TU WIEN Informatics

Dark Patterns in Austrian eCommerce

A Web Crawler-Based Study on Prevalence and Legal Compliance

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Philipp Limbeck, BSc
Matrikelnummer 01429344

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Thomas Grechenig
Mitwirkung: Christoph Wimmer

Wien, 7. Dezember 2023

Unterschrift Verfasser

Unterschrift Betreuung

TU WIEN Informatics

Dark Patterns in Austrian eCommerce

A Web Crawler-Based Study on Prevalence and Legal Compliance

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Philipp Limbeck, BSc
Registration Number 01429344

to the Faculty of Informatics

at the TU Wien

Advisor: Thomas Grechenig
Assistance: Christoph Wimmer

Vienna, 7th December, 2023

Signature Author

Signature Advisor

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Dark Patterns in Austrian eCommerce

A Web Crawler-Based Study on Prevalence and Legal Compliance

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Philipp Limbeck, BSc
Matrikelnummer 01429344

ausgeführt am
Institut für Information Systems Engineering
Forschungsbereich Business Informatics
Forschungsgruppe Industrielle Software
der Fakultät für Informatik der Technischen Universität Wien

Betreuung: Thomas Grechenig

Wien, 7. Dezember 2023

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Philipp Limbeck, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 7. Dezember 2023

Philipp Limbeck

Danksagung

Ein aufrichtiges Dankeschön gilt all den wunderbaren Menschen, die mich in meinem Studium und während dieser prägenden Zeit begleitet und unterstützt haben. Eure Ermutigung, Ratschläge und bedingungslose Unterstützung haben mein akademisches Abenteuer zu einer unvergesslichen Reise gemacht. In tiefer Dankbarkeit möchte ich nun die Namen all jener festhalten, die einen unverwechselbaren Abdruck in meinem Studienweg hinterlassen haben:
Nikolai, Joshua, Barbara, Manuela, Herbert, Julian, Lukas, Max, Martin, Faruk.

Acknowledgements

A sincere thank you goes out to all the wonderful people who have accompanied and supported me during my studies and throughout this formative time. Your encouragement, advice, and unconditional support have turned my academic journey into an unforgettable adventure. In deep gratitude, I now want to record the names of all those who have left an indelible mark on my academic path:
Nikolai, Joshua, Barbara, Manuela, Herbert, Julian, Lukas, Max, Martin, Faruk.

Kurzfassung

Im Forschungsbereich der Human-Computer-Interaction (MCI) wird der Begriff "Dark Pattern"verwendet, um Instanzen zu beschreiben, in denen Designerinnen und Designer ihr Wissen über menschliches Verhalten, Psychologie und die Wünsche der Nutzerinnen und Nutzer einsetzen, um Mechanismen der Täuschung zu implementieren, die nicht im besten Interesse der Nutzer liegen. Allerdings weist die aktuelle Forschung auf einen Mangel an Software-Tools zur automatisierten Erkennung von Dark Patterns zur Sammlung aussagekräftiger Daten hin. Darüber hinaus bleibt die rechtliche und regulative Lage in Österreich im Bezug auf Dark Patterns unklar.

In dieser Arbeit wurde ein Web-Crawler implementiert, der mithilfe von Regular Expressions eCommerce-Websites auf vordefinierte Dark Patterns untersucht und die Ergebnisse in einer Datenbank speichert. Semi-strukturierte Interviews mit Experten wurden durchgeführt, um die rechtliche und politische Perspektive von Dark Patterns zu durchleuchten. Etwa 16% der Top 250 (gemessen am jährlichen Umsatz) eCommerce-Unternehmen in Österreich verwenden derzeit mindestens ein Dark Pattern. Österreichische Gesetze bieten derzeit keinen Schutz vor UI-Designproblemen, es gibt aber Regulierungen für herkömmliche- und Online-Einkäufe, welche teilweise auch Dark Patterns betreffen. Zukünftige Forschung wird voraussichtlich auf die automatisierte Erkennung mittels künstlicher Intelligenz ausgerichtet sein und kann auf den Erkenntnissen dieser Studie aufbauen.

Keywords: *Dark Patterns, Onlinehandel, UI Design, Web-Crawler, Mensch-Computer-Interaktion, Gesetzliche Regulierung*

Abstract

In the research field of Human-Computer Interaction (HCI), the term "dark pattern" is used to describe instances where designers use their knowledge of human behavior, psychology and users' desires to implement deceptive features that are not in the users' best interest. However, current research indicates a lack of software tools to support automated detection of dark patterns to gather insightful data. Additionally, the legal and regulatory framework in Austria concerning dark patterns remains unclear. In this work a web crawler was implemented that uses regular expressions to examine eCommerce websites for predefined dark patterns and stores the results in a database. Semi-structured interviews with experts were conducted to explore the legal and political perspectives on dark patterns. Around 16% of the top 250 eCommerce companies in Austria (measured by annual revenue) currently employ at least one dark pattern. Austrian laws do not currently provide protection against UI design issues, although regulations exist for conventional and online purchases, which partly concern dark patterns. Future research is expected to focus on automated detection using artificial intelligence and could build upon the findings of this study.

Keywords: *Dark Patterns, eCommerce, UI Design, Web-Crawler, Human-Computer-Interaction, Legal Regulation*

Contents

CHAPTER 1

Introduction

In the field of human computer interaction (HCI) the term "dark pattern" is used to define instances where designers use their knowledge of human behavior (e.g. psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest [GKB+18][BLSD23]. Dark patterns commonly occur on digital platforms like social media, mobile apps, video games and eCommerce websites and can have negative effects such as for example frustrating the user (rather harmless), causing financial damage or leading to a loss of control over internet privacy [MAF+19a]. In this paper, a web-crawler based approach is used to investigate and measure the prevalence of dark patterns in Austrian eCommerce. Furthermore, the dark patterns found will be discussed with legal and policy experts in terms of legal compliance and whether the use of a web crawler to detect dark patterns is conceivable from a legal or policy perspective. This chapter provides an overview of the problem statement, the motivation behind the research, and the derived research questions. Additionally, it outlines the anticipated outcomes of the study.

1.1 Problem Statement

According to eMarketer data [es19], global eCommerce sales will reach a total of 6,310 billion US dollars in 2023. In 2019, this sum was still 3,351 billion US dollars, which means that the sales in the eCommerce sector have almost doubled in only 4 years. This rapid growth also illustrates the relevance of eCommerce for national and international companies.
The success factors of the eCommerce sector include the characteristics of the internet, such as the fact that content can be presented in multimedia form (audio, text, video, etc.), interactive access via the internet (in real time), multifunctional communication and, above all, spatial and temporal independence. Information can be transmitted and retrieved worldwide regardless of the physical presence of a communication partner

[Böi13].
With the popularity and success of online shops also comes fraud, cheating and manipulative behavior. One of these manipulative and deceptive tools are dark patterns, which are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions[MAF+19a]. Dark patterns exploit the limited cognitive abilities of humans by leveraging cognitive biases [MKM21]. Cognitive biases are systematic errors in thinking and perception that people make, influencing their decision-making [TK74].
Dark patterns are used in eCommerce to boost sales, increase conversion or any other success metric of eCommerce driven companies by the usage of manipulative and sometimes unethical design. While design is - by definition - a persuasive act and has the potential to manipulate the user, there are occasions where designers may abuse this power [MKM21]. There is only a fine line between subtle encouragement to purchase by design and harmful user manipulation with the usage of dark patterns.
With a rise in popularity of dark patterns there also arise questions about ethics and legal concerns. In particular the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States have brought new opportunities to define unethical or unlawful design decisions [BLSD23]. Recently dark patterns have attracted increasing legislative and regulatory attention in the EU and U.S but policymakers have a hard time creating regulations since dark patterns are not consistent in their definition and under-conceptualized [MKM21]. The current laws regarding the regulation of dark patterns in eCommerce do not yet encompass UI Design[SSS22]. They instead aim to draw comparisons between the traditional (offline) purchasing process and the digital buying process [MDSW21]. However, in the near future, the utilization of dark patterns is expected to be more effectively regulated through new EU regulations such as the Digital Services Act (DSA) and the Digital Markets Act (DMA), as well as other laws like the AI Act, Data Act, and Data Governance Act [Ger22][MDSW21].
So, at a time when eCommerce is becoming increasingly important in society, there is a growing use of dark patterns. There are several prevalent issues with dark patterns: they are not well conceptualized [MKM21], and there is no consistent definition or taxonomy, making it challenging to identify and protect against them[SSS22]. Even experts in the field of UI design struggle to identify dark patterns [HG21]. Furthermore, there are, to the best of our knowledge, currently no studies on the topic of dark patterns in the Austrian eCommerce market or for German-speaking websites or onlineshops in general. Due to this lack of large-scale studies, there is limited data available for further research in areas like the use of machine learning algorithms for automated detection (e.g., in the form of a warning system with a browser extension).
The legal status of dark patterns in eCommerce is currently unclear, and the use of a web crawler for collecting dark pattern data and related metadata has not been explored yet. Consequently, it is not yet known whether a web crawler can be used from a legal or political perspective, or what potential applications of such technology might be.

1.2 Motivation and Research Question

With the growing relevance of eCommerce and the associated online sales, the relevance of sales-promoting techniques and methods has also increased. Dark patterns are often used to optimize sales or to encourage certain actions by users (such as signing up for a subscription model or subscribing to a newsletter for targeted marketing campaigns). Despite the popularity of online stores, the growing eCommerce sector, and the increasing number of dark patterns, to the best of our knowledge, there has been only one study investigating the prevalence of dark patterns in eCommerce [MKM21]. However, this study was limited to English-language online stores and manual classification of HTML segments was performed to identify them as dark patterns. Consumers and online shoppers need to understand the concept of dark patterns in order to avoid becoming victims of such instruments [LS21].
The evolution of research on dark patterns by Kollmer [KE23] showed that initial research on this topic attempted to define and conceptualize dark patterns. So far, however, scholars have not been able to agree on a single definition or taxonomy. Subsequent studies have focused a lot on the issue of privacy on the internet and the loss of privacy control through the use of dark patterns (especially concerning cookie banners and the GDPR) [SSS22] [BEK+16]. In further consequence, research is required to develop software or supporting tools for dark pattern detection [LVBB+22]. Currently, it is also unclear which technologies or tools (such as web crawlers) are useful and operational from a legal or political perspective.
In this study, the current legal situation of dark patterns and the regulation of online stores can be evaluated through the opinion of experts. This can shed light on whether the current laws and measures are sufficient to protect users from dark patterns or not. A study of dark patterns in the eCommerce domain can also be beneficial for corporations. By understanding the impact of these tactics on consumer behavior, companies can rethink their design and marketing strategies and implement more ethical practices. In the long run, this can lead to a more positive perception of the brand and increase consumer trust. The above motivation of this study has led to the following research questions:

- How prevalent are dark patterns among the top Austrian eCommerce onlineshops when identified using a web crawler with regular expression detection?
- How reliable is the automated detection of dark patterns with regard to true-positive and false-positive, using a web crawler with regular expressions?
- According to legal experts, can automated methods for detecting dark patterns be used from a legal or policy standpoint?
- How do legal experts in Austria assess the effectiveness of current laws and regulations in addressing dark patterns in eCommerce, and what recommendations do they propose for enhancement?

Currently, it is unclear which types of dark patterns (defined by established taxonomies) are legally compliant and which of these dark patterns violate laws (at the national level). Previous studies have focused on the legal and ethical aspect of dark patterns, but did not compare them against specific (national) laws [GKB+18]. Furthermore, a web-crawler based approach is used to observe the prevalence and presence of different types of dark patterns in the Austrian eCommerce sector. This study also examines whether the use of a web crawler with automated textual recognition is applicable from a legal or policy perspective, and what the use cases are.
For the investigation of the research question, subpages of a large list of eCommerce websites are checked for the presence of dark patterns using an automated web crawling approach. When examining a website for dark patterns, pages that are directly related to the purchase process are more relevant than other pages, which is why the crawler must also be trained to select pages with more relevance as a data basis. The crawler simulates the browsing behavior of a user to collect relevant URLs which can be used to detect dark patterns using regular expression matching. With the help of these textual recognition methods, it is possible to classify the detected dark patterns using literature grounded taxonomies. Finally, this work will focus on dark patterns from a legal and political perspective. Possible law violations of detected dark patterns are discussed, where Austrian legislative texts will be used to investigate legal violations. Legal and policy experts will discuss the use of web crawler technology and what use cases there are for it.

1.3 Expected Result

The eCommerce sector is growing at a rapid pace and it is hard to imagine people's digital lives without it. The rise in online shop popularity and success brings with it instances of fraud, deception, and manipulative conduct. Prior work in the research field of dark patterns focused on the definition of taxonomies to describe the existing types of dark patterns, but there is to the best of my knowledge currently only one work examining the automated large-scale detection of dark patterns in the eCommerce domain [MAF+19a]. Other works focusing on the automated detection of dark patterns either utilize preprocessed datasets for a classification problem (is it a dark pattern or not) [SSS22], describe tools that could be developed to warn and educate users [RWA+22], exclusively concentrate on recognizing (privacy) dark patterns in cookie banners using machine learning approaches [HG21], or classify dark patterns based on whether they can be detected manually, semi-automatically, or fully automatically [COG+21].
This work will refine and extend the methodology and crawling approach of prior work in order to detect dark patterns on top-selling eCommerce websites that operate in Austria and compare it with national legal restrictions. Using the collected data from the web crawler a quantitative analysis of the data is performed.
In this study a web-crawler will be constructed that helps to analyze the presence of dark patterns by text-based detection mechanisms. Lastly, the dark patterns found and the application of web-crawler technology is examined by a qualitative analysis of

semi-structured interviews with legal experts.

The answering of the research questions will be divided into the following phases:

- literature review about dark pattern definitions, taxonomies, its context in eCommerce and automatic detection
- data selection and corpus creation for web crawler - which eCommerce websites shall be considered for crawling and detection
- crawl onlineshops to find URLs or product pages
- use regular expression matching to detect dark patterns on previously crawled product pages and store them in database
- quantitative analysis of crawled and scraped dark pattern data
- semi-structured expert interviews with legal experts about legal violations on dark patterns found and web-crawler usage from a legal and political viewpoint
- qualitative analysis of interviews

By answering the research question, it shall be evaluated whether a systematic approach can be developed to use a web crawler in the eCommerce domain to automatically detect dark patterns. Furthermore, existing dark patterns will be investigated and pointed out for possible violations of Austrian national law and legislative texts. In a discussion with legal experts, it is then jointly evaluated whether web crawlers can be used to detect violations of the law or whether the use of such technologies makes sense and which areas of application could be considered.
The results of this study could benefit individuals in the eCommerce sector, consumers and companies that offer eCommerce. Both results (prevalence and legal compliance of dark patterns) can serve consumer protection associations and legislators as a basis for decisions on necessary measures to deal with dark patterns. Reviewing the legality of different dark patterns can help companies to make their eCommerce offerings legally compliant. Through studying the prevalence of dark patterns consumers can be educated and consumer awareness can be raised. Furthermore, a study on the prevalence of dark patterns can be relevant for policy stakeholders to show that there is a need for action regarding the regulation of dark patterns in the eCommerce sector. Last and most importantly, the examination of the prevalence of dark patterns and the data generated in this process can provide valuable data for future research. The data can be used to better understand dark patterns in eCommerce (also in German), which helps to conceptualize and define them more accurately. Furthermore, the found dark patterns and the associated meta-data can be analyzed in detail to understand the characteristics of dark patterns or to implement a detection using machine learning algorithms. Previous work in the field of dark pattern detection with machine learning showed first successes

with pre-trained language models [SSS22], but it also showed that the classification of dark patterns on the web is very difficult, because they are hard to capture and often context-dependent [HG21].

1.4 Structure

This chapter offers a concise overview of the thesis's structural organization. The initial section delves into the research problem's foundation and elaborates the thesis's motivation, which reinforces the relevance of the issue at hand. Following that, the chapter outlines the research questions that have emerged and defines the expected results.
Moving on to the second chapter, it comprises a comprehensive review of the fundamental theoretical concepts related to dark patterns. Drawing on recent research, diverse definitions and taxonomies, employed to categorize various types and characteristics of dark patterns, are presented. Subsequently, domain-specific dark patterns within the eCommerce domain are explained, accompanied by an exploration of current research in this domain. Moreover, the second chapter distinguishes between web crawling and web scraping, explicating how web crawling works and why it is increasingly employed in research contexts. Ultimately, the chapter delves into the legal challenges surrounding the regulation of dark patterns and outlines existing measures designed to address this issue.
The third chapter provides an extensive overview of the study's methodological approach and the study design. It explains the utilization of a web crawler-based research method and the execution of semi-structured interviews with legal experts to analyze the legal compliance of individual dark patterns. The research questions derived from this approach are explicated and presented, alongside a comprehensive explanation of the step-by-step process involved in conducting the web crawling investigation.
In the fourth chapter, the definition and taxonomy employed in this study are introduced. It will be explained why this particular definition was chosen and which dark patterns were selected for automated detection based on their textual content. Furthermore, it will be outlined the individual steps of implementation, followed by an examination of the challenges in the programming process.
In the subsequent fifth chapter, the study's results are detailed, encompassing the outcomes of dark pattern detection using the web crawler and the findings from expert interviews regarding the legal compliance of specific dark patterns. The crawler results are assessed quantitatively and exploratively, while the insights gleaned from the semi-structured expert interviews are evaluated qualitatively.
Chapter 6 delves into an in-depth discussion and interpretation of the results. The study's limitations are outlined and conclusions are drawn from the results leading to a prospective outlook for future research.
Finally, in Chapter 7, the thesis culminates with a comprehensive conclusion and interpretation, summarizing the key findings and contributions of the entire study.

CHAPTER 2

Theoretical Foundations and State-of-the-Art

This chapter provides an overview of the different definitions of dark patterns and their underlying taxonomies. The use of dark patterns in eCommerce is also discussed and a brief overview of their relevance is shown. It also presents the current state of research, theoretical models, definitions and taxonomies that are relevant to this study. Subsequently, web crawlers are considered and explained as a scientific method. Finally, dark patterns are examined from a legal point of view.

2.1 Literature Review

To gain a comprehensive understanding of the topic of dark patterns and the current state of research, a semi-systematic literature review was conducted. The chosen methodology involved a semi-systematic literature review following Snyder´s [Sny19] approach. semi-systematic literature reviews are particularly suitable when a topic or research area is observed and conceptualized by different research groups or fields [WGW+13] as it is the case with dark patterns, as they are of interest for behavioural scientists, UI designers and legal scholars. The procedure of a semi-systematic literature review was divided into the following four phases:

1. Phase 1: designing the review
2. Phase 2: conducting the review
3. Phase 3: analysis
4. Phase 4: writing the review

The following paragraphs describe the procedure of the literature search, which serves as the knowledge base for the underlying theory chapter of this thesis.

2.1.1 Phase 1 - designing the review

To design the literature review, basic questions about motivation and need were asked in advance. The literature review was conducted to get an overview of the interdisciplinary field of dark patterns (disciplines: HCI, behavioral sciences, law) and to find scientifically relevant definitions and taxonomies (for categorization) of dark patterns, which can be used in the context of this thesis. The choice of methodology fell on a semi-systematic literature review, as it offers the most benefit for the thesis by performing a systematic approach, but requires less resources than a full systematic literature review [Sny19].

To conduct the search, the following search strategy was developed:

Search Terms	"dark pattern" OR "dark patterns"
Sources and Databases	Google Scholar
Inclusion Criteria	language: English
	publication date after 2010 (after publication of the work of Brignull 2010 that defined the term dark patterns in HCI)
	only reviewed articles
	abstract refers to the definition of the term dark patterns, the development of a taxonomy, dark patterns from a legal or eCommerce perspective
Exclusion Criteria	focus other than definition of term or development of taxonomy/categorization, legal- or eCommerce persepctive (e.g. privacy or video games)
	dark patterns in the sense of illumination/light
	focus on one or some pre-defined dark patterns (e.g. focus on consent banners)

It was decided to use "Google Scholar" as the literature database – the reasons for this choice are explained in 2.1.2, in the description of changes made after the pilot test. The table below illustrates the settings employed in Google Scholar for conducting the search:

Literature Database	Google Scholar
Search Query	"dark pattern" OR "dark patterns" in the title of the article (since Google Scholar does not allow searches in abstracts)
Additional Settings	Custom Date Range: from 2010 Language: English Type: do not include citations or patents
Papers found	N=109
Filter	Order by relevance and select first 100 entries N=100

2.1.2 Phase 2 - conducting the review

Before the actual conduct of the review, a pilot test was carried out to test and refine the previously defined search strategy if necessary.

The decision was made to use "Google Scholar" as the search database instead of the previously designated databases: IEEE, dblp, Springer, and ACM. This choice is attributed to the limited resources for the literature review, which constitute only a small part of this study, and the complexity and diversity involved in an extensive search across different search engines. In some cases, it was not possible to perform the same search queries across different search engines.

Furthermore, a more intricate search query was avoided, as a "title search" was employed, and additional restrictions through a complex search query would overly limit the results. The inclusion criteria were adjusted to include not only articles that define dark patterns or create a taxonomy but also articles that examine dark patterns in the domains of eCommerce or law.

During the execution, the articles were classified into the following four categories. These categories were used to describe the focus of the work, aiming to simplify the effort in synthesizing the articles retrospectively:

- Category A: Definitions, taxonomies, or literature reviews (related to dark patterns)
- Category B: Dark patterns in eCommerce
- Category C: Dark patterns in the legal context (laws)
- Category D: Irrelevant for the study (excluded through manual review)

The semi-systematic literature review was conducted in four iterations. After each iteration, articles deemed non-relevant (Category D articles) were removed:

1. Selection of the 100 most relevant articles based on the search query (N=100)
2. Categorization into Categories A-D based on the title (N=48)
3. Categorization into Categories A-D based on the abstract and conclusion (N=23)
4. Reading the full text of the selected articles

Finally, 8 articles from category A (definition or taxonomy), 7 articles from category B (eCommerce dark patterns) and 8 articles from category C (dark patterns from a legal perspective) were selected for further analysis. In the course of the study, further insights were gained continuously and knowledge about the multi-layered nature of the phenomenon of dark patterns was acquired, which is why these 23 articles are not considered final, but merely served as a basis for a basic understanding of the subject area.

The literature utilized in this study extends beyond the articles selected in the literature review. As the study progressed, new insights were gained, prompting additional research in various directions. For instance, an exploration into the realm of cognitive biases from a behavioral psychology perspective was undertaken, as well as an investigation into mechanisms for detecting dark patterns.

2.1.3 Phase 3 - analysis

This phase consisted mainly of reading and understanding the previously selected articles and writing notes or summaries to synthesize the knowledge. Particular attention was paid to the individual definitions of dark patterns and also which studies are based on which definitions or even mention or cite them (which indicates an explicit relevance of these definitions). Focusing on taxonomies, attention was paid to the hierarchical depth of the taxonomies (are there superordinate categories of the individual types?) and how high the number of categories was. For the eCommerce dark pattern articles, attention was paid to how the framework of the studies was chosen and which market was considered (e.g. Alexa top pages or online stores of a certain region). For the legal articles, the specific laws mentioned were identified, where they apply, and how do these legal regulations impact the use of dark patterns.

2.1.4 Phase 4 - writing the review

The review itself, or the accumulated knowledge from this semi-systematic literature review, is presented in the following chapters.

2.2 Dark Patterns

Dark patterns can be understood as certain types of "design patterns". Design patterns in themselves can in turn be defined as a methodology for facilitating the reusability of good design practices [Mik98]. Design patterns thus provide templates for commonly encountered design tasks. Designers utilize these patterns to develop interfaces and interactive elements that are easily comprehensible to users. In the field of human computer interaction (HCI) the term “dark pattern” is used to define instances where designers use their knowledge of human behaviour (e.g. psychology) and the desires of end users to implement deceptive functionality that is not in the user’s best interest [GKB+18] [BLSD23]. Dark patterns are frequently encountered on digital platforms such as social media, mobile apps, video games, and eCommerce websites. These manipulative design tactics can result in adverse outcomes, ranging from minor user frustration to more severe consequences like financial harm or compromised internet privacy. [MAF+19a].

2.2.1 Definitions

Historically, misleading and manipulative design is probably much older than it is considered relevant by science. With digitization, design patterns emerged that were used

for recurring design problems. Design patterns are general, reusable solutions to help designers create efficient, usable, and aesthetically pleasing user interfaces. Dark patterns are certain types of design patterns that aim to trick the user into performing an action such as signing up for a newsletter, taking out a subscription, or agreeing to store cookies by using manipulative or misleading design. Thus, the original intention of design patterns was to improve the user experience and support user goals, whereas dark patterns influence users in undesirable ways to benefit digital system operators.
In order to better understand and conceptualize these particular design patterns, researchers in the discipline of Human Computer Interaction (HCI) have been working on definitions for dark patterns. With the help of these definitions, information about the properties and characteristics of dark patterns in general and their meaning should be provided. Recent research[MKM21] has shown that the concept of dark patterns is under-conceptualised, making it difficult for legislators and consumer advocates to establish rules and guidelines for dealing with dark patterns. Priority questions of current research are: what makes a user interface dark and why are some designs problematic for our society [MKM21].
Nevertheless, there are various definitions and taxonomies in the research literature that allow dark patterns to be defined and categorised in general or domain-specifically.

In 1998, Fogg [Fog98] conducted a significant study on "Persuasive Technology," exploring the influence of computer systems on human behavior. The paper also referenced design examples from the 1970s that aimed to persuade young individuals to adopt healthy lifestyles and enhance workplace productivity. It is stated that computers now functioned as "persuaders" much like teachers, doctors, therapists, and more. As the internet expanded, the use of persuasive technologies became more prevalent, and Fogg [Fog98] recognized the immense potential of these technologies, which, unfortunately, could be subject to abuse.
Fogg [Fog98] introduced the term "captology," short for "computers as persuasive technologies," defining it as the design, research, and analysis of interactive computing products intended to alter people's attitudes or behaviors. This concept laid the foundation for subsequent research on dark patterns, while also highlighting crucial aspects such as the enormous potential of persuasive technologies, the distinction between ethical and unethical design decisions, the importance of trust in digital systems, and the essential frameworks required to achieve effective persuasion.
Fogg's [Fog98] work serves as a fundamental stepping stone, shedding light on the ethical considerations surrounding the design and implementation of persuasive technologies, thus contributing to the ongoing discourse on dark patterns and the responsible use of technology to influence user behavior.

A pioneer in the HCI research field was Harry Bringnull [BLSD23], a User Experience (UX) designer and researcher who coined the term "dark pattern" in 2010 with the following definition: **"dark patterns are used to define instances where designers**

use their knowledge of human behaviour (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest". This definition was the first to mention that operators of digital systems deliberately use misleading design decisions to influence users in the operators' favor. Brignull's [BLSD23] original intention was to create a website, where dark patterns are defined and categorized in order to raise awareness and shame companies that use interface elements and unethical approaches to manipulate their users and buyers. Under the web link https://www.deceptive.design/ [BLSD23], the community itself could then upload instances of dark datterns (in the form of screenshots and a textual description) to publicly denounce and draw attention to companies that use manipulative design. With the launch of this website, Brignull created a platform to document and publicly denounce dark patterns. The site became an archive of documented types of dark patterns that were actually used in productive systems by companies (some of them international). The point behind the launch of the platform was relatively simple: companies that use such business practices should be ashamed of themselves and users should be made aware that such patterns exist in order to resist them. Figure 2.1 shows a screenshot of the website https://deceptive.design [BLSD23], where the community can upload found dark patterns, legal sayings, or articles about deceptive design.

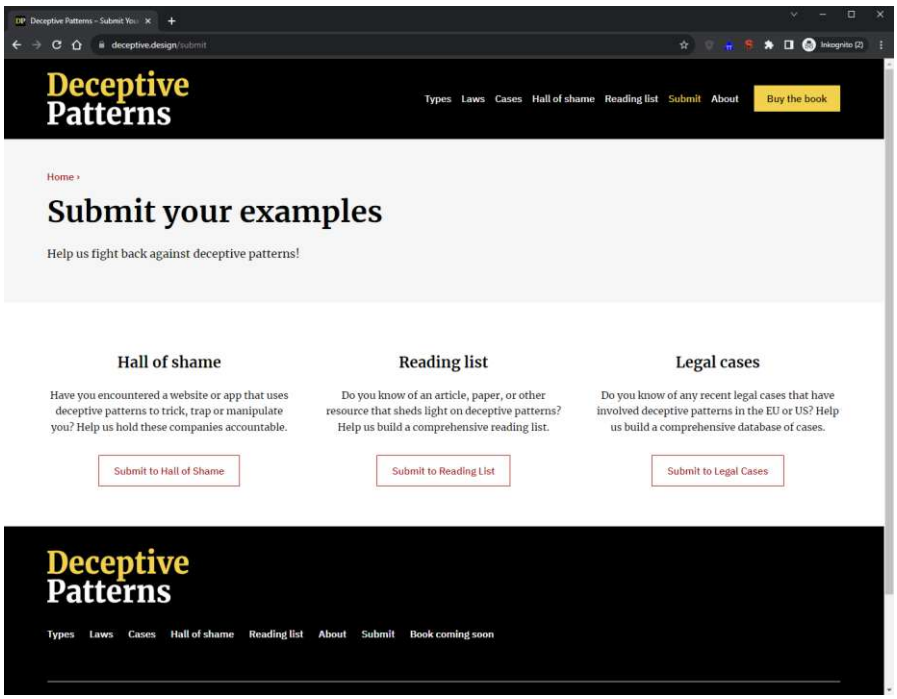

Figure 2.1: Screenshot of the website "www.deceptive.design" (formerly known as "www.darkpatterns.org") where the community and upload instances of dark patterns, relevant articles or legal cases concerning deceptive design

The naming "dark patterns" gave the research field a boost and many researchers from different disciplines showed interest. The interdisciplinary field of dark patterns was not only concerned with Human Computer Interaction (HCI), but also with human psychology, behavioral sciences, and subsequently with law. Thus, economists and behavioral scientists, such as Thaler [TSB13] in 2013, also became aware of the topic and published a paper on choice architecture. The key message of this work was that the way decisions are presented has a major impact on people's choices and how they behave when making decisions. Thus, to guide a decision, the decision environment can be changed so that with clever design of decision options, certain options are highlighted while other options are relegated to the background. With these changes, decisions can be steered in a certain direction by means of design. In Figure 2.2, an example of choice architecture is presented. The choice architecture is designed in such a way that the two choices presented are not equally preferred by users. The first choice is presented in a more prominent location whereas second choice is presented in a less prominent location. This design is likely to lead to more users choosing the first choice, even if the two choices are of equal value and even though the original intention of users is to cancel a subscription (which is represented in the second option).

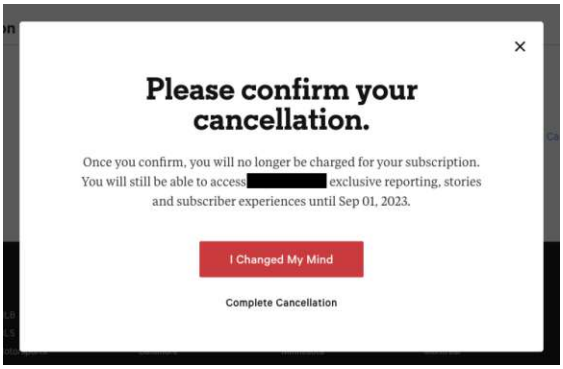

Figure 2.2: Example of a manipulated choice architecture of the subscription-based website. When canceling a subscription, the option to not cancel the subscription is highlighted.

