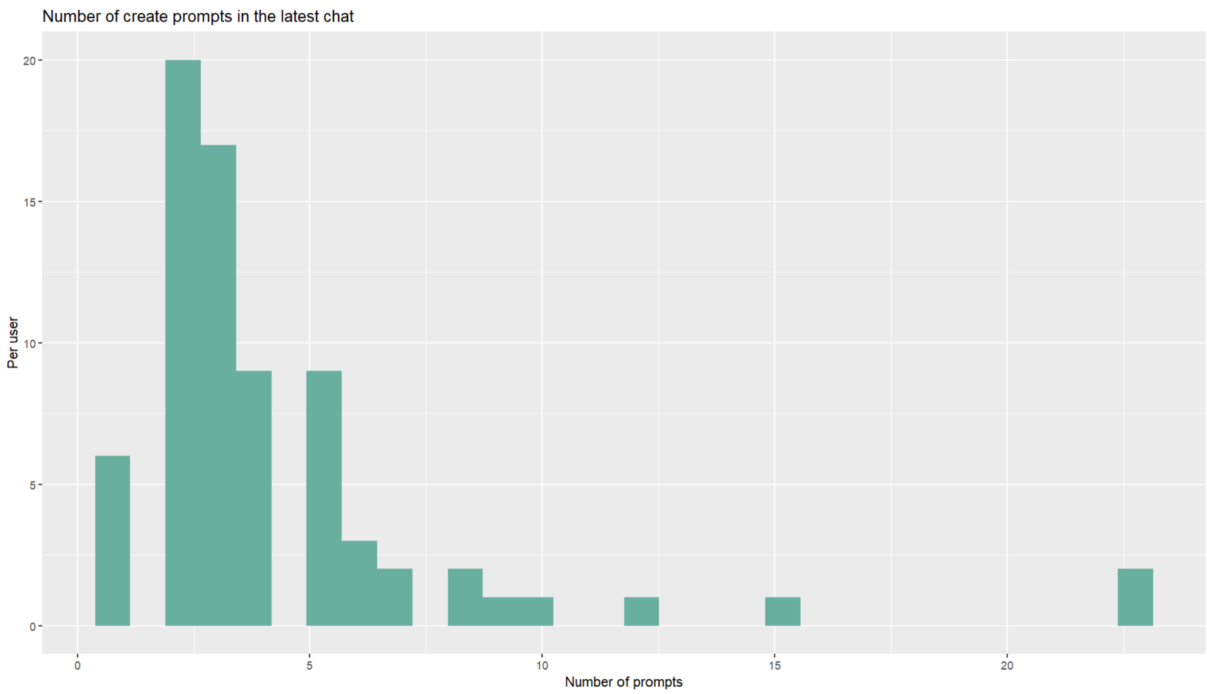

Figure 4. Distribution of Creation Prompts in the Latest Chat

Moreover, we coded the summary of the number of prompts created in the latest chat. As seen in Table C4 of Appendix C, we looked at the min, max and median number of prompts per participants' latest chats. From those findings, we can read that the minimum number of prompts made in the latest chat was 1, the maximum was 23 and the median was found to be 3. The result ended up correlating to Figure 4 further solidifying our results.

Level of User’s Awareness of the environmental impact of ChatGPT

To dive deeper into users' awareness of the environmental impact of ChatGPT, a key focus of our project, we analysed the responses from the questions designed to investigate it. Additionally, we examined users' awareness of climate change to compare the results.

The data sets used in this part are:

- In the Demographic section, question 5: “*How much do you care about the impact of climate change?”* and question 6: “*Have you ever considered the impact of climate change in your daily activities?”*
- In the Awareness of the Environmental Impact of ChatGPT section, question 22: “*How well do you believe you understand the CO2 footprint of ChatGPT”* and

question 23: "*Which type of energy consumption do you believe produces the most CO_2 emissions? Please rank them from 1 (highest) to 5 (lowest). Adjust the order by dragging the text up or down"* with 5 statements.

First, the results from the analysis of questions 5 and 6 in section 1 were found. Figure 5 shows a level of consideration for climate change. According to the result of question 5, shown in the second bar, 11% of respondents chose extreme care (rating 10) and 9% expressed great care (rating 9). On the other hand, for question 6, seen in the first bar, 8% rated it a 10 (always) and 20% rated it a 9. Overall, the respondents have a high level of awareness and concern for climate change with a median score of 7 for both questions (See Table C5 of Appendix C).

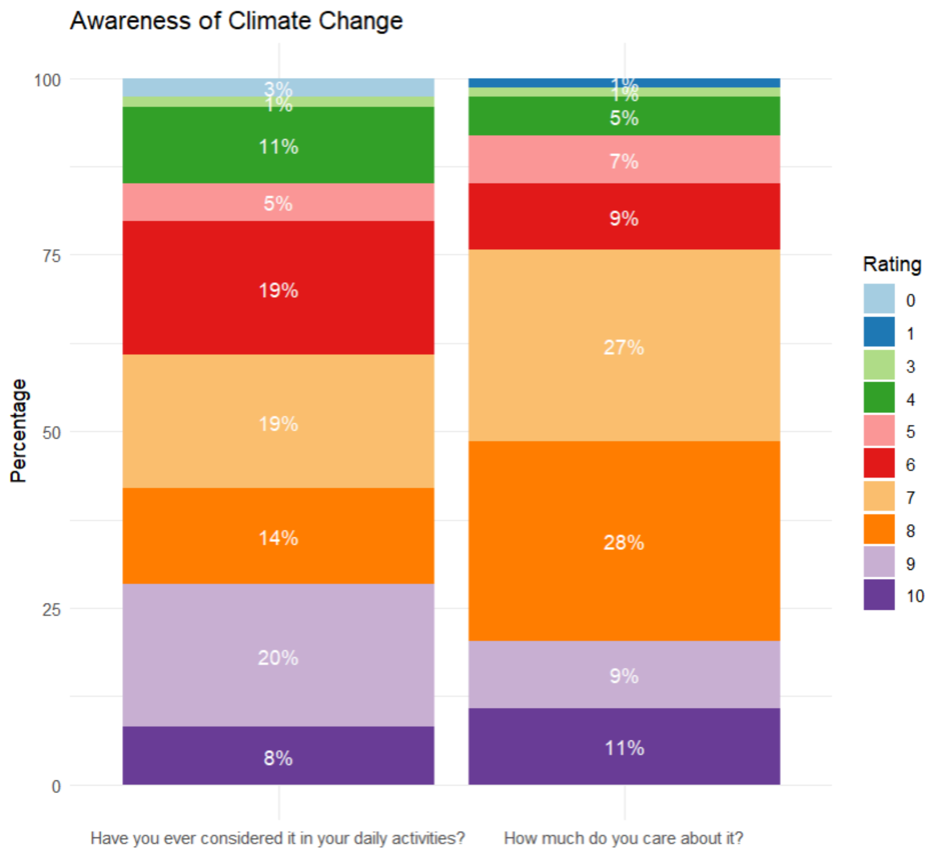

Figure 5. Awareness of Climate Change

Despite many participants raising concerns about climate change, only a small fraction stated that they understand the CO_2 footprint of ChatGPT. The respondents have little knowledge of the environmental impact, according to the results from question 22: *"How well do you believe you understand the CO_2 footprint of ChatGPT?"*

For the question, they were required to indicate between 0 and 10 on the Likert scale. On the Likert scale 0 indicated "*no knowledge*", while 10 indicated "*full understanding*" of the environmental impact of ChatGPT. Looking closely at Figure 6, we can see that 17 people, approximately 23% of respondents chose 0 (no knowledge). It was observed the median score is 3, implying that 50% of the respondents chose 3 or below (See Table C6 of Appendix C). The minimum number was 0 and the maximum was 9. These results correlated with the findings in the graph. Only one participant chose a rating of 9 and none chose a rating of 10 (extremely well). Moreover, a few participants chose a higher understanding between 7 and 9. It implies the lack of knowledge in this area of research and a need for more knowledge and awareness in this field.

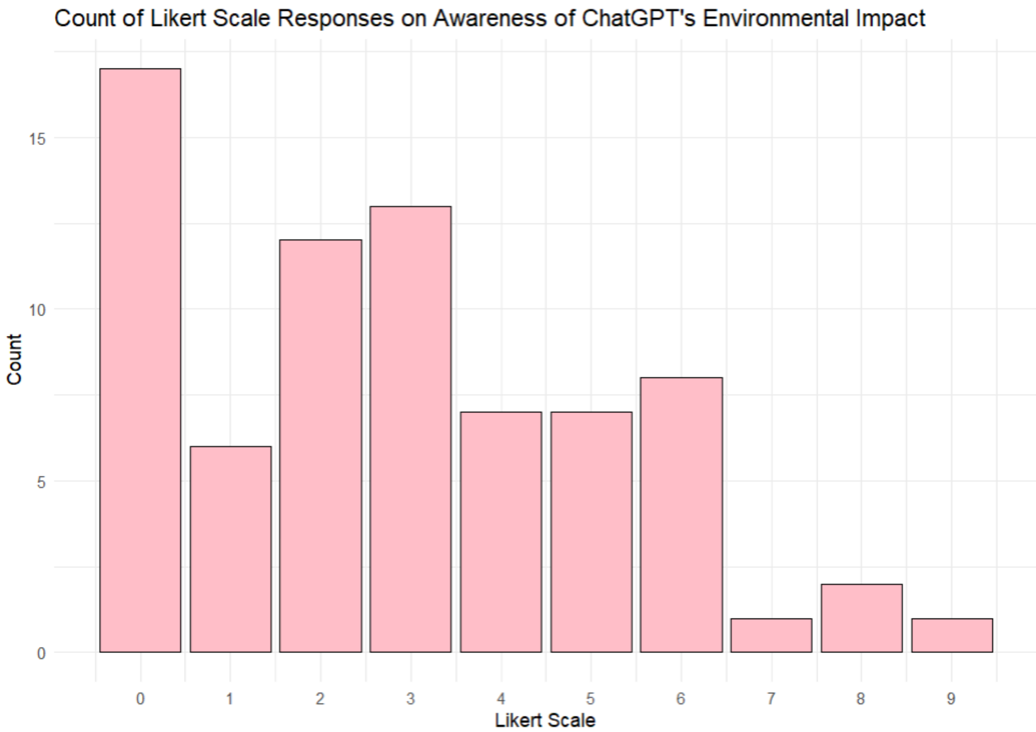

Figure 6. Awareness of ChatGPT's Environmental Impact

On the other hand, participants were asked question 23, a ranking question where they were required to rank types of energy consumption based on the amount of CO2 produced with five options. These options are ordered by the most CO2 produced energy type: *Watching videos online (2 hours per day), Using ChatGPT (20 prompts per day), Using a light bulb (3 hours per day), Google searching (50 times a day) and Email/Messaging (30 times per day)*.

However, interestingly, despite *"Using a light bulb (3 hours per day)"* being the third most CO2-intensive energy type in real life, it was chosen by 36 people, around 49%, as the most CO2-intensive type in the survey (See Figure C4 of Appendix C). The second largest group chose *"Google searching (50 times a day)"* as the most CO2-intensive energy type although it is the fourth most CO2-intensive energy type in real life. On the other hand, *"Email/Messaging (30 times per day)"* was chosen by only 3 participants as the most CO2-intensive energy type, aligning with the true statement.

When looking closely at Figure 7, showing the rank of each energy type, several interesting points were observed. On the *"Using light bulb"* option it was selected most as 1st rank and interestingly, 5th rank was selected as second most.

