Bridging the AI adoption gap with Microsoft Copilot: A comparative experiment of AI-generated SEO text

Daniel Friedrich Commentz

Dissertation written under the supervision of Dr. Michael König

Dissertation submitted in partial fulfilment of requirements for the MSc in International Management with specialization in Entrepreneurship and Innovation, at Universidade Católica Portuguesa and for the MSc in Strategy, Innovation and Management Control (SIMC) at WU University of Economics and Business, 18.8.2025.

ABSTRACT

Small and medium-sized enterprises (SMEs) face persistent challenges in adopting artificial intelligence (AI) due to limited resources, technical know-how, and data privacy concerns. On the contrary, especially within the intersect of AI and search engine optimization (SEO), while the literature is still in an embryonic phase, existing research has already acknowledged the potential of large language models (LLMs) for SEO purposes. While most studies focus on freely accessible tools like ChatGPT, other enterprise-grade solutions like MS Copilot are mostly overlooked.

To address this gap, this study evaluates whether Copilot-generated search engine optimized product texts can enhance product page performance across three dimensions: Algorithmic relevance (H1), user engagement (H2), and commercial impact (H3). For this purpose, a controlled online field experiment was conducted within a German SME in the beauty and healthcare sector. From a sample of 20 product pages, half of the product texts were exchanged with Copilot-generated content, while the other half served as a control group and tracked eight weeks before and after the intervention.

While statistical significance remained limited, descriptive patterns cautiously indicate that Copilot-generated content may improve search visibility and, in isolated cases, positively affect sales outcomes. Importantly, no evidence was found that such AI-generated content harms visibility or user experience. In doing so, this study contributes to the emerging embryonic research of applied AI in the field of SEO, offering an explorative foundation for future research. Furthermore, it offers first practical hints for SME practitioners evaluating the integration of AI-driven tools like MS Copilot.

Dissertation Title: Bridging the AI adoption gap with Microsoft Copilot: A comparative experiment of AI-generated SEO text

Author: Daniel Friedrich Commentz

Keywords: Artificial Intelligence (AI), Large Language Models (LLMs), Microsoft Copilot, Search Engine Optimization (SEO), on-page SEO, Small and Medium-Sized Enterprises (SMEs), Digital marketing, AI adoption, Digital Transformation

RESUMO

As pequenas e médias empresas (PMEs) enfrentam desafios persistentes na adoção de inteligência artificial (IA) devido a recursos limitados, conhecimento técnico e preocupações com a privacidade dos dados. Na intersecção entre IA e otimização para motores de busca (SEO), embora a literatura esteja em fase inicial, já se reconhece o potencial dos grandes modelos de linguagem (LLMs), sendo o Microsoft Copilot pouco explorado em comparação com ferramentas gratuitas como o ChatGPT.

Este estudo investiga se textos de produtos otimizados para SEO, gerados pelo Copilot, melhoram o desempenho de páginas de produto em três dimensões: relevância algorítmica (H1), envolvimento do utilizador (H2) e impacto comercial (H3). Conduziu-se um experimento de campo online controlado numa PME alemã do setor de beleza e saúde, com 20 páginas de produto, divididas igualmente entre grupo de tratamento e grupo de controlo, acompanhadas por oito semanas antes e depois da intervenção.

Os resultados, embora sem significância estatística robusta, indicam que o conteúdo gerado pelo Copilot pode melhorar a visibilidade nas pesquisas e, em casos isolados, influenciar positivamente as vendas, sem evidências de prejuízo à visibilidade ou à experiência do utilizador. O estudo contribui para a investigação emergente sobre a aplicação de IA em SEO e oferece orientações iniciais para PMEs que pretendam integrar ferramentas como o Microsoft Copilot em fluxos de trabalho de marketing de forma eficiente e escalável.

Título da Dissertação: Colmatando a lacuna de adoção de IA com o Microsoft Copilot: Uma experiência comparativa de texto SEO gerado por IA

Autor: Daniel Friedrich Commentz

Palavras-chave: Inteligência Artificial (IA), Modelos de Linguagem Grandes (LLMs), Microsoft Copilot, Otimização para Motores de Pesquisa (SEO), SEO on-page, Pequenas e Médias Empresas (PMEs), Marketing Digital, Transformação Digital

ACKNOWLEDGMENTS

As this master's thesis marks the end of my academic journey at WU Vienna and Católica Lisbon School of Business and Economics, I want to express my gratitude to all the professors, mentors, and peers who have supported and inspired me along this way.

First, I would like to express my heartful gratitude to my supervisor, Dr. Michael König, for his continuous guidance, patience, and support throughout every step of this thesis.

Another heartfelt thank you goes to Rebecca for her unlimited support, energy, and dedication to this project, always being there despite her own workload.

In addition, I sincerely thank Julian for his trust, his time, and his open-mindedness in offering me this research opportunity.

A very special and heartful thank you goes to Filipa for her radical honesty, her encouragement, her invaluable feedback, and her kindness, but especially for drawing for me.

Finally, I would like to thank all my friends and family who stood by me with advice, emotional support, and motivation through every high and low.