Around the time of Brignull's [BLSD23] publication in 2010, Conti and Sobiesk [CS10] also released their work titled "Malicious Interface Design: Exploiting the User," a study with a similar research focus, examining intentionally crafted user interfaces meant to harm and deceive users. While this work does not explicitly mention the term "dark patterns," the concept of "malicious interface design" largely aligns with the definition of dark patterns during that period. The study explored the impact on users, revealing that users weigh their own goals and values against the perceived "pain" inflicted by using such services through a cost-benefit analysis. The tolerance or pain threshold varies depending on the provider and domain. For instance, test subjects exhibited less tolerance towards

misleading interface elements when using search engines compared to adult content or shopping websites.
After the first general mentions of misleading design and dark patterns in particular, work on specific domains also followed, such as the work of Zagal [ZBL13], which dealt with the domain of video games. Similar findings were made as in the more general work of Brignull [BLSD23] or Conti [CS10], namely that game makers use their knowledge of design and the human psyche, to create interfaces that lead to actions which are against the interests of the players/users. Literally, the following definition was formulated: **"A dark game design pattern is a pattern used intentionally by a game creator to cause negative experiences for players which are against their best interests and likely to happen without their consent"** [ZBL13]. This definition thus refers specifically to the video game industry and is quite fuzzy, as the phrases "negative experience" and "likely to happen without consent" introduce a certain vagueness.

In their study, Bösch et al. [BEK+16] focused on the formulation of a comprehensive definition and taxonomy for privacy strategies and privacy dark patterns. Despite the well-established awareness of these tactics and their evident harmful effects, the researchers explored the reasons behind their persistent prevalence in today's world.
Bösch [BEK+16] described two distinct thinking processes using Kahneman´s [TK74] Dual process theory, illustrated through the example of deciding whether to agree to terms and conditions. The first is the System 1 thinking process, characterized by quick, intuitive, and automatic decision-making with minimal effort. In contrast, the second is the System 2 thinking process, which involves taking time to read and analyze the terms and conditions, considering their pros and cons or employing a cost-benefit analysis before making a deliberate and controlled decision on consent. While System 2 decisions are more time-consuming, they are well-thought-out and intentional. The graphic 2.3 below shows the characteristics of the two thinking processes System1 Thinking and System 2 Thinking from the perspective of designers.

The researchers pointed out that the System 1 thinking process provides opportunities for "darkness" to flourish. Privacy dark patterns and strategies find an effective environment in System 1, as users tend to react impulsively and are more susceptible to manipulative tactics. Consequently, this sheds light on the persistence of privacy dark patterns despite their exposure.
People tend to use a System 1 thinking process for two main reasons: (1) they lack the motivation to think and reason in an elaborate way, or (2) they have no opportunity to do so because they lack the knowledge, ability, or time [KCK16]. Further, it is noted that an important basic human need, namely the "need to belong to significant others" often runs counter to high privacy standards and that people care more about meeting basic needs than they do about privacy [Hig97]. Similar to Brignull [BLSD23], Bösch [BEK+16] also launched website with examples of privacy dark patterns and their explanation. The URL of the site is: https://dark.privacypatterns.eu/.

System 1	System 2
Does not require working memory	Requires working memory
Autonomous	Cognitive decoupling; mental simulation
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Biased responses	Normative responses
Contextualised	Abstract
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability

Figure 2.3: Characteristics of the two thought processes of the Dual Process Theory by Kahnemann and Tversky [TK74] illustrated in a paper by Kannengiesser [KG19].

Despite initial efforts to define dark patterns in general and at the domain level, no agreement or unified definition was reached. Subsequent work continued to sharpen and refine the term dark patterns. In 2018, Gray et. al [GKB+18] released a paper on the limits of dark patterns and expanded on Brignull's [BLSD23] definition, claiming that **"user value is being supplanted in favor of shareholder value"**. The initial 12 types/instances created by Brignull [BLSD23] were coded, resulting in five designer strategies. These strategies are not meant to describe the patterns per se, but rather to describe the motivation of the designers who created and used these patterns. These 5 categories are Nagging, Obstruction, Sneaking, Interface Interference, and Forced Action. They then looked at the categories and individual patterns from a user experience (UX) perspective and found that some patterns perform well on usability testing, but do so at the expense of user choice. Through their work, Gray et. al [GKB+18] showed that user interface elements can unbalance user and stakeholder value and what basic strategies designers use to get users to make certain decisions.

The study titled "Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites" conducted by Mathur et al. [MAF+19a] was published in 2019 and played a crucial role in laying the foundation for subsequent research in this area. The researchers employed an automated approach specifically targeted at examining eCommerce websites to identify the presence of dark patterns. This novel methodological approach proved to be valuable in shedding light on deceptive practices in the eCommerce domain.
During their investigation, the Princeton University researchers devised a taxonomy for eCommerce dark patterns, classifying them based on specific characteristics. Utilizing an automated crawler, they meticulously analyzed approximately 53,000 product pages from around 11,000 shopping websites and online stores. Through this process, they

discovered a total of 1,181 instances where dark patterns were deployed.
Their findings revealed that 183 of the examined websites employed at least one form of dark pattern. Additionally, they identified 22 third-party platforms, such as consent management systems or store systems, and digital systems that offered dark patterns to eCommerce websites.
To promote transparency and reproducibility, the researchers made their software source code and result data publicly available. As a result of their clear and valuable contributions, the study became a fundamental reference for numerous subsequent studies in this evolving field of research.

In their research, Mathur et al. presented their own comprehensive definition of dark patterns as follows: **"Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions."** [MAF+19a].

This definition aligns well with the existing understanding of dark patterns, as it encompasses several characteristics previously defined in the literature.
By describing dark patterns as "user interface design choices" the definition emphasizes the intent of user interface designers and online service operators in employing these tactics. It recognizes that dark patterns are deliberate choices made to influence user behavior.
The phrase "benefit an online service" points to the underlying motive behind implementing dark patterns, highlighting that these tactics are utilized to favor the interests of the online service provider over those of the users. This concept resonates with the notion of an imbalance between users and stakeholders, as described by Gray [GKB+18].
The reference to "unintended decisions" in the definition can be linked to the System 1 thinking processes, as explained by Bösch [BEK+16], which builds upon Kahneman's dual process theory [TK74]. System 1 thinking involves quick and intuitive decision-making, which can make users more susceptible to falling for dark patterns and making unintended choices without fully considering the consequences.
By incorporating these elements, Mathur et al.'s [MAF+19a] definition of dark patterns enriches our understanding of these manipulative design practices and establishes connections to previously established concepts in the field.

In alignment with Mathur et al.'s previous research [MAF+19a], our thesis adopts the identical definition of dark patterns, namely:
„Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions.“[MAF+19a]

In 2020, legal scholars Luguri and Strahilevitz [LS21] conducted a significant study titled "Shining a Light on Dark Patterns," which delved into the impact of dark patterns

through two large-scale experiments. Alongside these experiments, they performed a comprehensive literature review, aiming to categorize existing taxonomies related to dark patterns and examining their implications from a legal perspective, particularly considering the prevailing U.S. laws in 2020.
Leveraging their legal expertise, the researchers highlighted the actions taken by the Federal Trade Commission (FTC), the U.S. consumer protection agency, against dark patterns in court. They identified specific types of dark patterns that were already prohibited by the FTC, indicating an ongoing effort to address these deceptive practices in the digital realm.
Within their experiments, Luguri et al. [LS21] demonstrated that subtle dark patterns were potentially more dangerous as they resulted in lower perceived annoyance, making the manipulative tactics better concealed from users. This finding underscored the nature of subtle dark patterns, as they could influence users without raising immediate suspicions.
Furthermore, the researchers uncovered a noteworthy correlation between education level and susceptibility to dark patterns. Their experiments indicated that individuals with lower levels of education were more likely to be vulnerable to the effects of dark patterns compared to their more educated counterparts.
By presenting these empirical findings and contextualizing dark patterns within the legal landscape, Luguri and Strahilevitz's [LS21] work significantly contributed to the understanding of the impact and implications of dark patterns. Their study shed light on the need for increased awareness, regulation, and protection against these manipulative design tactics to safeguard consumers' interests and ensure fair practices in the digital domain.

In 2020 the article titled "Cognitive Biases, Dark Patterns, and the 'Privacy Paradox'" released by Ari Ezra Waldmann [Wal20], a law professor based in California, U.S., explored the influence of cognitive biases on people's perceived indifference to online privacy. The study also investigates how current laws can be exploited through manipulative techniques, particularly dark patterns. Drawing upon the definition established by Mathur [MAF+19a], the research primarily aims to explain the workings of cognitive biases and their exploitation in conjunction with dark patterns.
Cognitive biases, as described by Kahneman and Tversky [TK74], refer to systematic thinking patterns and decision-making that lead to deviations from normative or rational choices. Some well-known cognitive biases are the confirmation bias and default bias [HK99] [Wei20].
The confirmation bias manifests as the tendency for individuals to seek affirmation for their pre-existing beliefs. On the other hand, the default bias reflects the inclination of people to accept suggested possibilities readily.
Waldmann's [Wal20] work strongly suggests that the traditional rational choice model is outdated and inadequate, as humans do not consistently engage in purely rational thinking. The rational choice theory posits that individuals, within their circumstances, act rationally to maximize utility based on cost-benefit analysis, defining the concept of

"homo economicus" [Wei20].
The article sheds light on the phenomenon of people seemingly indifferent to their online privacy and attributes this behavior to cognitive biases. Additionally, it explores how dark patterns leverage these behavioral anomalies, such as the incorrect assessment of probabilities, to motivate or coerce decision-making. By exploring the interplay of cognitive biases and dark patterns, Waldmann's [Wal20] research provides valuable insights into the factors that influence decision-making in the digital realm. It urges a re-evaluation of traditional rational choice models and underscores the significance of understanding and addressing cognitive biases and their implications on online privacy and user behavior.

Mathur et al. [MKM21] noted possible key differences between dark patterns and traditional marketing, including low cost, large scale, and unprecedented sophistication. In their study "What Makes a Dark Pattern... Dark?", Mathur et al. [MKM21] classified dark pattern definitions of academic publications and governmental material by four main facets: characteristics of user interface (e.g. misleading, coercive), mechanism of effect on user (e.g. confuse, attack, exploit), role of user interface designers (abuse of designer knowledge or designer intent) and benefits and harms (benefit of service or harm to users). The findings of this categorization into facets showed there is significant variation between the facets and also within each facet, which shows the challenges of creating one final definition for dark patterns. The results of the study showed that nine dark pattern definitions do not involve any characteristic of the user interface, four definitions do not specify a mechanism of effect on users, eight definitions do not address the role of user interface designers, and ten definitions do not involve benefit or harm elements [MKM21]. The results of this study thus clearly show the inconsistencies between the individual definitions of different researchers. There are many reasons why these inconsistencies exist, such as the fact that dark patterns affect many fields of research (behavioural sciences, HCI, law, etc.) and thus the researchers' focus is on different topics.

Numerous research papers have explored the concept of dark patterns and emphasized the potential harm they can cause. They have urged the Human-Computer Interaction (HCI) community to develop tools, metrics, and measurement instruments to detect manipulative interface elements. Additionally, these studies have called for regulatory agencies to take action and establish concrete restrictions on the use of dark patterns.
Responding to these calls, the EU conducted a study titled "Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation" [LVBB+22]. This study revealed that dark patterns are widespread in the EU, with their prevalence varying across different domains. For instance, eCommerce websites commonly employ tactics like countdown timers and limited-time messages whereas fitness websites/apps more often use nagging patterns.
In conjunction with the study, researchers also conducted neurophysiological measurements to observe how participants reacted to frustrating dark patterns. The results indicated that encountering such patterns led to heightened heart rates, signaling in-

creased anxiety and alertness among the participants. Furthermore, this study also showed that the current EU measures only partially protect against dark patterns. In particular, the study focuses on the EU's Unfair Commercial Practices Directive (UCPD), which (as described in more detail in section 2.4) specifically prohibits some dark patterns on the basis of a list of unfair business practices.
The EU study by Lupinanez et al. uses the following definition of dark patterns: **"dark patterns" is a concept that is generally used to refer to practices in digital interfaces that steer, deceive, coerce, or manipulate consumers into making choices that often are not in their best interests** [LVBB+22]. This definition is very similar to the one by Mathur [MAF+19a], but the aspect that the operators of online services profit from the use of dark patterns was left out. The "potentially harmful decision outcome" has also been omitted.

As it turns out, there are many different approaches to defining dark patterns, but there is still no consistent definition, as there are several points of criticism of the previous definitions, such as the definition of what makes a design pattern "dark", how to recognise whether a dark pattern has been crafted intentionally or when the design is misused for one's own advantage. Abusiveness also does not automatically imply intentional action or even an intention to harm, although this is required by some definitions. Dark patterns can arise unintentionally due to reasons such as poor design or misinterpretation of the context. Poor design occurs when the user interface is not well-designed and does not clearly communicate the intended actions to the user. Misinterpretation of the context occurs when the user interface is designed for a specific user group, but is misinterpreted by another user group[GKB+18].

All the above definitions aim to define dark patterns in general. However, there are also academic works that define dark patterns in a specific context or domain. For example, there are several definitions of privacy dark patterns, as this topic area has received increased attention since the introduction of the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). However, there are also definitions for the domains of proxemic interaction, online gaming or eCommerce. Given that this research pertains to the field of eCommerce, chapter 2.2.3 contains in-depth elucidations of the definitions and classifications of dark patterns, particularly those relevant to the eCommerce domain.

2.2.2 Taxonomies

A taxonomy is a classification method which organizes things or concepts into predefined hierarchical structure. In this thesis, a taxonomy is used to categorize the different types of dark patterns in order to investigate their prevalence in the eCommerce sector and to examine the individual categories for legal compliance. Ever since the first definition of dark patterns by Brignull 2010 [BLSD23], several taxonomies have been created for different domains. This chapter provides a brief overview of existing academic work on taxonomies in the field of dark patterns. Afterwards a taxonomy is chosen which is

used for the categorization of dark patterns in this thesis. As described in the previous chapter, the term "dark patterns" was coined by Harry Brignull in 2010 [BLSD23]. With the help of this definition, a first general (i.e. not domain-specific) taxonomy was created to classify the different types of dark patterns. This taxonomy explained 12 different types of dark patterns, which was used and extended by many subsequent academic works. These 12 different types were defined as follows [BLSD23]:

1. Trick questions – questions that tricks you into giving an answer you didn't intend
2. Sneak into Basket – in eCommerce when you attempt to purchase something, but the site sneaks an additional item into your basket
3. Roach Motel – you get into a situation very easily, but it is hard to get out of it (e.g. a premium subscription)
4. Privacy Zuckering – you are tricked into publicly sharing more information about yourself than you really intended to (named after Meta CEO Mark Zuckerberg)
5. Price Comparison Prevention – the retailer makes it hard for you to compare the price of an item with another item, so you cannot make a well-informed decision
6. Misdirection – the design focuses your attention on one thing in order to distract your attention from another
7. Hidden Costs – in the last step of the checkout process you discover some unexpected charges have appeared, e.g. delivery charges, taxes, etc.
8. Bait and Switch – you set out to do one thing, but a different, undesirable thing happens instead
9. Confirmshaming – the act of guilting the user into opting into something (e.g. the word "decline" is framed as "I do not want this offer and continue to pay for my outdated tariff"
10. Disguised Ads - adverts that are disguised as other kinds of content or navigation, in order to get you to click on them
11. Forced Continuity – after your free trial ends your preferred payment method silently starts getting charged without any warning
12. Friend Spam – you are asked for your email or social media permissions under the pretense it will be used for a desirable outcome (e.g. finding friends), but then spams all your contacts in a message that claims to be from you

In 2018, Gray et al. [GKB+18] published a paper on the ethics of dark patterns and how they are used to supplant user value in favor of shareholder value. They collected "artifacts" (i.e. design instances) that other designers declared as "dark patterns",

"manipulative designs" or synonyms for these and used constant comparison method and document analysis to refine and adjust Brignulls original taxonomy. Further, they used an open coding approach to define a taxonomy based on the motivation that may have shaped the designer´s use of the pattern, splitting them into 5 primary categories/strategies [GKB+18]. This newly created taxonomy thus aims to categorize the individual dark patterns according to the designers' strategic motivators. These 5 new strategic motivators were used as the main categories used to classify the previously defined 12 original dark patterns [MAF+19a]. Figure 2.4 illustrates the 5 strategies and the dark patterns that belong to these categories.

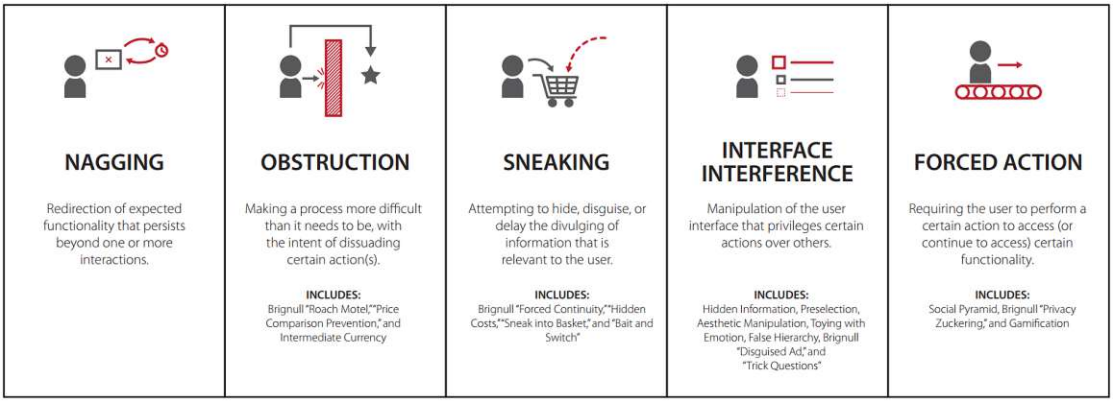

Figure 2.4: Summary of dark pattern strategies of Gray et al. where the focus of the taxonomy is based on the motivation that may have shaped the designer´s use of the pattern.

The results of Gray et al. [GKB+18] have shown that not all interactions that take on these strategies are necessarily equally "dark" in terms of design intent and motivation, they do have the potential to produce poor user outcomes, or force users to interact in ways that are out of alignment with their goals.

Gunawan et al. [GPC+21] studied the relationship between technology platforms and dark patterns in order to investigate the variety of dark pattern depending on the modality of the platform (mobile app, mobile browser, desktop web browser). Observing 105 popular online services they found 50 instances of dark patterns that could be categorized into 9 types (and 37 sub-types) of dark patterns of which 12 sub-types have not yet been documented. They further observed that the use of these dark patterns is highly inconsistent, as dark patterns were more often used in mobile modalities but they also differed in count, type and other traits [GPC+21]. They created their own codebook for the classification of dark patterns and in contrast to other studies they did not include the designer's intention, as they believe it to be highly subjective. Their study showed that users who only use one modality perceive the darkness of the service differently. The use of only one modality with comparatively few dark patterns can lead to deceptive

perceptions about whether a service is harmless or not.

Another taxonomy was developed by Luguri et al. [LS21] by conducting a literature review and aggregating or combining existing taxonomies. In their work on the effectiveness of dark patterns (i.e., a quantified measure of effectiveness), a custom aggregated taxonomy with 8 categories and 27 variants was created.

Mathur et al. [MKM21] addressed the issue of what makes a user interface a "dark pattern" and why some designs are problematic for users and even society. Their work compares many published definitions and taxonomies of dark patterns and illustrates that there are large inconsistencies between each definition and taxonomy. Different taxonomies are also contrasted, such as the taxonomy by Harry Brignull [BLSD23] mentioned above with 12 types of dark patterns or Conti and Sobiesk [CS10], who documented 11 problematic types of user interfaces (not named dark patterns in their work). In the video games domain Zagal et al. [ZBL13] used a domain to categorize seven types of dark patterns that are commonly used by video game designers. Greenberg et al. [GBVD14] investigated the topic of dark patterns from a rather unconventional direction, namely how designers can leverage their knowledge of proxemic interactions and exploit this knowledge by using dark patterns. Greenberg [GBVD14] considered systems that account for physical space and described a taxonomy of eight “proxemic” (i.e., physical proximity) interaction dark patterns. In recent years, some dark pattern taxonomies for the privacy and data security domain also appeared like Bösch et al. [BEK+16] who introduced a set of 7 privacy dark patterns, the NCC report that described 5 privacy dark patterns and the CNIL report [oIC19] where 18 distinct types of dark patterns are described.

There has been a substantial body of academic research on the different types and taxonomies of dark patterns. However, no single taxonomy has yet achieved widespread adoption.

All of the above taxonomies have either been defined in general terms (i.e. without being dedicated to a specific domain) or infer the generality from a specific domain. In the following paragraphs, taxonomies are presented that have been defined specifically for the eCommerce domain.

With their work published in 2019, Mathur et al. [MAF+19a] have done pioneering research on dark patterns in the eCommerce domain. In their work they present automated techniques that enable experts to identify dark patterns on a large set of websites using a web crawling approach. With their findings of the observation of 11.000 shopping websites they developed a dark pattern eCommerce taxonomy and uncovered 22 third-party entities (such as consent management platforms, website plugins/addons, etc.), that offer dark patterns.

The term dark patterns for the domain of eCommerce was defined as follows: Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of

selecting alternatives, they might not make [MAF+19a]. Their taxonomy consists of five dimensions, namely:

1. Asymmetric - Does the user interface design impose unequal weights or burdens on the available choices presented to the user in the interface?
2. Covert - Is the effect of the user interface design choice hidden from users or does the interface design steer users into making specific purchases without their knowledge?
3. Deceptive - Does the user interface design induce false beliefs either through affirmative misstatements, misleading statements, or omissions?
4. Hidden Information - Does the user interface obscure or delay the presentation of necessary information to the user?
5. Restrictive - Does the user interface restrict the set of choices available to users?

Based on these characteristics they created a taxonomy of 15 types of dark patterns contained in 7 broader categories. For each of the defined dark patterns, the main category to which it belongs, a brief description of how often this particular pattern was found in the study, the characterisation of the pattern and which cognitive bias is triggered by the pattern are given. Graphic 2.5 shows a screenshot of a visualization of Mathur et al.'s taxonomy of eCommerce dark patterns. In the graphic, the individual dark patterns are divided into superordinate categories and described. Furthermore, it shows how many instances of these dark patterns were found during the study and which of the previously described characteristics these dark patterns exhibit. Finally, cognitive biases of the dark patterns are listed.

Many following works have used the above taxonomies as a basis for further research or have tried to develop their own taxonomies.

Neem et al. [NL21] have created their own taxonomy based on the work of Brignull [BLSD23], Gray [GKB+18] and Mathur [MAF+19a]. Their taxonomy was chosen because the definitions of Brignull [BLSD23] and Gray et al. [GKB+18] form the basis for most academic work to date. Thus, it serves as an appropriate template to make the present study comparable with others. In addition, definitions from Mathur et al. [MAF+19a] were added as this work mainly covers the area of this study, namely eCommerce.

In order to make the results of this thesis comparable with other studies in the field of eCommerce and dark patterns, the dark patterns found are classified according to the taxonomy of Mathur et al. [MAF+19a]. In this way, the results can be made comparable with the directly preceding study by Mathur, but also with other studies that deal with the matter of dark patterns in eCommerce.

A study by the European Union developed its own behavioural taxonomy, which was used to categorize manipulative digital business practices in its own study [LVBB+22]. The

Table 1. Categories and types of dark patterns along with their description, prevalence, and definitions. Legend: ●= Always, ◐= Sometimes, ○= Never

Category	Type	Description	# Instances	# Websites	Asymmetric?	Covert?	Deceptive?	Hides Info?	Restrictive?	Cognitive Biases
Sneaking	Sneak into Basket	Adding additional products to users' shopping carts without their consent	7	7	○	○	◐	●	○	Default Effect
	Hidden Costs	Revealing previously undisclosed charges to users right before they make a purchase	5	5	○	○	◐	●	○	Sunk Cost Fallacy
	Hidden Subscription	Charging users a recurring fee under the pretense of a one-time fee or a free trial	14	13	○	○	◐	●	○	None
Urgency	Countdown Timer	Indicating to users that a deal or discount will expire using a counting-down timer	393	361	○	◐	◐	○	○	Scarcity Bias
	Limited-time Message	Indicating to users that a deal or sale will expire will expire soon without specifying a deadline	88	84	○	◐	○	●	○	Scarcity Bias
Misdirection	Confirmshaming	Using language and emotion (shame) to steer users away from making a certain choice	169	164	●	○	○	○	○	Framing Effect
	Visual Interference	Using style and visual presentation to steer users to or away from certain choices	25	24	◐	●	◐	○	○	Anchoring & Framing Effect
	Trick Questions	Using confusing language to steer users into making certain choices	9	9	●	●	○	○	○	Default & Framing Effect
	Pressured Selling	Pre-selecting more expensive variations of a product, or pressuring the user to accept the more expensive variations of a product and related products	67	62	◐	◐	○	○	○	Anchoring & Default Effect, Scarcity Bias
Social Proof	Activity Message	Informing the user about the activity on the website (e.g., purchases, views, visits)	313	264	○	◐	◐	○	○	Bandwagon Effect
	Testimonials	Testimonials on a product page whose origin is unclear	12	12	○	○	◐	○	○	Bandwagon Effect
Scarcity	Low-stock Message	Indicating to users that limited quantities of a product are available, increasing its desirability	632	581	○	◐	◐	◐	○	Scarcity Bias
	High-demand Message	Indicating to users that a product is in high-demand and likely to sell out soon, increasing its desirability	47	43	○	◐	○	○	○	Scarcity Bias
Obstruction	Hard to Cancel	Making it easy for the user to sign up for a service but hard to cancel it	31	31	○	○	○	◐	●	None
Forced Action	Forced Enrollment	Coercing users to create accounts or share their information to complete their tasks	6	6	●	○	○	○	●	None

Figure 2.5: Table of dark patterns and their categorization by Mathur et al. [MAF+19a] as well as their prevalence in the study, dimensions and cognitive biases

proposed taxonomy categorizes well-known dark patterns by Mathur et. al [MAF+19a]

and Luguri et. al [LS21] according to two axes: the "choice-architecture" (i.e. design or structure in which information is presented to the consumer) and the "decision-making-process". Graph 2.6 shows the aforementioned taxonomy, in which the categorization of dark patterns defined by Mathur is based on components of the decision-making process (shown on the X axis) and components of the choice architecture (shown on the Y axis).

		Choice architecture		
		Attribute complexity	Cost complexity	Choice complexity
Decision-making component	Affect budget constraint	• Hidden information • Visual interference • Limited time message • Friend spam/social pyramid/address book leeching • Private Zuckering	• Price comparison prevention • Intermediate currency • Hidden cost • Immortal account	• Roach motel • Sneak into basket • Hidden subscription • Preselection • Forced enrolment • Gamification • Nagging
	Shape preferences (consumer valuation)	• Activity messages, testimonials • False hierarchy • Low stock message • High demand message • Cuteness • Manipulative personalised communications and advertisements	• Bait and switch • Misleading reference pricing • Pressured selling • Manipulative personalised pricing	• Confirmshaming • Toying with emotions • Trick questions • Countdown timer with messages • Disguised ads • Manipulative personalisation choices (in search results ranking, recommender systems etc.)

Figure 2.6: Dark pattern taxonomy by the European Union in their study on “Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalization” [LVBB+22]

2.2.3 The Current State of Dark Patterns in eCommerce

This chapter deals with the current state of knowledge on dark patterns, especially in the eCommerce sector. Current, scientific work on the topic of dark patterns in eCommerce

will be presented and its results and findings discussed. Furthermore, in this chapter it is elaborated why dark patterns are so popular and widespread in the eCommerce sector and why they are relevant to many companies.

ECommerce can be defined generally as the sale or purchase of goods or services, whether between businesses, households, individuals or private organizations, through electronic transactions conducted via the internet or other computer-mediated (online communication) networks. The term covers the ordering of goods and services which are sent over computer networks, but the payment and the ultimate delivery of the goods or service may be conducted either on- or off-line [Uni19].

In recent years, it has become apparent that eCommerce and online trading became an integral part of companies and that eCommerce accounts for a large part of the turnover of many companies. This trend can also be seen from a scientific point of view, as the popularity of eCommerce has led to the publication of many scientific articles dealing with this topic. The shift towards eCommerce is based on many factors such as technological progress, changing user behavior, improved internet infrastructure, spatial and temporal independence of users and many more.

The research questions of past studies on the topic of dark patterns mainly deal with the conceptualization of the term "dark patterns" and the development of a taxonomy to be able to assign the individual dark patterns to a clear group.

In the field of eCommerce, academic literature is already one step further and examines corporate websites (especially online shops) for the presence of dark patterns [MAF+19a] [NL21] or for their effectiveness [SHNB22]. The study by Sin et. al [SHNB22] made it clear by conducting an experiment that the use of (in this case 3 selected) dark patterns significantly increased the purchase intention and impulse buying behavior of the study participants in comparison the control group. Although the study refers to the eCommerce domain, the definition of dark patterns was used by Sin et al. was held relatively general. They defined dark patterns as "design interfaces or features that subtly manipulate people into making suboptimal decisions". Despite the generality of this definition, it is relatively fuzzy because subjective terms such as "subtly" and "suboptimal decisions" are applied.

The study of Mathur et al. [MKM21] showed that around 11% of the most popular shopping websites around the world make use of the concept of dark patterns to increase their online sales. In a study in Sweden, a manual approach was applied by Neem [NL21] to analyze Swedish shopping websites for the presence of dark patterns. The result of the study showed that around 60% of the websites examined used dark patterns. Unfortunately, the studies by Mathur [MAF+19a] and Neem [NL21] are not comparable, as different definitions and taxonomies of dark patterns were used. Furthermore, the manual approach is much more time-consuming but can be carried out more accurately, and the difference in time of 2 years could also explain the percentage difference in presence due to the rapid development of technologies in the web sector and the dynamic nature of web content. However, a trend can be discerned that makes clear that the use of dark patterns is increasing rather than decreasing.

Luguri et. al [LS21] also demonstrated the effectiveness of dark patterns in their study.

Mild and aggressive dark patterns (which were self-categorized by the researchers) were used to persuade users to sign up for a dubious service. The use of mild dark patterns resulted in twice as many people subscribing to the service as the control group and the use of aggressive dark patterns resulted in four times as many people subscribing. Moreover, it was also found that people who had already been educated about dark patterns in advance were significantly less susceptible to dark patterns compared to their counterparts.

As mentioned earlier, the work of Brignull [BLSD23], Gray [GKB+18] and Mathur [MAF+19a] has formed the basis for many scientific studies and experiments. Mathur [MAF+19a] published the results of their study on Github [MAF+19b] and included a dataset of individual textual segments with a binary encoding created to indicate which text segments are dark patterns and which are not. This dataset was used and extended by Yada et. al [YFM+22a] to create an automated dark pattern detection using machine learning systems. Different machine learning models for natural language processing were used and among them the language model RoBERTa was the best model with an accuracy of 97.5% for the detection of dark patterns. Their intention was to create a solid baseline for machine learning models on the research topic of automated dark pattern detection and so dataset and baseline source codes are available on Github too [YFM+22b].

But not only in sales-driven companies are dark patterns increasingly used to increase conversions. Nonprofits also have a hard time fulfilling partly unprofitable social missions that depend on financial resources. Due to this conflicting dilemma, nonprofit organizations are tempted to use the same tools as ordinary companies and therefore also resort to dark patterns [Ban21]. Many non-profit organizations use donation forms or subscription models, which in turn are structured similarly to eCommerce sites (with a similar purchase process and deposit of payment methods). However, the ethical component is viewed even more critically by these organizations than by conventional companies.