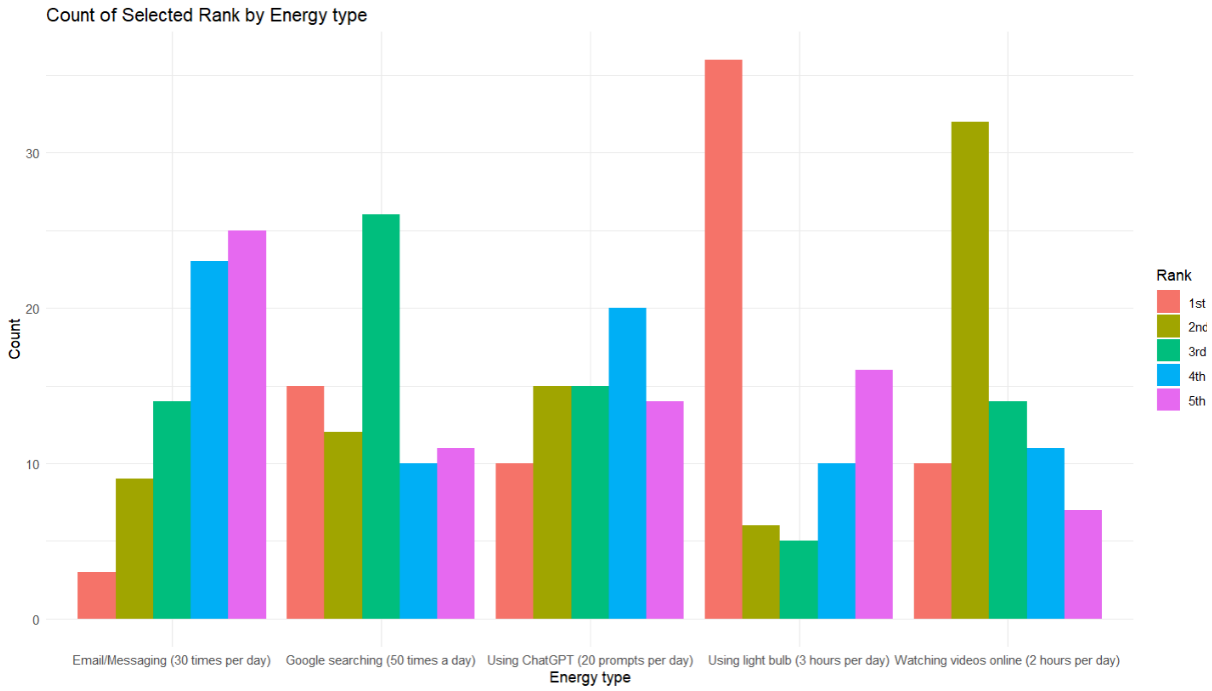

Figure 7. Ranked Energy Type

Notably, the "*Using ChatGPT*" option was most frequently ranked fourth and least frequently ranked first. It indicates the participant's lack of knowledge in regard to ChatGPT's environmental impact and the impact of ChatGPT's CO2 footprint which is not widely recognised. This finding highlights that the environmental impact of ChatGPT was underestimated; the participants' perception of the environmental impact of daily energy consumption may not align with real-life context.

On the other hand, the *"Watching videos online"* option, which is the most CO2-intensive energy type in real life, was chosen as the second one in the survey. On the other hand, the *"Email/Messaging"* option was most frequently chosen as the 5th rank, which aligns with real-life context.

4.1.2. Text Mining Analysis

Text mining was implemented to identify the most frequently occurring words in the context of using GenAI. Through the responses from the second section, Utilisation of Generative AI, we found that a vast majority of participants (89%) are using Generative AI for text generation purposes and ChatGPT was chosen as the most popular tool, with 93% of users indicating it is their primary choice among other options. Besides the data collected from close-ended questions, we have significant qualitative data gathered from the open-ended question, question 10: *"In what contexts do you use Generative AI? Please provide further details on the usage of Generative AI."* To get insight into the context of using GenAI, which is directly related to RQ2, a hybrid approach of text mining was conducted. As mentioned in Chapter 3.3.3. Content Analysis, the hybrid approach of text mining helped us to interpret the extracted keywords deeply regarding the context of using AI and the responsibility of reducing ChatGPT's CO2 footprint.

Context of Using GenAI

Figure 8 visualised in word clouds format represents the overview of the top keywords of responses from question 10. The most notable word is "*use*" which is the biggest size situated in the centre of the word cloud. It shows that "*use*" was the most mentioned word in the responses. Apart from that, it was observed that the words *need, text, writing, help, information, generative, work, ideas, and ChatGPT* surround the most frequent word, "*use*".

Figure 8. Context of Using GenAI

On the other hand, the bar chart displays the most frequently occurring words detected from the same data set, ranked from 1 to 20 (See Figure C5 of Appendix C). It helped us investigate specific words deeply with specific amounts of frequency. For example, the word "*use*" was the most frequently occurring one, appearing 31 times.

Notably, there is a significant gap between the most frequently mentioned word and the second most common one, "*need*", which appeared 14 times. Apart from them, there is no significant gap between the rest of the words.

The third most frequent word, "*text*" was found, appearing 13 times as many times as "*help*" appeared. It corresponds to the findings from question 9 that the vast majority (89%) of participants answered that they are using Generative AI for *"text generation"*. Additionally, ChatGPT was clearly observed in the bar chart, appearing 9 times, whereas other tools were not mentioned in the top 20. It aligns with the findings from the questionnaire that ChatGPT was chosen as the most popular choice, with 60% of users indicating its primary usage.

The words, *information, writing, work and ideas* had a similar amount of frequency, appearing around 10 times. In terms of the context of using GenAI, some specific words were uncovered such as *research, questions, studies and searching*. It can be interpreted that users easily link GenAI with these words to the context of using GenAI.

Context of Responsibility of Reducing CO2 Footprint of ChatGPT

Text mining was used for question 27, which is a follow-up question to question 26: "*In your opinion, who is primarily in charge of limiting the environmental impact of ChatGPT?"* Question 27 required the respondents to specify the reasons for their answer to question 26. The participants were divided with around 30% votes about responsibility going into *Companies, Governments, and Individuals* without a significant gap. The largest group selected the "*Companies*" option at 36%, followed by the "*Individuals*" and "*Government*" at 31% and 30%, respectively.

Figure 9 presents the context behind the respondents' choice. According to the results from text mining, the word *"companies"* ranked as the most frequently occurring word. On the other hand, the word *"government"* was the third most frequent word and *"individuals"* ranked 8th.

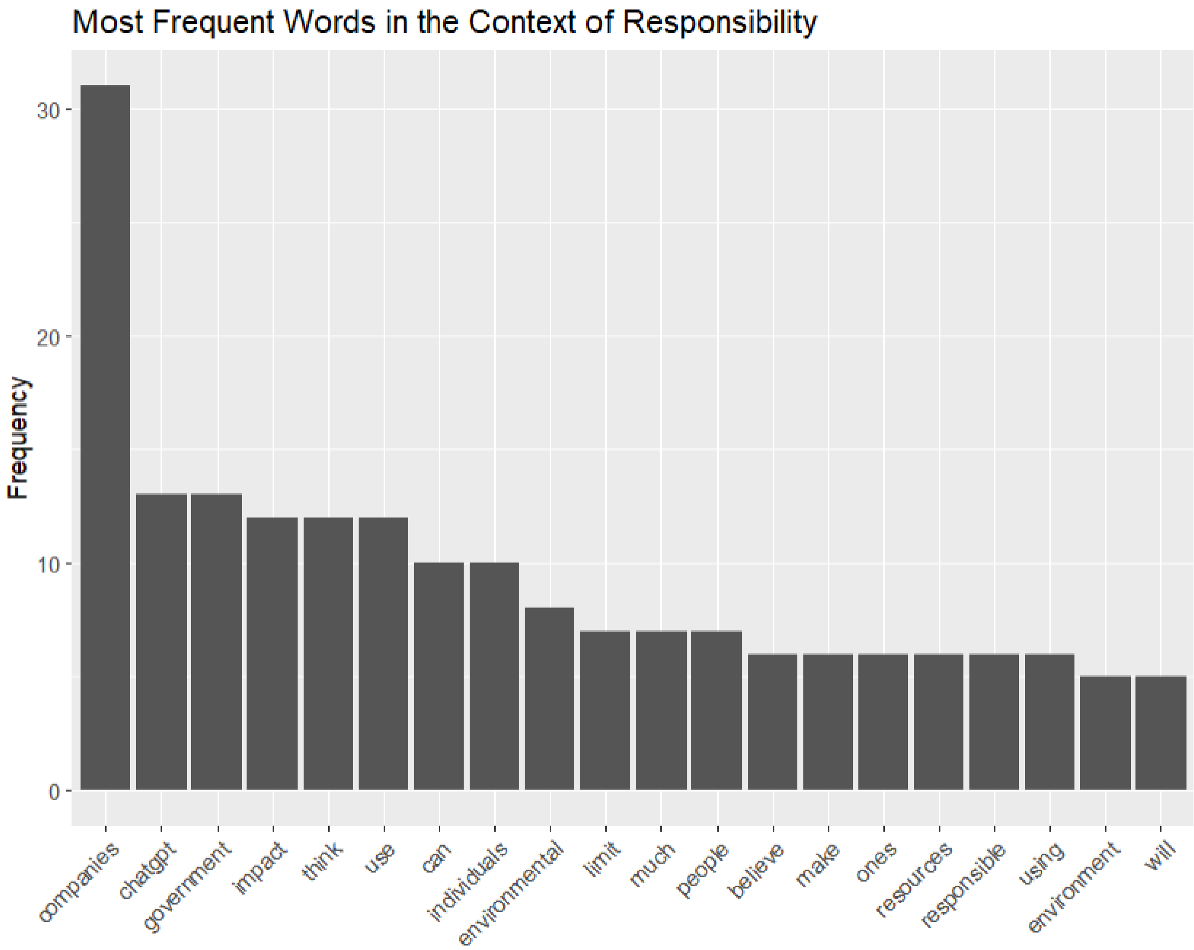

Figure 9. Most Frequently Occurring Words Regarding Responsibility

“*ChatGPT”* was observed as the second most frequently occurring word. It can be interpreted in two ways. First, ChatGPT was mentioned significantly as the word was included in the question. On the other hand, there would possibility that respondents connected *“Companies”* to *“ChatGPT”*, indicating companies providing ChatGPT have the responsibility of limiting its environmental impact as the largest group selected the *“Companies*” option.

4.1.3. Sentiment Analysis

Based on our hybrid approach, sentiment-content analysis enhanced our understanding of participants' perspectives on questions concerning changes in user behaviour regarding ChatGPT. This specific fusion allowed us to deepen our insights into the participants' emotional states and the structural analysis of their responses. It helped us gather more accurate and better-quality results.

Additionally, as mentioned in the previous chapter 3.3.3.3 Sentiment Analysis, we are using the data set acquired by the questionnaire, specifically question 24: *“Research indicates that ChatGPT has 180.5 million users and 1.6 billion visits (January 2024) and hundreds of millions of prompts on ChatGPT costs around 1 GWh each day, which is equivalent to the energy consumption of 33,000 U.S. households. Would this influence your behaviour using ChatGPT?”* and question 25: *“Please specify the reasons for your response”.*

In our study, it was crucial to detect users’ willingness to change their behaviour with ChatGPT as we aim to develop users’ sustainable practices. Results from sentiment analysis revealed interesting insights into participants' attitudes towards ChatGPT's significant energy consumption and their willingness to change their behaviour.

Users’ willingness to change their behaviour

Regarding question 24, 64% of the participants answered “*No*”, which means 64% reported that they would not change their behaviour even though generating prompts has a significant CO2 footprint.

The findings from sentiment analysis show deeper insights beyond their answer. The analysis showed that for Coder 1, the number of positive responses surpassed the negative ones by two points, with 28 positives, 26 negatives, and 18 neutral comments evaluated. In contrast, for Coder 2, the number of negative responses outweighed the positives, with 22 positives, 32 negatives, and 18 neutral comments out of the 72 responses analysed (See Figure 10).