THE BEST PLACE TO HIDE A DEAD BODY,

IS PAGE 2 OF

Google

SEARCH RESULTS

- Unknown Author

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

1.1 RELEVANCE AND PROBLEM STATEMENT

Since the attention for Artificial Intelligence (AI), and specifically large language models (LLMs), skyrocketed with the launch of ChatGPT 3.5 in November 2022, many companies are searching for ways to implement this technology for optimizing business tasks and processes. A closer look at the adoption rate of AI in the business landscape shows that especially small and medium enterprises (SME's) are lagging behind in this field. In 2024, only 11.21% of small enterprises in the EU reported using AI technologies, compared to 41.17% among large enterprises, highlighting a growing digital divide across firm sizes (Eurostat, 2025).

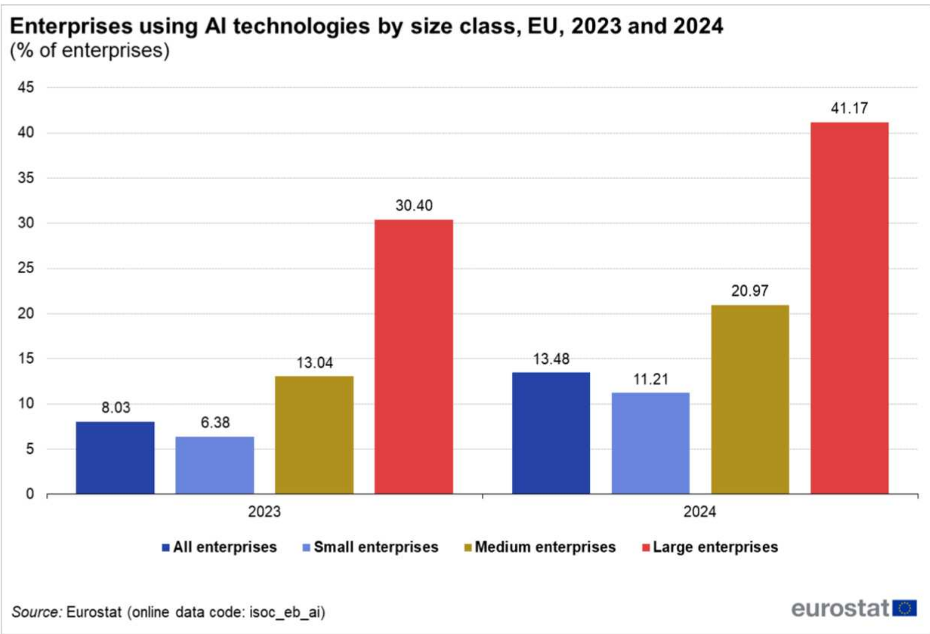


Figure 1. Share of EU Enterprises using AI (Eurostat, 2025)

Recent literature dwelling on this observation identifies that the major hurdles for SMEs consist of high investment costs, a lack of knowledge, and inadequate technical infrastructure, also connected to data privacy concerns (Empl & Pernul, 2021; Govori & Sejdija, 2023; Iyelolu et al., 2024; Oldemeyer et al., 2024). At the same time, the SME segment resembles one of the most important pillars of the German economy, with over 99% of all German firms falling into the category of SMEs, corresponding to around 3.4 million companies. In 2022, SMEs generated over 27% of total turnover in Germany despite the economic consequences of the coronavirus pandemic, contributing to more than 55% of the total net value added of all companies. Moreover, 54% of all dependent employees in Germany are employed in SMEs, including 70% of all trainees (IFM Bonn, 2023). Given their significant economic contribution, it is crucial to find ways to overcome the barriers preventing SME's from adopting AI in their daily business processes, ensuring they remain competitive in an increasingly digitalized world and reducing the risk of widening the gap between them and larger corporations.

One emerging topic with high potential of optimization through AI is the usage of LLMs in digital content marketing (DCM), or more specifically, for search engine optimization (SEO) purposes. The importance of digital marketing has increased massively over the past 20 years. A look at the changes in advertising expenditure in the USA over the last 30 years shows that online advertising grew from virtually 0% in 1995 to approximately 20% in 2010 and from there on in only ten years to a total add spending of over 50% in 2020, overtaking Television, Print and all other accumulated channels (Lammenett, 2025; The Economist, 2020). In Europe, this trend is equally evident. Total digital advertising expenditure has risen significantly from around 42% of total advertising expenditure in 2017 to about 67% in 2024 (Statista, 2024b). The development of this gap is projected to widen even further, with estimated digital advertising expenditure of around €167 billion in 2029, making up 74% of the estimated €226 billion total advertising expenditure. Since 2017, this growth has been driven by the increasing shift from traditional media, like print or TV, towards new digital media channels, like search engine advertising, influencer marketing, or digital video advertising. Within the digital advertising landscape, search engine advertising represents the largest share of investment, accounting for 48% out of the total digital advertising expenditure as of 2024 (Statista, 2024b). This highlights the growing relevance of search engine marketing and search engine optimization, which is also reflected by consumer behaviour.