In the eCommerce sector, research is being conducted on the conceptualization, diffusion, and automated detection of dark patterns. Initial tools and approaches have been developed that can be used for both research and policy purposes. Studies have shown the impact of eCommerce dark patterns on users [SHNB22], the prevalence of dark patterns in Sweden [NL21], and a machine learning baseline has been created that can be used for automated detection of dark patterns[YFM+22a]. The internet as a sales channel and thus the online stores of companies has been indispensable for years. Despite recognizing the challenges coming from the high number of suppliers and intense competition in eCommerce, it is crucial to avoid forcing companies to resort to unfair practices like dark patterns to maintain competitiveness. Achieving this necessitates the establishment of fair and transparent conditions to foster competition in online sales.

2.3 Web Crawling and Web Scraping

Web crawlers have been around since the creation of the World Wide Web and can be described as software code or computer programs written with the objective to download web pages, extract various data (e.g. hyperlinks, attributes, HTML-code, etc.) or create lists of URLs and store them in a structured form in a local database [NP11]. One prominent application of web crawlers is their utilization in search crawling and indexing for search engines. As early as 1998, the founders of Google themselves conducted a scientific study to assess the efficacy of crawlers in the context of search engines [BP98].

In nearly all scientific research domains, theoretical data is commonly tested against reliable and extensive datasets. Consequently, a recurring challenge in research is to construct a well-structured database containing real data that aligns with the research question while being cost-effective and up to date. By employing web crawlers and scrapers, it is possible to download and store online data that is readily available, current, and rich in information, almost in real-time [NP11].

Web crawling or web scraping is the process of extracting and storing data or information from web pages on the internet. Crawlers simulate the behaviour of real users when accessing and browsing a website. They follow links, remain on pages and interact with them. The crawler is thus responsible for following the hyperlinks from a certain starting point in a human-like user behaviour. The scraper in turn extracts relevant data from the pages indexed by the crawler. By using these technologies, the process of data extraction and collection can be significantly facilitated and shortened in contrast to the manual approach.

Data extraction with the help of a scraper therefore requires two software artifacts: a crawler and a scraper. The crawler indexes and downloads data from the internet and the scraper then extracts the most relevant information, encodes it and stores it in a database (or similar storage system) in a user-defined format. This processed data can then be evaluated and analyzed in a way that was not possible in its original form on the internet [Khd21].

The web scraping process can be divided into 3 stages: fetching, extraction and transformation [Per19]. Figure 2.7 visualizes these 3 stages accordingly. In the fetching phase (1), the desired website is accessed using the HTTP protocol, where web browsers and tools like “curl” and “wget” send an HTTP GET request to the target URL and receive the HTML page as a response. During the extraction stage (2), crucial data is obtained by the use of regular expressions, HTML parsing libraries, and XPath queries, which enable the identification of information in documents using the XML Path Language (XPath). In the final transformation stage (3) the collected data may be converted into a suitable format for storing and analyzing [Khd21].

In the domain of Human-Computer Interaction (HCI), data collection has often employed crawlers. One noteworthy study in the realm of dark patterns is "Dark Patterns at Scale:

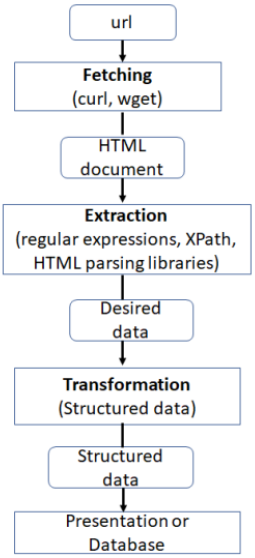

Figure 2.7: The three stages of the web scraping process as described by Persson [Per19]

Findings from a Crawl of 11k Shopping Websites" conducted by Mathur [MAF+19a]. This study used a web-crawler approach to analyze 100k websites, out of which 11k shopping websites were specifically selected for a more in-depth investigation to identify the presence of dark patterns.
Likewise, other researchers in the field have also utilized web crawlers to gather data. For instance, Nouwens et al. [NLV+20] examined consent pop-ups (commonly known as "cookie banners") from 680 websites in the UK. Their research focused on studying the impact of eight most prevalent design variations on users. The visual elements, interaction design, and text content of each consent pop-up were recorded in a database for subsequent analysis.
A different research paper that utilizes a web crawler for investigating web cookies is titled "An Empirical Study of Web Cookies" [CABM16]. In this study, a web crawler accessed web pages through Mozilla Firefox and utilized a Firefox extension to record the cookies set by each website. A total of 3.2 million cookies were collected and subjected to analysis in two separate crawl passes.
Di Geronimo et al. [DGBF+20] conducted a study focusing on user perception of dark patterns in mobile applications and also constructed a database of the apps to be analyzed using a web crawler. The web crawler was employed to download the most popular apps, along with their metadata, from each category in the Google Play Store. Subsequently, an online experiment involving 589 users revealed that the majority of users do not readily recognize dark patterns, and when informed about them beforehand, they become more adept at identifying and responding to such patterns. Furthermore, the study revealed that 95% of the 240 apps examined incorporated dark patterns.
In 2021, Curley et al. [COG+21] conducted a study focusing on the development of a

framework for detecting Web-Based dark patterns. Brignull's taxonomy [BLSD23] was employed for this purpose, and each dark pattern identified by Brignull was assessed to determine whether it could be automatically detected (partially or fully), manually detected (partially or fully), or not detected at all. The findings of Curley's study [COG+21] revealed that out of the 12 existing dark patterns, 3 patterns could only be detected manually, 5 patterns were undetectable, 1 pattern could be fully automated, and 3 patterns could be partially automated. In this study, we used a similar approach to find out which dark patterns can be detected in a full automated approach using text-based pattern matching. Section 4.1 provides more details about the methodology and explains which dark patterns from Mathur's taxonomy [MAF+19a] can be detected with a web crawler and regular expression pattern matching, and which cannot, and why.

TABLE I. Dark Patterns and Their Detection

Category	*Pattern*	*Detection*	*Rationale*
Sneaking	Sneak into Basket	Manual (fully)	Highlight changes in cost
	Hidden Costs	Manual (fully)	Highlight changes in cost
Misdirection	Trick Questions	Automated (partially)	Look for phrases like "opt-in" and "opt-out", as well as pre-ticked checkboxes
	Misdirection	Cannot be detected	There is too much variation in how this pattern is implemented.
	Confirmshaming	Cannot be detected	There is too much variation in how this pattern is implemented.
	Disguised Ads	Automated (partially)	Look for buttons (noting colour and size) and see which ones link to external sites.
Obstruction	Roach Motel	Automated (fully)	Look for sites with "activate" or "subscribe" links or buttons but with no "deactivate" or "unsubscribe"
Forced Action	Forced Continuity	Cannot be detected	There is too much variation in how this pattern is implemented.
Variegations	Privacy Zuckering	Cannot be detected	There is too much variation in how this pattern is implemented.
	Price Comparison Prevention	Manual (fully)	Highlight if products are displayed with different units of the product
	Bait and Switch	Cannot be detected	There is too much variation in how this pattern is implemented.
	Friend Spam	Automated (partially)	Check if the site asks for email or social media permissions, and notify users.

Figure 2.8: Results of Curley´s [COG+21] study on which dark patterns can be automatically, manually or not detected using Brignull´s [BLSD23] taxonomy of dark pattern

These examples in HCI research show that the use of web crawlers for empirical studies is practical and already form a good basis for a methodical approach to web crawling and scraping. The rapid availability of real-time data on the internet and the collection of real-world data are major advantages of using web crawlers. For many research areas related to the WWW, web crawlers will probably be used more frequently in the future.

2.4 Legal Implications

In the context of Austrian law, dark patterns are examined in this chapter, providing a general understanding of their implications. Due to the relatively recent emergence of the term "dark patterns" and its lack of a definitive definition, there are currently no explicit Austrian or supranational laws specifically addressing dark patterns. Nevertheless, existing legal frameworks and statutes can be invoked to combat manipulative practices and safeguard the welfare of consumers and users. These regulations serve as a means to prevent or mitigate the negative effects of dark patterns and ensure the protection of individuals' interests.

Numerous scientific papers explore the legal compliance of dark patterns in a general context. A significant portion of these papers focuses on EU directives, which serve as overarching legislation within the European Union and represent a political call to action for the community. It is mandatory for member states' national parliaments to incorporate these directives into their respective national laws within a specified timeframe. Conversely, EU regulations exist as laws that automatically and directly apply to all member states without the need for additional implementation procedures.

Typically, the process of implementing prohibitions against deceptive design patterns involves the inclusion of explicit bans in both EU and national legal frameworks, as advised by scientific policymakers [Wei20].

In 2022, a study on "Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation" [LVBB+22] was conducted on behalf of the European Union. As part of the study, a legal assessment was carried out to investigate whether EU legislation correctly regulates the use of dark patterns. The results showed that the regulation of unfair business practices in the digital domain are subject to the laws of consumer protection, data protection, and other relevant instruments in the EU legal framework, including new and forthcoming legislation such as the Digital Services Act, Digital Markets Act, AI Act and Data Act [LVBB+22]. The result of the legal investigation also showed that "some legislative adjustments may be necessary" [LVBB+22] - a sobering conclusion of a study carried out by the European Union on its self-enacted laws.

2.4.1 Challenges in Legal Regulation

Despite the term "dark patterns" being recognized for approximately 15 years and holding significance from both scientific and legal standpoints, concrete regulatory approaches in the form of specific laws and jurisdictions are still lacking. This deficiency can be attributed to various challenges in formulating such laws. The development of clear regulations for dark patterns is complicated by the existence of multiple definitions of the term "dark patterns," each describing its characteristics, manifestations, and features. Within the academic sphere, pioneers in dark pattern research, such as Brignull [BLSD23], Gray [GKB+18], and Mathur [MKM21], each use their own developed definitions and

taxonomies for their research. Moreover, the EU itself has conducted a study on the prevalence, impact, and legal implications of dark patterns, led by the research group headed by Lupinanez [LVBB+22]. However, even this EU study, grounded in behavioral science, has introduced a new taxonomy to categorize dark patterns, further complicating the establishment of standardized regulations.

Another challenge is the line between good persuasion by design and manipulation. Dark patterns are often very subtly constructed. For instance, web pages may incorporate background designs or carefully designed decision architectures that can significantly influence users without causing confusion [LS21]. Establishing a clear boundary between ethical persuasion and manipulation proves to be exceedingly complex. Manipulation occurs when individuals deviate from their preference without changing it and without the presence of anything other than behaviouralist market failure [Wei20]. A (legitimate) persuasion, on the other hand, takes place through an adjustment of preferences based on information or offers provided. However, in the case of spontaneous decisions such as a consent decision, it is difficult to determine or measure whether there were preferences and whether they were adjusted.

A pioneering work in the field of research on dark patterns is the paper "What Makes a Dark Pattern...? Dark?: Design Attributes, Normative Considerations, and Measurement Methods" by Gray [GKB+18]. This work presents a comprehensive overview of the existing research on dark patterns up to that point, highlighting its interdisciplinary nature. The research draws from various fields, including psychology, economics, ethics, philosophy, law, and human-computer interaction. This convergence of multiple disciplines and the relatively under-developed conceptualization of dark patterns pose significant challenges to their regulation, as it requires a solid grasp of diverse concepts and subject areas. Gray [GKB+18] concludes their paper by urging human-computer interaction (HCI) researchers to create metrics and methodologies that can be adopted by regulatory agencies. By doing so, it can facilitate the regulation of dark patterns and contribute to a better understanding and control of these deceptive design practices.

Website operators often find it relatively simple to design interface elements in a manner that avoids obvious violations of broadly formulated legal norms [Ger22]. Small alterations in wording or sentence structure are often sufficient to achieve the same intended effect with language while avoiding these broadly stated legal norms. Additionally, dark patterns manifest in various forms, allowing many variants to be substituted with more subtle manifestations. For instance, one subtle pattern known as "confirmshaming" [BLSD23] involves presenting the option of refusal in a way that makes users feel uncomfortable about declining.To illustrate, an online retailer could employ this pattern to persuade a customer to choose a more expensive shipping option. The button or option might be phrased as "Yes, I am happy to bear the costs of environmentally friendly shipping for my goods" or "No, I do not care about the environment when shipping my order" [Ger22]. By utilizing this (rather discreet) pattern, a similar effect can be achieved as when a checkbox for environmentally friendly shipping is pre-selected.

The regulation of dark patterns is also difficult or at least critical from the point of view of the behavioural sciences. Due to the heterogeneity of individual human reaction patterns,

legal norms are difficult to generalise or formulate in concrete terms [MDSW21].Each person is individual and thus has different characteristics and backgrounds that influence the decision process.
When litigation actually occurs, the preservation of evidence presents another significant challenge for the plaintiff. Not only does it demand a high level of technical expertise, but the intricate reproducibility of evidence also proves to be a major hurdle. These challenges often stem from subjective and process-related factors that are not easily captured through visual or verbal means [MDSW21]. Furthermore, the dynamic nature of the web contributes to frequent changes in web pages, requiring real-time evidence gathering. Even then, there is no assurance that all aspects of the decision-making process or emotions influencing it can be adequately captured. Engaging in dark pattern litigation is a rare occurrence and is often lengthy and highly uncertain in its outcome, which dissuades many plaintiffs from pursuing legal action against (often international) companies [Ger22]. The prospect of exorbitant court costs, public attention, and the aforementioned uncertainties typically deter individuals from initiating legal proceedings.
To conclude, the regulation of dark patterns is a complex endeavor, far from straightforward at first glance. It operates at the intersection of multiple disciplines, demanding substantial expertise to grasp the concept and effectively address it. Challenges encompass undefined and non-standardized definitions, ambiguous formulations, and the intricacies of distinguishing between persuasion and manipulation. As a result, the HCI community faces the ongoing challenge of developing measures, methods, and metrics that can serve as valuable tools for legislative institutions in tackling dark patterns. Further research and collaboration are crucial to advance understanding and effectively mitigate the adverse effects of these deceptive design practices.
Despite all these challenges and difficulties, there are regulatory approaches from the legislature to limit dark patterns. Although dark patterns are never explicitly mentioned in legal texts, they are sometimes described in relatively concrete terms and thus also regulated. Some directives and regulations already existed before the term "dark patterns" became known and others were only enacted afterwards. The following subsections contain an excerpt from various directives, ordinances and laws that directly or indirectly deal with the regulation of dark patterns.

2.4.2 UCPD (Unfair Commercial Practices Directive) – European Union

The EU-wide Directive 2005/29/EC of 11 May 2005 on competition- and consumer protection generally prohibits "aggressive commercial practices" in Art. 8. In Austria (and Germany) this provision was implemented in the "Gesetz gegen unlauteren Wettbewerb (UWG)" [Wei20]. Furthermore, a "black list" was defined in the UWG, which contains clauses and practices that are considered unfair and therefore may not be used in business transactions. The list contains various prohibitions and restrictions to protect consumers and to keep competition fair and transparent.
In Germany, the "black list" of the Annex to Section 3 (3) UWG 112 covers some dark

patterns and thus prohibits the use of some dark patterns in commercial transactions. Although these patterns are not mentioned by name (as it was not possible to agree on a definition or taxonomy in law as well as in science), they describe relatively clearly known manifestations of individual dark patterns [MDSW21].
The “black list” may differ in Austria and Germany, but many similarities can be found in both lists. The following points of the "black list" of the Austrian UWG describe directly or partly indirectly certain manifestations of dark patterns:

Point 6:
Die Aufforderung zum Kauf von Produkten zu einem bestimmten Preis und dann
a) Weigerung, dem Umworbenen den beworbenen Artikel zu zeigen, oder
b) Weigerung, Bestellungen dafür anzunehmen oder innerhalb einer vertretbaren Zeit zu liefern, oder
c) Vorführung eines fehlerhaften Exemplars
in der Absicht, stattdessen ein anderes Produkt abzusetzen („bait-and-switch“-Technik).

In this context, the "bait-and-switch" technique is explicitly mentioned, which is also a term coined by Harry Brignull [BLSD23] to describe a specific category of dark patterns. Brignull [BLSD23] defined "bait and switch" as follows:
You set out to do one thing, but a different, undesirable thing happens instead. Brignull's [BLSD23] term is thus broader in scope and does not specifically pertain to product purchases.

Point 7:
Die unrichtige Behauptung, dass das Produkt nur eine sehr begrenzte Zeit oder nur eine sehr begrenzte Zeit zu bestimmten Bedingungen verfügbar sein werde, um so den Verbraucher zu einer sofortigen Entscheidung zu verleiten, so dass er weder Zeit noch Gelegenheit hat, eine informierte Entscheidung zu treffen.
Here, one refers to so-called "scarcity patterns", i.e. design patterns that use an artificial scarcity of resources (in this concrete case time) to create buying pressure.
Point 7 of the Black List describes relatively clearly the use of two dark patterns from the "Urgency" category described by Mathur [MAF+19a], namely "Countdown Timer" and "Limited-time Message". A countdown timer is self-explanatory in terms of its name, a "Limited-time Message" is an indication that an offer is about to expire without defining an end time.
When employing these design elements, it may not be readily apparent to the user whether these elements qualify as dark patterns, as the user lacks knowledge or the ability to examine the pre-offer price or the post-offer price. Only the current price is displayed, leaving the user uncertain about potential changes once the timer expires (or even how it was before the offer was introduced). Similarly, from a legal standpoint, there is no visual cue to substantiate the transparency of these offers.

Point 20:

Die Beschreibung eines Produktes als „gratis“, „umsonst“, „kostenfrei“ oder ähnlich, obwohl der Umworbene weitergehende Kosten als die Kosten zu tragen hat, die im Rahmen des Eingehens auf die Geschäftspraktik und für die Abholung oder Lieferung der Ware unvermeidbar sind.
The prohibited practices outlined in this ban fall under the dark pattern category of "sneaking," as defined in Mathur's taxonomy [MAF+19a]. Specifically, it refers to the "hidden costs" type, in which costs that were previously undisclosed to users are revealed immediately before they make a purchase, as described by Mathur [MAF+19a].
However, it is important to note that the dark pattern itself encompasses a broader scope than the prohibition outlined in the UWG. According to the definition of the hidden cost pattern, it includes all costs that were not clearly indicated beforehand and are added afterwards. In contrast, the prohibition explicitly pertains to products that are advertised as "free, free of charge, free of cost," etc. Therefore, the prohibition of hidden costs may not be applicable in these cases.

Point 21:
Die Beifügung einer Rechnung oder eines ähnlichen Dokuments mit einer Zahlungsaufforderung zu Werbematerialien, die dem Umworbenen den unrichtigen Eindruck vermittelt, dass er das beworbene Produkt bereits bestellt habe.
One might mistakenly think that this prohibition is also a paraphrase of the hidden cost pattern, but the pattern clearly refers to the time "shortly before the conclusion of the purchase", whereas the prohibition refers to the time after the conclusion of the purchase (i.e. when the invoice is received).

2.4.3 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) was officially released by the European Union (EU) in 2018. Functioning as a regulation, it holds the status of direct law and has been obligatory for all EU member states since its inception. The primary objective of the GDPR is to safeguard the privacy of users during the processing of their data [EC16]. Consequently, the implementation of the GDPR also led to the introduction of "cookie banners" on websites. These banners seek the user's consent to enable the storage, processing, and potential sharing of their data through cookies. Cookies serve as a means to retain user information and enable online services to track users across multiple websites [HIB22]. Dark patterns are not expressly or explicitly prohibited in the GDPR, but the GDPR aims to protect privacy and thus undermine the effect of some privacy dark patterns.
According to Brignull [BLSD23], "Privacy Zuckering" is defined as the act of deceiving individuals into publicly sharing more information about themselves than they intended. This definition encompasses all deceptive schemes aimed at coercing users into disclosing sensitive personal data. However, due to its broad and generalized nature, the Privacy Zuckering (named after Mark Zuckerberg - CEO of Meta Platforms, Inc.) dark pattern as defined by Brignull [BLSD23] does not fall under the protection of the GDPR.
Three elements are central to the effectiveness of the GDPR: consent must be given

specific, informed and unambiguous. These three elements can be used to argue against some dark patterns or to explain why certain dark patterns (or individual manifestations of a dark pattern) are not GDPR-compliant [Wei20]. "Unambiguousness" can be violated by the interface elements with an opt-out design or pre-selection patterns (and thus also contradict the principle of privacy-by-default of the General Data Protection Regulation)[BLSD23][Wei20]. Furthermore, unambiguousness can potentially also be violated by the use of misdirection patterns [MAF+19a] or trick-questions [GKB+18][MAF+19a][Wei20].

For some online services, consent to the processing of data is mandatory in order to use certain services, although the personal data are not necessary for this service. The GDPR defines horizontal and vertical purpose limitation ("horizontale und vertikale Kopplung") [EC16][Wei20]. The horizontal prohibition of tying describes that separable data processing operations require separate consent (i.e. checkboxes). The vertical prohibition of tying states that consent is not voluntary if the performance of a contract depends on the granting of consent [EC16]. The vertical tying prohibition thus refers to voluntariness (freely given consent) and thus prohibits conditional patterns of the forced action category [BLSD23]. A fictitious example of forced action with vertical tying would be a calculator app that you can only use if you give permission for location tracking.

Furthermore, the nagging pattern or the trick-question pattern could speak against a voluntary action or freely given consent. In the nagging pattern, the user is repeatedly and sometimes aggressively asked to perform a certain action. Trick questions are questions with a deliberately imprecise formulation (e.g. by double negation).

Hidden information patterns conflict with the central element of "informed consent" of the GDPR [EC16], as their visual design leaves users in the dark about which permissions providers are specifically asking for - for example, by hiding options. Trick question patterns can cause confusion about the scope of data processing due to their wording or overly complex structures [Wei20].

Dark patterns of manipulation, such as the countdown timer and confirmshaming, conflict with the fundamental elements that ensure the effectiveness of the GDPR. However, it is important to note that not all dark patterns can be effectively addressed or undermined by these elements. An illustrative example of a dark pattern that remains resistant to the aforementioned elements is the countdown timer [MAF+19a]. As per Mathur's taxonomy [MAF+19a], this dark pattern falls under the urgency category. Although the decision to consent may be voluntary, it is influenced by time pressure, and the countdown timer is not constrained by the requirement of informedness. Another dark pattern that operates outside the scope of the GDPR is confirmshaming [MAF+19a], which employs language and emotional manipulation to coerce users into performing a specific action (e.g., "No, I want to take the full risk," "Yes, I am smart and take insurance"). This pattern utilizes framing, a stylistic technique that appeals to human psychology but is not necessarily subjected to rational evaluation and processing [Ger22].

Article 25 of the GDPR [EC16] emphasizes the concept of "data protection by design." This provision mandates that companies incorporate technical and organizational measures to safeguard user privacy at the earliest stages of information system development.

However, as highlighted by Martini et al. [Wei20], this article presents a conflict with default effects, which manifest in patterns such as pre-selection.
To summarize, the General Data Protection Regulation (GDPR) primarily focuses on the protection of the private sphere, including patterns that pose a risk to it. However, it is important to acknowledge that the regulation, like many other laws, operates under the assumption of a flawed understanding of human behavior, specifically the concept of homo oeconomicus. Homo economicus assumes that individuals, within their respective circumstances, always strive to maximize their own benefits and make decisions based on rational cost-benefit analysis [Wei20]. Nonetheless, it is well-established that human behavior is not always driven by rationality, as people often exhibit irrational tendencies and cognitive biases.

2.4.4 Contract Law and Consmer Contract Law | Vertrags- und Verbrauchervertragsrecht

In addition to the General Data Protection Regulation (GDPR) and the Unfair Competition Law (UCDP), there are other legal instruments regulating dark patterns at the national level and at the level of the European Union.
Contract law is a component of civil law, which governs the legal relationships between individuals and legal entities, whereas public law deals with the regulation of relationships between individuals or legal entities and the state. In Austria, contract law is primarily based on the General Civil Code (ABGB - Allgemeines bürgerliches Gesetzbuch)[Ös23], enacted in 1811 and subject to ongoing updates. It encompasses the validity and formal requirements of contracts, contractual freedom, obligations of the contracting parties, legal ramifications of contract breaches, and various aspects of contract formation and execution.Apart from the ABGB, additional legal underpinnings for contract law in Austria are found in the Austrian Commercial Code (UGB - Unternehmensgesetzbuch) and the Consumer Protection Act (KSchG - Konsumentenschutzgesetz). The main objective of contract law is to ensure the enforcement of the mutual intent of both parties when entering into a contract. However, the use of dark patterns can disrupt this intent, particularly from the consumer's perspective. Contracts established on the basis of deceptive information can be contested due to the element of deception. Moreover, contracts founded on misleading information may be rendered invalid [MDSW21].These regulations mainly refer to untrue scarcity patterns, e.g. low-stock messages, activity messages or countdown timers, if these correspond to the untruth. Untrue scarcity patterns in this context mean that a shortage of resources (stock, time restriction) is display on a website, but these shortages do not exist in reality. Verifying these circumstances poses significant challenges, as it is difficult to ascertain the true stock level or determine if an offer maintains the same price after a timer expires (or if it matches a previous price from two weeks ago, for instance). Additionally, the utilization of dark patterns falling under the categories of Bait & Switch or Hidden Costs may lead to further breaches of contracts [MDSW21].

Consumer contract law (Verbrauchervertragsrecht) represents a specialized subset of contract law aimed at regulating agreements between consumers and businesses. Its purpose is to establish rules that promote transparent information and fairness while safeguarding consumers from unfair terms. In Austria, consumer contract law finds its legal foundation in the Consumer Protection Act (KSchG - Konsumentenschutzgesetz) and the Distance and Foreign Transactions Act (FAGG - Fern- und Auswärtsgeschäfte Gesetz). It encompasses aspects related to information requirements, transparency obligations, and withdrawal rights for the involved parties.
Through the application of this law, certain dark patterns, such as Hidden Information patterns or Price Comparison Prevention patterns, can be regulated or restricted. These patterns involve concealing or omitting product properties/characteristics or vital information and making price comparisons more challenging or even impossible, thereby disregarding information and transparency obligations. Additionally, other patterns, partly governed by consumer contract law, or leading to the nullification of contracts include Preselection patterns, Forced Continuity, Hidden Costs, Hidden Subscriptions, or Sneak into Basket [MDSW21].

2.4.5 Digital Services Act (DSA) and Digital Market Act (DMA)

In the future, the two EU regulations, the Digital Services Act (DSA) and the Digital Markets Act (DMA), will also be relevant for the regulation of dark patterns. The Digital Services Act and the Digital Markets Act have two main goals, according to the EU [Uni23]:

1. To create a safer digital space in which the fundamental rights of all users of digital services are protected.

2. To establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.

The DSA was published on October 27, 2022, will enter into force on February 17, 2024, and must be implemented by all member states in the form of national, concrete laws at that time. In addition, each member state must provide a Digital Services Coordinator at the time of the entry into force of the DSA, who are responsible for the implementation, monitoring, and enforcement of the DSA.

The Digital Markets Act (DMA) was published on November 1, 2022, and came into effect on May 2, 2023. The DMA pertains to very large companies that are classified as "gatekeepers." Gatekeeper companies are those with more than 45 million active users (i.e., more than 10% of the entire EU population), and as such, they are to be regarded with special scrutiny due to their market power and influence on society. Online platforms and search engines are obliged to disclose their user numbers for the purpose of categorizing companies as gatekeepers.

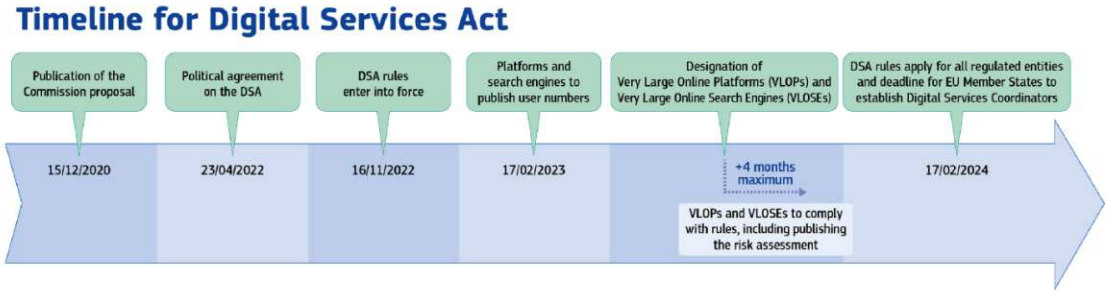

Figure 2.9: Timeline of the Digital Services Act published by the European Union [Uni23]

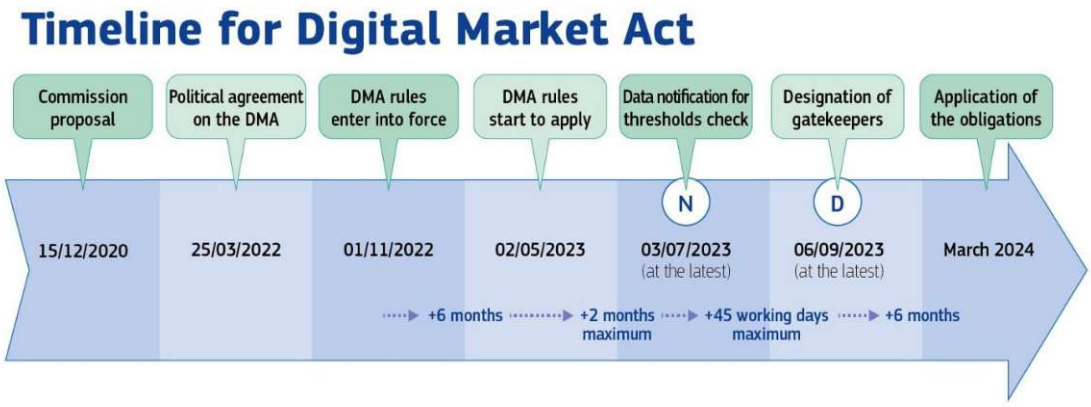

Figure 2.10: Timeline of the Digital Market Act published by the European Union [Uni23]

The Digital Services Act explicitly defined and mentioned dark patterns for the first time. Recital 67 defines dark patterns as follows: *Dark patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them. [Uni22]*
Furthermore, in Recital 67, it is described that interface design should not compel a decision that does not align with the user's interests, highlight specific options in a manner that creates an imbalance, or influence the user's autonomy and decision-making. Certain dark paterns are delineated in some detail, such as the Roach Motel Pattern, where it is relatively easy to sign up for a service but the process of unsubscribing is made very difficult or nearly impossible. Additional patterns outlined in the Digital Services Act (DSA) include the Preselection Pattern (prefilling certain parameters, such as selecting the most expensive combination of a variable product) and the Nudging Pattern, where the user is persistently presented with the same question even though the user's decision has already been expressed by answering the initial question. It is important to note,

however, that this description in Recital 67 should not prohibit interaction with users in the form of advertising or information presentation.
Article 25 of the DSA is titled "Online interface design and organization" and refers to the design of manipulative interfaces:
Providers of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions. [Uni22]
Other articles of the Digital Services Act (DSA) pertain to the transparent design of advertising and recommendation systems, which have been positively mentioned by Martini [MDSW21]. However, Martini et al. tend to adopt a more critical perspective toward the DSA because some design patterns are specifically described, defined, and prohibited, while others are not mentioned at all. Particularly, dark paterns that target perceived pressure are not regulated at all. Examples of pressure patterns include the Countdown Timer or the Low-Stock Message. For a clear regulation, it is therefore necessary to describe all types and categories of dark paterns [MDSW21].
The Digital Services Act thus presents some noteworthy approaches to regulating certain dark paterns. Additionally, the fact that dark paterns are explicitly and literally mentioned (rather than vaguely described) will be relevant for future legislation. How these specific laws are implemented and enforced at the national level will only become evident in 2024.