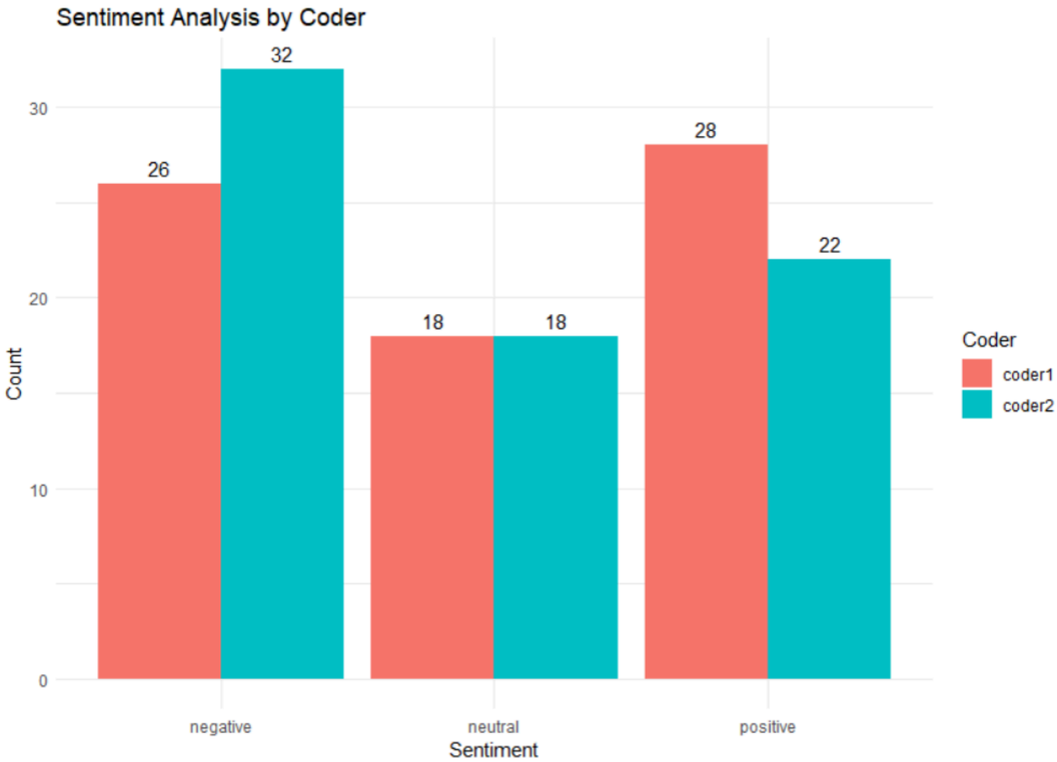

Figure 10. Sentiment Analysis Per Coder

These findings suggest that participants had mixed and negative feelings about changing their behaviour based on ChatGPT's energy consumption. Thus, sentiment analysis gave us a better understanding of how the participants view and feel about the energy consumption of ChatGPT and their thoughts on changing their behaviour.

Moreover, the analysis shed light on the reasons behind participants' willingness or reluctance to change. For instance, there are 11 responses that we both marked as positive regarding how knowing the facts about energy consumption would impact their usage of ChatGPT. Additionally, there were 3 of the same responses marked positive about looking into how to rephrase prompts for more efficacy while using ChatGPT. However, a significant portion of negative comments expressed concerns about not hindering technological advancement and emphasised the greater usefulness of ChatGPT compared to the environmental issues it poses.

Figure 11 intuitively shows the most prominent words from responses to question 25. The most striking words are *“use”, “energy”* and *“ChatGPT”*, which are the biggest and located in the middle of the word cloud. This indicates that these three words were the most mentioned words in participants' responses. The verbs related to action and environment were shown as *don’t, will, change, using, think, help and impact*. These words can also be seen as the second most visible ones in the word cloud. Additionally, we can also see some significant words related to the question such as “*environment*” and “*climate*”.

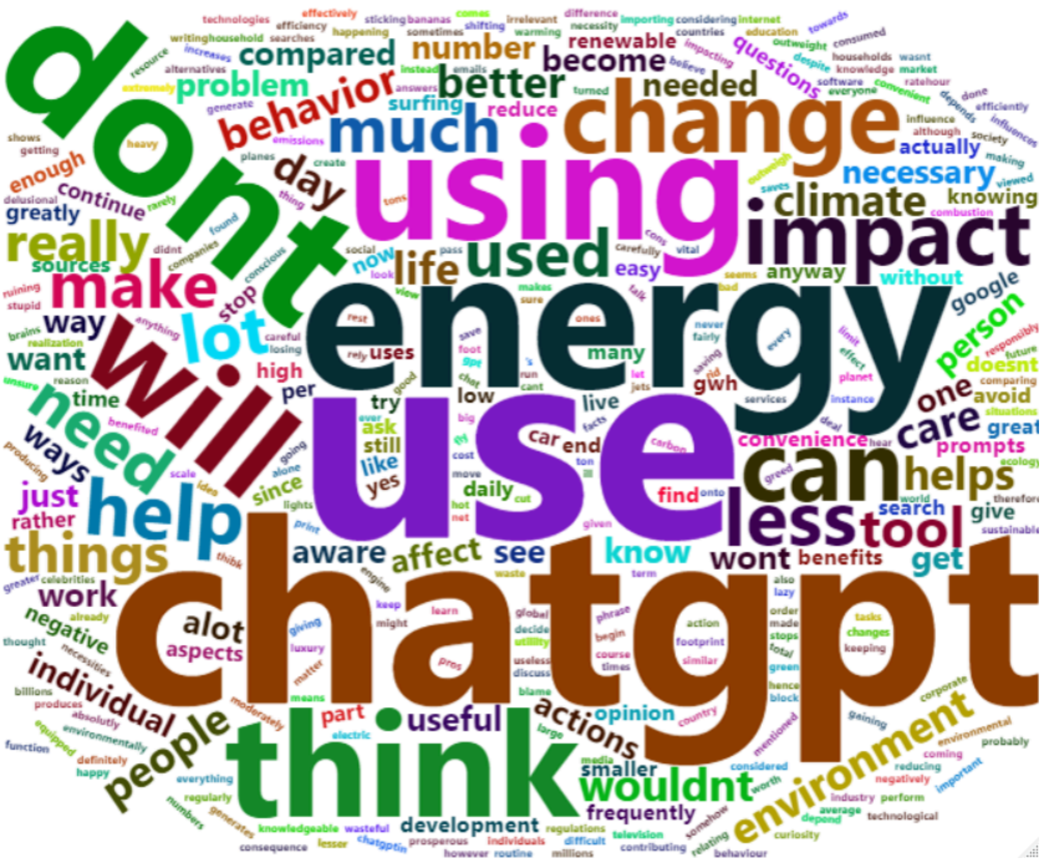

Figure 11. Most Mentioned Words in Sentiment Analysis

In summary, sentiment analysis provided valuable insights into participants' sentiments, highlighting both areas of potential acceptance and resistance towards behaviour change regarding ChatGPT's energy consumption. These findings contribute to a deeper understanding of user perspectives and can help develop strategies and ideas for promoting sustainable usage of ChatGPT.

Cohen's Kappa

Cohen's kappa analysis further solidifies our understanding of the sentiment classification process. It reveals a substantial agreement between us two raters regarding the sentiment expressed in the responses.

As shown in Table 1, the unweighted Cohen's kappa coefficient, which measures agreement beyond chance, is estimated at 0.66, with confidence boundaries ranging from 0.52 to 0.80. This indicates a substantial level of agreement between us the raters, reinforcing the reliability of our findings. Moreover, the weighted Cohen's kappa coefficient, which considers the degree of disagreement among sentiment categories, is estimated at 0.85, with confidence boundaries between 0.80 and 0.90. This specifies an even higher level of agreement when considering the degree of disagreement among the different sentiment categories.

```
> R_columns <- cbind(Sentiment$R1, Sentiment$R2)
> cohen.kappa(x = R_columns)
Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)

Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
                 lower estimate upper
unweighted kappa  0.52     0.66   0.8
weighted kappa    0.80     0.85   0.9

 Number of subjects = 72
```

Table 1. Result of Cohen's Kappa

With a sample size of 72 subjects, these results demonstrate a strong level of agreement between us. This reliability in coding enhances the credibility of our sentiment analysis findings and of the concussion from them.

4.1.4. Thematic Analysis

The data set analysed by Thematic Analysis was the responses to question 19: *"What positive experiences have you had while interacting with ChatGPT in your most recent chats?* and question 20: *"What poor experiences have you had while interacting with ChatGPT in your most recent chats?"* in section 4 of the questionnaire. Through thematic analysis, five distinct themes were discovered that describe the factors contributing to the positive experience of ChatGPT, whereas four distinct themes were identified that represent the factors influencing the negative experience of ChatGPT. The whole process of the Thematic Analysis we conducted can be found in this link. In this chapter, the findings for each factor and the comparison between the two factors were discussed.

Factors contributing to positive experience

The data collected from the open-ended question, question 19: *"What positive experiences have you had while interacting with ChatGPT in your most recent chats?"* included 74 responses, matching the number of survey participants. During the initial coding phase, 59 out of 74 responses were coded besides invalid responses such as irrelevant responses.

Before concluding 5 distinct themes of positive experience, 8 different themes were categorised in the first place by examining the codes. The 8 themes are as follows: *1) Provide a starting point for new or complex topics and creative ideas, 2) Save time and accurate answers, 3) In-depth information based on the understanding of instructions, 4) Summarise, 5) Professional looking, 6) Grammar correction, 7) Unbiased interaction and 8) Interaction* (See Figure 12).

Among the themes, it turned out that the vast majority of factors contributing to positive experience came from the theme of "*Providing a starting point for new or complex topics and creative ideas*". For example, some remarkable codes related to the theme are *"It helped me get an overall idea of what to do"* and *"It gave me at least an idea of where I should start with my research"*. On the other hand, under the theme of saving time and accurate answers, interesting codes were observed. For instance, *"straightforward answers", "quicker and easier to create a report from start to finish"* and *"good outcome when it comes to requesting prompts*" were shown in the theme.

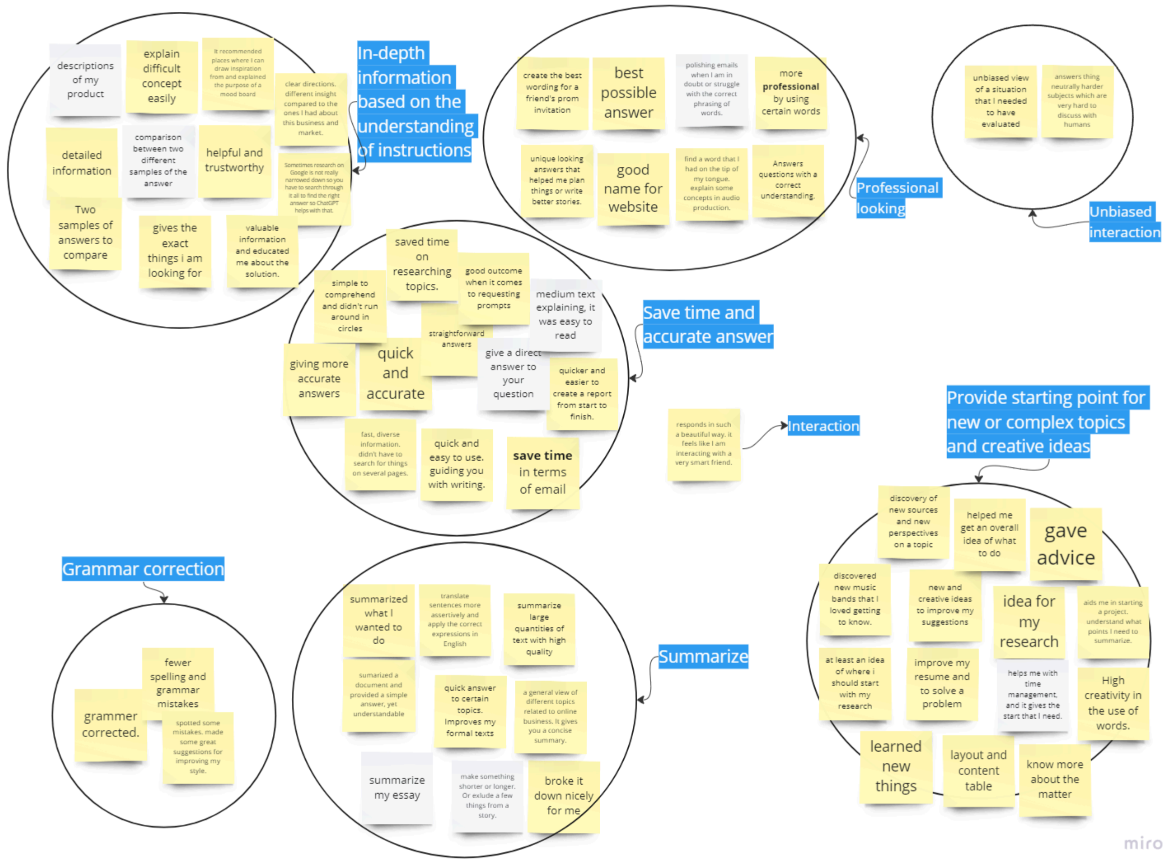

Figure 12. Initial Themes of Positive Factors

Through the reviewing themes phase, the codes under 8 themes were reviewed to see whether they could be merged with each other to get concrete results. For instance, the cluster titled "*Grammar correction*" was seen as closely associated with the cluster titled "*Professional looking*" so it was integrated into the "*Professional looking*" theme.