2.4.6 Other Regulative Approaches

In addition to the aforementioned legislative and regulatory instruments, there exist additional legal provisions utilized for the regulation of dark patterns. Dark patterns have implications across various legal domains, including criminal law and the regulations found within the realm of telemedia, as established by the Telemedia Act (Telemediengesetz - TMG). Furthermore, the Platform-to-Business Regulation (P2B Regulation) of the European Union employs specific phrasings to emphasize the obligation to inform and the imperative of transparency. Similar formulations are also present in the ePrivacy Regulation of the European Union, determining the requirement for information to be presented in a manner that is readily perceivable, comprehensible, and unambiguously clear [MDSW21].

However, it is important to note that the main focus of this thesis is not exclusively on the legal aspects of dark patterns, but on the concept and detection of dark patterns from a technical perspective. In light of this specific focus and considering the constraints of resources, the research has been delimited to a selection of relevant laws and regulations. It is possible that additional legal frameworks may exist that have not been encompassed within the scope of this thesis.

CHAPTER 3

Methodology

In this chapter, comprehensive overview of the research methods employed is provided. These research methods encompass a methodological approach integrating web crawlers and semi-structured interviews. Following this, the research questions will be articulated. Lastly, we will expound on the implementation of the study and elaborate on the data analysis procedures.

3.1 Study Design

In this study a web-crawler is implemented to observe the prevalence of dark patterns in the Austrian eCommerce sector. Furthermore, the dark patterns found were examined from a legal perspective by conducting semi-structured expert interviews. To address the research questions, empirical-quantitative and empirical-qualitative analyses were conducted. The entire research process was implemented following Hevner's [HCHC10] Information Systems Research Framework to ensure scientific quality and relevance. The methodological approach used can be broken down into two main tasks: (1) conducting a web crawler study to observe the prevalence of dark patterns in the Austrian eCommerce sector, and (2) reviewing the detected dark patterns from a legal perspective using semi-structured interviews (SSI) with legal experts.

3.1.1 Design-Science Research in Information Systems Framework

This thesis was implemented using the Design-Science Research Framework in Information Systems as proposed by Hevner [HCHC10]. This framework aims to integrate the complementary paradigms of behavioral sciences and design science to achieve high-quality research in the field of information systems. The following sub-sections provide a detailed explanation of the application of the framework in this work and elucidate

the individual elements involved. Figure 3.1 shows the design science research framework according to Hevner [HCHC10] and the individual components of the framework, which were implemented and are explained in this chapter.

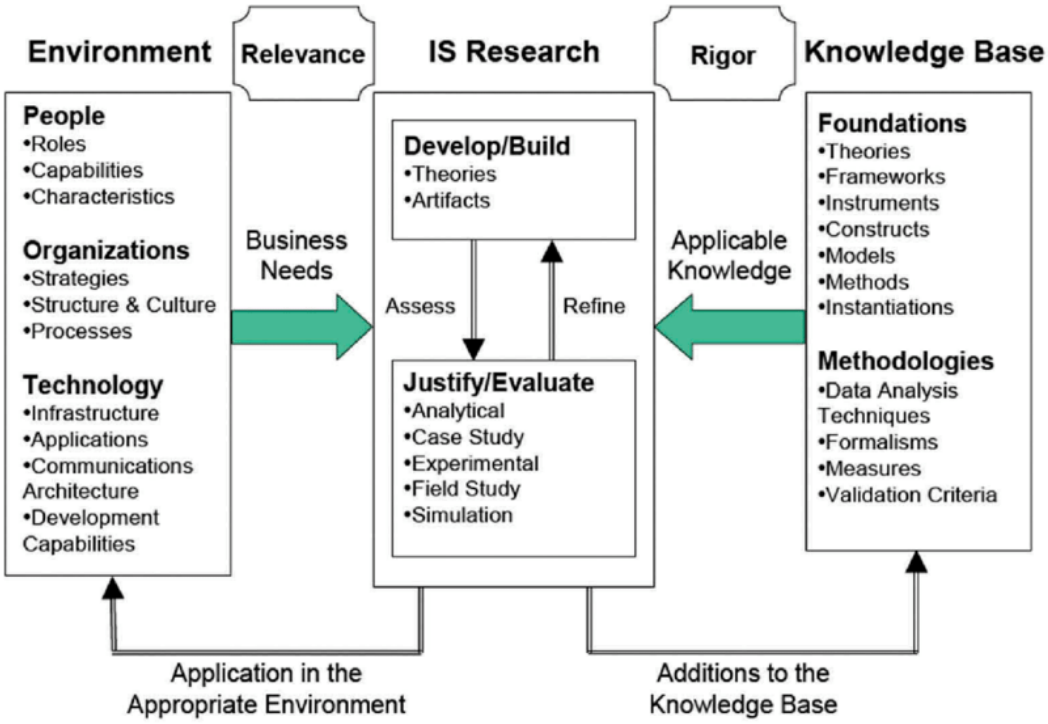

Figure 3.1: Information System Research Framework by Hevner [HCHC10]

The basic idea of this framework is that the process of high quality research in the Information Systems field consists of 3 cycles:

1. Relevance Cycle
2. Design Cycle
3. Rigor Cycle

The Relevance Cycle serves the purpose of incorporating requirements from the contextual environment into the process and feeding the resulting artifact back into the environment. The Rigor Cycle provides the underlying literature and expertise in the research area, ensuring that the emerging artifact is reintegrated into the existing knowledge base, meaning that the artifact is relevant to the research. In the central Design Cycle, an artifact or process is constructed and evaluated in a loop [Hev07].

Environment – People, Organizations & Technology

The environment specifies the problem space of the phenomenon under investigation in the study [HCHC10].

The contextual environment of this work comprises users and operators of online shops. Users are affected by dark patterns, while the operators of online shops are responsible for the content of the websites, including the creation and use of dark patterns. Additionally, the environment includes experts in the field of law and politics. Within the scope of this work, an examination is conducted to determine whether the use of a crawler is feasible from legal and political perspectives and whether an assessment of the legal conformity of dark patterns is possible. Furthermore, the research explores potential use cases for web crawlers in detecting dark patterns.

The organizations involved in this context include companies utilizing online shops as a distribution channel, as well as legal departments of these companies. Legal departments of political and legislative institutions are also impacted, along with political stakeholders in the eCommerce sector.

From a technological perspective, the environment consists of the web crawler used to detect dark patterns and the underlying technologies employed to construct this tool. This includes the programming languages used, database systems, regular expressions for pattern matching, and analysis tools and resources for data evaluation.

In this thesis, semi-structured interviews are conducted with experts in the legal field who are the stakeholders of the environment according to the information system research framework.

IS Research - Develop/Build and Justify/Evaluate

Within the scope of this thesis, a web crawler was implemented to examine product detail pages of online shops for the presence of dark patterns. Additionally, regular expressions were created to identify predefined dark patterns based on commonly used phrases. The collected data is stored in a database and can be utilized for further research and analysis.

To evaluate the results of the created artifact (the web crawler), an iterative process was employed in the development of regular expressions and the web crawler. Manual searches for dark patterns were conducted, and regular expressions were adjusted to detect as many manipulative patterns as possible. Furthermore, the parameters and settings of the crawler were modified and tested until a sufficiently effective crawling process was achieved.

Conducting interviews with legal and political experts aimed to verify whether the use of web crawlers is justified and meaningful from both political and legal perspectives.

The Design Cycle was iteratively traversed until satisfaction was reached with the quality of the research results and the process. This led to the creation of an appropriate solution for the challenges defined in the pre-established environment, one that could also contribute to the scientific knowledge base.

Knowledge Base - Foundations & Methodologies

To establish a knowledge base in the field, a semi-systematic literature review was conducted as part of this study. The objective was to better understand the definitions, taxonomies, and scope of dark patterns in eCommerce. Additionally, legal aspects related to dark patterns were incorporated to prepare the interviewer for discussions with legal experts. This process resulted in an overview of the legal regulation of online shops, the prevalence and legal landscape of dark patterns in the eCommerce market, and current research on dark patterns.

The methodologies employed included the execution of a semi-systematic literature review following Snyder [Sny19], the development of a web crawler based on an existing methodology by Mathur [MAF+19a], and the conduct of semi-structured interviews following the approach outlined by Adams [Ada15]. An interview guide was created for questioning legal experts. The study also investigated current methods for identifying dark patterns [COG+21][SSS22][YFM+22a], assessing their effectiveness and efficiency, and examined existing regulatory measures for dark patterns and their efficacy [Ger22][MDSW21].

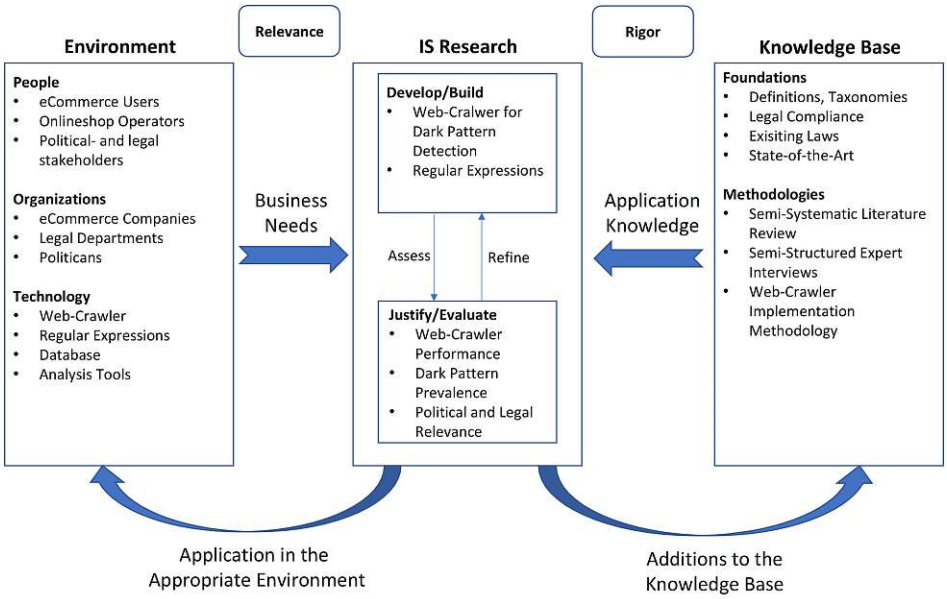

Figure 3.2: Implementation of the Information System Research Framework by Hevner [HCHC10] in this study

3.1.2 Web-Crawler

This section provides a concise overview of the methodological steps that were performed to implement the web crawler. The underlying theory of web crawler technology is described in chapter 2.3 and further details of the implementation and execution of the crawl are outlined in 4.2.

In the first task a web crawler was implemented to do research on prevalence and distribution of dark patterns in the Austrian eCommerce sector. In the context of this study, a "focus crawler" [NP12] was implemented, i.e. a crawler that crawls only certain, specialized pages and extracts important information on the basis of a defined set of rules (in contrast to a "general purpose crawler" [NP12], which can be understood as a high-performance crawler that crawls millions of pages for search engines). The focus of the "focus crawler" implemented in this study were the product detail pages of online shops. In order to implement a crawler, you need to have a solid understanding of how users behave on the website and what happens on the website from an information processing point of view. Based on these concepts, the idea behind the implementation of a crawler is to (1) imitate the actions of a user visiting a webpage, (2) extract the information that is needed, and (3) repeat these steps. The focus crawler can be further categorized as an exploration crawler [NP11], since many (partly unknown) pages of a certain domain (eCommerce) are explored and objects are identified and stored for further analysis. The goal of an exploratory crawler is to illuminate important characteristics of unexplored fields - in the case of this thesis, certain types of dark patterns are stored and their characteristics are subsequently analyzed.

Even though there is to the best of my knowledge currently no standardized procedure for the use of web crawlers in scientific studies, the process was implemented to be transparent, comprehensible and systematic.

An exploration web crawler [NP11] was implemented using a proposed framework of Mathur et al. [MAF+19a] to break down the complexity of the methodological approach in three phases.

1. Phase: Corpus Creation
2. Phase: Data Collection
3. Phase: Data Analysis

In the first phase, the corpus, i.e. a list of relevant web pages (URLs), was constructed. For the creation of the corpus the top-selling eCommerce websites that operate in Austria (and are therefore obliged to be compliant with Austrian law) as well as a list of all

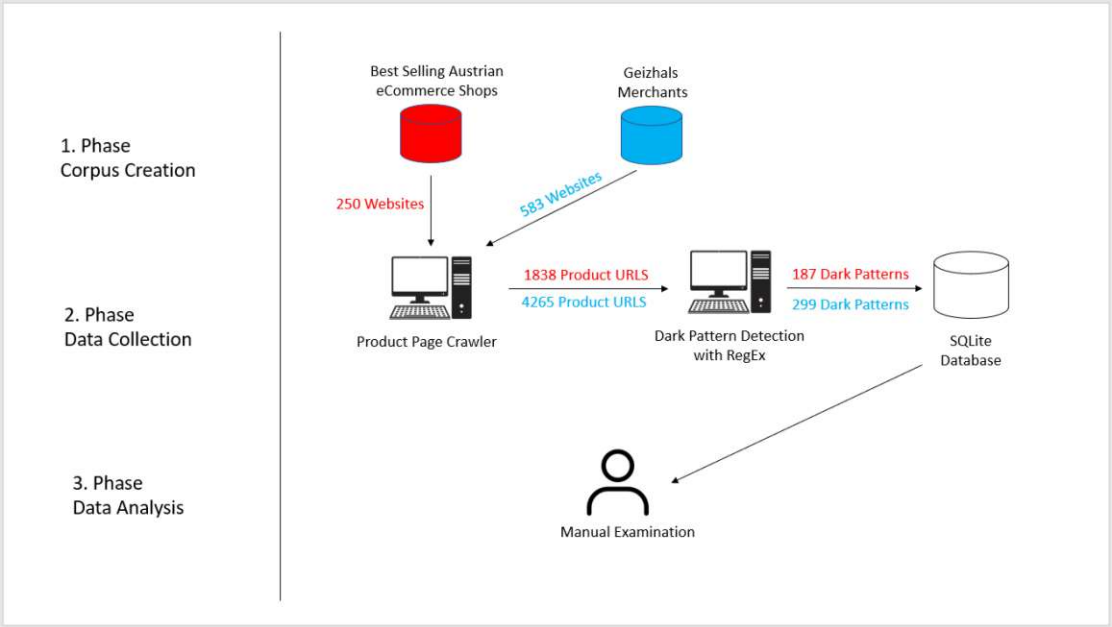

Figure 3.3: Overview of the 3 phases of the crawling methodology: corpus creation, data collection and data analysis

merchants of the price comparison platform "Geizhals" [AG23] were selected as a sample to represent the Austrian eCommerce Market. The eCommerceDB dataset was selected because the goal of this study was to examine the top eCommerce companies in Austria for the presence of dark patterns. As a second dataset, the Geizhals retailer list was chosen because the list offers a good cross-section of online stores in Austria and the number of websites included could not be too large due to hardware restrictions. The main reason why a second dataset was chosen at all is for the validation of the crawler and to test if the crawler can be applied to other datasets (than the eCommerceDB dataset) in general.

During second phase the product page web crawler was implemented using Python as a programming language and supporting libraries to imitate the behavior of real users (as opposed to robots) to collect relevant data for the study [NP11]. In the research conducted by Mathur et al. [MAF+19a], they employed a product page crawler to gather a list of product URLs, and subsequently utilized a checkout crawler to collect data during the checkout process of online stores. In contrast to their methodology, this study diverges by not incorporating a checkout crawler. Instead, in a secondary crawling phase, the product URLs are examined for the existence of dark patterns using regular expressions. The dark patterns found are then stored in a database together with the associated meta data (such as CSS classes, HTML tag, etc.) and subsequently checked manually for their truth content (whether the dark pattern found by the crawler is really a dark pattern).

In the third and final step, the data stored in the database is analyzed and evaluated. This occurs during a manual examination phase, where the dark patterns discovered by the crawler are verified to determine how many of them qualify as actual dark patterns. Descriptive statistics are then extracted from the data using SQL statements.

The detailed execution of these steps can be found in section 4.2.

3.1.3 Semi-Structured Interviews

In the second and final task of this study semi-structured interviews with legal experts were conducted, aimed at answering the research questions from a legal perspective on the topic of dark patterns. After gaining comprehensive knowledge about the subject and a clear and precise definition of the research questions, the semi-structured interviews were designed. Here, the methodological approach of Adams [Ada15] was followed. The process of conducting SSIs that was implemented can be divided into four phases that were followed. In the first phase, the respondents for the interview were selected. The respondent sampling was done using a mixture of purposive sampling (subjects are selected on researcher's personal judgement) and convenience sampling (sample is taken from a group of people that is easy to reach or contact) as the respondents must have legal expertise and the financial and temporal means for conducting this study were restricted. The sample consists of three legal experts with expertise in the field of eCommerce. It is important to note that qualitative researchers typically prioritize the observation of individual experiences and perceptions rather than striving for statistical representativeness in respondents or situations.[Web10].
In the second phase of conducting SSIs according to Adams [Ada15] questions were drafted and the interview guide was created – a list of questions, which directs conversation towards the research topic during the interview [KPJK16]. The main questions (directly related to the research topic) and follow-up questions (to better understand the viewpoint of the test subjects) were based on the previously mentioned literature review on existing laws that can potentially be violated by well-defined dark patterns and other legal issues that emerged throughout the study. With regards to best practice solutions the interview questions were drafted using well-formulated, participant oriented, not leading, clearly worded, single faceted and open-ended questions [Ada15] [Sea99] [KPJK16]. After a pilot test and minor adjustments of the interview guide, the third phase started, where the semi-structured interviews were conducted in a professional and serious manner [Ada15][Sea99]. The interviews were audio-recorded (and subsequently transcribed) and additional notes were taken by pen and paper during the interview in order to capture emotions and expressions. The interview guide was regarded as work in progress and was refined and adjusted between single interview sessions.
In the fourth and final task of Adams [Ada15] approach, the collected qualitative data was analyzed using qualitative data analysis to answer the research questions and critically examine the dark patterns found from a legal perspective. For the analysis of the qualitative data, Ravindran's [R+19] approach was followed, in which the data was refined, coded, and categorized after processing, and finally theories or concepts were

developed. The specific steps of Adams'- [Ada15] approach and the execution of the semi-structured interviews are explained in more detail in section 3.3.

3.2 Research Questions

This thesis explores the prevalence of dark patterns in the context of Austrian eCommerce and considers dark patterns and the systematic use of web crawlers for dark pattern detection from a legal perspective. The research seeks to address the following research questions:

- **How prevalent are dark patterns among the top Austrian eCommerce onlineshops when identified using a web crawler with regular expression detection?**
- **How reliable is the automated detection of dark patterns with regard to true-positive and false-positive, using a web crawler with regular expressions?**

The study of the prevalence of dark patterns was conducted with an automated approach using web crawling technology. With the aim of questioning the use of the technology from a legal and policy perspective, the following research questions emerged:

- **According to legal experts, can automated methods for detecting dark patterns be used from a legal or policy standpoint?**
- **How do legal experts in Austria assess the effectiveness of current laws and regulations in addressing dark patterns in eCommerce, and what recommendations do they propose for enhancement?**

3.3 Semi-Structured Interviews

Semi-structured interviews with legal experts were conducted to answer the research question on the legal context of dark patterns. In order to better understand the topic of dark patterns from a legal perspective, a semi-systematic literature review (as described in chapter 2.1.2 was conducted. Based on the information gained from this research, a general understanding of the legal basis of dark patterns was created in order to create an interview guide for a semi-structured interview.
To conduct a semi-structured interview, the methodological approach of Adams [Ada15] was followed, which consists of four phases for conducting semi-structured interviews: respondent sampling, drafting questions and creating an interview guide, pilot testing and conducting the interviews and finally analyzing and reporting the results. These four phases are described in the following subsections of this chapter.

3.3.1 Participants

To initiate semi-structured interviews, the initial phase involved the careful selection of interview respondents. Employing a blend of purposive sampling (where subjects are chosen based on the researcher's discernment) and convenience sampling (drawing from a readily accessible group), participants meeting specific criteria of legal expertise were identified. Given the constraints on financial and temporal resources for this study, the sample comprised three legal experts well-versed in eCommerce. It's noteworthy that in qualitative research, statistical representativeness is typically not a primary concern, as the focus lies on exploring individual experiences and perceptions [Web10].
In advance, careful consideration was given to the selection of interview participants to ensure that they possess expertise in the fields of law and eCommerce. Individuals with the relevant qualifications were recruited from the study author's network. The detailed characteristics and attributes of the interview partners can be found in section 5.2.1 of this paper.

3.3.2 Interview Guide and Question Draft

The second phase according to Adams [Ada15] is drafting questions and the interview guide – a list of questions, which directs conversation towards the research topic during the interview [KPJK16]. The main questions (directly related to the research topic) and follow-up questions (to better understand the viewpoint of the test subjects) were based on the previously mentioned literature review on potential law violations and the legal context of dark patterns. With regards to best practice solutions the interview questions were drafted using well-formulated, participant oriented, not leading, clearly worded, single faceted and open-ended questions [Ada15] [Sea99] [KPJK16].
A structured interview guide was developed, beginning with a polite greeting and a thank-you to the study participants/interviewees for their participation. The interviewees were assured that there are no right or wrong answers, emphasizing that the interview aims solely to capture their experiences and assessments to gain new insights into the topic of dark patterns from a legal perspective.
Before commencing the actual interviews, a brief verbal explanation was provided about dark patterns and their purpose, which is to influence users through design using psychological tactics. Following this, a brief explanation was given about what a web crawler is and how it was utilized within the framework of this study.
Participants were informed that they are not obligated to answer any questions, and that they do have the right to refuse to answer any question for any reason. Furthermore, participants were advised that the interview will not exceed one hour in duration, a timeframe that was pre-evaluated and tested through pilot testing.
Subsequently, participants were asked for their consent regarding the recording of the conversation, and if the participant agreed, the recording was initiated.
In formulating the questions, care was taken to maintain a certain structure to support and guide the flow of conversation and encourage participants to speak openly. Initially, simpler and more general questions were posed to break the ice and create a supportive

conversational atmosphere. After that, more critical questions could be asked, with attention given to wording to avoid intimidating or influencing the interviewee. The most controversial, complex, or personal questions were saved for the end of the interview, as by this point, the interviewer was no longer considered to be a stranger [Ada15].
To conclude, demographic questions were addressed (which are typically collected in advance through a pre-survey in other scientific methods). Demographic questions served as a moderation tool and were used in conjunction with a forward-looking perspective to round off the conversation. The interview guide that was used in this study is presented in the appendix 7. Since the semi-structured interviews were conducted in German, the interview guide is also written in German. For the sake of readability, colored markings were removed from the interview guide (these colored markings served the interviewer for better orientation and to indicate the importance of individual topic blocks).

3.3.3 Pilot Test and Conducting the Interviews

To test the interview guide, a pilot test was conducted. The aim of the pilot test was to determine whether the interview questions were clear and understandable, even for individuals without technical background knowledge. Additionally, it assessed the logical structure of the questions to create a conducive conversational atmosphere. Furthermore, the goal of the pilot test was to refine, add, or remove questions to streamline and make the interview questions more relevant to the research topic.
The pilot test was carried out with a single male participant aged 29. The interview process followed the interview procedure described in the previous chapter. After introducing the topic and obtaining consent for recording, the conversation was recorded. The recording served the purpose of assessing the functionality of technical aids and the quality of the recording for subsequent transcription (including volume and clarity of the recording). The pilot test was conducted, and then improvements were collaboratively addressed.

The insights gained from the pilot test were as follows:

1. Reduce interdependencies between questions to allow for optional skipping (e.g., if there's limited time or a question no longer needs to be answered).
2. Prioritize questions to facilitate more precise probing during the conversation or to skip questions if necessary.
3. Provide a concrete example of a dark pattern (preferably in printed form to avoid distractions).
4. Correct minor typos and grammar errors.
5. Formulate follow-up questions more precisely.

These findings were subsequently integrated into the interview guide. As mentioned earlier, the interview guide was considered a work in progress and could be adjusted not

only after the pilot test but also after individual interviews were conducted.
After the pilot test and minor adjustments of the interview guide, the third phase started, where the semi-structured interviews were conducted.

3.3.4 Analyzing and Reporting

In the fourth and final task of Adams [Ada15] approach, the collected qualitative data was analyzed to answer the research question on the legal perspective of dark patterns in Austrian eCommerce.
To analyze qualitative data, the following steps, based on Ravindran [R+19], were followed:

1. Preparation of data
2. Reading and reflecting
3. Coding, categorising and memoing
4. Developing themes/conceptual models or theory

The data preparation was done by cleaning and transcribing the recorded interviews. Sufficient time was allocated for transcription, as one hour of spoken data can take approximately 4-6 hours to transcribe [R+19]. Handwritten notes and data were cleaned immediately after the interview, and if possible, transcription was performed right after the interviews. The data preparation step was considered an iterative process that could be changed at any time. Each iteration, that is, each time the recording was revisited or reread, data could be modified or supplemented.
In the second step, the transcribed texts were read and reread to reflect on and understand what study participants tried to express. The focus of this phase lied on understanding the perspectives, experiences, and expertise of the study participants. While reading, data could also be expanded, supplemented, or questions that arose were noted. In this step, commonalities between interviews could be identified and recorded. It is important to note that with each reading of a text, the subsequent interview could be adjusted to either improve it or delve deeper into the subject by integrating new insights into the follow-up interview.
In the third step, the data was coded, categorized, and memoed. This process was carried out using MAXQDA [MAX23] as a supporting software tool. A latent content analysis was performed in which the data was coded and categorized to identify patterns and themes. Coding was done in individual phases or iterations to control the granularity of coding. Initially, open coding was used to create initial codes that could be used for the following analyses. After that, connections and relationships between the codes were established, and a basic structure between the codes was created. Categorization followed a conceptual coding approach, where not every line was coded and categorized, but the most important and meaningful data was summarized. It was also considered to be also

possible for a hierarchy of categories and subcategories to emerge if the data allowed. Memoing involved documenting the researcher's thoughts and explanations, which also influenced the data analysis.
The final step involved understanding and summarizing fundamental concepts to comprehend connections, relationships, and associations, and interpreting them accordingly.

CHAPTER 4

Implementation and Execution

This chapter provides an overview of the chosen definition of dark patterns and the taxonomy utilized in this study. It also explains which dark patterns were selected for automatic detection using a web-crawler and the reasons behind these selections. Subsequently, it outlines the systematic approach for implementing the crawler.

4.1 Dark Pattern Definitions and Taxonomy

As mentioned in Chapter 2.2.1, the definition of dark patterns lacks consensus within the scientific community. Therefore, in order to maintain consistency with Mathur et al.'s [MAF+19a] study on the occurrence of dark patterns in the eCommerce sector, the same definition was adopted for the purpose of this thesis. ***"Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions."***[MAF+19a]

Moreover, there is no unified or single, correct taxonomy of dark patterns in the eCommerce domain within the scientific community. Some taxonomies are designed to be more general and encompassing all types, while others focus on specific domains such as mobile games, eCommerce, or gambling. To ensure comparability with Mathur's study [MAF+19a], the same taxonomy that was developed and used in their study was employed in this research for the categorization of types of dark patterns.
The taxonomy of Mathur et. al [MAF+19a] contains 15 dark patterns from 7 larger categories. Not all of these dark patterns can be processed and recognized by a crawler with the methodology used (the extraction of text using regular expressions). Thus, it had to be systematically determined in advance which dark patterns can be detected with a web crawler based on textual patterns and which cannot. It was found that some dark patterns cannot be detected in an automated manner (but only manually),

such as the testimonial pattern (where testimonials advertise a product or service whose origin is unknown). Another example of a pattern that cannot be detected using web-crawler technology and text-pattern matching is the Hard to Cancel pattern, where it is easy to sign up for a service, but it is very difficult or even impossible to cancel that service.

Due to the aforementioned constraints, this research focuses solely on identifying specific dark patterns that can be detected using regular expression pattern matching. The study is restricted to the following 5 dark patterns of Mathur´s [MKM21] taxonomy: Countdown Timer, Limited-time Message, Activity Message, Low-stock Message and High-demand Message.

When selecting these patterns, a similar approach was followed as in the work "The design of a framework for the detection of web-based dark patterns" [COG+21]. In their work Curley et. al. evaluated which dark patterns (from Brignull's taxonomy [BLSD23]) can be detected semi-automatically, fully automatically, manually or not at all. A similar approach to Curley's [COG+21] study was used in this study to find out which eCommerce dark patterns can be automatically detected using regular expressions according to Mathur's [MAF+19a] taxonomy. Each dark pattern was coded as follows:

1. **Yes**, the pattern can be detected automatically with RegEx text recognition on the single product detail page
2. **No**, the pattern cannot be detected automatically with RegEx text recognition on the single product detail page

The following table shows the results of the evaluation of every single eCommerce dark pattern that was defined by Mathur et. al. [MAF+19a] in their work:

Category	**Type**	**Cognitive Bias**	**Can be detected using RegEx Approach?**	**Rationale**
Sneaking	Sneak into Basket	Default Effect	No	requires full checkout process to check changes in cart/basket

Sneaking	Hidden Costs	Sunk Cost Fallacy	No	requires full checkout process to compare before and after checkout price
Sneaking	Hidden Subscription	None	No	requires at least one payment action and observation over longer time period
Urgency	**Countdown Timer**	Scarcity Bias	Yes	look for phrases like “only 3 products left”, etc.
Urgency	**Limited-time Message**	Scarcity Bias	Yes	look for phrases like “only available for a short time”, etc.
Misdirection	Confirmshaming	Framing Effect	No	requires natural language processing
Misdirection	Visual Interference	Anchoring & Framing Effect	No	requires interpretation of style (contrast, sizes, placement, etc.)
Misdirection	Trick Questions	Default & Framing Effect	No	requires natural language processing

Misdirection	Pressured Selling	Anchoring & Default Effect, Scarcity Bias	No	requires mechanism to compare all possible product variations to find the most expensive one
Social Proof	**Activity Message**	Bandwagon Effect	Yes	look for phrases like "you and 3 others are currently viewing this product", etc.
Social Proof	Testimonial	Bandwagon Effect	No	highly individual text based on testimonial, length of text and other characteristics
Scarcity	**Low-Stock Message**	Scarcity Bias	Yes	look for phrases like "only 3 left on stock", "low on stock", etc.
Scarcity	**High-Demand Message**	Scarcity Bias	Yes	look for phrases like "was purchased 3x this week", etc.

Obstruction	Hard to Cancel	None	No	requires scan of all (sub-)pages of a website and all hyperlinks and look for potential phrases like “sign-in”, “login”, etc. but no "sign-out", "unsubscribe", etc.
Forced Action	Forced Enrollment	None	No	too many possibilities for automated detection (check for hidden pages, privacy settings, etc.)

Table 4.1: Manual evaluation of Mathur´s[MAF+19a] dark pattern types to find out which dark patterns can or can not be automatically detected using RegEx pattern matching on single product pages

The following paragraphs describe the 5 selected dark patterns (highlighted in table 4.1 above) in more detail.

4.1.1 Countdown Timer

"A countdown timer is indicating to users that a deal or discount will expire using a counting-down timer" [MAF+19a].

The utilization of countdown timers raises significant ethical concerns, particularly when the offer persists even after the timer expires.

The countdown timer exploits the **Scarcity Bias**, a cognitive bias, in users [MAF+19a]. The Scarcity Bias states that things or opportunities seem more valuable when they are less available. People seem to be more motivated by the thought of losing something than

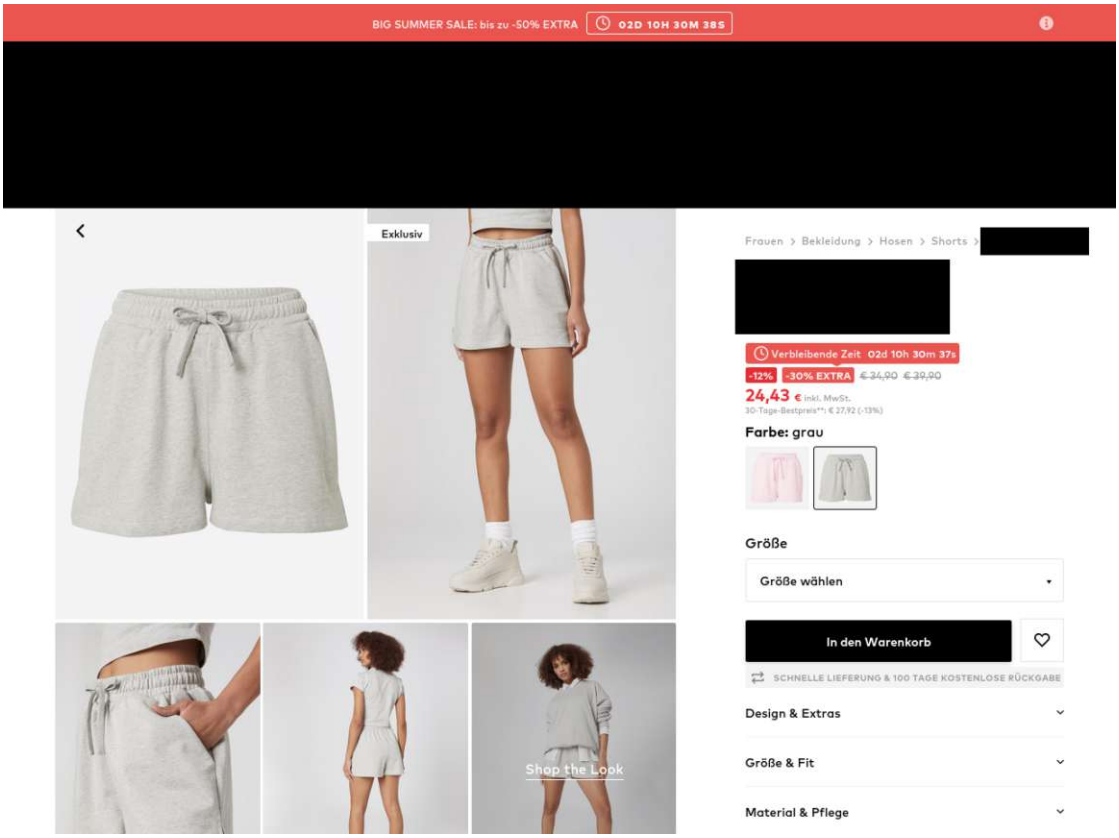

Figure 4.1: Countdown Timer placed above the main navigation section of the page (top bar) and close to the price display of a product

by the thought of gaining something of equal value [CG04]. Especially under conditions of risk and uncertainty, the threat of potential loss plays a powerful role in human decision making [TK74].
Cialdini [CG04] observed that sellers often deceive their customers by claiming that a product is only available in limited quantity or at limited time even though there are no actual/real limitations. To further investigate the concept of a Countdown Timer, we can distinguish between a "true" and a "false" countdown timer. The "true countdown timer" signifies that a resource is actually available under the stated conditions (or at all) until the timer expires. In contrast, the "false countdown timer" merely pretends scarcity to enhance the attractiveness of the offer. The Scarcity Bias is triggered regardless of whether the claim about resource scarcity is true or not. In both instances, scarcity is created, thereby increasing the perceived value of the product in the eyes of the customers. From a psychological point of view, the Scarcity Bias triggers a fear of loss and a desire due to scarcity [CG04]. Furthermore, the (often artificial) scarcity caused by Countdown Timers also deprives people of the freedom to make a free decision.
Countdown Timers also exploit a cognitive bias known as **Hyperbolic Discounting**.

This phenomenon refers to the tendency of individuals to increasingly choose smaller, but immediate rewards over larger, delayed rewards [Lai97]. People generally show a preference for short-term gains rather than considering the long-term consequences. For instance, when selling a car, individuals might opt to receive 2,000€ immediately instead of waiting for 3,000€ after six months. The allure of Hyperbolic Discounting lies in the desire for immediate rewards, as people perceive the present as certain and the future as uncertain. This creates a sense of urgency or greed to obtain immediate gratification. A clear example illustrating the implications of Hyperbolic Discounting is cigarette smoking. Individuals choose the short-term pleasure of smoking while neglecting the long-term health consequences it may cause.
Tversky and Kahneman [TK74] identified numerous situations where consumers do not act rationally or consistently with the aim of maximizing their utility when confronted with cognitive biases. Another cognitive bias that can be triggered by Countdown Timers is the **Loss Aversion Bias**, which states that potential losses are weighted much more heavily than potential (equivalent) gains. Thus, individuals tend to avoid risks in order to prevent potential losses, even when the potential gain is equally large or even greater. The consequence of the Loss Aversion Bias is that people make decisions aimed at minimizing losses rather than maximizing opportunities. For Countdown Timers specifically, this means that users perceive a missed offer as a loss (e.g. you could have saved 10€). This loss should be avoided by accepting the offer and buying the product.
Countdown timers are prominently featured on eCommerce platforms, typically situated close to the product price or the "Add to Cart" button. They are also commonly placed near the main navigation of the website for general promotions. These timers are designed to capture attention, using eye-catching visuals, bright colors (often including signal colors like red), and animated visuals, making them highly noticeable to users.
Detecting Countdown Timers can be achieved by regular expression pattern matching, as the similar phrases and textual patterns are often used by website operators, such as "09:27:52" or "9h 27min remaining". However, detecting a change of product characteristics (e.g. price) after the timer has expired can be technically challenging. For this purpose, one would have to examine the same page over and over again over a long period of time to see how the product (and its characteristics and conditions) change. This way, it could determined whether the displayed Countdown Timer was true or false pattern (that is, whether the price was really reduced while the timer was running or whether the product was really only offered at special conditions during this specified period).

4.1.2 Limited-time Message

"A Limited-time Message is indicating to users that a deal or sale will expire soon without specifying a deadline" [MAF+19a].

In line with the Countdown Timer, the Limited-time Message represents another dark pattern falling under the "Urgency" category, as classified in Mathur's [MAF+19a] taxonomy. Unlike a ticking clock, the Limited-time Message lacks a specific date or time to specify the end of a deal, introducing uncertainty for users. Intentionally vague expressions such

as "limited offer" or "only now" are employed, leaving ample room for user interpretation , which could encompass hours, days, or even months for the expiration of the deal or offer.
The use of a Limited-time Message doesn't guarantee the presence of an actual time constraint; it might be deliberately engineered to create a sense of scarcity. Consequently, the message might either disappear without having any effect on the product characterisitcs or conditions (e.g. price) or persist indefinitely. Such tactics add to the user's uncertainty and may influence their decision-making.

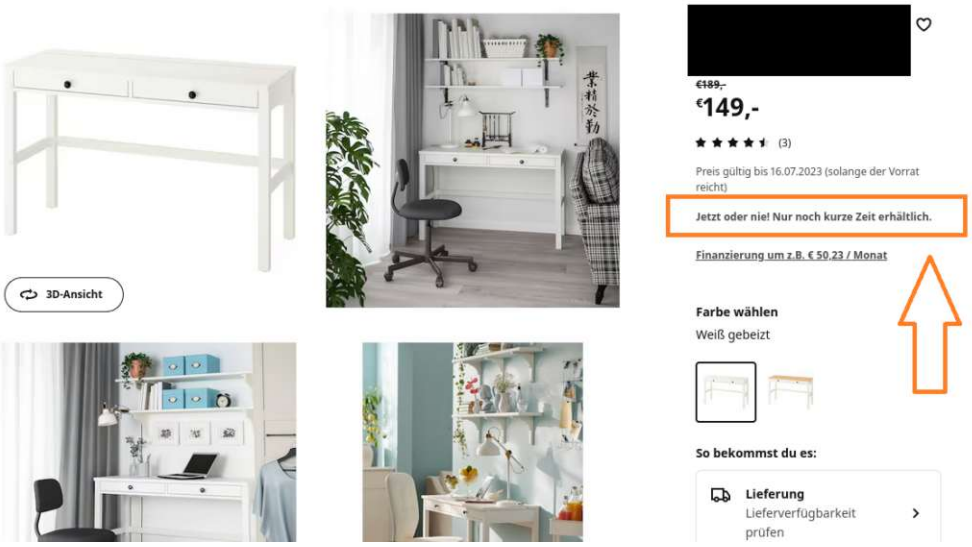

Figure 4.2: Limited-time Message placed close to the price display of a work desk

Similar to the Countdown Timer, the Limited-time Message (along with other dark patterns) leverages the **Scarcity Bias** [MAF+19a] and the **Hyperbolic Discount** [CG04] bias to influence user behavior. Interestingly, the presence of increased uncertainty in Limited-time Messages can amplify the effects of these cognitive biases on users compared to when specific time frames are provided [Lai97].
Limited-time Messages are typically positioned near the "Add to Cart" button or close to the product's shipping conditions. Frequently, these interface elements featuring a Limited-time Message are also placed in proximity to the price display, further accentuating the perceived urgency for users.
Detecting Limited-time Messages primarily relies on identifying specific keywords or phrases like "for a short time only", "while supplies last", or "today only." These explicit cues signal the limited availability or temporal constraint of an offer, aiming to trigger a sense of urgency and encourage immediate action.

4.1.3 Activitiy Message

"An Activity Message is informing the user about the activity on the website (e.g., purchases, views, visits)" [MAF+19a].

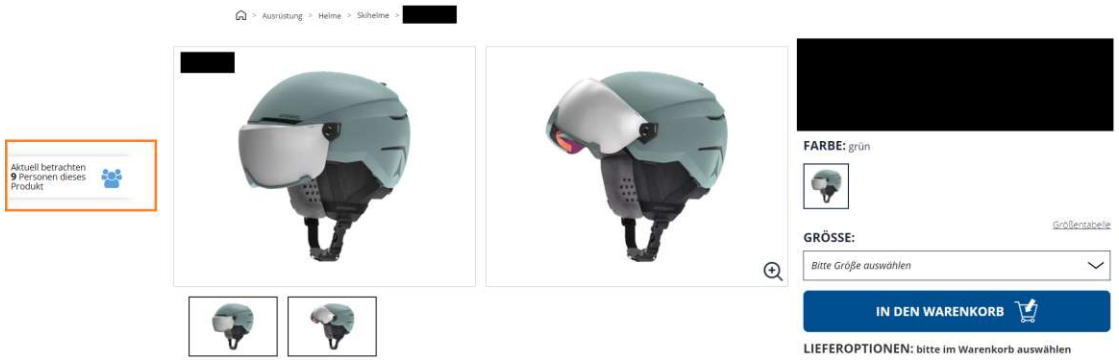

Figure 4.3: Activity Message on the left side of the screen showing the current viewers of a ski-helmet on a product detail page of a sports equipment online shop

Activity messages exploit the **Bandwagon Effect**[MAF+19a] in the user, a cognitive bias that was first described by Sherif [She37] in 1936. The Bandwagon Effect defines the tendency of individuals to value something more because others seem to value it, which means that products or things are perceived as valuable because they are/were perceived as valuable by others.
The metaphorical label of this phenomenon dates from late 19th-century American politics and alludes to the wagon in a parade that carries the band and attracts a large crowd of followers marching behind it to enjoy the music [SB15]. There are many reasons why the bandwagon effect works for humans, such as the desire to belong to a group [Nie18]. Everyone wants to be on the winning team and thinks "if everyone does this, it will be correct" or in the case of eCommerce: "if many people buy this product, it must be good". The bandwagon effect has been studied in politics and marketing for decades now and has also been used in a targeted way. For example, political campaigns are often designed in such a way that as many voters as possible publicly show which party they are voting for (with social media filters on profile pictures, stickers on cars, election advertising by celebrities in public media, etc.). In marketing, for example, it is used on books and shelves with "bestseller" stickers to promote products by showing that many customers bought certain products.
However, it should also be noted that the reverse conclusion, namely "everything many people do is bad" is also not true. There are many examples of popular things (i.e. activities that are done by a large amount of people) that are actually harmful, such as drinking alcohol or eating unhealthy foods. In order not to fall for the Bandwagon Effect, one should become aware of this phenomenon and rationally decide for oneself whether

a certain decision is appropriate or not - independent of the social environment or the decisions of others.
Activity messages, Testimonials and Low-stock/High-demand patterns are often found on tourism websites or booking platforms [LVBB+22]. However, verification of the authenticity of the statements can only be established in the rarest of cases.
As with the other dark patterns of the Social Proof category, it cannot be determined (especially automatically) whether the information about the activity on the page is real or not.
The Activity Message pattern placement is often not too conspicuously, rather subtly.

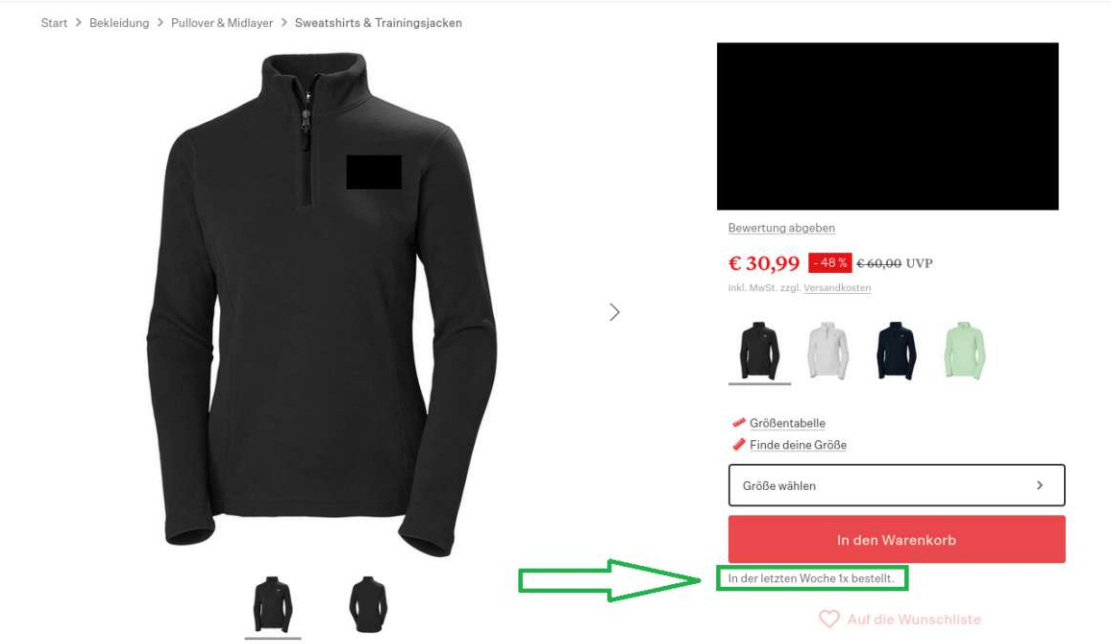

Figure 4.4: Subtle placement of an Activity Message showing that the product has already been sold this week

It is often found below the product description or as a short (often with an animation) insertion at the bottom of the screen.
Activity Messages can be faked with low effort using small code-snippets, for example by selecting a random name and city from a list and then displaying on the page "<randomName> from <randomCity> just purchased <titleOfProduct>". Often Activity Messages indicate the sale of a product (or how many pieces have already been sold) or how many people are currently viewing the page/product. Even the current viewers can be generated automatically by the website operator with a simple script: "You and <randomNumberBetween2and100> other people are currently viewing this product".
Detecting activity messages can be accomplished by employing particular phrases like "this product has <number> observer" or "<firstname> is currently viewing this product."
As previously stated, it can be challenging to verify the legitimacy of Activity Messages

in order to find out whether they are deceptive or not. Both true/real and false/fake Activity Messages can trigger cognitive biases.
Furthermore, it is feasible to automate the detection of Activity Messages generated by random scripts. For instance, this can be achieved by checking whether the script responsible for rendering the message operates on the client-side (e.g., client-side JavaScript) and is therefore not connected to any information about the activity on the page from a server side perspective.

4.1.4 Low-Stock Message

"A Low-Stock Message is indicating to users that limited quantities of a product are available, increasing its desirability" [MAF+19a].

In a true Low-stock Message, the product's stock level is publicly disclosed, indicating the current quantity available for purchase. However, certain companies may intentionally provide false stock figures to create an artificial scarcity. Even when the stock level data is real, there is rarely transparent information about what happens when a product goes out of stock. Questions arise, such as whether the product will be promptly reordered and become available again in substantial quantities after a day or a week, or if there will be changes in the product itself, its batch, or only the price upon the next reorder. Additionally, uncertainty surrounds whether a product will be permanently removed from stock once it becomes unavailable for purchase.
This ambiguity, uncertainty and scarcity contribute to triggering the **Scarcity Bias**

Figure 4.5: Low-Stock Message placed above the price of a foldable sleeping bag presented on the website of an sports equipment reseller

[CG04], which is also triggered by Countdown Timers or Limited-time Messages as described in previous sections of this thesis.
Typically, information concerning stock levels and Low-stock Messages are positioned close to the "Add to Cart" button or within the shipping options section, often adjacent to the product's short description. Detection of various instances of this pattern can

be automated using a web crawler and textual pattern recognition, as they exhibit a similar textual structure, such as "only <number> pieces/products left in stock" or "low on stock."
However, the challenge lies in monitoring how the product evolves over time or understanding the subsequent actions after the product runs out of stock. Even with re-crawling, it may be difficult to ascertain what occurs after the stock level is exhausted, primarily due to the dynamic nature of the web, e.g. if the structure of the product page changes completely. Additionally, the product's URL is sometimes partially constructed dynamically, leading to potential changes in the URL itself, further complicating the exact repetition of a crawling process.

4.1.5 High-Demand Message

"A High-Demand Message indicates to users that a product is in high demand and likely to sell out soon, increasing its desirability" [MAF+19a].

The High-demand Message is commonly employed alongside Low-stock Messages to convey both low stock levels and frequent product demand. Similar to the uncertainty surrounding Low-stock Messages, it remains unclear what actions will follow once the stock level is depleted. When High-demand Messages are utilized, one might assume that the products in high demand are known and thus restocking would occur promptly. However, the lack of transparency in this restocking process and the resulting uncertainty lead to the recurrence of the **Scarcity Bias**[MAF+19a], akin to the cognitive biases triggered by the Countdown Timer, Low-stock Message, and Limited-time Message patterns described earlier.
Automated identification of High-demand Messages is not as straightforward, as the phrasing tends to be more customized and less frequently reused. Nonetheless, certain phrases and keywords, such as "sold out quickly," can be identified using web crawlers with regular expressions, or other textual recognition methods.

4.2 Crawler Implementation

This section describes the methodological framework used when implementing the web crawler for automated dark pattern detection, relying on RegEx pattern recognition. Initially, the process of generating the dataset (referred to as corpus creation) is expounded upon. Subsequently, the formulation of the regular expressions utilized in the detection process is explained. Lastly, the procedures for data storage and the methods employed for data evaluation are detailed.

4.2.1 Corpus Creation

The first step in the implementation of the web crawler is to find and prepare a data set that provided starting points for the web crawler. In order to make the generation of the corpus as transparent and comprehensible as possible, an approach similar to that used

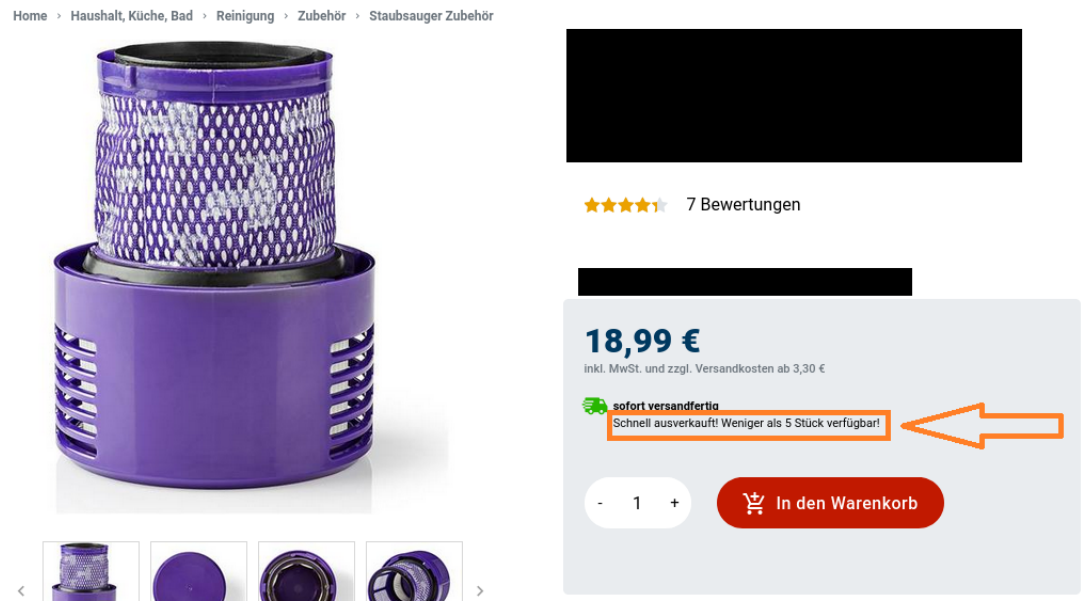

Figure 4.6: High-Demand Message in combination with a Low-Stock Message on a product page for a filter of a vacuum cleaner

in the study by Mathur et. al. [MAF+19a] was followed. To generate a meaningful list of shopping websites, Mathur et. al [MAF+19a] used the Alexa Top Site API to find the most popular websites in the world according to the Alexa ranking. Then, only websites with an eCommerce categorisation were selected and language detection was applied through the home page of the selected website to detect whether the main language of the website is English (as other languages were ignored in the study). Figure 4.7 visualizes the process of data selection Mathur et. al. [MAF+19a] which was implemented in an adapted form in the course of this study.

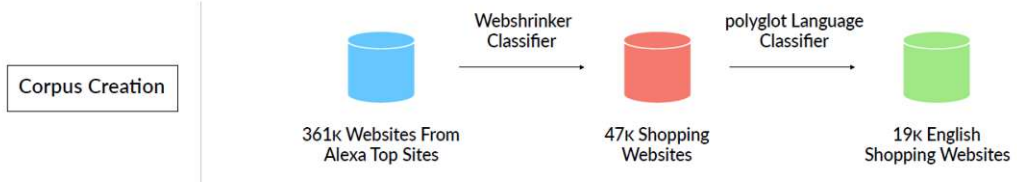

Figure 4.7: Corpus creation process of Mathur [MAF+19a]

For the implementation of this study, the best-selling online shops in the Austrian eCommerce sector (based on turnover) were examined. The data on the top-selling online shops operating in Austria was collected using data from ecommercedb.com [es22]. ECommerceDB is a gathered list of the 250 top-selling online shops in Austria made

available for research purposes free of charge for researchers or university students. During the crawling process a language recognition algorithm was applied in order to identify non-German-language websites to be skipped by the crawler. A secondary dataset was chosen to assess the web crawler's performance on an alternative dataset, thereby verifying its capacity for generalization. To accomplish this objective, geizhals.at [AG23], a prominent online price comparison platform in Austria, was employed. Geizhals maintains a publicly accessible registry of collaborating merchants, comprising companies that interface with Geizhals by furnishing automated price comparison capabilities, primarily through product feeds in XML format. This registry was extracted and reformatted to a suitable structure for integration with the crawler.
These two datasets thus make up the corpus used by the web crawler in this study. In a scientific context, a corpus, as opposed to the simple concept of a data set, is referred to when the data is representative of a particular problem and certain properties are met, such as a selection criterion, balance and homogeneity within the data [JGOTV16]. The described data set qualifies as a corpus according to Jacomy's definition[JGOTV16] as the three properties of selection criterion, balance and homogeneity are fulfilled. Websites with online shops were chosen as the selection criterion, the balance is given by the cross-sectoral selection of all online shops (i.e. not only shops of one industry were selected) and the homogeneity is given because the selection refers to online stores that operate in Austria and are therefore subject to Austrian legislation.

4.2.2 Product-Page Crawling

The afore mentioned corpus (i.e. list of homepages of shopping websites) was used to extract a maximum of 10 product pages for each of these websites/start pages.To accomplish this, a web crawler initiates its search at the homepage and traverses the hyperlinks present on the page until it encounters a product page.
To find a product page, the crawler must first find and select the hyperlinks on the page that are most likely to be product pages. To rank the list of hyperlinks found on the page by their characteristics (length of URL, number of slashes, etc.) according to their probability of being a product URL, a logistic regression classifier was trained - a statistical model for predicting probabilities. The logistic regression classifier is a statistical technique to predict probabilities given input data as numeric features and binary output data (1=true and 0=false for belonging to a certain group). The classifier was trained using the SGD (stochastic gradient descent) classifier from scikit-learn [PVG+11].
The logistic regression classifier is trained on a set of features that are known to be associated with product pages, such as the length of the URL, the number of slashes ('/') in the URL, or the presence of certain words such as "product". The classifier learns the weights of the features that are most predictive of a hyperlink being a product page. Once the algorithm is trained, it can be used to predict the probability of a hyperlink being a product page for any new page.
The logistic regression classifier is able to learn the features that are most predictive of

product pages, even in the presence of noise and other factors that can affect the accuracy of the prediction. This classifier is a valuable tool for web crawlers as it can help crawlers to find product pages more efficiently and accurately. This can save time and resources, and it can also improve the quality of the results that are returned by the crawler.
The source code of Mathur [MAF+19b] published on Github was used as a starting point for the creation of the algorithm. The training dataset was created manually by saving 5 product URLs and 5 other URLs from each of 50 known shopping websites and preparing them in a format that can be processed by the program.
Each URL visited was saved in a list to avoid the crawler visiting the same URL more than once. This allowed alternative paths of the website to be searched, avoided infinite loops and no duplicates were saved.
In the previous steps, hyperlinks were ordered based on the probability of being a product URL. If the crawler decided to open a potential product URL, a Mozilla Firefox instance in the browser would open the selected URL. Subsequently, it must be examined whether the selected URL is actually a product URL. As in the preliminary study by Mathur [MAF+19a], it is assumed that a product URL has (only) one add-to-cart button and that it is placed prominently on the page. For verification purposes, the HTML source code of the opened page is checked and analyzed for the following characteristics:

- Allowed HTML tags (e.g. <a>, <button>)
- Placement of the button (it is assumed that the add-to-cart-button is displayed in the main viewport of the webpage
- Size
- Textual content of the HTML-Element (using Regular Expressions)

These above characteristics were used by the crawler to find out whether the opened web page is actually a product URL.

A large number of parameters have been defined for fine-tuning the web crawler. These parameters are used to keep the efficiency of the crawler high and also to avoid loops. For example, it was defined that a maximum of 100 URLs per web page are examined, the page load time is a maximum of 60 seconds, the crawl depth (measured from the start URL) of 5 URLs, the waiting time on the page before the next link was visited (to minimize automatic bot detection) of 10 seconds.
For every starting point, representing an individual online shop, a dedicated output folder was generated to store the following data:

- A list containing all visited URLs during the crawling process
- A list comprising all identified product URLs
- The HTML source code corresponding to each identified product URL

- Screenshots capturing the visual representation of each identified product URL

Finally, a small script was employed to consolidate all product URLs into a single unified file. This aggregated file served as input for subsequent stages, specifically dark pattern recognition and the storage of data in the SQLite database. Graphic 4.8 shows the procedure of the product page crawler in simplified form. When reading the diagram, please note that implementation details such as some parameter settings or algorithms have been deliberately omitted to make the diagram easier to read.

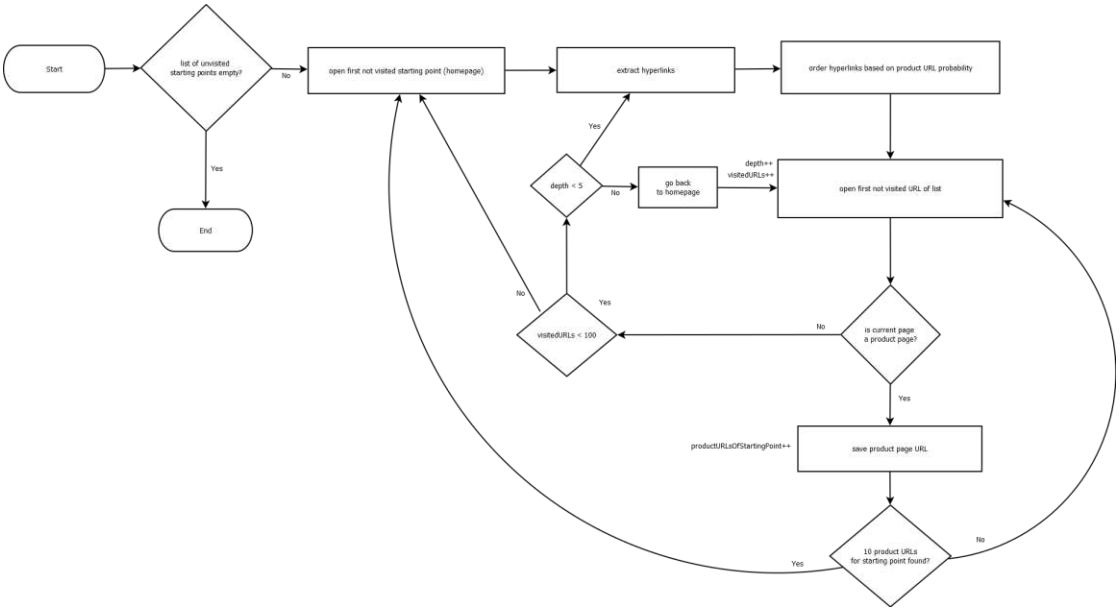

Figure 4.8: Simplified flow chart of the product page crawler

4.2.3 Dark Pattern Detection

In the second phase of the crawler implementation, an approach was adopted that relies on regular expressions to identify and extract text-based elements as required. To prepare the regular expressions, product pages were manually inspected to identify and document their presence. The identified dark patterns, following the taxonomy of Mathur et al. [MAF+19a], along with their corresponding URLs and textual content, were recorded in a list. For instance, a low stock message might be identified by text such as "less than 3 available" or "only a few left in stock."
Once the product pages are collected by the web crawler, the next step was to check these pages for dark patterns. To do this, a segmentation algorithm was used to divide each page into smaller sections, to break down the page into small units that could be observed in isolation. Following that, the resulting sections were examined for the presence of specific patterns using predefined rules. For the creation of these patterns, a

rule set of regular expressions was created in advance. Five different RegEx rules were defined, each designed to detect a specific type of dark pattern. These patterns include Limited-time Messages, Countdown Timers, Low-stock Messages, Activity Messages and High-demand Messages.
The process of creating a regular expression (RegEx) that is suitable for a particular task was considered to be iterative. This means that the RegEx was repeatedly refined and adapted based on the results of testing it on data sets and websites. The iterations of creation for the regular expression rules were as follows:

1. Manual search of the dark patterns under investigation in the top 250 online stores according to eCommerceDB [es22] (5 product URLs per page) and documentation of dark patterns found
   - Origin URL
   - Inner text of the HTML Element (e.g. "only 2 left in stock")
   - Dark pattern type (e.g. "Low-stock Message")
2. Creation of 5 regular expressions (one expression per dark pattern type)
3. Testing the regular expressions using the online tool regex101.com [Dib20] as visualized in the screenshot 4.9

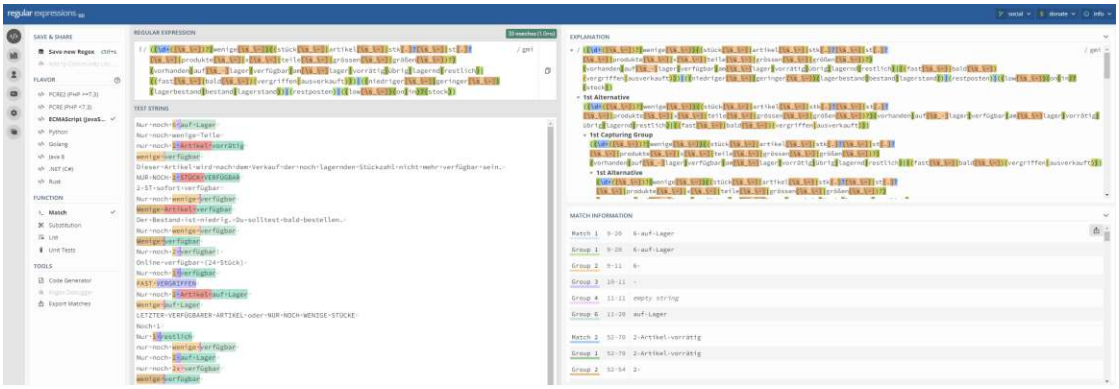

Figure 4.9: Testing regular expressions with the online tool regex101.com illustrated by an example ruleset for Low-stock Messages

Over a period of four weeks, an iteration was run once a week to check possible new releases (in the form of products or even software releases) in the online stores for dark patterns. By repeating this process and continuous testing with supporting tools, a robust RegEx could be developed that could accurately match the desired patterns in the data.
The result of this iterative approach was the creation of 5 regular expressions that could

be used to examine a single HTML element for the presence of certain phrases indicating dark patterns.