At the same time, the clusters having fewer codes than others were discussed as well such as the clusters named "*Interaction*" and "*Unbiased Interaction*". The code under the Interaction theme was *"Response in such a beautiful way. It feels like I am interacting with a very smart friend."* On the other hand, there were codes, *"Unbiased view of a situation that I needed to have evaluated"* and *"Answers thing neutrally harder subjects which are very hard to discuss with humans"* under Unbiased interaction. Despite a few codes, it had significant meaning since the positive factor regarding the unbiased interaction was not detected in the related works.

During the defining themes phase, 5 themes finally came out and the most relevant codes were written to represent each theme (See Table 2). Moreover, while defining the themes, the title of the themes was advanced. For example, the initial name of the cluster, "*In-depth information based on the understanding of instructions*", was changed to "*Accurate and detailed information*" to deliver findings in a clearer way. Lastly, in the theme of "*Providing ideas to understand topics*" the codes were categorised into two sub-themes. The first sub-theme is *"Suggest new and creative ideas"*; ChatGPT influenced positive experience since it suggested great ideas that users would not come across in terms of creativity. The second sub-theme, *"Assist users to solve problems"* indicates that users had a positive experience with ChatGPT as it helped them understand topics and solve problems quickly and easily at the starting point.

Help with Summarising	Provide ideas to understand topics	Professional looking	Save time	Accurate and detailed information
Summarise my essay. Summarise what I wanted to do. Summarise a document and provide a simple answer, yet understandable. Summarise large quantities of text with high quality.	- **Suggest new and creative ideas** Idea for my research. Discovery of new sources and new perspectives on a topic. New and creative ideas to improve my suggestions helped me get an overall idea of what to do. **- Assist users to solve problems** Aids me in starting a project. Understand what points I need to summarise. Explain difficult concepts easily. Sometimes research on Google is not really narrowed down so you have to search through it all to find the right answer, so ChatGPT helps with that.	More professional by using certain words. Spotted some mistakes and made some great suggestions for improving my style. Polishing emails when I am in doubt or struggle with the correct phrasing of words. Unique-looking answers that helped me plan things or write better stories.	Quicker and easier to create a report from start to finish. Quick and easy to use and guiding you with writing. Saved time on researching topics. Save time in terms of email Fast, diverse information. didn't have to search for things on several pages.	Answer questions with a correct understanding. Give the exact things I am looking for. Two samples of answers to compare. Detailed information.

Table 2. Defined Themes of Positive Factors

Factors contributing to negative experience

74 responses were included as responses to question 20: *“What poor experiences have you had while interacting with ChatGPT in your most recent chats?”*, and 52 responses were coded during the initial coding phase besides invalid data. As conducted in the former thematic analysis, the process to detect factors impacting poor experience followed the same procedure. After the initial coding phase, the codes were sorted out into clusters. As shown in Figure 13, 8 clusters were categorised based on codes: *1) Inaccurate outcomes, 2) Unhuman like responses, 3) Vague information, 4) Repetition of prompts, 5) Invalid information, 6) Technical issues, 7) Reaction to sensitive topics and 8) Image generation.*

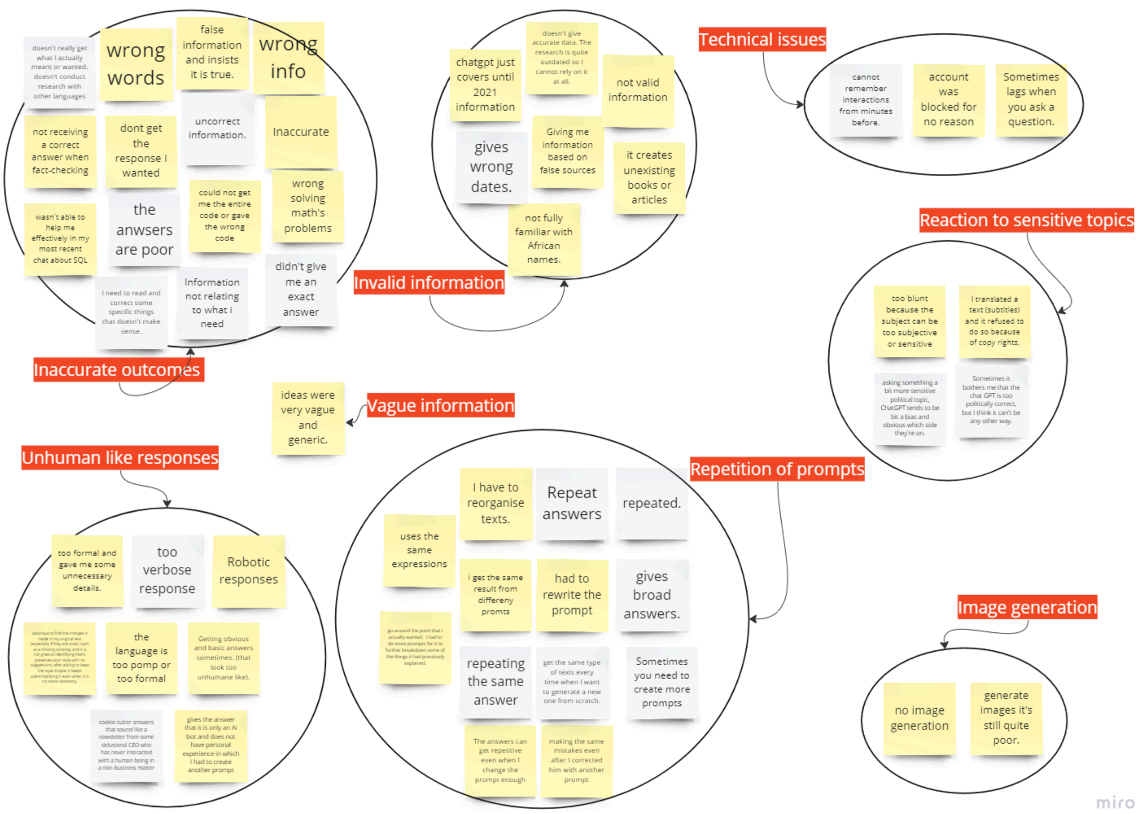

Figure 13. Initial Themes of Negative Factors

During the reviewing and defining theme phases, it was detected that the vast majority of opinions on negative experiences were caused by wrong information. Users were not satisfied with ChatGPT because of inaccurate and invalid information. The cluster, “*Inaccurate outcomes”* means that ChatGPT did not provide the right answers that users desired. For example, in the survey, one of the participants described *“It doesn't really get what I actually meant or wanted. It doesn't conduct research with other languages”*. Another participant illustrated that *“False information and insists it is true”*. On the other hand, some participants described *“It doesn't give accurate data. The research is quite outdated so I cannot rely on it at all”* and *“ChatGPT just covers until 2021 information”* regarding invalid information.

Apart from that, the cluster of "*Unhuman-like responses"* contains meaningful codes to be analysed such as *"Too formal and gave me some unnecessary details"* and *"It gives the answer that it is only an AI bot and does not have personal experience in which I had to create another prompt"*. It can be interpreted that users expected more human-like engagement when they had a chat with ChatGPT. The theme of "*Repetition of prompts"* has two sub-themes: "*Repetitive responses from ChatGPT"* and "*Repetitive creation of prompts'*. The first sub-theme tells that users had a poor experience as ChatGPT provided the same answers that users did not desire and kept making the same mistakes after several corrections. The latter sub-theme illustrates the situation when users had to create prompts repeatedly as they did not get the desired outcomes at once. The codes telling about what technical issues users have had when interacting with ChatGPT were disclosed. Furthermore, it was revealed that users had poor experience with sensitive topics. The main 5 themes drawn by the implementation are shown in Table 3.

Inaccurate outcomes	**Invalid Information**	**Unhuman-like responses**	**Repetition of prompts**	**Technical issues**
Wrong solving maths problems. Not receiving a correct answer when fact-checking. False information and insists it is true. I need to read and correct some specific things that don't make sense. Information not relating to what I need.	Giving me information based on false sources. ChatGPT covers until 2021 information. Doesn't give accurate data. The research is quite outdated so I cannot rely on it at all. Not valid information. Gives the wrong dates.	Gives the answer that it is only an AI bot and does not have personal experience in which I had to create another prompt. Robotic responses. Too formal and gave me some unnecessary details. Repeating the same answer.	**- Repetitive responses from ChatGPT** I get the same result from different prompts. Making the same mistakes even after I corrected him with another prompt. **- Repetitive creation of prompts** Sometimes you need to create more prompts. Go around the point that I actually wanted. I had to do more prompts for it to further break down some of the things it had previously explained.	Sometimes lags when you ask a question. The account was blocked for no reason. Cannot remember interactions from minutes before.

Table 3. Defined Themes of Negative Factors

Comparison findings between positive and negative factors

The comparison of findings between positive and negative factors provides several interesting observations. First, it was seen that there were more codes related to positive experiences than negative ones during the initial coding phase. It indicates that the participants had more valid positive experiences with ChatGPT. On the other hand, there were notable conflicts in a certain topic. For instance, while accurate and detailed information was identified as a factor contributing to positive experiences, the participants also reported that they had poor experiences with ChatGPT due to inaccurate and invalid information. It shows that users primarily expect to gain accurate and reliable information when interacting with ChatGPT. These observations shed light on both the strengths and areas for improvement within ChatGPT's functionality.

4.2. Findings from Contextual Inquiry

This section presents the findings obtained from the contextual inquiries conducted with two participants. The aim of these sessions was to gain insights into the users' behaviour, satisfaction with the system's responses, and preferences within ChatGPT. Each participant engaged in separate sessions, allowing for different perspectives to be observed. The findings are divided by each participant for a better perspective of the differences between the participants. Lastly, we are concluding the main findings from both CI sessions.

Demography

Before delving into the findings, we wanted to provide a brief overview of our participants' demographics:

- **Participant 1:** age: 25-35, occupation: student, technological proficiency: basic understanding of ChatGPT
- **Participant 2:** age: 25-35, occupation: student, technological proficiency: advanced understanding of ChatGPT

Participant 1

During the first task, asking whether the first of May was a public holiday in Denmark, participant 1's approach was to formulate prompts, with precise grammar and sentence structure to obtain the desired answer. Their thought process was that it would have a better accuracy in response. When presented with ChatGPT's answer, participant 1 did not doubt the correctness of the response, although the response was inaccurate. The response they received was that Denmark was celebrated the same as many other countries, when in fact the 1st of May is not a public holiday in Denmark. Since the question was straightforward the participant was satisfied with just one prompt. After a brief discussion, the participant made one more prompt question asking ChatGPT how the holiday came to be. In total, two prompts were made.

For the second task, writing an email for a job position, participant 1 once again focused on the correct grammar and sentence structure in order to gain the best possible answer. Upon receiving the initial response, participant 1 expressed their satisfaction with the well-organised structure of the response. They further showed how they utilise it for acquiring more complex words to use in the email. Once more as we discussed the participant shared that they mostly used ChatGPT for grammar correction, acquiring more complex words and for the structure of emails. Finally, in the end, participant 1 was satisfied with ChatGPT wording. The total number of prompts used was three.

In the interview, when asked if the participant would rather use ChatGPT or Google for a simpler question as the first task, participant 1 talked about mostly using ChatGPT for answering all his questions, even the simple questions that Google could easily answer. The reason is that ChatGPT immediately summarises all the information and is convenient for doing further follow-up questions.

When it came to the question about going through the history or creating new chats, they expressed frustration about not being easily able to find the chat they had previously created. So, for that reason they did say they try to search for them but more times than not they create a new chat. Additionally, they provided us with an excellent idea of incorporating a history-categorising system. That way the users can sort their chats which are important and have an easier way of finding them.

Moreover, participant 1 expressed that on average they create five prompts per chat depending on what the chat is about. For general chats, it would be five but if the chats are about coding it could be up to ten.

Participant 2

It was observed that ChatGPT provided users with in-depth information related to the topic even though the prompt was quite short and simple. For instance, participant 2 created the initial prompt, *“Is a labour day a holiday in Sweden?”* and it gave her the definition of labour day in general and how labour day is celebrated in Sweden.