4.2.4 Data Collection with Web Crawler

During the data collection phase, previously stored product pages undergo a thorough examination for the presence of dark patterns. To facilitate this process, a custom web scraper was developed that opens and analyzes each URL using a Firefox browser instance. Initially, the scraper attempts to close or dismiss any modals or dialogue windows through JavaScript, ensuring that the resulting screenshot does not solely display pop-ups or cookie bars. The crawler divides HTML elements into smaller units and examines the text contained within these elements for matching phrases based on the predefined regular expressions. When a matching element is identified (indicating the presence of a dark pattern), relevant data along with associated meta information are stored in a database. The database structure is visually illustrated in Figure 4.10.

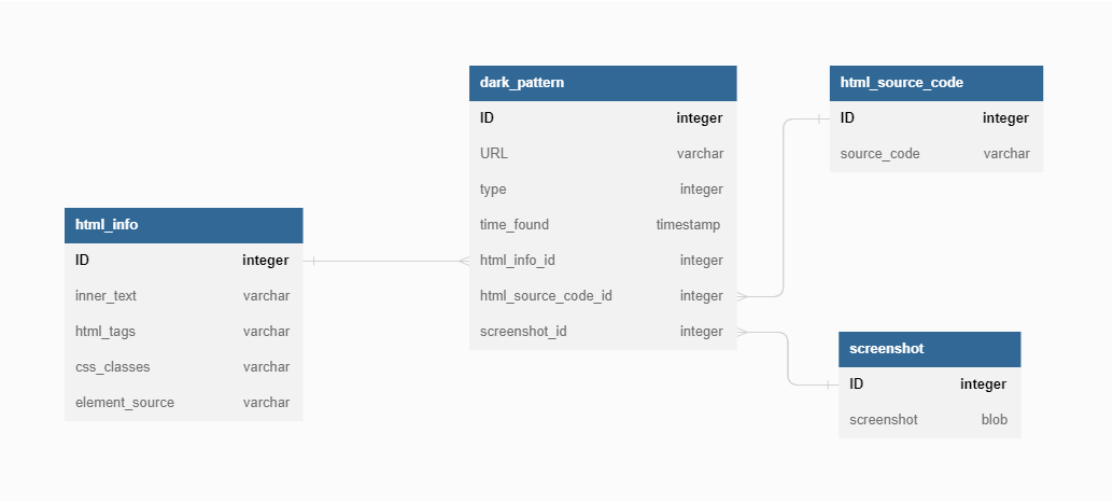

Figure 4.10: ER diagram of the database used to store the data that is used by the web scraper for the extraction phase of the web crawling process

4.2.5 Quantitative Data Analysis

The dark pattern detection process results in a SQLite database that contains information about the detected dark patterns and relevant meta-information. To evaluate the performance of the detection process, each potential dark pattern was manually reviewed to confirm whether it was actually a dark pattern. The database was then updated to include this information.
The database was used to quantitatively analyze the results of the dark pattern detec-

tion process using SQL queries. Exploratory data analysis was conducted to identify meaningful patterns and trends in the data that are represented in chapter 5.1.

4.2.6 Challenges of Software Development

This work expands upon previous scientific research conducted on dark patterns and utilizes the code base developed by Mathur [MKM21] [MAF+19b] to create a web crawler. However, the implementation process of the web crawler encountered various unexpected challenges, leading to difficulties.
The code base lacked proper documentation and lacked instructions on compiling and running the software. Additionally, no specific software versions or environment were provided, requiring significant effort to set up the software. It appears that the primary focus of the project on Github [MAF+19b] is to make the results and data collection accessible rather than to facilitate further studies using the code base. To compile the code by Mathur et al. [MAF+19a], a thorough examination and analysis of the code were conducted to determine a potential starting point. The first step involved identifying the version of Python used. Given that the work was published in 2019, it was assumed that Python 3 (likely version 3.7) was utilized since Python 3.7 was released in 2018 [Fou23] and was considered the most advanced Python version at that time. However, through manual code inspection, it was discovered that a mixture of Python 2 (specifically Python 2.7) and Python 3.7 were employed, deduced from specific code passages. The definitive evidence for the Python version used was the presence of print statements within the code.

Mathur's project [MAF+19a] involved the utilization of multiple Python versions. Specifically, Python 2 was employed for generating the cluster and product URL lists, while Python 3 was used for the checkout crawler, which relied on the OpenWPM framework [EN16]. Consequently, comprehending and interpreting the code posed considerable challenges due to the presence of inconsistencies that lacked explicit explanations or predefined definitions.
To proceed with the project deployment, the following step involved establishing an appropriate development environment. This entailed creating a Python environment and acquiring and compiling all necessary software libraries with the correct versions. However, challenges arose when some software libraries lacked specified version numbers. In such cases, the latest versions were automatically downloaded during Python development. Regrettably, these latest versions often lacked compatibility with the version utilized in the research project. Consequently, a manual search was conducted to identify the latest version of each respective library that was compatible with Python 2.7. During the process of establishing the Python development environment, it was observed that certain libraries were exclusively compatible with Unix systems. Initially, an attempt was made to utilize a virtual machine running Ubuntu, but after a brief testing phase, it was determined that setting up a physical machine with Linux would be more suitable. Ubuntu 16.04 [Ltd18] was installed to provide a Unix system with long term support. However, Mathur's work [MAF+19a] did not specify precise hardware requirements, only

mentioning the usage of two “off-the-shelf-PCs” equipped with quad-core CPUs and 16GB RAM. Due to limited financial resources for this study, a device provided free of charge, an “HP ProDesk 600 G6 Desktop Mini PC” with a dual-core processor and 16GB RAM, was utilized.

Hardware Specifications
HP ProDesk 600 G6 Desktop Mini PC

CPU	Intel Core i5-10500T; 2.3 GHz; 12 MB Cache; 6 Cores and 12 Threads
RAM	16GB
Graphics	Intel UHD Graphics 630
Storage	512GB SSD
OS Type	64-Bit
OS	Ubuntu 16.04 LTS

Efforts were made to identify an entry point to initiate the software in order to complile and subsequently run it. After creating the corpus, the follow-up task involved generating a product URL list. However, during the compilation of the corresponding code, it was discovered that the provided logistic regression classifier was already pretrained using a SGD classifier which was unsuitable for this study’s needs. The SGD classifier’s purpose is to assign weights to hyperlinks on a website based on specific features such as URL length, character usage, etc., based on the probability of them being a product URL. Unfortunately, the pretrained data could not be utilized since it was trained specifically for English URLs, while the websites used in this project predominantly employed German as their primary language. To address this, a new classifier needed to be trained, necessitating the manual creation of a CSV-formatted input file. The CSV file consisted of three columns: the first column being the domain of the website, the second column contained the product URLs, while the third column contained a binary value indicating whether the listed URL was a product URL or not. To accomplish this, 5 pages each from 50 websites were searched for both product URLs and non-product URLs, resulting in a CSV list comprising 500 labeled URLs. This dataset served as the training data for the machine learning code associated with the SGD classifier. The following screenshot shows a section of the generated CSV file for the training of the Logistic Regression Classifier. The first column contains the domain of the website (which is usually also the homepage), the second column contains the URL of a subpage of the website and the third column contains a label with "1" or "0", where the value "1" means that it is a product URL (or a product detail page) and the value "0" means that it is not a product URL.

CHAPTER 5

Results

In this chapter, the results of the study are presented and explained, which will subsequently be used to answer the research question. The results of the quantitative analysis on the performance of the two crawlers (product URL crawler and dark pattern detection crawler) and the qualitative results from the semi-structured interview with experts on the topic of legal compliance of selected dark patterns in the eCommerce sector are discussed.

5.1 Quantitative Results - Dark Pattern Prevalence

To examine the prevalence of dark patterns in the Austrian eCommerce market using a web crawler, the following two datasets were selected:

1. **eCommerceDB Dataset** - a list of the 250 most successful eCommerce enterprises in the Austrian market (based on sales volume), sourced from eCommerceDB [es22]
2. **Geizhals Merchants Dataset** - a list of merchants in Austria obtained from Geizhals [AG23], an online platform for price comparison, encompassing 583 merchants

While appraising the outcomes, it's important to bear in mind that the eCommerceDB dataset served as the foundation for constructing the web crawler. This involved manual scanning of eCommerceDB web pages to facilitate automatic detection of product pages and formulation of regular expressions. Therefore, the crawler's construction is rooted in the eCommerceDB data, akin to a quasi-training dataset. Nonetheless, it's not a conventional "training" dataset in the sense of machine learning, as it doesn't employ machine learning algorithms, and its dynamic nature, where web pages can change with each crawl-run, sets it apart.

5.1.1 Analysis of the Performance of the Product Page Crawler

The task of the product page crawler was to find ten product URLs per website based on a specific start URL, usually the home page of an eCommerce website. The crawler starts at a given website and follows the hyperlinks on that page. For each page that it visits, it checks to see if the page is a product page. If it is in fact a product page, the crawler stores the URL in a list. The crawler continues to follow hyperlinks on each page until it has found 10 product URLs or if other exit parameters are met (e.g. time spent on site exceeded, more than 100 URLs checked, etc.). The implementation of the product page crawler is based on the code base published by Mathur et. al. [MAF+19a] on Github [MAF+19b]. When analysing the crawler for specific performance metrics, the hardware and technology used must also be considered.
The performance of the product URL crawler is summarized in the following table:

	ecommerceDB	**geizhals**
dataset size (n)	250	583
possible URLs to be extracted	2500	5830
URLs extracted	1838	4265
false-positives (absolute)	330	669
false-positives (relative)	17.95%	15.69%
true-positives (absolute)	1508	3596
true-positives (relative)	82.05%	84.31%
true-positives relative to possible URLs	60.32%	61.68%
crawl time in minutes	914.75	1806.67
avg. crawl time per website in minutes	3.659	3.099
number of URLs visited	10823	22310

Table 5.1: Performance of the product page crawler for both datasets (eCommerceDB Top 250 & Geizhals Merchant List)

The eCommerceDB dataset comprises a list of 250 websites or online shops that were used as starting points for the product page crawler. As described in the methodology chapter 4.2, the crawler attempted to extract and store 10 product URLs from each online shop. This results in the total number of theoretically possible URLs for extracting from 2500 pages (250 online shops with 10 product URLs each).
Out of the 2500 theoretically possible URLs, 1838 potential product URLs were extracted. When interpreting this number, it should be taken into consideration that there are websites that cannot be crawled or have more complex mechanisms for accessing the webpage, such as mandatory logins or specific routing based on country or shop.
These 1838 product URLs were then individually and manually examined to determine whether they were indeed product URLs. Through this manual analysis, it was found that out of the 1838 potential product URLs, 1508 actual product URLs were identified. This results in a success rate of 82.05% (actual URLs / potential URLs). The Product

Page Crawler mistakenly identified 330 URLs as product pages, accounting for 17.95% of the potential URLs. The success rate of the web crawler in finding and retrieving product URLs was 60.32%. This was calculated by dividing the number of product URLs that were found and actually correct (1508 URLs) by the number of theoretically possible product URLs (2500 URLs).
The total runtime of the crawler for the eCommerceDB dataset amounted to 914.75 minutes, which translates to 15.25 hours. The crawler spent an average of 3.66 minutes per page (total runtime / number of examined pages) and visited 10823 pages while crawling the dataset.

Similarly, the product page crawler was also applied to the Geizhals dataset, which comprises a list of 583 merchants from the online price comparison platform Geizhals[AG23]. The list encompasses 583 online shops, and 10 product URLs were sought from each, resulting in a theoretically possible count of 5830 pages (583 pages with 10 product URLs each).
A total of 4265 potential product URLs were extracted, and they were manually verified for accuracy. This led to the identification of 3596 correct product URLs and 669 (15.69%) false positives, which were erroneously identified by the crawler as product URLs. Thus, the success rate amounts to 84.31% (actual URLs / potential URLs). The true positive product URLs in relation to all theoretically possible product URLs yield: 61.68%.
The total runtime of the crawler for the Geizhals dataset was 1806.67 minutes, equivalent to 30.11 hours. The crawler expended an average of 3.1 minutes per page (total runtime / number of examined pages) and visited 22310 pages while crawling the dataset.

5.1.2 Analysis of Dark Pattern Detection Crawler

The task of the dark pattern detection crawler was to check the product URLs found in the previous step by the product page crawler for the presence of the in section 4.1 selected dark patterns. The crawler opened each page from the afore created product URL list with an instance of a Mozilla Firefox Web-Browser [Cor98] and checked whether certain text segments are present on the page that indicate dark patterns by using regular expression pattern matching. In addition to regular expression pattern matching, other criteria are used to detect dark patterns, such as the positioning of the element, the visibility within the viewport of the user (using a 1980x1080px resolution screen), the presence of predefined exceptions, etc.
The result of the dark pattern detection crawler was stored in an SQLite database, which contains valuable information about the dark patterns, such as which HTML element was used, CSS classes of the element, screenshots as proof of the presence of the pattern at the time of the crawl and other attributes that can be used for evaluation. With the help of this database, SQL statements were used to analyse the data. The following key figures were calculated:

Prevalence Results - absolute numbers

	eCommerceDB	**Geizhals**
dataset size (n)	250	583
websites with >=1 DP	40	61
websites without DP	210	522
websites with >=2 types of DP	8	6

Table 5.2: Prevalence analysis of dark patterns in absolute numbers

Prevalence Results - relative numbers

	eCommerceDB	**Geizhals**
websites with >=1 DP	16%	10.46%
websites without DP	84%	89.54%
websites with >=2 types of DP	3.20%	1.03%

Table 5.3: Prevalence analysis of dark patterns in relative numbers

As part of the process of identifying dark patterns using web crawlers, two datasets were examined: firstly, the dataset comprising the top 250 revenue-generating Austrian online shops from ecommercedb.com [es22], and secondly, the Geizhals merchant list, consisting

of 583 merchants listed on the online price comparison platform geizhals.com [AG23]. The subsequent paragraphs present the results regarding the prevalence of dark patterns, followed by the outcomes related to the crawler's performance. The following chart 5.1 visualizes the data from the table with the results on the prevalence of dark patterns in relative numbers, which are then elaborated.

Websites using Dark Patterns -
websites where at least one instance of dark patterns was found

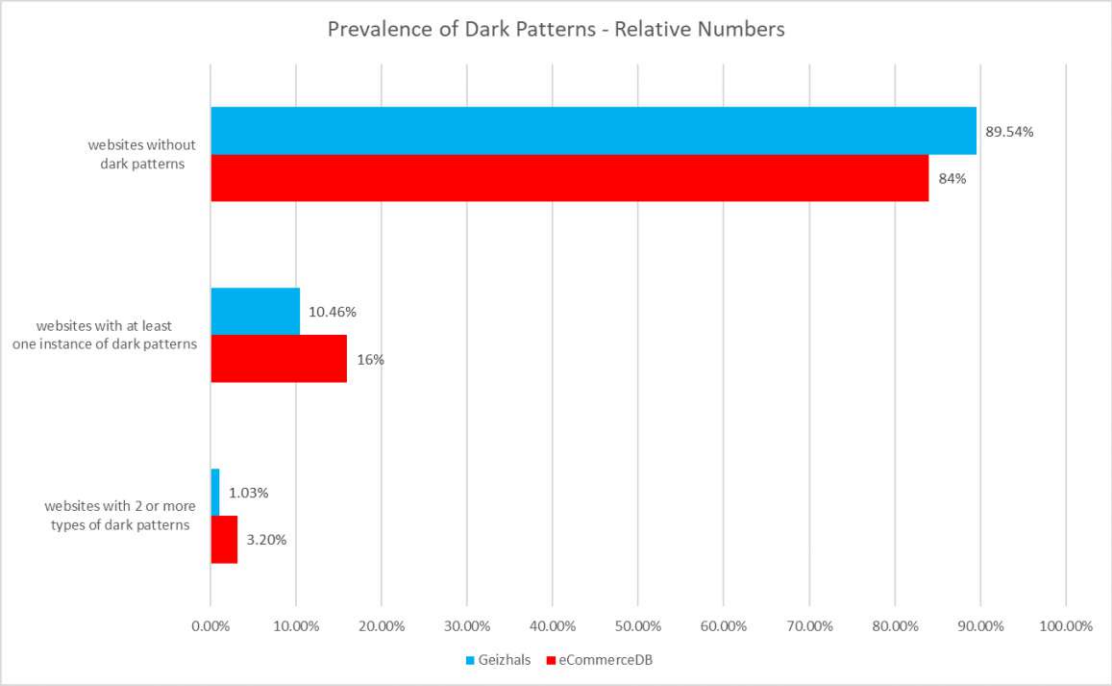

Figure 5.1: Visual representation of the prevalence of dark patterns relative to the number of websites investigated.

Among the top 250 revenue-generating online shops in Austria, 40 online shops incorporate at least one instance of the examined dark patterns. This implies that 16% of the scrutinized online shops employ dark patterns, while 84% of them showed no indications of dark patterns through automated crawler methodology. Moreover, out of these 40 online shops, 8 utilize at least 2 variations of dark patterns (such as a Low-Stock Message and a Countdown Timer) – this accounts for 3.2% of all examined online shops in the eCommerceDB dataset.

In the Geizhals merchant list dataset, 61 online shops integrate at least one instance of the examined dark patterns. This translates to approximately 10.46% of them implementing dark patterns on their websites, while the remaining 89.54% do not utilize any dark

patterns. Among these online shops, 6 were found to incorporate a minimum of two variations/types of dark patterns – representing 1.03% of the total.

Tables 5.4 and 5.5 below explains the performance of the dark pattern detection crawler. The performance is evaluated by two main metrics: (1) true-positives, i.e. the number of dark patterns detected by the crawler that are actually dark patterns after manual examination and (2) false-positives, i.e. the number of dark patterns detected by the crawler that turned out not to be dark patterns after manual examination.

Dark Pattern Detection Crawler Performance Results - absolute numbers

	eCommerceDB	**Geizhals**
dataset size (n)	250	583
total number of potential DPs detected	187	299
true-positives	163	233
false-positives	24	66

Table 5.4: Performance analysis of dark patterns detection in absolute numbers

Dark Pattern Detection Crawler Performance Results - relative numbers

	eCommerceDB	**Geizhals**
true-positives	87.17%	77.93%
false-positives	12.83%	10.46%

Table 5.5: Performance analysis of dark patterns detection in relative numbers

In the eCommerceDB dataset, a total of 187 potential instances of dark patterns were identified. Potential instances in this context refer to cases where the automatically detected dark pattern instances were subsequently verified manually for accuracy. Out of these 187 potential dark patterns, 163 recognized instances proved to be accurate (constituting 87.17%). The remaining 24 (12.83%) instances were false-positives, meaning that the web crawler erroneously identified them as dark patterns.

From the Geizhals merchant list dataset, 299 potential instances of dark patterns were identified, with 233 of them correctly recognized (amounting to 77.93%). This resulted in 66 false-negative outcomes (10.46%), wherein the web crawler mistakenly identified them as dark patterns when they were not.

The following chart 5.2 visualizes the performance of the dark pattern detection crawler. The left bar visualizes the evaluation of the crawler's performance on the eCommerceDB

top 250 eCommerce companies in Austria (based on annual sales), whereas the right bar contains the list of all Geizhals merchants. The green area of the bar shows the true-positives, i.e. the data that was detected by the crawler as dark patterns and subsequently also manually confirmed as dark patterns. The red area, on the other hand, visualizes the false-positives, i.e. the data that was marked as a dark pattern by the crawler but turned out to be no dark pattern after manual examination.

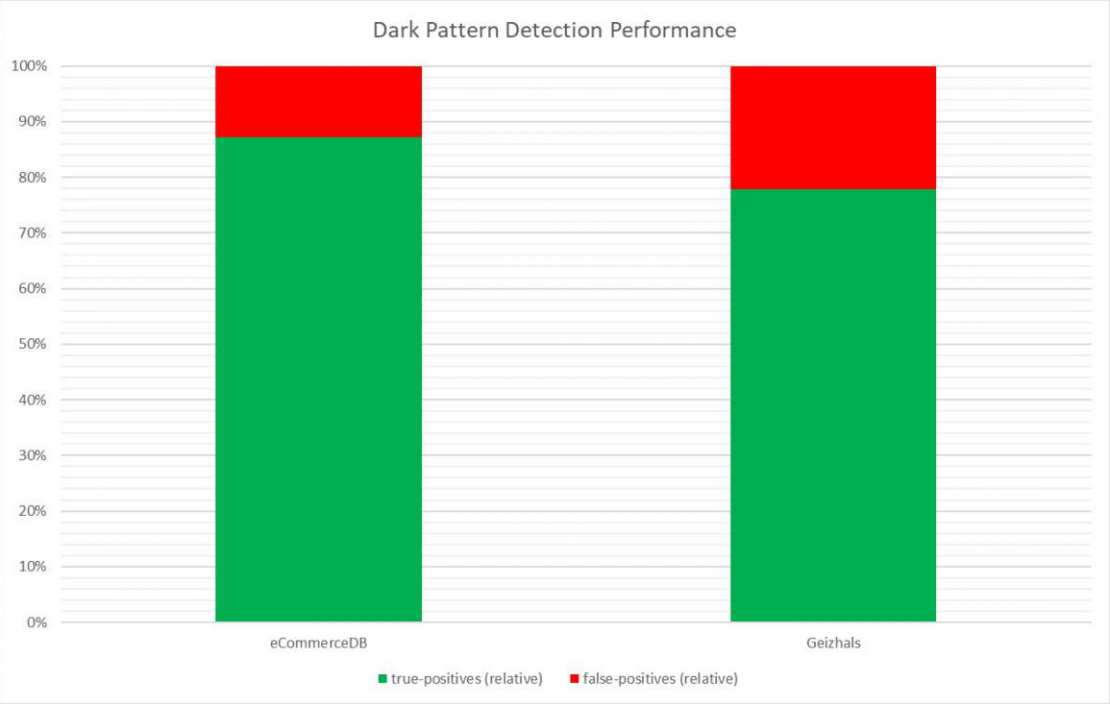

Figure 5.2: Visualisation of true-positive (green) and false-positive (red) detected dark patterns by the web-crawler for both datasets

Furthermore, an analysis was conducted to determine the industries employing dark patterns. For this purpose, each online store that contains at least one instance of dark patterns has been assigned a category/industry. Finally, the number of online shops using dark patterns was counted for each industry, and both the absolute and relative figures were documented. Table 5.6 presents the results of this procedure in tabular form, wheres chart 5.3 visualizes these results in a bar chart (eCommerceDB data bars are visualized in red and Geizhals data bars are visualized using blue color).
The industries with the highest prevalence of dark patterns are Electronics & Technology, Sports & Leisure and Home & Furniture. When interpreting the data, it's essential to consider that there is no equal distribution of industries, especially in the Geizhals dataset. The fact that the eCommerceDB list, as well as the Geizhals merchant list, lacks information on the industries of the online stores makes the analysis by industry difficult. In the context of this study, each of the more than 800 online stores was not categorized manually, which is why the percentage of online stores in each industry that

use dark patterns is not indicated. Therefore, it is possible that a significantly larger number of companies from the Electronics & Technology category are included in the dataset compared to, for example, online shops categorized under Jewelry & Accessories. This fact should be considered when interpreting the absolute numbers of detected dark patterns per industry.

Industries that incorporate dark patterns

	Geizhals Merchants	eCommerceDB
Fashion & Clothing	0	7
Electronics & Technology	25	12
Home & Furniture	8	2
Groceries & Beverages	2	1
Sports & Leisure	8	5
Jewelry & Accessories	1	2
Health & Beauty	2	2
Toys & Games	6	2
Crafts & DIY	1	1
Automotive Accessories & Parts	1	0
Online Marketplace	1	5
Garden & Garden Tools	1	0
Other	5	2

Table 5.6: Industries that incorporate dark patterns in absolute numbers per dataset

Distribution of Dark Pattern Types

Furthermore, the distribution of the individual instances of the examined dark patterns was evaluated. This evaluation showed how many instances (in absolute and relative numbers) of Low-stock messages, Countdown Timers, High-demand Messages, Activity Messages and Limited-time Messages were found.

	eCommerceDB		Geizhals	
	absolute	relative	absolute	relative
Low-stock Message	116	62.03%	227	75.92%
Countdown Timer	52	27.81%	36	12.04%
High-demand Message	10	5.35%	29	9.70%
Activity Message	8	4.28%	5	1.67%
Limited-time Message	1	0.53%	2	0.67%

Table 5.7: Distribution of dark pattern types by type for both datasets

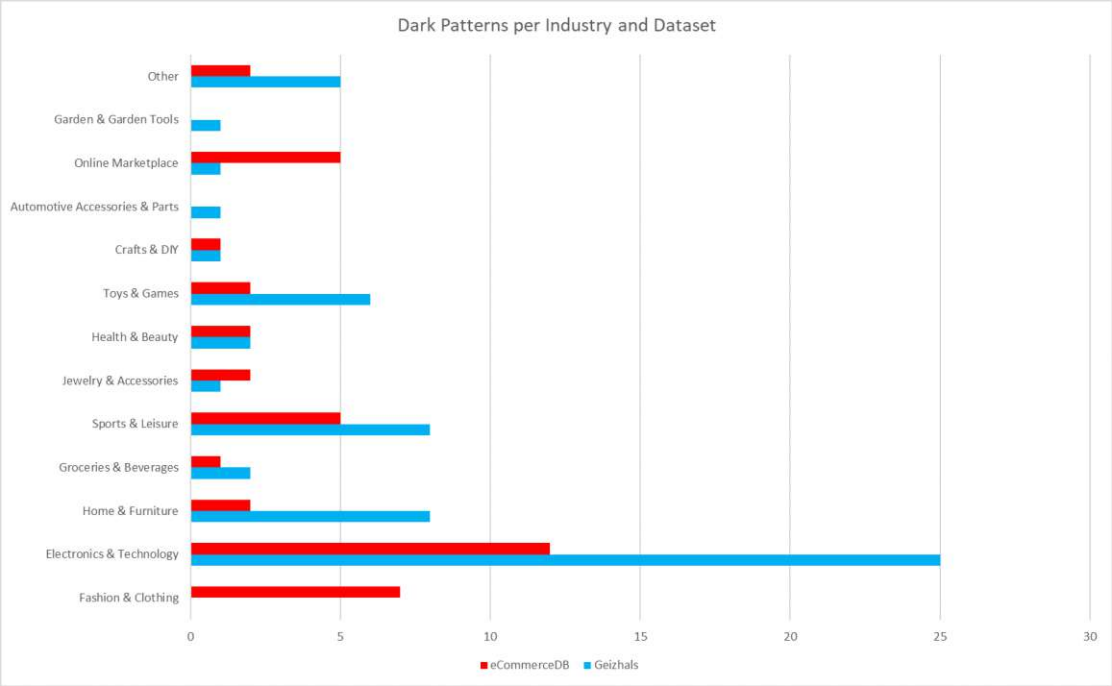

Figure 5.3: Instances of dark patterns detected by automated web crawler approach categorized by industries for both datasets

In the eCommerceDB dataset, as mentioned above, 187 instances of dark pattern were found. Of these, 116 instances of Low-stock Messages, 52 instances of Countdown Timers, 10 instances of High-demand Messages, 8 instances of Activity Messages and one instance of a Limited-time Message were found.

A total of 299 instances of dark patterns were found in the data set of the merchant list of the price comparison platform Geizhals. Of these, 227 instances of Low-stock Messages, 36 instances of Countdown Timers, 29 instances of Activity Messages, 5 instances of Limited-time Messages and 2 instances of High-demand Messages were found.

Thus, over 60% of the types of dark patterns found were low-stock messages on both datasets. The second most common type of dark patterns detected were countdown timers.

5.2 Qualitative Results - Semi-Structured Interviews

In the context of semi-structured interviews, three legal experts were interviewed to gather their assessments and opinions regarding the issue of dark patterns in Austria. This section describes the application of the methodological approach of semi-structured interviews with legal and policy experts described in chapter 3.3. The interviews primarily

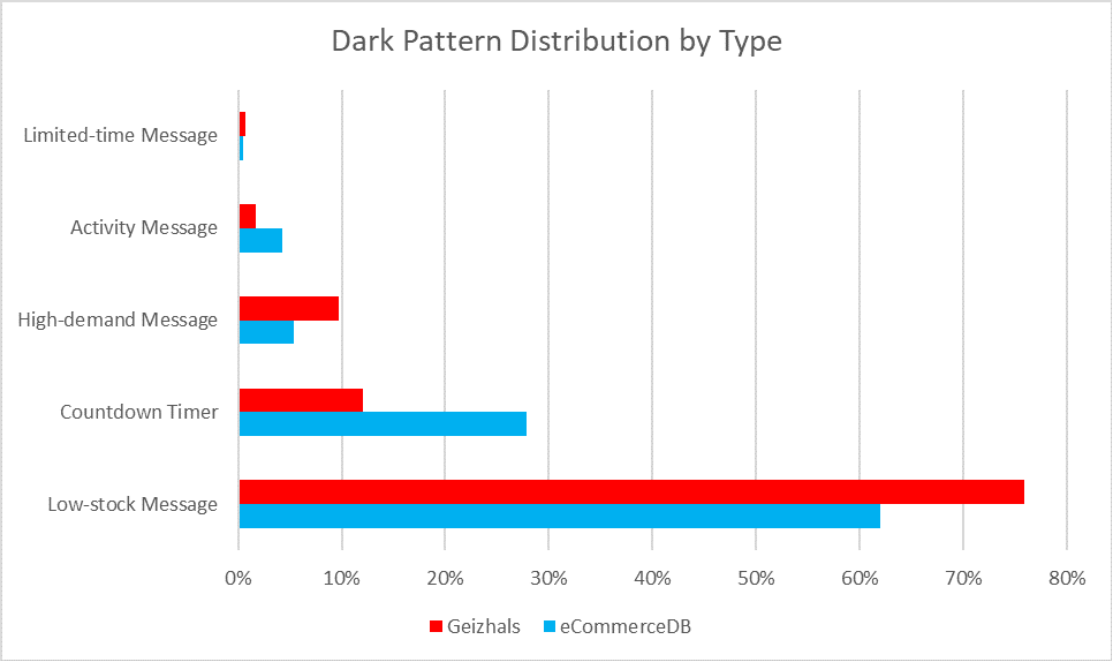

Figure 5.4: Distribution of dark patterns of both data sets in relative comparison

focused on legal expert opinions regarding the current state of regulation concerning dark patterns in Austria, the utilization of technology (specifically web crawlers) for automated detection and a prospective analysis from legal and political perspectives. Below, the results of the qualitative analysis of the semi-structured interviews are presented.

5.2.1 Interview Partcipants - Legal Experts

As described in detail in chapter 3.3.1, interviewees were selected with a mixture of purposive and convenience sampling to discuss the legal and political status of dark patterns.
In selecting interview participants, care was taken to ensure that all respondents possess experience in dealing with eCommerce and have received university-level education in legal studies. Additionally, attention was given to covering various legal domains, as participants had different specializations during their academic studies and work in diverse industries. This approach aimed to provide a comprehensive overview of the current landscape from diverse perspectives.The table below illustrates the characteristics of the interview partners.

5.2.2 Current Situation and Legal Framework

All interview participants indicated that the regulation of online shops is a current, legally relevant, and prevalent topic. In universities, legal fields with a digital focus, such as private IT law, domain law, and eCommerce law, are increasingly being taught. In the

	Participant A	**Participant B**	**Participant C**
age-range	20-29	20-29	30-39
highest level of completed education	LLM (WU)	Mag. iur.	Mag. iur.
legal specialization	civil right	public law	data protection
current employment status	employed, full-time	employed, full-time	employed, full-time
citizenship	Austrian	Austrian	Austrian
online shopping frequency	several times a week	several times a week	once a week
professional area	legal department in the hospitality industry	legal department in a public office at the state (Bundesland) level	data protection in a public office at the national level

private sector, the subject of online shops is highly prominent, leading to a growing demand for legal experts in areas such as data protection and eCommerce law. Moreover, the experts in the public positions have claimed that there is currently a significant shift in the public sector at the EU level due to the release of new directives and laws, namely the AI Act, Data Governance Act, Data Act, Digital Service Act, and Digital Market Act.

While all legal experts may not have been familiar with the term "dark patterns," they were well acquainted with the concept behind it. All participants were able to spontaneously list examples of websites employing dark patterns. Participants notably observed dark patterns, such as Activity Messages, High-demand Messages, and Countdown Timers, predominantly on booking platforms. Additionally, manipulative techniques, such as pre-selected checkboxes on multiple pages and the linking of multiple conditions to a single checkbox (prohibited by the GDPR's prohibition on bundling), were mentioned.

The regulation of online shops in Austria, including regulations concerning eCommerce dark patterns, can be subject to various laws. Throughout the interviews, the following laws were identified as potentially violated by the use of dark patterns in the context of online shops:

- Allgemeines Bürgerliches Gesetzbuch (ABGB)
- Gesetz gegen unlauteren Wettbewerb (UWG)
- E-Commerce-Gesetz (ECG)
- Konsumentenschutzgesetz (KSchG)
- Fern- und Auswärtsgeschäfte-Gesetz (FAAG)
- Datenschutzgesetz (DSG) and Datenschutzgrundverordnung (DSGVO)
- Allgemeines Unternehmensgesetzbuch (UGB)

This list provides an overview of laws dealing with the regulation of purchase contracts, fair competition, data protection, and other legal areas that could potentially be violated

by dark patterns. This list is a collection of all laws that were mentioned during the interviews with all participants regarding dark patterns and this list may be incomplete. Another reason why this list may not be complete is the fact that due to the rapid development and proliferation of online shops, new laws are continuously emerging.

Participant A (who was the first to be interviewed) noted that not all law texts are generally formulated; however, specific laws may be written in a manner that allows for broad applicability to a wide range of situations or cases. To apply them, individual circumstances must be examined, and the law is then assessed or interpreted in relation to the specific case. This involves the use of so-called "general clauses." During the subsequent interviews, participants 2 and 3 were asked for their opinions on this statement. They confirmed that the broadly formulated language in legal discourse is a common stylistic device aimed at ensuring a wide applicability.

In the course of a follow-up research on this topic, the following statement was found: "General clauses enable adaptation to changing needs and perspectives, as well as changes in constitutional law, without the need to alter the literal wording of the law. At the same time, significant powers are delegated to the individual judge: ultimately, they become a legislator and must fill in open legislative evaluations"[MFS17]. These general clauses make it difficult to make a clear statement about whether a specific dark pattern violates a particular Austrian law.

Decisions by the Austrian Supreme Court (Oberste Gerichtshof - OGH) or the European Court of Justice (Europäischer Gerichtshof - EUGH) create precedents for specific situations (or in this case, phenomena like specific types/variations of dark patterns). These precedents are used for future decisions and are understood as a form of "legislation" [MFS17].

General clauses create gray areas and loopholes that can be exploited. Furthermore, all respondents also reinforced the assertion that the internet is always faster than the law and that the emergence of new laws also leads to new loopholes. Participant C noted that "loopholes have always existed, and it is the profession of lawyers (and, for example, tax consultants) to exploit these loopholes to achieve the best possible outcome for their clients".

5.2.3 Pressure on Purchasing Decisions in eCommerce

This thesis investigates Austrian online shops using a web crawler to identify the presence of five selected eCommerce dark patterns. Four out of these five dark patterns under consideration exploit the Scarcity Bias, a cognitive bias indicating that people perceive something as more valuable when it is scarce or rare [TK74]. This scarcity creates pressure on the user or potential buyer. To examine the legal aspect of pressure in purchasing decisions, legal experts were consulted.

In the case of Low-stock Messages, a distinction must be made between "true" and "false" Low-stock Messages. True Low-stock Messages are those where the indication of low

stock accurately reflects the reality. False Low-stock Messages, in this context, mean that the message about low stock is not true, and the product is still available in sufficient quantity.

All three experts agreed that, true low-stock messages, or true patterns in general, have relatively less legal significance. However, false Low-stock Messages or false patterns are legally more complex. In response to the question of whether low-stock messages violate item 7 of the "black list" (Annex UWG), all of the experts indicated that it could be a violation under certain circumstances (if the indication of low stock is untrue and other criteria is met). Point 7 of the black list of the UWG states (in German): "The incorrect claim that the product will only be available for a very limited time or for a very limited time only under certain conditions to induce the consumer to make an immediate decision, so that he has neither time nor opportunity to make an informed decision."

A case-by-case examination in court would determine what "very limited time" and "immediate decision" mean in this context. This interpretation depends heavily on the circumstances of the purchase process. Participant B explained in more detail that factors such as the product's price, target audience, and type of product (utility item, luxury good, etc.) play a role. Decision-making time for purchasing a car, for example, is considered to be longer than for buying a chocolate bar. Furthermore, Participants A and C claimed that the entrepreneur's intention is also relevant to the case – whether the entrepreneur was aware that a false claim was made to induce customers to purchase.

For the actual examination of the violation of the UWG and a point from the black list, specific examination schemes exist, which, however, is not detailed here, as it would exceed the scope of this study.

Participants B and C indicated that the misrepresentation of a stock level could possibly (after closer examination considering all factors) be interpreted as "fraud" (§ 870 ABGB). For the examination of fraud exists a specific examination scheme consisting of the following examination steps:

1. Objective factual situation
2. Subjective factual situation
3. Illegality
4. Causality

In both cases (violation of point 7 of the black list or fraud under ABGB), it is very difficult or nearly impossible for the plaintiff to prove that the statement (regarding stock status) is incorrect. All legal experts did not want to make a blanket statement about a violation in this regard.

5.2.4 Criticism of the Legal Regulation of Onlineshops in Austria

In dark pattern literature within a legal context, the legal system is criticized, particularly the way laws are created, emphasizing the slow process and the assumption of a rational decision-maker. For this reason, participants in interviews were asked about their opinion on the following quote: "The law is stuck to the image of the rational decision-maker. Currently, it does not protect rationality – it presupposes it."

In this regard, all legal experts agreed that a kind of baseline – an average consumer – must be defined as a neutral standard for our legal system to function. This average consumer must possess the ability to act and transact business, making thoughtful decisions – rationality, in this sense, can be interpreted as "reasonable". The principle behind this is similar to the principle of trust in the Austrian Road Traffic Regulations ("Vertrauensgrundsatz in der Straßenverkehrsordnung") – without the principle of trust, traffic could not function, as stated by Participant A.

The respondents expressed dissatisfaction with the current state of regulation of online shops in the Austrian legal system, as onlineshop operators (or Austrian citizens in general) only take action when they are (usually financially) penalized. Even if online shop operators are aware of the legal situation and knowingly violate it, they only act when they face (financial) harm or are explicitly instructed by a higher authority. Participant C drew a comparison to a court decision regarding cookie banners, which currently shows no effectiveness because there are (still) no financial penalties for the unlawful implementation of cookie banners.

However, for penalties to occur, precedents must be set, meaning decisions must be made by the Austrian Supreme Court (OGH) or the European Court of Justice (EUGH). During the interviews, each participant mentioned how important precedents in the field would be to follow. For this to happen, a case must actually go to court. Many cases of harm caused by dark patterns do not go to court because a legal process is costly, and the burden of proof lies with the plaintiff (and proving harm caused by dark patterns can be challenging).

According to the three interviewed legal experts, it is conceivable that a dark pattern case could go to court if class-action lawsuits (e.g., by consumer protection organizations) are taken against specific online shop operators who knowingly use dark patterns to maliciously manipulate their users.

5.2.5 How to take legal action against dark patterns?

There are various actions that consumers can take if they believe they have been targeted by dark patterns (or encountered fraud in online shops). The most straightforward action is to exercise the 14-day right of withdrawal for distance contracts (§ 11 para 1 FAGG - Fern- und Auswärtsgeschäfte-Gesetz). "The 14-day withdrawal period is fully harmonized throughout the EU! Therefore, neither a longer nor a shorter withdrawal

period can be legally stipulated. Contractually, this period can be extended, but under no circumstances shortened"[Ös22]. Some exceptions exist, such as for sealed goods.

In situations where the 14-day withdrawal period has lapsed or the right of withdrawal does not apply, additional recourse is available. Consumers can seek assistance from consumer protection organizations, including:

- Consumer Information Association (Verein für Konsumenteninformation - VKI)
- Chamber of Labor (Arbeiterkammer - AK)
- Austrian Society for Consumer Studies (Österreichische Gesellschaft für Verbraucherstudien - ÖGVS)
- Austrian Trade Union Federation (Österreichischer Gewerkschaftsbund - ÖGB)
- Regional Consumer Protection Offices

The list above emerged during the interviews and may not be complete, but it provides a good overview of what can be done to fight back against dark patterns.

These organizations can offer guidance and may even initiate collective legal actions if multiple individuals report similar issues. Such collective actions can have a significant impact, often garnering media attention and potentially proceeding to court. If legal decisions are reached in these cases, they may establish precedents. Consumer protection organizations and collective legal actions thus play a vital role in shaping legal practices within the eCommerce sector.

Additional actions may include consulting a lawyer specializing in eCommerce, initiating a civil lawsuit, submitting online complaints to the Austrian Advertising Council, and more.

5.2.6 Web-Crawler and Technology Deployment for Dark Pattern Detection

The study participants were questioned regarding their personal assessment of the use of web-crawlers concerning dark patterns in eCommerce. They were asked whether it would be legally or politically feasible to employ web-crawlers to examine the prevalence of dark patterns or observe the evolution of their dissemination over an extended period. Furthermore, insights were sought into potential applications of web crawler technology in relation to dark patterns.

The participants unanimously agreed that a sole reliance on a crawler for a legal evaluation is inadequate since dark patterns are diverse, requiring a legal examination by lawyers to assess the situation properly. Nevertheless, a crawler could serve as a kind of "preliminary check," potentially alleviating legal personnel or authorities. As mentioned earlier, legal

texts are often generally formulated, and individual assessments are made by lawyers and judges, making a purely crawler-based evaluation inconceivable.

However, the use of web crawler technology is certainly conceivable and beneficial for evidence gathering as Paticipants B and C stated. A web crawler can repeatedly crawl web pages or online shops over an extended period, storing data that can be used as evidence for lawsuits or court proceedings. For example, it could demonstrate that the stock status for a product is consistently declared as "low" over several weeks, months, or years. A crawler can be set to crawl the same URL at a self-defined frequency (e.g., once a week) over a specific period (e.g., 6 months) and store the data in the form of source code and screenshots.

From a political perspective, Participant C stated that - a web crawler could be used to verify whether enacted laws (related to dark patterns) are effective, i.e., whether the dissemination of dark patterns affected by the laws decreases. Additionally, web crawler technology could be employed for the detection of dark patterns, highlighting the urgency of the problem and serving as a starting point for further studies or research in the field, urging lawmakers to take action.

Participant C in this context also mentioned the Austrian eCommerce Quality Seal ("Österreichisches E-Commerce Gütezeichen" [zFdkNdI23]), an online certificate awarded by the association for the promotion of fairness in sales over the internet in Austria. The Quality Seal is operated by the Austrian Chamber of Commerce, the Federal Ministry of Labor and Economics, and the Federal Chamber of Labor. It was suggested that within this quality seal or the online platform associated with it, a self-check for companies could be offered. With this self-check, one could crawl their own website or onlineshop and check for the presence of dark patterns. With an extension of the crawler, automated detection could be further expanded to recognize any (prohibited) content that can be identified textually.

According to Participants A and C, the use of technology and digital tools for support is essential for the future. "Due to the flexibility of the internet and the constant evolution of eCommerce, refusing to use technology would be negligent, as it can assist us supportive", stated Participant C.

If web crawlers are used for legal and political purposes in the future, it was claimed that there should be a clear framework for the implementation of web-crawlers. It was proposed to create an ISO certification for the implementation of web crawlers. Awareness should be raised regarding the use of web-crawlers, including where they can be used and what their limitations are.

5.2.7 The Future of Dark Patterns in Law

The future of regulating dark patterns in eCommerce remains challenging and partially uncertain. The difficulty arises from various factors such as the versatility of dark patterns, their flexible design, the fine line between persuasive and manipulative design, and other considerations. All the legal experts interviewed agreed on these statements.

Participant C is professionally exposed to the Digital Services Act and the Digital Market Act and commented as follows: "The Digital Services Act (DSA) and the Digital Markets Act (DMA) will bring changes to the regulation of online shops, including the regulation of specific dark patterns. The aim of the DSA is to establish equal competition conditions, fairness, and a framework for equal fundamental rights." It was asked why the prohibitions mentioned in the DSA are not simply blacklisted in the UWG, to which Participant C responded that, in contrast to the Unfair Competition Act (UWG), the DSA explicitly focuses on digital content. This allows for a more precise examination of the characteristics of certain (purchase) processes, as a clear distinction can be made between "offline" and "online" purchases.

However, all experts still see a challenge for the future, as the burden of proof remains with the plaintiff, and many phenomena are challenging to prove from a consumer's perspective. For instance, proving that the stock status provided by an online shop operator is an incorrect statement or that it is precisely this misrepresentation (e.g. of a low stock level) that has led to the purchase (causality). The DSA is intended to be an instrument for resolving such cases, helping clarify matters by prescribing certain implementation details or making documentation publicly accessible. Participant B stated that, reversal of the burden of proof is also considered a conceivable solution to this problem: the accused must prove that they are not acting unlawfully. By 2024, the DSA must be implemented in all EU member states in the form of national laws. Therefore, online shop operators can proactively gather information and plan ahead, as it might be necessary to revise or remodel online shops due to new laws. To ensure sufficient resources are available for these measures, it is recommended to gather information and prepare accordingly in advance.

It is anticipated that more laws and regulations will be established in the future for the regulation of eCommerce, and a significant transformation is expected in this legal domain.

CHAPTER 6

Discussion

In this chapter, the previously elaborated results are interpreted and compared with or linked to the previous state of knowledge in research. The new findings of this study are explained and a prospect for future work in the research area is given. Furthermore, the limitations of this study and methodology used are explained.

6.1 Answering the Research Questions

Within the scope of this study, a systematic approach was employed to implement a web crawler designed to search for product pages within Austrian online shops and examine them for the presence of selected dark patterns (Low-stock Message, High-demand Message, Limited-time Message, Activity Message and Countdown Timer). This crawler facilitates a comprehensive examination of eCommerce websites for the utilization of manipulative interface elements. Furthermore, the results obtained from the web crawler study, in conjunction with an extensive literature review of dark patterns from a legal perspective, served as the foundation for conducting semi-structured interviews with legal experts. These interviews aimed to explore the legal and political viewpoints of legal experts regarding the use of web crawlers to detect dark patterns and to assess their perspectives on current regulatory measures pertaining to dark patterns.

6.1.1 Web-Crawler Study - Dark Pattern Prevalence

In order for the web crawler to recognize dark patterns based on textual elements, regular expressions were created in advance to recognize dark patterns based on pattern matching. Subsequently, the web crawler was applied to two data sets: (1) eCommerceDB [es22] top 250 eCommerce shops from Austria (based on revenue) and (2) Geizhals Merchant List, a list of all retailers of the online price comparison platform geizhals.at[AG23].

Q1: How prevalent are dark patterns among the top Austrian eCommerce onlineshops when identified using a web crawler with regular expression detection?

Regular expressions were employed to identify 5 selected dark patterns on a list of onlineshops and their respective product URLs. These 5 selected dark patterns were:

- Low-stock Message
- High-demand Message
- Limited-time Message
- Activity Message
- Countdown Timer

The quantitative analysis of the web-crawler results revealed that in the eCommerceDB dataset, approximately 16% of all examined onlineshops employ at least one instance of these 5 variations of dark patterns, while in the Geizhals dataset, 10.5% of the observed onlineshops employ them. Considering that only 5 selected dark patterns were examined out of the possible 15 according to Mathur's taxonomy [MAF+19a], this percentage is notably high.

In the study titled 'Dark Patterns at Scale' [MAF+19a] a similar methodology was applied to assess the presence of dark patterns in online shops, revealing that approximately 10% of 11,000 online shops employed dark patterns. However, in this study, all 15 variations of dark patterns were investigated, not just 5 selected eCommerce dark patterns as in this thesis. Furthermore, in their work not only the product detail page but the complete checkout process was examined and no regular expressions were used, but previously extracted text segments (referred to as text clusters in their work) were manually examined and classified.
In order to make the results of this thesis better comparable to the previous study, the results were downloaded from Github[MAF+19b] and subjected to further analysis. The findings were stored in an SQLite database, and subsequently, only the 5 dark pattern types relevant to this thesis were analyzed. In Mathur's work, approximately 9.8% of the examined websites were found to utilize the 5 selected dark patterns that are relevant to this study. The results of this study yielded similar findings: 16% of the top 250 eCommerce companies in Austria employ dark patterns, and approximately 10.5% of the Geizhals merchants employ them. These discrepancies (especially compared to the eCommerceDB dataset) can be explained by various factors. In the present thesis, 10 product URLs were examined (not only 5 URLs), the parameters of the crawler were changed (to search more efficiently), regular expression were applied for automated detection (compared to manual detection in preceding study) and other small technical changes in the implementation and detection. Comparing the results of prevalence (considering the aforementioned differences) of the top 250 German-language online shops

from Austria at the crawl time of August 2023 (16%) with the results of the top 10,000 English-language online shops (9.8%)[MAF+19a], an upward trend in prevalence becomes apparent. Despite the awareness of dark patterns for nearly 15 years and increasing media attention, in the year 2023, dark patterns are employed more frequently than in 2019.

Neem et al. [NL21] manually scrutinized 96 websites for the existence of dark patterns, taking into account not only textual content but also visual elements such as style, colors, and images. Within the scope of their research, Neem developed their own eCommerce dark pattern taxonomy consisting of 20 dark pattern types and found that 60% of all examined Swedish online shops utilized dark patterns. Other scientific works have also addressed the detection of dark patterns. Di Geronimo et al. [DGBF+20] manually examined 240 popular mobile apps for the presence of dark patterns and found that 95% of all examined apps contained at least one instance of dark patterns. Utz et al. [UDF+19] investigated 1000 consent notices (cookie banners) and discovered that 57.4% of the examined consent notices included nudging or elements that encourage users to choose privacy-unfriendly options (described as dark patterns in their work).

Although the cross-section of industries is different in the two data sets of this study, the most common industry sector for dark patterns is "Electronics & Technology". In the eCommerceDB dataset, 12 independent websites of the industry Electronic & Technology employed dark patterns, and in the Geizhals dataset 25 websites employed them.

Q2: How reliable is the automated detection of dark patterns with regard to true-positive and false-positive using a web crawler with regular expressions?

To assess the reliability and effectiveness of the web crawler and the automated detection methodology using regular expressions, an additional analysis of the crawler's performance was conducted. The results indicate that for the eCommerceDB dataset, approximately 87% of detected dark patterns were correctly identified (true-positive), while for the Geizhals dataset, the true-positive rate was approximately 78%. In this context, true-positive means that the crawler has recognized an element as a dark pattern and that this element was also declared as a dark pattern after manual examination. Unfortunately, the data could not be evaluated with false-negatives and true-negatives, as this would require manually checking every single product URL (5104 URLs in total) for correctness, which would go beyond the scope of this paper.

In the first step of the crawling process, the web-crawler attempted to discover 10 product URLs per eCommerce website by simulating the behavior of a real user. The results of the product page crawler are satisfactory and robust, with of more than 80% of correctly identified product pages for both tested datasets. The average crawl time per website/online store, approximately 3 minutes, is challenging to interpret. This is because it involves considering not only the hardware specifications of the computer/server but also the software quality, especially parallelization, and the influence of different crawler parameters, such as the time to wait on a page before opening the next page, to simulate user browsing behavior effectively (and not be blocked by web servers).

When all correctly identified product URLs are compared to the theoretically possible product URLs (which is the number of online shops multiplied by 10 URLs), it becomes evident that approximately 60% of all theoretically possible product URLs were indeed found. At first glance, this percentage may not appear particularly high; however, it is essential to consider that, on the one hand, not all websites can be crawled due to security mechanisms and bot detection, leading to blocks. On the other hand, there are also online shops that are highly customized and not easily generalized, requiring substantial technical effort to crawl (or categorize) using automated methods. Furthermore websites with languages other than German as their main language were skipped for the purpose of this process.

Other works in the field of dark pattern detection primarily focus on the application of Machine Learning algorithms. Soe [SSS22] attempted to automatically recognize 5 selected dark patterns from Gray et al.'s taxonomy [GKB+18] using a feature-based Machine Learning approach (as opposed to text and image recognition). However, they expressed dissatisfaction with the results, as the best accuracy score was 72%, and the worst was 50% (whereas random classification into 3 possible classes would be 33%). Their work aims to lay the foundation for further research on dark pattern detection using Machine Learning. Additionally, it elucidates the challenges posed by Machine Learning and dark pattern detection, including the issue that unclear definitions may hinder even humans from accurately classifying dark patterns for the creation of a balanced dataset. Furthermore, it underscores that purely text-, image-, or feature-based classification might not capture the entire context, leading to potentially inaccurate solutions.

Another study in the realm of machine learning is the research by Yada et al. [YFM+22a], where a text-based classification was conducted using the Transformer-based pre-trained language model RoBERTa (large). The exclusive reliance on text for classifying the dataset from Mathur [MAF+19b] resulted in an accuracy of 0.975 – a notable achievement that holds promise for future research projects in this domain. The dataset created in this thesis can also serve as input for the machine learning problem of Yada. To achieve this, an export from the result database to a CSV file must be performed, negative examples need to be generated (can be done using the source code of Yada and the product URLs from the Product Page Crawler), and then the dataset needs to be balanced to have an equal number of positive and negative samples.

6.1.2 Semi-Structured Interviews - Legal Compliance

The following two subsections represent responses to the research questions on dark patterns from legal and policy perspectives. The results of the semi-structured interview with legal experts are discussed.

Q3: According to legal experts, can automated methods for detecting dark patterns be used from a legal or policy standpoint?

Legal experts agree that web crawlers cannot be used to assess the legal compliance of online shops regarding dark patterns, as dark patterns are too flexible, many legal texts

are too broadly formulated, and the legal framework for dark patterns is not yet clearly enough defined to take this step (and it is unsure if it ever will be). A conceivable use case from a legal perspective would be the preservation of evidence with a web crawler that collects data over a longer period of time, at a predefined frequency. The automated detection of dark patterns based on text using web crawler technology might be applied in the future, but more likely in political contexts. Possible scenarios envisioned by legal experts include:

1. Prevalence review (assessing the effectiveness of laws or highlighting the relevance of dark pattern regulation)
2. Self-check tool for onlineshop operators on plattforms operatated by political institutions
3. Support for political institutions (preliminary assessment for online complaints; evaluation for the awarding of e.g. eCommerce quality seals)

Above all, web-crawler-based systems for detecting dark patterns can provide crucial statistical information which can, in turn, serve as a foundation for policy-makers to formulate effective countermeasures. Furthermore studies like this offer essential insights into the prevalence and characteristics providing valuable knowledge for future studies. The publication of studies on the prevalence of dark patterns can also generate public interest in the subject, thereby educating users about unethical practices. This, in turn, might lead to increased awareness among individuals, potentially reducing the number of victims of Dark Patterns in the future.

There is a generalized interest in the deployment of technology (including artificial intelligence) in law and its application in politics. There is an opinion that if politicians and lawyers resist the supportive use of technologies such as web crawlers, they will soon encounter their resource-related limitations.

In legal literature, there are calls for research and the IT community to develop software tools to counteract dark patterns. Mathur [MAF+19a] proposes, based on the identified data, the creation of a browser extension that alerts users in real-time to the presence of dark patterns during browsing. To emphasize the significance of dark patterns, it is advocated that information campaigns, specific guidelines, and software tools should be created to raise awareness among end-users [Ger22]. Thus, the use of web crawlers to detect the prevalence of dark patterns is accordingly scientifically relevant and in demand.

Due to the popularity of web crawlers, there is a call for regulations governing their use. Clear guidelines should be established regarding what a web crawler is allowed and not allowed to do. A proven method for website operators to define what a web crawler can index/crawl is the use of a robots.txt file. This file specifies which parts/paths of the website are allowed to be indexed (and by which bots) and which are not. Currently,

however, it is up to the crawler itself whether it wants to adhere to these specifications and rules. Additionally, regulations should mandate how a web crawler needs to identify itself (user agent string) and where information about the collected data is to be documented to ensure transparency. To demonstrate effectiveness, a mechanism for liability in case of abusive use of web crawlers should be introduced.

The use of web crawlers and web scrapers is well-established in various domains such as Business Intelligence, Artificial Intelligence, Data Science, Big Data, Cloud Computing, and Cyber Security [Khd21]. For instance, web scrapers are employed to extract prices from eCommerce websites, by onlineshop operators in order to gain a competitive advantage by undercutting rival´s prices. In the past, social media posts have also been extracted and stored to make predictions and forecasts regarding election results [Khd21]. This work presents suggestions for additional application scenarios, particularly in the fields of law and politics. It underscores the versatility and flexibility of a web crawler, emphasizing the potential of this technology in collecting data through textual recognition.

Q4: How do legal experts in Austria assess the effectiveness of current laws and regulations in addressing dark patterns in eCommerce, and what recommendations do they propose for enhancement?

While the topic of regulating online shops is as relevant and prominent as ever, legal experts express dissatisfaction with the current legal and political landscape. Although there are laws addressing the regulation of online shops and the purchasing process, such as the UWG (Gesetz gegen unlauteren Wettbewerb), E-Commerce Act, FAAG (Fern- und Auswärtsgeschäfte-Gesetz), and others, these laws are often broadly formulated to apply to a wide range of cases. Consequently, there have been few court decisions (at the OGH or ECJ), leading to a lack of precedent cases essential for advancing the legal regulation of dark patterns.

Furthermore, according to legal experts, the situation in Austria is such that action is taken only after facing financial penalties. From a political perspective, the approach is more inclined towards "warning instead of penalties." This results in online shop operators adhering to laws only to a limited extent, as warnings are not taken seriously (due to the absence of expected penalties).

Researchers in the legal field acknowledge that the internet evolves rapidly, while the legislative process is comparatively slow. Nevertheless, regulations must be established. With each enacted law, new loopholes and gray areas emerge that are exploited — an acknowledged fact in legal circles. A quote from the interviews was, "It is the profession of lawyers (and other professions like tax consultants) to exploit loopholes. They will always exist."

These findings from the interviews are also consistent with the current state of research in the legal field on the topic of dark patterns, namely that there are frameworks for regulating online stores, but that they are only partially effective [LVBB+22].

Furthermore, as described in Chapter 2.4, website operators often find it easy to design their websites in a way that avoids using overly aggressive or blatant elements, yet remains effective in manipulating users [DGBF+20]. This assertion is reinforced by the deliberate generalization of some legal texts, aiming not to self-restrict in regulation. This interpretative flexibility gives rise to gray areas that can be exploited. Dark patterns must also be considered from a behavioral science perspective, which states that cognitive biases are exploited to control the user to behave in a way that is disadvantageous to them [Wei20]. This fact makes regulation with conventional regulatory instruments particularly difficult.

In order to improve the current situation regarding the regulation of dark patterns, cases of abuse through dark patterns must be brought to court, and decisions must be made there to establish legal precedents. Furthermore, stricter measures must be taken at the government level, as in Austria, action is only taken when financial penalties are threatened (and are actually executed). In the future, it is essential to continue to respond promptly to new phenomena on the internet to provide the best possible protection for online shoppers.

6.1.3 Additional Findings

This subsection presents unexpected findings from the Web Crawler Study and from the results of the Semi-Structured Interviews with legal experts.

- Low-stock Messages were by far the most prevalent dark patterns in both datasets (62% of all dark patterns found in the eCommerceDB dataset, 76% of all dark patterns found in the Geizhals dataset)
- The industry where dark patterns were most prevalent (in both tested datasets) was "Electronics & Technology"
- Number of websites with 2 or more types of dark patterns was unexpectedly low (1% for Geizhals and 3% for eCommerceDB)
- Decisions of the Austrian Supreme Court (OGH) and the European Court of Justice (CJEU) are groundbreaking for how dark patterns will be dealt with in the future
- Generally formulated laws make it difficult to make a blanket statement about violations of the law by dark patterns (need case-by-case examination)
- Class action lawsuits are probably the most effective way to advance the legal situation of dark patterns
- Legal experts expect to see increased use of technology (including web crawlers) for political and legal concerns

6.2 Future Research Outline

As demanded by other scientific works, there is a need for unified concepts and definitions for dark patterns to advance research in this field [MKM21][GKB+18][Fog98]. Currently, a challenge exists as the concept of dark patterns has not been fully developed, clearly, and unambiguously formulated, making the detection of dark patterns difficult [SSS22]. As demonstrated by Soe [SSS22], people find it (still) challenging to manually identify dark patterns, leading to disagreements among reviewers. If the concept is not clearly understood by experts in the field, it cannot be defined clearly enough to understand dark patterns through supervised machine learning algorithms [HG21].