It says labour day is not a public holiday in Sweden, but Sweden celebrates its own labour day which is a public holiday. However, participant 2 misinterpreted the answer in the beginning, saying *“I got an answer that labour day is not a public holiday in Sweden”* since the sentence, *“labour day is not a public holiday in Sweden”* was written in the first line with the definition of labour day. After reading the answer once again they realised that the labour day is a public holiday in Sweden. Lastly, they were happy with the initial prompts and said they would not ask further questions regarding the first task since they already got an answer to this Yes or No question.

For the second task, participant 2 created a prompt, asking what should be included in an email to a hiring manager and the initial answer gave a very detailed structure with some advice. Participant 2 was happy with the long and detailed answer, mentioning that it gave them good advice that they would not think about it by themselves, and went over the answer thoroughly. As a follow-up prompt, participant 2 asked more specific questions such as replacing the word and improving the wording based on what she took out of the sentences from the initial prompts. Additionally, in terms of writing like the second task, participant 2 mentioned they usually use ChatGPT in this quite concrete structure: *Ask examples, Write her own thing, Get assistance with refining vocabulary and grammar repeatedly and Evaluate the final writing*. Moreover, regarding the question of how many participant 2 creates prompts usually, they said they form at least 5 times at a time and it really depends on what they ask.

In the interview, when we asked whether they would rather use Google for the first task which is kind of a simple question, participant 2 said they would insist on ChatGPT instead of Google. The reason is that it is easier to use and provide a lot of information at once and they do not need to visit several links to get the desired information. Furthermore, it is beneficial to ask follow-up questions in ChatGPT since it already learnt the topic. On the other hand, participant 2 said they might be able to contribute to reducing CO2 emissions by cutting down the number of prompts, for instance combining questions.

Regarding the usage of "*New chat*" and the history, participant 2 normally opens a new chat for a simple question. However, in case the topic is quite heavy and complicated participant 2 retrieves what was created on the same topic before since it kept learning and provided better outcomes; it helped them remember what was asked before, participant 2 said. When it provides wrong information, participant 2 experienced that it gave wrong information regarding the Excel functions and in this situation, participant 2 tends to ask differently and does not use the edit function at all.

Participant 2 said they regularly clean up the chat which is not used frequently. As a heavy user, participant 2 had some great ideas about the improvement of ChatGPT in terms of function. What was suggested was a favourite function in the chat history so that the most frequent chat is pinned and users can find it easily since it is tough to find specific topics in the heavily stacked history.

Cross-Participant Analysis

For the cross-participant analysis, we decided on a few key points we wanted to focus on, such as formulation of initial prompts, follow-up prompts, suggestions for improvement of ChatGPT, and patterns and trends.

Formulation of Initial Prompts

When it comes to the first task, whether the first of May is a public holiday, both participants wrote straightforward questions with precise grammar and simple structure. The interesting finding was that while participant 1 received inaccurate results, participant 2 received correct results but misinterpreted them as participant 2 did not read the full summary. Nonetheless, both participants showed satisfaction with the provided results and had created just one to two prompts.

Furthermore, in the case of the second task, writing an email for a job position, both of the participants went for different formulations of the question, yet both received similar responses. Participant 1 once more started with a straightforward question with a focus on grammar and sentence structure. Participant 2 asked what should be included in the email.

Both of the questions resulted in the ChatGPT providing them with a detailed and structured job position email.

Follow-Up Prompts

For the first task, we could say that neither of the participants had a need to create a follow-up question due to being content with the initial response. However, in the second task, both participants created follow-up questions. With the initial question they received a detailed structure of the email, but while participant 1 only asked for more complexity in the vocabulary for the follow-up question, participant 2 asked specific questions to refine vocabulary and improve the wording, leading to multiple iterations of feedback and refinement.

Suggestions for Improvement of ChatGPT

Throughout the contextual inquiry, both of the participants had thought of a possible improvement of ChatGPT. Participant 1 as they expressed their frustration with finding previous chats in history, suggested a history-categorising system to sort and find important chats more easily.

However, participant 2 proposed a feature to pin or favourite specific chats for easier access to frequently used conversations. They also mentioned the potential for environmental impact reduction by combining questions to minimise the number of prompts.

Patterns and Trends

Lastly, we wanted to mention some of the consistencies we saw in both users' behaviour. For instance, regarding prompt formulation they both emphasised the importance of precise and well-structured initial prompts. Follow-up prompts were also constructed in a similar manner although participant 2 was more detailed they both shared the same thought, which was to refine or verify the information provided. For the chat history, both participants expressed difficulties in managing them and suggested improvements such as categorisation, pinning, or favouriting chats to improve navigation and efficiency. Moreover, both participants preferred using ChatGPT over Google for answering questions, mentioning convenience and the ability to ask follow-up questions as key benefits.

5. Discussion and Conclusion

This section integrates the findings presented in the previous chapters, addressing the research questions that guided our project. The discussions explore the environmental impacts of Generative Artificial Intelligence (GenAI) technologies, particularly focusing on their CO2 emissions footprint (RQ1). We examine the most commonly utilised GenAI tools and the specific contexts in which they are employed (RQ2). Furthermore, we delve into the usage patterns and preferences of ChatGPT users (RQ3) and identify factors contributing to both positive and negative user experiences (RQ4). The analysis also highlights how users utilise and optimise prompts to achieve desired outcomes with ChatGPT (RQ5). Additionally, we assess the extent of user awareness regarding the environmental impact of ChatGPT on the digital carbon footprint (RQ6) and propose strategies for enhancing sustainable practices to mitigate CO2 emissions (RQ7). This thorough discussion integrates our findings with existing literature, offering insights into the suggestions and potential future directions regarding ChatGPT. The conclusion combines these insights, emphasising the significance of the study and its contributions to understanding the environmental impacts of ChatGPT.

5.1. Reflection on the Research Questions

RQ1: What are the environmental impacts of Generative Artificial Intelligence (GenAI) technologies, specifically regarding their CO2 emissions footprint?

Carbon footprint has been traditionally applied to sectors such as transportation, manufacturing and airlines. Recently, however, growing concerns have emerged regarding the CO2 emissions associated with the Generative Artificial Intelligence (GenAI) technologies sector as the usage of GenAI increased dramatically across various sectors.

GenAI technologies have a dual role in the context of the climate crisis. On one hand, GenAI supports the development of low-emission infrastructure and enhances climate predictions across various industries. However, the creation and operation of these AI systems contribute to significant carbon emissions as they rely on traditional electricity sources due to the limited availability of renewable energy.

Notably, the training and inference phases of GenAI are substantial contributors to its CO2 footprint. For instance, the CO2 emissions from training a single large language model are comparable to the emissions from 125 round-trip flights between New York and Beijing. Furthermore, it is estimated that the CO2 emission caused by inference accounts for 60% of AI energy usage. To address the environmental impact of GenAI, various movements are underway to develop sustainable frameworks for AI practitioners and engineers. However, current efforts and research are limited to specific engineering practices.

RQ2: What is the most commonly utilised Generative AI tool, and in which specific contexts is it used?

Our analysis revealed that ChatGPT is the most frequently used GenAI tool with 93% of respondents selecting it as their primary tool, and respondents primarily use GenAI to generate text with 89% of responses. It correlates with the statistics discovered in Innovation of Generative AI in Chapter 2.2.2 where we found the most popular GenAI application is chatbots and ChatGPT is the most used tool, reaching one million users in five days and 100 million monthly users in three months. Overall, this indicates the most generalised usage of GenAI is text generation and ChatGPT is the most preferred tool for it.

Respondents said that they had used ChatGPT at least once before in their life at 60% of respondents, followed by Gemini (Bard) at 14%, DALL-E and Copilot at 11%. Additionally, 5% of respondents said they had used other types of GenAI tools such as Suno AI (music creation), Deep AI (chat, image and video generation), WhatsApp AI (AI assistant in messenger app, WhatsApp) and Aria AI (AI in Opera browser). This demonstrates that besides ChatGPT, a variety of GenAI tools are integrated into users' daily lives. Specifically, the broad adoption of GenAI tools recalled some concerns regarding climate change, demonstrated in the Related Works chapter: 1) High-carbon energy sources like coal and natural gas are consumed by the majority of ML models as their primary energy source (Luccioni and Hernandez-Garcia, 2023) and 2) Compared to the phase of training the model, Inference eventually has enormous sessions used by millions of users, leading the carbon footprint becomes significant (Kumar & Davenport, 2023).

On the other hand, while examining the specific context in which GenAI is used, text mining revealed significant keywords such as searching, information, questions, writing, work, ideas and research. GenAI is generally used for searching information and asking questions. Additionally, it is beneficial for assisting with writing and providing ideas for both work-related tasks and studies and academic (studies) or research-related tasks.

RQ3: What are the common usage patterns and preferences of ChatGPT users?

The purpose and frequency of using ChatGPT and the preference for its usage were unveiled in our study. First, regarding the type of purpose, it was found that over half of respondents reported using ChatGPT for both personal and work-related activities. It aligns with the observation of Skjuve et al. (2023) described in Chapter 2.2.3, illustrating that most participants had a good experience with ChatGPT in both work and daily life in terms of creating activities and knowledge.

Investigating the frequency of ChatGPT usage, the largest number of respondents use it a few times a week, taking up 34% of responses, followed by those who use it a few times a month at 22%. The third largest group uses ChatGPT daily. Interestingly, the number of respondents using ChatGPT multiple times per day, the most frequent option, is similar to those who rarely use it, the least frequent option, with 9 and 8 respondents respectively. It is highlighted that our findings echo the findings by Skjuve et al. (2023), displaying that users use ChatGPT quite frequently, once a week or more. Frequent use of ChaGPT leads to significant CO2 emissions. According to our findings in related works, it was estimated ChatGPT produces 11 million requests/hour which causes 12.8k metric tons of CO2/year.

Users mostly prefer to use ChatGPT for assistance with writing tasks, summarising content and asking simple questions. For instance, approximately 60% of respondents indicated that they use ChatGPT either most frequently or frequently for getting help with writing. As a second preferred purpose, summarising content was ranked, with 57% of respondents indicating its frequent use. Notably, there is only a 2-point difference between the percentage of respondents using ChatGPT for summarising content and those using it for asking simple questions.

Interestingly, there is a significant difference between our results and the findings by Eiden (2023), drawing that people tend to use this largest LLM model instead of a smaller model for trivial stuff and playing with it without a clear objective, described in Chapter 2.2.3. Unlike Eiden's findings, our study revealed users interact with ChatGPT for specific purposes such as assistance with writing and summarising content. Additionally, the analysis of the latest chat topics in our questionnaire data set shows that users used ChatGPT with clear objectives related to personal or work-related tasks.

RQ4: What factors contribute to positive and negative experiences among ChatGPT users?

To address which positive and negative experiences participants or users of ChatGPT come across, we have been looking throughout this project. According to our findings, there were more positive experiences than negative ones. However, certain topics overlapped with each other, for instance, accurate and detailed information in the positive experience and inaccurate and invalid outcomes in the negative experiences. This indicates that users mainly expect to receive accurate and reliable information when interacting with ChatGPT. This outcome correlated to the study conducted by Skjuve et al. (2023) where it was found that most poor experiences were caused by getting irrelevant and useless outcomes to queries.

Through the thematic analysis, we found out that participants had a good experience with ChatGPT when seeking *help with summarising, providing ideas to understand topics, professional looking, saving time, and accurate and detailed information*. We can say that the positive factors in ChatGPT are influenced by the great ideas it provides, helping users understand topics and solve problems quickly and easily. Moreover, similar findings can also be seen in the results from the Skjuve et al. (2023) questionnaire. Skjuve et al. (2023), mentioned that ChatGP assists users in completing their tasks more quickly, improving the quality of their work for school or professional settings, and helps in understanding new topics. Additionally, it was noted that they used ChatGPT to generate inspiration for ideas.

When looking at the thematic analysis regarding poor experiences we came up with five themes which are *1) inaccurate outcomes, 2) invalid Information, 3) unhuman-like responses, 4) repetition of prompts and 5) technical issues.* From these themes, we can interpret that the participants have come across some inaccurate information and invalid outcomes which resulted in creating more prompts and eventually more time lost. As well as ChatGPT providing the same prompt outcome multiple times. Moreover, the participants stated that they would like a more human-like response when having a chat with ChatGPT. The more human-like response can be described as ChatGPT communicating with users less formally and giving users precise answers based on their input instead of too broad information covering not requested information.