Research in the field of dark pattern detection is clearly moving towards the use of artificial intelligence [SSS22][HG21]. Traditional machine learning tasks consider data in isolation and classify them. However, for many dark patterns, the isolated consideration of data is insufficient, as they depend on context (HTML DOM-Tree) or are built across multiple pages (as in the Nudging-Pattern [MAF+19a], where users are repeatedly prompted to answer the same question, even if it has already been answered). Therefore, research should focus on the combination of detection methods: text recognition with natural language processing (NLP), image recognition (pixel-to-pixel correlation), and feature-based detection (HTML tags, CSS classes, structure, text length, etc.) [SSS22]. A machine learning pipeline is needed that works for the complex requirements of dark patterns.

Works like this thesis can help to extract real data from the internet on a large scale, which can be used for further research. The use of web crawlers and simple detection algorithms can help to collect massive amounts of data needed to construct an algorithm or a machine learning pipeline for a classification problem.

From a legal perspective, a shift regarding dark patterns is expected in 2024. With the release of the DSA (which must be implemented by all EU member states by February 17, 2024), legal frameworks for regulating dark patterns will change. The specific implementation by member states will provide information on which dark patterns are affected and how precisely. In some scientific works on dark patterns (mainly concerning cookie banners), it is criticized that concrete (newer) laws such as the GDPR do not focus on UI design [SSS22][HG21]. Regulatory measures in this area are expected sooner or later.

At the EU level, many laws are currently being enacted, such as the Digital Service Act (DSA), Digital Market Act (DMA), Data Act (DA), and the AI Act. The effectiveness of these EU measures in practice will be revealed in the future. If regulatory measures take effect, court decisions are also expected, creating precedent-setting cases that, in turn, will determine the future of dark patterns in the legal field.

6.3 Limitations

This thesis also has limitations that need to be acknowledged and are to be discussed in this section.

Initially, it is crucial to acknowledge the overall constraints associated with studies relying on web crawlers. Such studies are inherently limited since not all content available on the World Wide Web (WWW) can be indexed or examined by crawlers. The inability to access certain content stems from various factors, such as the implementation of security mechanisms designed to block crawler access. Additionally, content that is restricted only to logged-in or premium users remains beyond the crawler's reach. Furthermore, an overly aggressive crawl strategy (e.g. sending too many request in short time) may inadvertently trigger security mechanisms on web pages, leading to IP or user-agent blocking and thus obstructing any further access to the website.

Other general limitation of web crawlers include the observation period and the dynamic nature of the web. The content available on the internet, particularly on eCommerce websites, undergoes frequent adaptation and updates, occurring daily or even more frequently. As a result of these updates, not only textual content changes but sometimes entire URLs, navigation elements, and more. Consequently, a product URL that was found during an initial crawl pass may no longer be located during a subsequent crawl pass. The changes in eCommerce product listings can be attributed to various factors, such as the end of a product's life cycle, seasonal variations and offers, depleted stock, among others. Due to these fluctuations, it becomes crucial to consider data preservation and evidence retention when conducting web crawls. This ensures that the data captured during the crawling process remains reliable and reflective of the web content at the time of the crawl.

These are the general limitations of crawler-based studies. Furthermore, there are limitations specific to this study and the study design and methodology used, which will be elaborated in the following paragraphs.

A limitation of this research lies in its exclusive focus on German-language websites. This selection was made due to the prevalence of German content on websites offering goods for sale in Austria. The decision to use a web crawler employing regular expressions to extract text-based content necessitated the restriction to a single language for the purposes of this study.
Moreover, the automated approach undertaken in this study does not identify all dark patterns from Mathur's taxonomy [MAF+19a]. As described, a subset of patterns was selected and a crawler was used to search just this subset of patterns (not all possible patterns). The reason for this was a preliminary investigation of which dark patterns can be detected "only" using text and regular expression matching. For instance, patterns like Confirmshaming [MAF+19a] require an exceptionally intelligent and well-trained language model for automated detection. Other patterns like Nudging [MAF+19a] need

to observe at least two (at best way more) pages to check if the same question is repeated over time and pages.
Another limitation of the study design is the inability to precisely capture the context of the web pages or fully understand the meaning of the data due to technical limitations. Context in web development is a broad term that encompasses both the visual (style), linguistic (text) and structural aspects of a page and the whole website.
The primary limitation of the study lies in its utilization of a static approach for text recognition using regular expressions, in contrast to a dynamic approach using advanced technologies (such as the employment of a machine learning pipeline using several algorithms). Adopting dynamic methods would necessitate a larger dataset and the application of various machine learning techniques and algorithms. The drawback of regular expression recognition is its reliance on known textual patterns, meaning that only previously identified text elements can be detected. To address this, a substantial number of product pages were manually scanned for dark patterns, forming the basis for creating regular expressions. However, this approach may overlook certain text phrases that were not discovered during the manual search and, consequently, remain unrecognized by the crawler.

These limitations mentioned above refer to the limitations of the study in using a web crawler. In the following paragraphs, the other limitations of the study design will be discussed, especially in the literature search and in conducting the semi-structured interviews.

Since a literature review for the underlying theoretical foundations could not be carried out in a fully systematic way, there is a possibility that not all existing definitions or taxonomies of dark patterns were examined. In the semi-systematic literature review, often cited works were preferred and considered in more detail as opposed to more recent and not yet (often) cited scientific publications.
Additional limitations pertain to the implementation and execution of semi-structured interviews, which inherently introduce a degree of subjectivity as participants share information based on their personal experiences, values, and memories. The presence of interviewer bias must also be acknowledged, as the interviewer's actions can influence the participants' responses. Every reaction, behavior, and even body language can impact the answers provided. Moreover, the relatively small number of participants (three legal experts) in the study and general often in semi-structured interviews results from the extensive time required for planning, conducting, and analyzing each interview and the fact that it is difficult to find legal experts for eCommerce who are willing to participate voluntarily in studies. While the flexibility of semi-structured interviews offers numerous advantages by enabling an exploratory approach and uncovering unforeseen insights, it simultaneously poses a challenge in terms of methodological standardization, making data comparison difficult.
Despite these limitations, this thesis provides valuable insights and lays the foundation for future research projects in this field.

CHAPTER 7

Conclusion

E-Commerce sales have almost doubled from 2019 to 2023 [es19], and this growth presents an opportunity for fraud and manipulation. Dark patterns are user interface design choices that coerce, guide, or deceive users into unintended and potentially harmful decisions, benefiting online services [MAF+19a]. Since individuals often don't make rational decisions and can be influenced by emotions and cognitive biases, dark patterns exploit these limitations, like the Scarcity Bias, which suggests that things or opportunities seen to be more valuable when they are less available [TK74]. Employing dark patterns, online shop operators can unethically boost their sales.

This thesis implemented a web-crawler to investigate online shops for predefined dark patterns. While initial large-scale studies on dark patterns exist for English-language websites [MAF+19a], the classification of dark patterns was carried out manually and therefore no automated detection mechanism was employed as it was in this study. Currently, the use of machine learning algorithms is not yet effective because dark patterns are underconceptualized and there is a lack of datasets that are large and rich enough in data and metadata for solving a classification problem at high quality [HG21].

The web crawler implemented in this study simulated user behavior, examining five types of dark patterns: Low-stock Message, High-demand Message, Activity Message, Limited-time Message and Countdown Timer. The crawler was applied to two datasets: the top 250 Austrian online shops by revenue in 2021 and the Geizhals Merchant list (consisting of 583 onlineshops). The results indicated that 16% of eCommerceDB's top 250 list and approximately 10.5% of Geizhals merchants used dark patterns, with the Low-stock Message being the most common. In the "Electronics & Technology" industry, the majority of dark patterns were found. The success rate for collecting product URLs exceeded 82% in both datasets, while detection with regular expressions yielded about 87% true positives for the eCommerceDB dataset and approximately 78% for the Geizhals dataset.

Semi-structured interviews with legal and policy experts revealed that making blanket statements about the legality of dark patterns is challenging, as they require case-by-case evaluation. Currently, the best protection against dark patterns is the 14-day right of withdrawal or assistance from consumer protection organizations. From a political perspective, web crawlers for dark pattern detection have potential as supporting tools for assessing the effectiveness of laws. However, automated compliance checks are not suitable from a legal standpoint as they require closer examination by lawyers.

In summary, in this work a web crawler as a tool to collect dark pattern data from German-language websites was developed, enhancing the understanding of dark patterns. It also highlighted their prevalence, the most common types and the industries where they are found. For the legal development of the regulation of dark patterns, it would be beneficial if researchers could agree on a uniform definition of dark patterns so that precise countermeasures can be implemented. Regarding automated detection, current research is moving towards artificial intelligence and the classification of dark patterns using images, texts, or features. Efforts like these could help to provide real-world data for developing machine learning models for automated detection. With the introduction of new EU laws such as the Digital Services Act (DSA), which must be implemented by all EU member states in 2024, the regulation of dark patterns will change according to legal experts. Nevertheless, for an enhancement of the legal framework concerning dark patterns and the promotion of fair competition in te eCommerce sector in Austria, it is essential to establish precedent cases. It is important to recognize that achieving this goal demands personal commitment and substantial resources, including the necessary knowledge, time, and financial capacity. This is particularly true when navigating legal proceedings against often large online retailers or tech giants. Hence, the dedicated efforts of individuals or consumer protection organizations play a crucial role in creating these precedent cases and preventing the misuse of dark patterns in the online domain.

List of Figures

List of Tables

Bibliography

[Ada15] William C Adams. Conducting semi-structured interviews. *Handbook of Practical Program Evaluation*, pages 492–505, 2015.

[AG23] Preisvergleich Internet Services AG. Geizhals - Preisvergleichsplattform Österreich. Website, 2023. Retrieved November 30, 2023 from `https://geizhals.at/`.

[Ban21] Matilda Bankel. Exploring the use of dark patterns in the donation processes of nonprofit eCommerce. In *Conference in Interaction Technology and Design*, page 65, 2021.

[BEK+16] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: privacy dark strategies and privacy dark patterns. *Proc. Priv. Enhancing Technol.*, 2016(4):237–254, 2016.

[BLSD23] H Brignull, M Leiser, C Santos, and K Doshi. Deceptive patterns – user interfaces designed to trick you. Website, 2023. Retrieved November 30, 2023 from `https://www.deceptive.design/`.

[Böi13] Christian Böing. *Erfolgsfaktoren im Business-to-Consumer-E-Commerce*, volume 38. Springer-Verlag, 2013.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

[CABM16] Aaron Cahn, Scott Alfeld, Paul Barford, and Shanmugavelayutham Muthukrishnan. An empirical study of web cookies. In *Proceedings of the 25th International Conference on World Wide Web*, pages 891–901, 2016.

[CG04] Robert B Cialdini and Noah J Goldstein. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55:591–621, 2004.

[COG+21] Andrea Curley, Dympna O'Sullivan, Damian Gordon, Brendan Tierney, and Ioannis Stavrakakis. The design of a framework for the detection of

web-based dark patterns. *ICDS 2021: The 15th International Conference on Digital Society, Nice, France, 18th – 22nd, July 2021 (online)*, 2021.

[Cor98] Mozilla Corporation. Mozilla firefox web-browser. Website, 1998. Retrieved November 30, 2023 from `https://www.mozilla.org/de/firefox/`.

[CS10] Gregory Conti and Edward Sobiesk. Malicious interface design: exploiting the user. In *Proceedings of the 19th International Conference on World Wide Web*, pages 271–280, 2010.

[DGBF+20] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. Ui dark patterns and where to find them: a study on mobile applications and user perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[Dib20] Firas Dib. regex101.com. Website, 2020. Retrieved November 30, 2023 from `https://regex101.com/`.

[EC16] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council. Website, 5 2016. Retrieved November 30, 2023 from `https://data.europa.eu/eli/reg/2016/679/oj`.

[EN16] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of ACM CCS 2016*, 2016.

[es19] eMarketer statista. Retail e-commerce sales worldwide from 2014 to 2026(in billion U.S. dollars). Website, 7 2019. Retrieved November 30, 2023 from `https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/`.

[es22] eMarketer statista. eCommerceDB - eCommerce insights for your need. Website, 2022. Retrieved November 30, 2023 from `https://ecommercedb.com/`.

[Fog98] Brian J Fogg. Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 225–232, 1998.

[Fou23] Python Software Foundation. Status of Python Versions. Website, 23. Retrieved November 30, 2023 from `https://devguide.python.org/versions/`.

[GBVD14] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, pages 523–532, 2014.

[Ger22] Torsten J Gerpott. Reichen Gesetze gegen trickreiche digitale Nutzerschnittstellen? Politischer Handlungsbedarf bei Dark Patterns. *Wirtschaftsdienst*, 102(9):688–693, 2022.

[GKB+18] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. The dark (patterns) side of ux design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[GPC+21] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark paterns across mobile and web modalities. *Proceedings of the ACM International Conference on Human-Computer Interaction*, 5:1–29, 2021.

[HCHC10] Alan Hevner, Samir Chatterjee, Alan Hevner, and Samir Chatterjee. Design science research in information systems. *Design Research in Information Systems: Theory and Practice*, pages 9–22, 2010.

[Hev07] Alan R Hevner. A three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2):4, 2007.

[HG21] Philip Hausner and Michael Gertz. Dark patterns in the interaction with cookie banners. *arXiv preprint arXiv:2103.14956*, 2021.

[HIB22] Veronica Hoth, Maria Ivanova, and Stefan Brandenburg. UX Design Pattern für Datenschutz und Vertrauen. Mensch und Computer 2022 - Workshopband, 2022.

[Hig97] E Tory Higgins. Beyond pleasure and pain. *American Psychologist*, 52(12):1280, 1997.

[HK99] Jon D Hanson and Douglas A Kysar. Taking behavioralism seriously: The problem of market manipulation. *NYUL rev.*, 74:630, 1999.

[JGOTV16] Mathieu Jacomy, Paul Girard, Benjamin Ooghe-Tabanou, and Tommaso Venturini. Hyphe, a curation-oriented approach to web crawling for the social sciences. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10/1, pages 595–598, 2016.

[KCK16] Alfred Kobsa, Hichang Cho, and Bart P Knijnenburg. The effect of personalization provider characteristics on privacy attitudes and behaviors: An Elaboration Likelihood Model Approach. *Journal of the Association for Information Science and Technology*, 67(11):2587–2606, 2016.

[KE23] Tim Kollmer and Andreas Eckhardt. Dark patterns: Conceptualization and future research directions. *Business & Information Systems Engineering*, 65(2):201–208, 2023.

[KG19] Udo Kannengiesser and John S Gero. Design thinking, fast and slow: a framework for Kahneman's dual-system theory in design. *Design Science*, 5:e10, 2019.

[Khd21] Moaiad Ahmad Khder. Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3), 2021.

[KPJK16] Hanna Kallio, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing*, 72(12):2954–2965, 2016.

[Lai97] David Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478, 1997.

[LS21] Jamie Luguri and Lior Jacob Strahilevitz. Shining a light on dark patterns. *Journal of Legal Analysis*, 13(1):43–109, 2021.

[Ltd18] Canonical Ltd. Ubuntu 16.04.7 LTS (Xenial Xerus). Website, 2018. Retrieved November 30, 2023 from `https://releases.ubuntu.com/16.04/`.

[LVBB+22] Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation.* Publications Office of the European Union, 2022.

[MAF+19a] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proceedings of the ACM International Conference on Human-Computer Interaction*, 3(CSCW):1–32, 2019.

[MAF+19b] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Github - Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. Website, 2019. Retrieved November 30, 2023 from `https://github.com/aruneshmathur/dark-patterns`.

[MAX23] MAXQDA. MAXQDA - Die Software für qualitative und Mixed-Methods-Datenanalyse. Website, 2023. Retrieved November 30, 2023 from `https://www.maxqda.com/de/`.

[MDSW21] Mario Martini, Christian Drews, Paul Seeliger, and Quirin Weinzierl. Dark patterns. *Phänomenologie und Antworten der Rechtsordnung, Zs. für Digitalisierung und Recht*, 1(2021):47–74, 2021.

[MFS17] Wedrac Stefan Meissel Franz-Stefan. Beiträge zur Rechtsgeschichte Österreichs. *Zeitschrift der Kommission für Rechtsgeschichte Österreichs der Österreichischen Akademie der Wissenschaften*, 2, 2017.

[Mik98] Tommi Mikkonen. Formalizing design patterns. In *Proceedings of the 20th International Conference on Software Engineering*, pages 115–124. IEEE, 1998.

[MKM21] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2021.

[Nie18] Małgorzata Niesiobędzka. An experimental study of the bandwagon effect in conspicuous consumption. *Current Issues in Personality Psychology*, 6(1):26–33, 2018.

[NL21] William Neem Laahanen. Dark patterns in Swedish Ecommerce Websites. Master's thesis, KTH, School of Electrical Engineering and Computer Science, Stockholm, Sweden, 2021. Available at `https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1589156&dswid=-4151`.

[NLV+20] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[NP11] András Nemeslaki and Károly Pocsarovszky. Web crawler research methodology. *22nd European Regional ITS Conference, Budapest 2011*, 2011.

[NP12] András Nemeslaki and Károly Pocsarovszky. Supporting e-business research with web crawler methodology. *Society and Economy*, 34(1):13–28, 2012.

[oIC19] National Commission on Informatics and Liberty (CNIL). Ip report: Shaping choices in the digital world. Website, 2019. Retrieved November 30, 2023 from `https://linc.cnil.fr/ip-report-shaping-choices-digital-world`.

[Per19] Emil Persson. Evaluating tools and techniques for web scraping. Master's thesis, KTH, School of Electrical Engineering and Computer Science, Stockholm, Sweden, 2019. Available at `https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1415998&dswid=9311`.

[PVG+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[R+19] Vinitha Ravindran et al. Data analysis in qualitative research. *Indian Journal of Continuing Nursing Education*, 20(1):40, 2019.

[RWA+22] S Hrushikesava Raju, Saiyed Faiayaz Waris, S Adinarayna, Vijaya Chandra Jadala, and G Subba Rao. Smart dark pattern detection: Making aware of misleading patterns through the intended app. In *Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021*, pages 933–947. Springer, 2022.

[SB15] Rüdiger Schmitt-Beck. Bandwagon effect. *The International Encyclopedia of Political Communication*, pages 1–5, 2015.

[Sea99] Carolyn B. Seaman. Qualitative methods in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, 25(4):557–572, 1999.

[She37] Muzafer Sherif. *The Psychology of Social Norms*. Harper and Row, New York and London, 1937.

[SHNB22] Ray Sin, Ted Harris, Simon Nilsson, and Talia Beck. Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? *Behavioural Public Policy*, pages 1–27, 2022.

[Sny19] Hannah Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339, 2019.

[SSS22] Than Htut Soe, Cristiana Teixeira Santos, and Marija Slavkovik. Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way. *arXiv preprint arXiv:2204.11836*, 2022.

[TK74] Amos Tversky and Daniel Kahneman. *Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.*, volume 185/4157. American Association for the Advancement of Science, 1974.

[TSB13] Richard H Thaler, Cass R Sunstein, and John P Balz. Choice architecture. *The behavioral foundations of public policy*, 25:428–439, 2013.

[UDF+19] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (un) informed consent: Studying gdpr consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 973–990, 2019.

[Uni19] European Union. Glossary:e-commerce. Website, 2 2019. Retrieved November 30, 2023 from `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:E-commerce`.

[Uni22] European Union. Digital services act (dsa) - directive 2000/31/ec. Website, 2022. Retrieved November 30, 2023 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065.

[Uni23] European Union. The digital services act package. Website, 2023. Retrieved November 30, 2023 from https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

[Wal20] Ari Ezra Waldman. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current opinion in psychology*, 31:105–109, 2020.

[Web10] Lisa Webley. 926 Qualitative Approaches to Empirical Legal Research. In *The Oxford Handbook of Empirical Legal Research*. Oxford University Press, 11 2010.

[Wei20] Quirin Weinzierl. Dark patterns als Herausforderung für das Recht: Rechtlicher Schutz vor der Ausnutzung von Verhaltensanomalien. *Neue Zeitschrift für Verwaltungsrecht: NVwZ*, 39(15 extra):1–11, 2020.

[WGW+13] Geoff Wong, Trish Greenhalgh, Gill Westhorp, Jeanette Buckingham, and Ray Pawson. Rameses publication standards: Meta-narrative reviews. *Journal of Advanced Nursing*, 69(5):987–1004, 2013.

[YFM+22a] Yuki Yada, Jiaying Feng, Tsuneo Matsumoto, Nao Fukushima, Fuyuko Kido, and Hayato Yamana. Dark patterns in e-commerce: a dataset and its baseline evaluations. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 3015–3022. IEEE, 2022.

[YFM+22b] Yuki Yada, Jiaying Feng, Tsuneo Matsumoto, Nao Fukushima, Fuyuko Kido, and Hayato Yamana. Github - Dark patterns in e-commerce. Website, 2022. Retrieved November 30, 2023 from https://github.com/yamanalab/ec-darkpattern.

[ZBL13] José P Zagal, Staffan Björk, and Chris Lewis. Dark patterns in the design of games. In *Foundations of Digital Games 2013*, 2013.

[zFdkNdI23] Verein zur Förderung der kundenfreundlichen Nutzung des Internet. Österreichisches E-Commerce Gütezeichen - Vertrauen und Sicherheit für Ihr Online-Angebot. Website, 2023. Retrieved November 30, 2023 from https://www.guetezeichen.at/.

[Ös22] Wirtschaftskammer Österreich. E-commerce: Rücktrittsrecht beim Warenkauf im Internet B2C. Website, 2022. Retrieved November 30, 2023 from https://www.wko.at/service/wirtschaftsrecht-gewerberecht/Ruecktrittsrecht_bei_Warenkauf_im_Internet.html.

[Ös23] Bundesgesetzgeber Österreich. Allgemeines Bürgerliches Gesetzbuch (ABGB). Website, 2023. Retrieved November 30, 2023 from https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001622.

Appendix - Interview Guide

Interview Guide

Dark Patterns in Austrian eCommerce

Einleitung

Vielen Dank, dass Sie sich für dieses Interview Zeit genommen haben. In meiner Diplomarbeit geht es um die **Verbreitung** von **Dark Patterns in österreichischen Onlineshops**.

Dark Patterns sind Benutzeroberflächen, die so gestaltet sind, dass sie Benutzer zu Handlungen verleiten, die sie normalerweise nicht ausführen würden, indem sie Tricks und Manipulation einsetzen (Beispiel Cookie Banner, Countdown Timer, Buchungsplattformen 10 Zimmer werden angesehen).

Dark Patterns nutzen **Verhaltensanomalien**, also das **Verhalten, dass Menschen nicht (immer) rational entscheiden**, sondern manchmal emotional oder nach anderen Kriterien.

In meiner Arbeit habe ich **mit Hilfe eines Web Crawlers**, also einem **Computer-Programm, das mit Webseiten interagieren kann**, automatisch Onlineshops auf das Vorhandensein von bestimmten Dark Patterns untersucht.

Diese Befragung hilft mir das **Thema** Dark Patterns und Web-Crawler **aus rechtlicher Sicht besser zu verstehen** und es besser einzugliedern. Es gibt also **keine richtigen oder falschen Antworten** auf meine Fragen, ich bin lediglich an Ihrer **Einschätzung** und Ihrer **Erfahrung** zu diesem Thema interessiert. Das Interview sollte **nicht länger als eine Stunde** dauern. Mit Ihrer Erlaubnis würde ich unser **Gespräch gerne aufzeichnen**, damit mir keine Antworten oder Kommentare entgehen. Ihre **Antworten** werden ausschließlich **im Rahmen dieser Studie** verwendet und Ihre **persönlichen Daten werden nicht veröffentlicht**. Sie können die **Beantwortung** jeder Frage auch **verweigern**, oder das Interview jeder Zeit aus jeglichem Grund **abbrechen**.

Gibt es noch **Fragen**?

Darf ich mein **Aufnahmegerät starten** und das Interview ab jetzt aufzeichnen?

Allgemeinere Fragen

- Ist die **Regulierung von Onlineshops aus rechtlicher Sicht ein relevantes bzw. präsentes Thema**?
    - Gibt es Gesetze, die Ihnen dazu spontan einfallen?
    - Hatten Sie beruflich schon einmal mit Dark Patterns zu tun? Wenn ja, wo?

- Wie würden Sie aus rechtlicher Sicht den **Druck bei Kaufentscheidungen** einschätzen? Beispielsweise kann die Nachricht, dass z.B. nur noch 2 Stück auf Lager oder die Gültigkeitsdauer von 5 Minuten für ein Angebot Druck auf den potenziellen Käufer erzeugen.
    - Gelten hier die gleichen Regeln „offline“ (also im Geschäft) wie online?
      Wenn ja, sollte hier unterschieden werden? Wenn nein, was sind die Unterschiede?

- Im Internet findet man zum Thema Recht und UI Design im eCommerce Bereich folgendes Zitat: **„Das Recht hängt dem Bild des rationalen Entscheiders an. Gegenwärtig schützt es Rationalität nicht – es setzt sie voraus“**. Würden Sie diese Aussage unterstützen? Warum?

Demografische Zwischenfragen:

- Wie oft betreiben Sie ca. online Shopping? Einmal pro Woche, Monat oder mehrmals pro Woche, …?
    - Welche Produkte kaufen Sie hier am häufigsten: Mode, Elektronik, Gesundheit, …?

Spezifische Fragen über Dark Patterns

- Ist Ihnen der Begriff „Dark Patterns“ bekannt bzw. war er das vor dem Beginn dieses Interviews?
    - Wenn ja, woher kennen Sie den Begriff?
    - Fallen Ihnen spontan Dark Patterns, trickreiche Elemente oder Webseiten ein, die Sie für manipulativ halten? (Hinweis: spontan, privat, nicht aus rechtlicher Sicht)

- **Bei meiner Recherche bin ich auch auf die „schwarze Liste“ im Anhang des UGB (Gesetz gegen unlauteren Wettbewerb) gestoßen.**
    - Hast Sie von dieser Liste gehört? Können Sie mir kurz erklären worum es sich bei dieser Liste handelt?
    - **Ein Punkt in dieser Liste passt gut zum Thema Dark Patterns und Druck bei Kaufentscheidungen. Ich lese Ihnen diese Punkt aus der schwarzen Liste vor:**

Die unrichtige Behauptung, dass das Produkt nur eine sehr begrenzte Zeit oder nur eine sehr begrenzte Zeit zu bestimmten Bedingungen verfügbar sein werde, um so den Verbraucher zu einer sofortigen Entscheidung zu verleiten, so dass er weder Zeit noch Gelegenheit hat, eine informierte Entscheidung zu treffen.

- Wie könnte diese **unscharfe Formulierung „sehr begrenzte Zeit“** interpretiert werden? 1 Stunde, 5 Minuten oder Sekunden?
- Wie kann **„sofortigen Entscheidung“** interpretiert werden?
- Wie ist die Beweislast bei solchen Fällen? Muss der Geschädigte bzw. Käufer beweisen, dass die Stückzahl die im Onlineshop angegeben ist nicht der Wirklichkeit entspricht? Muss zusätzlich bewiesen werden, dass die die Zeit, um eine Kaufentscheidung zu treffen, zu kurz war?

- Bei einigen Dark Patterns kann unterschieden werden **zwischen „wahren“ und „falschen“** Patterns, also ob z.B. die Behauptung, dass nur noch 3 Stück auf Lager sind stimmt (=wahres Pattern) oder nicht (=falsches Pattern).

    - Egal ob das Pattern wahr oder falsch ist, löst es beim Menschen Druck aus, der sich auf die Entscheidung auswirkt
        - Etwas wird als wertvoller angesehen, wenn es nicht mehr viel davon gibt
        - Verlustangst → Der Verlust von 10€ ist für die meisten Menschen wesentlich emotionaler als der Gewinn von 10€
    - **Wie kann man rechtlich mit „wahren Mustern“ umgehen?** Also Mustern, die in diesem Fall Druck erzeugen, auch wenn die Behauptung stimmt, dass z.B. nur noch 3 Stück auf Lager sind?

- Im **AGBG** werden auch **Täuschung oder List** erwähnt, um einen Vertragspartner zu einem Abschluss zu bewegen.
    - Könnte eine **falsche Angabe vom Lagerstand** als List interpretiert werden?
    - Wenn ja, was wären Strafen hierfür? Oder wird der Vertrag aufgelöst und muss rückabgewickelt werden?

- Welche **rechtlichen Schritte können Verbraucher ergreifen**, wenn sie glauben Opfer von Betrügern bzw. Dark Patterns geworden zu sein?

- Welche Rolle spielen **Verbraucherschutzorganisationen** (Konsumentenschutzorganisationen) beim Schutz der Konsumenten in Bezug auf Dark Patterns?

Spezifische Fragen über Web Crawler und Technologieeinsatz

- Haben Sie den Begriff **Web Crawler schon einmal gehört**?
    - Web Crawler sind, vereinfacht ausgedrückt, neugierige Programme, die Teile des Internets durchsuchen um Informationen zu sammeln. Das wohl bekannteste eines Web-Crawlers ist Google. Der Google Web-Crawler muss alle Seiten besuchen, verarbeiten und Daten sammeln bevor sie für die Google-Suche verwendet werden können.
    - Wenn ja, wo?

- Web Crawler sammeln Daten, die sie auf Webseiten finden. Kann das **aus Datenschutz-Sicht problematisch sein?**
    - Dürfen persönliche Daten wie **der Name und die Anschrift von Unternehmern im Impressum**, gespeichert werden?

- **Spielt das Urheberrecht in Bezug auf Web-Crawler eine Rolle?**
    - Ist es legal **Screenshots von Webseiten** zu machen und diese z.B. im Rahmen dieser Studie zu veröffentlichen?
    - Darf ich Texte von einer Webseite speichern oder **wie finde ich heraus ob diese Texte oder Bilder urheberrechtlich geschützt sind**?

- In meiner Arbeit wird ein Crawler eingesetzt um **großflächig Webseiten auf den Einsatz von Dark Patterns** zu überprüfen.
    - Denken Sie, dass diese **Technologie aus rechtlicher oder politischer Sicht relevant** sein könnte?
    - Was wären mögliche Anwendungsgebiete?
        - Beispiel: Studien über die Verbreitung von der Politik um die Wirksamkeit von Gesetzen zu prüfen
        - Beispiel: Tool für Selbst-Check für Unternehmen

Ausblick in die Zukunft

- In einer Expertenrunde zum Thema Recht bei Dark Patterns wurde folgende Aussage getätigt: „das Netz ist immer schneller als das Gesetz, sobald ein Gesetz veröffentlicht ist, werden neue Schlupflöcher gefunden“ – Wie finden Sie diese Aussage?
    - Können Gesetze überhaupt vor Design schützen?

- Haben Sie von den DSA / DMA EU Verordnungen gehört?
    - 2022 veröffentlicht und treten 2024 in Kraft
    - Kann man sich hierüber schon Gedanken machen als Webseiten-Betreiber oder sollte man abwarten bis die Gesetze in den Mitgliedsstaaten umgesetzt wurden?

- Der Digital Service Act enthält erstmals wortwörtlich den Begriff „Dark Patterns“
    - Es werden einzelne Praktiken relativ konkret verboten
        - Hervorheben von Auswahlmöglichkeiten (bzw. keine neutrale Präsentation der Möglichkeiten)
        - Wiederholtes Fragen, obwohl die Frage schon beantwortet wurde
        - Stornierung eines Dienstes „erheblich umständlicher“ gestaltet als die Anmeldung

- Und noch einige weitere
- → Warum werden diese konkreten Praktiken nicht in die schwarze Liste aufgenommen?

- Wie sehen Sie die Zukunft der rechtlichen Regulierung von Dark Patterns? Welche Entwicklungen sind in diesem Bereich zu erwarten?