Once again our findings correlate to those of Skjuve et al. (2023), who state that the poor experiences they notice are: *irrelevant and useless outcomes to queries, failure to understand their queries and technical issues*.

RQ5: How do users utilise prompts and optimise their use of prompts to obtain desired outcomes, interacting with ChatGPT?

To find out how users use prompts and optimise their usage, we looked into our questionnaire results, contextual inquiry, and related works. We found that users' interactions with ChatGPT are diverse, with various ways and quantities of prompt creation. However, the findings from contextual inquiry also highlight the need for system enhancements to better support users dealing with complex queries and to improve the management of chat history for easier navigation.

Looking back at Chapter 4.1.1 Quantitative Analysis we can see that users created mostly 1-5 prompts per session. This indicates that most users are relevantly satisfied with ChatGPT’s provided answers and do not require additional rephrasing. However, the second highest is 6-10 which can suggest the participants need more assistance. With these findings, we can say that participants who need to rephrase more than 5 times are likely to create more complex queries or look for specific information. Additionally, the participants selected higher numbers such as 15 and even over 20 prompts. Although the numbers are low this can still be seen as an issue which needs to be addressed.

Furthermore, in Chapter 4.2 Findings from Contextual Inquiry, we can see that some of the participants' choices correlate with the quantitative analysis. For example, both of the participants made one to two prompts for the first task which was a simple question. For the second task which was a more complex question participant 1 created three prompts while stating that they would create more prompts if they required additional information. On the other hand, participant 2 said they would create a minimum of 5 but likely more because of all the possibilities that ChatGPT can provide.

Moreover, during the CIs both participants specified that they create clear and straightforward instructions when creating prompts. The follow-up prompts are usually related to the vocabulary, improving the wording and grammar. Lastly, both participants said they tried to look through the history but ultimately could not find the chats they wanted resulting in them creating new chats. This can be problematic because starting new chats leads to generating more prompts. We assume that if participants need to ask a follow-up question to an existing chat but cannot locate it, they will need to create additional prompts in a new chat to reach the point they had previously.

According to our research prompts can be created in many ways and depending on the way they are phrased, different outputs are provided. With our participants, we saw two ways of creating prompts. The explicit prompts, with the first task, require clear and straightforward instructions for a more accurate and better-quality output. The second is the implicit prompts which have fewer directions in the instructions providing the AI a chance to interpret the desired outcome (Data Science Horizon, n.d.). At the end, in ChatGPT’s Prompt Engineering for Developers course, they also mentioned the same instruction in order to receive the desired outcome, as our participants had made. That would be the formulation of clear and specific instructions, stating that the clearer the prompts are the greater the quality of output results are provided (Takyar, n.d.).

RQ6: To what extent are individuals aware of the environmental impact of ChatGPT on the digital carbon footprint?

Our study found that individuals' awareness of the environmental impact of ChatGPT was notably low. While participants demonstrated a high level of consideration for climate change in their daily lives, the vast majority were unaware of the ChatGPT's environmental impact.

Specifically, as shown in Chapter 4.1.1, 20% of respondents reported caring about the impact of climate change with either extreme care (rating 10) or great care (rating 9), and 20% indicated that they have considered the impact of climate change in their daily activities either always (rating 10) or almost always (rating 9). However, regarding the environmental impact of ChatGPT, 23% of respondents stated they were unaware of it at all. Furthermore, while the results of questions related to the awareness of climate change indicate that respondents have a moderate or high level of awareness and concern for climate change, there was a notable lack of awareness regarding ChatGPT's environmental impact, with 81% of responses ranging from rating 0 to rating 5 out of 10.

It shows individuals do not consider the environmental impact of ChatGPT as significant as that of other industries. The results from the ranking question in the questionnaire also support this issue. Participants were asked to rank types of energy consumption based on the amount of CO_2 produced. According to the results, they did not know ChatGPT's significant environmental impact compared to other energy types more closely related to their daily lives.

Interestingly, the most selected energy type as first rank was Using a light bulb (3 hours per day), followed by Google searching (50 times a day), Using ChatGPT (20 prompts per day), Watching videos online (2 hours per day) and Email/Messaging (30 times per day). Particularly, Using ChatGPT (20 prompts per day) was ranked in third place, tied with Watching videos online (2 hours per day) and followed by Email/Messaging (30 times per day).

RQ7: How can sustainable practices, in terms of user behaviour, be enhanced to prevent CO2 emissions?

Enhancing sustainable practices to prevent CO_2 emissions can be achieved in many ways, including promoting energy-efficient behaviours, raising awareness, optimising the use of ChatGPT, and implementing a more user-friendly interface. Through our research, several key points and ideas came out such as a history-categorising system, a pin function, a more human-like interface and introducing features to teach users.

First, as suggested functions in ChatGPT, a history-categorising system and a pin function were drawn by the results from CIs which provided deeper insights into the participants' interactions with ChatGPT in Chapter 4.2. The participant experienced frustration when trying to find previous chats. To address this, they suggested a history-categorising system that sorts and finds important chats more easily, reducing not only the time and energy spent on looking for the chats but also the chances of creating new chats from scratch in terms of sustainable practices.

Additionally, participants expressed a need for a feature to pin or favourite specific chats for easier access to frequently used conversations. This would reduce repetitive prompt creation and contribute to decreased energy consumption. Lastly, the potential area that can be looked into regarding reducing prompts is the combination of related questions into a single prompt. This practice can minimise the number of interactions needed, conserving energy and reducing CO_2 emissions.

Furthermore, developing a more human-like interface can contribute to reducing the CO_2 footprint of ChatGPT. The results from the thematic analysis revealed that respondents have poor experiences with ChatGPT due to unhuman-like interaction; for example, ChatGPT kept providing the same answer although they tweaked the prompts and as well as providing unnecessary details. Additionally, similar opinions were found in the response to question 21: *"How do you think ChatGPT could be enhanced to better meet your needs?"* Several participants highlighted the need for more human-like responses from ChatGPT; for instance, there were responses*: "It needs more 'humanity'"* and *"I've noticed that whenever I ask them to try and expand on topics, they tend to repeat the same suggestions over again but with different wording, so I would like them to try to avoid that."* Enhancing the human element of ChatGPT can improve user satisfaction and efficiency, thereby reducing the time spent on tasks and reducing prompt creation.

Another notable suggestion from the participants was the inclusion of features to teach them how to use prompts more efficiently. For example, we found a meaningful response in the answers to question 21, which is "*maybe there should be a feature to teach how to use prompts in the most effective way."* Efficient prompt usage can minimise unnecessary interactions and reduce energy consumption.

Moreover, participants recommended adding categories, topics, or actions to better the search results and response quality, enabling more targeted and efficient information outcomes, saying *"It could implement tools like other Generative AI that already have categories, topics and actions you can select. And you can put by default "someone" who helps you to use ChatGPT and ask you what you want to use it for so it gets sort of "programmed" in the way the user chooses."*

Integrating these insights can greatly improve sustainable practices in ChatGPT. Implementing features that educate users on efficient prompt usage and sustainable practices can considerably reduce unnecessary energy consumption. This includes tutorials, tips, and real-time feedback on phrasing queries effectively. Improving interaction design by adding more options for answers and categorisation of information can lead to more efficient user interactions. Furthermore, the ability to pin or favourite important chats can improve workflows and reduce repetitive prompt creation. Lastly, creating a categorising section in history can improve user frustration and save time. This system should allow users to easily find and revisit important conversations, reducing the need for creating new chats and more queries.

5.2. Conclusion

This report has explored the environmental impacts and user awareness of Generative AI, focusing specifically on ChatGPT. By examining related works, it was found that GenAI, particularly ChatGPT, has a significant environmental impact. The CO2 emission from the inference phase was notably highlighted. Additionally, the review of related literature revealed that while there are rising concerns about the substantial usage of GenAI, research and efforts towards sustainable practices remain quite limited and are primarily focused on engineering solutions.

This study emphasised the importance of understanding individual awareness of the environmental impact of GenAI, specifically ChatGPT which is currently the most used GenAI tool, due to the dramatic increase in its integration into daily activities. Through the questionnaire and contextual inquiries, this study revealed frequent use of ChatGPT and usage patterns by individuals. The vast majority of respondents use ChatGPT a few times a week, primarily for tasks such as writing assistance, text summarisation and information searching in personal and work-related contexts.

Moreover, the findings indicate a significant gap between user awareness of climate change and the environmental impact of ChatGPT. Most users are unaware of the carbon footprint associated with their AI usage. This points to a critical need for raising awareness in regard to this issue.

Several potential sustainable practices are proposed in this thesis, including raising awareness, optimising prompt usage, and improving user interface features to reduce unnecessary interactions. Educating users on efficient prompt usage, such as creating clear and specific prompts, providing details and giving examples could significantly reduce the environmental impact.

In summary, this master thesis investigated user awareness and behaviour, highlighting the need for increased awareness and practical strategies to reduce the environmental impact of GenAI technologies.

Bibliography

Advice. (n.d.). AI Innovation for Decarbonisation's Virtual Centre of Excellence. AI and decarbonisation: innovation. 12-17. Retrieved March 14, 2024, from https://www.turing.ac.uk/sites/default/files/2023-12/advice_-_ai_for_decarbonisation_ecosystem.pdf.

Advice. (n.d.). AI Innovation for Decarbonisation's Virtual Centre of Excellence. Introduction. 5-12. Retrieved March 14, 2024, from https://www.turing.ac.uk/sites/default/files/2023-12/advice_-_ai_for_decarbonisation_ecosystem.pdf.

Ahmad, M., Shabib, A., Syed M., and Sarfraz, A. (2017). Machine Learning Techniques for Sentiment Analysis: A Review. Int. J. Multidiscip. Sci. Eng 8(3): 27–32.

Albrecht, M. & Aliaga, S. (2023). The transformative power of generative AI. J.P.Morgan Asset Management. 1-26. Retrieved March 16, 2024, from https://am.jpmorgan.com/content/dam/jpm-am-aem/global/en/insights/The%20transformative%20power%20of%20generative%20AI.pdf.

Almeida, F., Faria, D. & Queirós, A. (2017). Strengths and Limitations of Qualitative and Quantitative Research Methods. European Journal of Education Studies. 3. 369-387. 10.5281/zenodo.887089.

Amatriain X. (2024). Prompt Design and Engineering: Introduction and Advanced Methods. Cornell University. 1-26. https://doi.org/10.48550/arXiv.2401.14423.

Aqlan, A.A.Q., Manjula, B., Lakshman Naik, R. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. vol 28. Springer, Singapore. 147-162. https://doi.org/10.1007/978-981-13-6459-4_16.

Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. International Journal of Contemporary Applied Researches, 6(5), 76-89.

Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., & Olivetti, E. (2024). The Climate and Sustainability Implications of Generative AI. An MIT Exploration of Generative AI. 1-45. https://doi.org/10.21428/e4baedd9.9070dfe7.

Bednar, P. (2010). Contextual Inquiry as a Critical Perspective in Research. In J. Esteves (Ed.), Proceedings of ECRM 2010. 61-69. Academic Publishing.

Benchimol, J., Kazinnik, S., & Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. Machine Learning with Applications, 8, 100286. 1-19.

Beyer, H. & Holtzblatt, K. (1997). Contextual Design: Defining Customer-Centered Systems. Principles of Contextual Inquiry. Morgan Kaufmann Publishers Inc. 41-66. https://dl.acm.org/doi/book/10.5555/2821566.

Bhattarai, P. (2022, January 25). Introduction to Text Mining with R. RPubs. https://rpubs.com/vipero7/introduction-to-text-mining-with-r.

Bhonde, S.B., & Prasad, J.R. (2015). Sentiment Analysis - Methods, Applications & Challenges. International Journal of Electronics Communication and Computer Engineering. Volume 6, ISSN (Online): 2249–071X. 634-640. https://ijecce.org/administrator/components/com_jresearch/files/publications/IJECCE_3633_Final.pdf.

Bolderston, A. (2008). Writing an Effective Literature Review. Journal of Medical Imaging and Radiation Sciences. 39. 86-92. 10.1016/j.jmir.2008.04.009.

Brady, D., (2023, April 7). What developers need to know about generative AI. GitHub. https://github.blog/2023-04-07-what-developers-need-to-know-about-generative-ai/.

Brink, H.I.L. (1993). Validity and reliability in qualitative research. Curationis. 16. 35-38. 10.4102/curationis.v16i2.1396.

Cantero, C. (2019). How to Write a Literature Review. San José State University Writing Center. 1-7.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226. 1-44.

Caruth, G. D. (2013). Demystifying mixed methods research design: A review of the literature. Online Submission, 3(2), 112-122.

Chetty, P. & Thakur, S. (2020, January 27). How to establish the validity and reliability of qualitative research? PG. https://www.projectguru.in/how-to-establish-the-validity-and-reliability-of-qualitative-research/.

Chien, A.A., Lin, L., Nguyen, H., Rao, V., Sharma, T., & Wijayawardana, R. (2023). Reducing the Carbon Impact of Generative AI Inference (today and in 2035). Proceedings of the 2nd Workshop on Sustainable Computer Systems. 1-7. https://doi.org/10.1145/3604930.3605705.

Cho, R., (2023, June 9,). AI's Growing Carbon Footprint. State of the Planet. https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/.

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., & Carrillo, P. M. (2009). The needs and benefits of Text Mining applications on Post-Project Reviews. Computers in Industry, 60(9), 728-740.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 10, Self-completion questionnaires. In Bryman's social research methods (Sixth edition ed.). 210-233. Oxford University Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 11, Asking questions. In Bryman's social research methods (Sixth edition ed.). 234-254. Oxford University Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 13, Content analysis. Bryman's social research methods (Sixth edition ed.). 270-292. Oxford University Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 23, Qualitative data analysis. In Bryman's social research methods (Sixth edition ed.). 523-551. Oxford University Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 24, Mixed methods research. In Bryman's social research methods (Sixth edition ed.). 555-576. Oxford University Press.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Chapter 3, Research designs. In Bryman's social research methods (Sixth edition ed.). 38-67. Oxford University Press.

Clarke, V., & Braun, V. (2017). Thematic analysis. The Journal of Positive Psychology, 12(3), 297-298. 10.1080/17439760.2016.126261.

Cohan, P. (2023, November 9). As ChatGPT And Other AI Tools Increase Energy Demand, Here's What Investors Need To Know. Forbes.

https://www.forbes.com/sites/petercohan/2023/11/09/equinix-and-vertiv-stock-prices-could-rise-on-generative-ais-energy-use/.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46. https://doi.org/10.1177/001316446002000104.

Cooper, H.M. (1988). Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. Knowledge in Society, 1, 104-126.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. The public opinion quarterly, 64(4), 464-494.

D'Andrea, A., Ferri, F., Grifoni, P. & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications. 125. 26-33. 10.5120/ijca2015905866.

Das, A., & Modak, A. (2023). The Carbon footprint of Machine Learning Models. International Journal on Emerging Research Areas (IJERA), 1-4.

Data Science Horizon. (n.d.). Mastering Generative AI and Prompt Engineering: A Practical Guide for Data Scientists. 1-41. https://datasciencehorizons.com/pub/Mastering_Generative_AI_Prompt_Engineering_Data_Science_Horizons_v2.pdf.

Denzin, N., & Lincoln, Y. (2011). The SAGE handbook of qualitative research. Location the Filed. Thousand Oaks, CA: SAGE. 70-86.

Determ. (2023, September 26). Top 5 Benefits of Sentiment Analysis for Businesses. Determ. Retrieved April 03, 2024, from https://www.determ.com/blog/top-5-benefits-of-sentiment-analysis-for-businesses/.

Dhar, P. (2020). The carbon impact of artificial intelligence. Nature Machine Intelligence. 2. 423-425. 10.1038/s42256-020-0219-9.

Dilmegani, C. (2024, January 3). Sentiment Analysis: How it Works & Best Practices in 2024. AI Multiple Research. Retrieved April 03, 2024, from https://research.aimultiple.com/sentiment-analysis/.

Duarte, F. (2024, March 1). Number of ChatGPT Users (Mar 2024). Exploding Topics. Retrieved March 16, 2024, from https://explodingtopics.com/blog/chatgpt-users#.

Duarte, F., (2024, March 27). Number of ChatGPT Users (Apr 2024). Exploding Topics. Retrieved April 20, 2024, from https://explodingtopics.com/blog/chatgpt-users.

Duda, S., Warburton, C., Black, N. (2020). Contextual Research. In: Kurosu, M. (eds) Human-Computer Interaction. Design and User Experience. HCII 2020. 33-49.Lecture Notes in Computer Science(), vol 12181. Springer, Cham. https://doi.org/10.1007/978-3-030-49059-1_3.

Eiden, M. (2023, August 28) Generative AI: A Conversation with the Future. AMPLIFY. CUTTER. 40-50. https://www.cutter.com/sites/default/files/Amplify/2023/ADL_CUTTER_Generative%20AI_Conversation_Future_0.pdf.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. Sage Open, 4(1). 1-10. https://doi.org/10.1177/2158244014522633.

Eugenie, P., & Gelles-Watnick, R. (2023, August 28). Most Americans haven't used ChatGPT; few think it will have a major impact on their job. Pew Research Center.

https://www.pewresearch.org/short-reads/2023/08/28/most-americans-havent-used-chatgpt-few-think-it-will-have-a-major-impact-on-their-job/.

Fahmi, U., Wibowo, C. & Yudanto, F. (2018). Hybrid Method In Revealing Facts Behind Texts: A Combination Of Text Mining And Qualitative Approach. 132-148. 10.2991/aapa-18.2018.14.

Feinerer, I. (2013). Introduction to the tm Package Text Mining in R. 1-8. Accessible en ligne: http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf.

Feuerriegel, S., Hartmann, J., Janiesch, C., Zschech, P. (2024). Generative AI. Bus Inf Syst Eng 66, 111–126. https://doi.org/10.1007/s12599-023-00834-7.

Fleming, J., & Zegwaard, K. (2018). Methodologies, methods and ethical considerations for conducting research in work-integrated learning. International Journal of Work-Integrated Learning. 19. 205-213.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. International Journal of Computer Applications, 85(17). 42-45.

García-Peñalvo, F., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. 1-10.

Gaur, L., Afaq, A., Arora, G. K., & Khan, N. (2023). Artificial intelligence for carbon emissions using system of systems theory. Ecological Informatics, 102165. 1-11.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. The Qualitative Report. 8. 597-607. 10.46743/2160-3715/2003.1870.

Goodman, E., Kuniavsky, M., & Moed, A. (2012). Surveys. In Observing the User Experience: A Practitioner's Guide to User Research. 2nd ed. 327–383.

Grewal ,A., Kataria, H., Dhawan, I. (2016). Literature search for research planning and identification of research problem. Indian J Anaesth;60(9):635-639. doi: 10.4103/0019-5049.190618. PMID: 27729689; PMCID: PMC5037943.

Grimshaw, J. (2014). SURGE (The SUrvey Reporting GuidelinE) (chapter 20). In Moher, D., Altman, D. G., Schulz, K. F., Simera, I., and Wager, E. (Eds.), Guidelines for Reporting Health Research: A User's Manual. 206-213. Oxford: John Wiley & Sons, Ltd.

Han, H., Zhang, Y., Zhang, J., Yang, J. & Zou, X. (2018) Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. PLoS ONE 13(8): e0202523. 1-11. https://doi.org/10.1371/journal.pone.0202523.

Haradhan, M. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. Published in: Annals of Spiru Haret University , Vol. 17, No. 4(24 December 2017). 56-82.

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. Big Data and Cognitive Computing, 4(1), 1. 1-34.

Hatzivassiloglou, V. & McKeown, K.R. (1997). Predicting the Semantic Orientation of Adjectives. In 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics. 174–181. Association for Computational Linguistics.

Hearst, M. (1992). Direction-based text interpretation as an information access refinement, in Text-Based Intelligent Systems. Lawrence Erlbaum Associates. 257-274.

https://www.ischool.berkeley.edu/research/publications/1992/direction-based-text-interpretation-information-access-refinement.

Hao, K., (2019, June 6). Training a single AI model can emit as much carbon as five cars in their lifetimes. MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

Holtzblatt, K. & Beyer, H. (2015). Contextual Design: Evolved. Synthesis Lectures on Human-Centered Informatics. 7. 11-21. 10.2200/S00597ED1V01Y201409HCI024.

Hölzle, U. (2009, January 11). Powering a Google search. Google. https://googleblog.blogspot.com/2009/01/powering-google-search.html.

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. Journal for Language Technology and Computational Linguistics, 20(1), 19-62.

Howarth, J. (2024, May 3). Most Visited Websites In The World (May 2024). Exploding Topics. https://explodingtopics.com/blog/most-visited-websites.

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. Psi Chi Journal of Psychological Research, 21(3), 138-151.

IPCC. (2014). Annex II: Glossary [Mach, K.J., S. Planton and C. von Stechow (eds.)]. In: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland. 117-130.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. Journal of mixed methods research, 1(2), 112-133.

Jusoh, S., & Alfawareh, H. M. (2012). Techniques, applications and challenging issue in text mining. International Journal of Computer Science Issues (IJCSI), 9(6), 431-436.

Singh, K. (n.d.). Principles of Generative AI A Technical Introduction. Carnegie Mellon University. 1-12. Retrieved March 16, 2024, from https://www.cmu.edu/intelligentbusiness/expertise/genai-principles.pdf.

Kibirige, H. & DePalo, L. (2000). The Internet as a Source of Academic Research Information: Findings of Two Pilot Studies. Information Technology and Libraries. 19. 11-16. 10.6017/ital.v19i1.10069.

Kim, G.-W., Lim, J., Choi, H., & Yun, M.-H. (2012). Adopting Network Analysis Methods for Contextual Inquiry: the Keyword Structure Representation of a Web Behavior. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56(1), 1609-1613. https://doi-org.zorac.aub.aau.dk/10.1177/1071181312561321.

Kip, H. & Beerlage-de J. N. & Wentzel, J. (2018). The contextual inquiry. Routledge, UK, USA. 167-186. 10.4324/9781315385907-8.

Kolesnyk, A.S., & Khairova, N.F. (2022). Justification for the Use of Cohen's Kappa Statistic in Experimental Studies of NLP and Text Mining. Cybern Syst Anal 58, 280–288. https://doi-org.zorac.aub.aau.dk/10.1007/s10559-022-00460-3.

Koot, M. & Wijnhoven, F. (2021). Usage impact on data center electricity needs: A system dynamic forecasting model. Applied Energy, 291, 116798. 1-13.

Kronberg, M. (2020). A Comparison Between Field User Research Methods in a Redesign Context : A Comparison Between Field User Research Methods in a Redesign Context (Dissertation). 3-9.Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-404655.

Kumar, A. & Davenport, T., (2023, July 20). How to Make Generative AI Greener. Harvard Business Review. https://hbr.org/2023/07/how-to-make-generative-ai-greener.

Li X, Liu J, Ni P. (2021) The Impact of the Digital Economy on CO2 Emissions: A Theoretical and Empirical Analysis. Sustainability. 13(13):7267. 1-15. https://doi.org/10.3390/su13137267.

Li, M., Gao, Q. & Yu, T. (2023) Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. BMC Cancer 23, 799. https://doi.org/10.1186/s12885-023-11325-z.

Li, Y., Yang, X., Ran, Q., Wu, H., Irfan, M., Ahmed, M. (2021). Energy structure, digital economy, and carbon emissions: evidence from China. Environ Sci Pollut Res 28, 64606–64629. https://doi.org/10.1007/s11356-021-15304-4.

Lietz, P. (2010). Research into questionnaire design: A summary of the literature. International journal of market research, 52(2), 249-272.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. 5. Sentiment Analysis: A Fascinating Problem. 7-16. 10.2200/S00416ED1V01Y201204HLT016.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. 5. The Problem of Sentiment Analysis. 16-29. 10.2200/S00416ED1V01Y201204HLT016.

Luccioni, A. S., & Hernandez-Garcia, A. (2023). Counting carbon: A survey of factors influencing the emissions of machine learning. 1-19. arXiv preprint arXiv:2302.08476.

Luccioni, A. S., Jernite, Y., & Strubell, E. (2023). Power hungry processing: Watts driving the cost of ai deployment?. 1-20. arXiv preprint arXiv:2311.16863.

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. All Ireland journal of higher education, 9(3). 1-14.

Malmodin, J., and Dag L. (2018). The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015. Sustainability 10, no. 9: 3027. https://doi.org/10.3390/su10093027.

Marr, B., (2023, Jan 24). The Difference Between Generative AI And Traditional AI: An Easy Explanation For Anyone. Forbes. https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/.

Martín, G. Q. & Ortega, A. (2021, May 26). Digitalisation with decarbonisation. Real Instituto elcano Royal Instituto. Retrieved March 14, 2024, from https://media.realinstitutoelcano.org/wp-content/uploads/2022/02/wp8-2021-martin-ortega-digitalisation-with-decarbonisation.pdf.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276–282. http://dx.doi.org/10.11613/BM.2012.031.

McQuate, S. (2023, July 27). Q&A: UW researcher discusses just how much energy ChatGPT uses. University of Washington. https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/.

Miles, M. B., & Huberman, A. M. (1994). Qualitative Data Analysis: An Expanded Sourcebook. Ethical Issues in Analysis. Thousand Oaks, CA: Sage Publications. 288-297.

Mirza, H., Bellalem, F., & Chahrazed, M. (2023). Ethical Considerations in Qualitative Research: Summary Guidelines for Novice Social Science Researchers. 11. 441-449.

Narula D. K. (n.d.). The carbon footprint of online digital performances. The University of Strathclyde business school. Introduction. 2-5. Retrieved March 29, 2024, from https://civicdigits.com/wp-content/uploads/2022/01/The-carbon-footprint-of-online-digital-performances.pdf.

Niklander, S. & Niklander, G. (2017). Combining Sentimental and Content Analysis for Recognizing and Interpreting Human Affects. In: Stephanidis, C. (eds) HCI International 2017 – Posters' Extended Abstracts. HCI 2017. Communications in Computer and Information Science, vol 713. 465–468. Springer, Cham. https://doi.org/10.1007/978-3-319-58750-9_64.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International journal of qualitative methods, 16(1), 1-13. 1609406917733847.

NVIDIA. (n.d.). Generative AI. How Does Generative AI Work? https://www.nvidia.com/en-us/glossary/generative-ai/.

Paré, G., Kitsiou, S. (2017). Chapter 9. Methods for Literature Reviews. In: Lau F, Kuziemsky C, editors. Handbook of eHealth Evaluation: An Evidence-based Approach [Internet]. Victoria (BC): University of Victoria; 2017 Feb 27. 157-181. Available from: https://www.ncbi.nlm.nih.gov/books/NBK481583/.

Patel, P. (2009). Introduction to Quantitative Methods. Empirical Law Seminar. 2-6.

Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M. & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. Computer, 55(7), 18-28.

Popping, R. (2015). Analyzing open-ended questions by means of text analysis procedures. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 128(1), 23-39.

Privitera, M. B. (2015). Contextual inquiry for medical device design. Contextual Inquiry Methods. Elsevier Science & Technology. 47-72. https://ebookcentral.proquest.com/lib/aalborguniv-ebooks/detail.action?docID=2060808#.

Prolific. (2024). About us. Prolific. Retrieved April 28, 2024, from https://www.prolific.com/about.

Rahman, M. (2016). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language "Testing and Assessment" Research: A Literature Review. Journal of Education and Learning. 6. 102-112. 10.5539/jel.v6n1p102.

Ramdhani, A., Ramdhani, M. & Amin, A. (2014). Writing a Literature Review Research Paper: A step-by-step approach. International Journal of Basic and Applied Science. 3. International Journal of Basic and Applied Science, 3 (1). 47-56. ISSN 2301-4458.

Ramer S.L. (2005). Site-ation pearl growing: methods and librarianship history and theory. J Med Libr Assoc. 2005 Jul;93(3):397-400. PMID: 16059431; PMCID: PMC1175807.

Rana, J., Luna, G.P. & Oldroyd, J. (2021). Quantitative Methods. 1-6. 10.1007/978-3-319-31816-5_460-1.

Randolph, J. (2019) "A Guide to Writing the Dissertation Literature Review," Practical Assessment, Research, and Evaluation: Vol. 14, Article 13. DOI: https://doi.org/10.7275/b0az-8t74 Available at: https://scholarworks.umass.edu/pare/vol14/iss1/13.

Rau, J. L. (2004). Searching the literature and selecting the right references. Respir Care. 49(10):1242-1245. PMID: 15447811.

Revelle, W. (2024). Package ‘psychTools’. Tools to Accompany the 'psych' Package for Psychological Research. 1-3. https://cran.r-project.org/web/packages/psychTools/psychTools.pdf.

Rhoades A.E. (2011). Literature Reviews. The Volta Review. 111. 354-369. 10.17955/tvr.111.1.677.

Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019).Chapter 8, Designing a Content Analysis. Analyzing Media Messages: Using Quantitative Content Analysis in Research (4th ed.). Routledge. 148-167. https://doi-org.zorac.aub.aau.dk/10.4324/9780429464287.

Roberts, P., & Priest, H. (2006). Reliability and validity in research. Nursing standard (Royal College of Nursing (Great Britain): 1987), 20(44), 41–45. https://doi.org/10.7748/ns2006.07.20.44.41.c6560.

Roopa, S., & Rani, M. S. (2012). Questionnaire designing for a survey. Journal of Indian Orthodontic Society, 46(4_suppl1), 273-277.

Rose, S., Spinks, N., & Canhoto, A.I. (2014). Applying quantitative and qualitative research designs. Management Research: Applying the Principles (1st ed.). Routledge. 114-141. https://doi.org/10.4324/9781315819198.

Ruel, E., Wagner III, W. E., & Gillespie, B. J. (2015). Chapter 6, Pre-testing and Pilot Testing. The practice of survey research: Theory and applications. Sage Publications. 101-119.

Saleh, Ziyad. (2019). Artificial Intelligence Definition, Ethics and Standards. 1-10. Retrieved March 12, 2024, from https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). Green ai. corr abs/1907.10597 (2019). 1-12.arXiv preprint arXiv:1907.10597.

Sharma, P. & Dash, B. (2022). The Digital Carbon Footprint: Threat to an Environmentally Sustainable Future. 14. 19-29. 10.5121/ijcsit.2022.14302.

Shen, Y., Yang, Z. and Zhang, X. (2023). Impact of digital technology on carbon emissions: Evidence from Chinese cities. Front. Ecol. Evol. 11:1166376. doi: 10.3389/fevo.2023.1166376.

Sipe, T. A. & Stallings, W.M. (1996). Cooper's Taxonomy of Literature Reviews Applied to Meta-Analyses in Educational Achievement. ED398275, 1-22.

Skjuve, M., Følstad, A., & Brandtzaeg, P. B. (2023, July). The user experience of ChatGPT: Findings from a questionnaire study of early users. In Proceedings of the 5th International Conference on Conversational User Interfaces (pp. 1-10).

Srinivasan, R., Lohith, C. P., Srinivasan, R., & Lohith, C. P. (2017). Pilot Study—Assessment of validity and reliability. Strategic marketing and innovation for Indian MSMEs, 43-49.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. ArXiv, abs/1906.02243. 1-6. 10.48550/arXiv.1906.02243

Su, S. (2023, Dec 11). 5 Ways Companies Can Promote More Sustainable AI. Forbes. https://www.forbes.com/sites/zendesk/2023/12/11/5-ways-companies-can-promote-more-sustainable-ai/.

Sukanya, M., & Biruntha, S. (2012). Techniques on text mining. In 2012 ieee international conference on advanced communication control and computing technologies (icaccct). IEEE. 269-271.

Sürücü, L., & Maslakçi, A. (2020). Nicel Araştirmada Geçerlilik Ve Güvenilirlik. Business & Management Studies: An International Journal, 8(3), 2694–2726. https://doi.org/10.15295/bmij.v8i3.1540.

Taherdoost, H. (2016) Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research (August 10, 2016). 28-36. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3205040.

Takyar A. (n.d.). A comprehensive guide to prompt engineering. LeewayHertz. Retrieved May 02, 2024, from https://www.leewayhertz.com/prompt-engineering/.

Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text mining: techniques, applications and issues. International journal of advanced computer science and applications, 7(11), 414-418.

Tan, A. H. (1999). Text mining: The state of the art and the challenges. In Proceedings of the pakdd 1999 workshop on knowledge disocovery from advanced databases. Vol. 8. 65-70.

Taube, A. (2010). Kappa — A Critical Review. Department of Statistics, Uppsala University. Introduction. 1-3. https://www.semanticscholar.org/paper/Kappa-%E2%80%94-A-Critical-Review-Taube/9b63c13f4f1eea48d64d67a26b2956d80c46b8fa.

Taube, A. (2010). Kappa — A Critical Review. Department of Statistics, Uppsala University. The problems of Kappa. 4-10. https://www.semanticscholar.org/paper/Kappa-%E2%80%94-A-Critical-Review-Taube/9b63c13f4f1eea48d64d67a26b2956d80c46b8fa.

The University of New Castle Australia. (n.d.). Learning Development Centre for Teaching & Learning (CTL). Writing a Literature Review. 1-13.

Tilman, M.D. (2016). The book of R : a first course in programming and statistics. Chapter 1: Getting started. San Francisco : No Starch Press, [2016] xxxi, 3-17. QA76.73.R3 D38 2016 ISBN: 97815932765151593276516.

Transforma Insights. (n.d.). AI & Machine Learning. Echo Nomad Limited t/a Transforma Insights. Retrieved May 14, 2024, from https://transformainsights.com/ai-machine-learning.

Turilli, M. & Floridi, L. (2009). The ethics of information transparency. Ethics Inf Technol 11, 105–112. https://doi.org/10.1007/s10676-009-9187-9.

TutorialsPoint. (n.d.). Prompt Engineering. Prompt Engineering – Introduction. 1-9. https://www.tutorialspoint.com/prompt_engineering/prompt_engineering_tutorial.pdf.

Umar, M., Aliyu, M. & Modi, S. (2022). Sentiment Analysis in the Era of Web 2.0: Applications, Implementation Tools and Approaches for the Novice Researcher. Caliphate Journal of Science and Technology. 4. 1-9. 10.4314/cajost.v4i1.1.

Umbach, P. D. (2004). Web surveys: Best practices. New directions for institutional research, 121, 23-38.

University of Guyana. (2022). Ethical Considerations. Polonski Chapter 5. SAGE Publications. 53-75. https://www.studocu.com/row/document/university-of-guyana/field-methods-and-techniques/polonski-chapter-5/41206701.

University of York. (n.d.). Cohen's Kappa. Department of Health Sciences. 1-11. https://www-users.york.ac.uk/~mb55/msc/clinimet/week4/kappa_text.pdf.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & health sciences, 15(3), 398-405.

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. WIREs Data Mining and Knowledge Discovery, 13(4), e1507. 1507. 1-26. https://doi.org/10.1002/widm.

Walkley, S. (2023, September). The Carbon Cost of an Email: Update! The Carbon Literacy Project. https://carbonliteracy.com/the-carbon-cost-of-an-email/.

Wen, J., Zhang, R., Niyato, D., Kang, J., Du, H., Zhang, Y. & Han, Z. (2024). Generative AI for Low-Carbon Artificial Intelligence of Things. 1-10. https://arxiv.org/html/2404.18077v1.

White, M.D., & Marsh, E.E. (2006). Content Analysis: A Flexible Methodology. Library Trends 55(1), 22-45. https://doi.org/10.1353/lib.2006.0053.

Whittemore, R., Chase, S. & Mandle, C. (2001). Validity in Qualitative Research. Qualitative health research. 11. 522-37. 10.1177/104973201129119299.

Wickham, H. (2023). Package ‘tidyverse’. Easily Install and Load the 'Tidyverse'. 1-6. https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf.

Wiebe, J.M., Bruce, R.F. & O’Hara, T.P. (1999). Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics. 246–253. Association for Computational Linguistics.

Ye, Q., Axmed, M., Pryzant, R. & Khani F. (2024). Prompt Engineering a Prompt Engineer. University of Southern California. pp. 1-31. https://doi.org/10.48550/arXiv.2311.05661.

Young, M. (2017). Quality of literature review and discussion of findings in selected papers on integration of ICT in teaching, role of mentors, and teaching science through science, technology, engineering, and mathematics (STEM). Educational Research and Reviews, 12(4), 189-201. 10.5897/ERR2016.3088.

Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. Qualitative Report, 16(3), 730-744.

Zhang, Y., & Wildemuth, B.M. (2005). Qualitative Analysis of Content. 1-12. https://www.semanticscholar.org/paper/Qualitative-Analysis-of-Content-by-Zhang-Wildemuth/b269343ab82ba8b7a343b893815a0bae6472fcca.