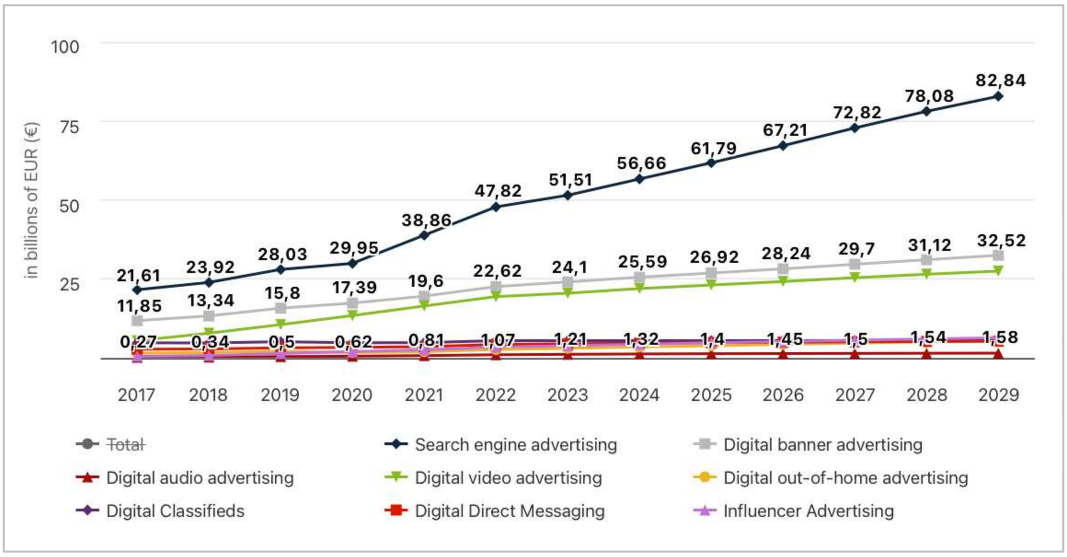


Figure 2. Digital Advertising Expenditure in the EU (Statista, 2024b)

Recent data testifies that search engines such as Google are by far the most frequently used source of information when consumers seek out new products, with 59% of respondents indicating to rely on search engines, placing them well ahead of other digital or offline sources like online shops (40%), physical stores (30%), or social media (19%) (Statista, 2024a). This strong consumer dependence on search engines for product discovery underscores even more that organic search visibility plays a pivotal role in shaping customers' buying decisions and influencing customer journeys. Especially for SMEs, this implies the clear imperative to remain visible and competitive in a digital-first purchasing environment.

In light of this customer behaviour shift, also the economic relevance of search engine optimisation has grown significantly in recent years. SEO plays a crucial role in helping businesses improve their online visibility and attract potential customers (Ahmad et al., 2022; Bhandari & Bansal, 2018; Chen & Sénéchal, 2023; Erlhofer, 2020; Sharabati et al., 2024; Usmany et al., 2024). It involves optimizing web content to rank higher in search engine results, increasing organic traffic, and enhancing a company's digital reach (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025). Especially for SMEs, effective SEO strategies are particularly important, as they often lack large advertising budgets of major corporations and must compete strategically for online attention. However, SEO is a complex and evolving field that requires consistent content creation, keyword optimization, and technical expertise, posing a challenge for many SMEs with limited marketing resources and knowledge in this field (Mou et al., 2022; Sharabati et al., 2024; Zerres & Israel, 2016). AI-driven tools could provide a significant advantage by streamlining SEO content production, reducing costs, and ensuring that smaller firms can effectively stay competitive in digital markets. But to be used in a safe and efficient way, those tools would need to be implemented into the firm's technical infrastructure, which is, as of today, still connected with high barriers like resource and knowledge constraints (Empl & Pernul, 2021; Govori & Sejdija, 2023; Iyelolu et al., 2024; Oldemeyer et al., 2024).

1.2 IDENTIFIED RESEARCH GAP AND RESEARCH OBJECTIVES

Despite the growing body of research examining the use of large language models for SEO purposes, current studies primarily focus on general capabilities such as content generation or keyword research, often using freely accessible tools like ChatGPT. While the potential benefits of those LLMs for on-page SEO are getting more and more explored, these findings remain largely theoretical for SMEs, facing prevalent practical adoption barriers. Microsoft Copilot, however, the LLM-based tool from Microsoft, introduced a potentially game-changing alternative: An enterprise-grade AI technology, seamlessly embedded within Microsoft Office

tools and protected by robust data governance. Unlike fragmented third-party SEO tools, Copilot offers an accessible and integrated solution that could make AI-powered SEO feasible even for resource-constrained SMEs. Yet, despite its wide availability and unique value proposition, no academic study has examined Copilot's actual effectiveness for SEO tasks, especially when it comes to a holistic on-page optimization approach. To address this research gap, the present study investigated the following research question:
To what extent can the performance of product pages be improved by using Microsoft Copilot to generate SEO-optimized content?

To answer this question comprehensively, the study evaluated page performance across three interrelated dimensions, with each level corresponding to a distinct hypothesis:

H1 – Algorithmic Relevance (Search Engine Level):
Product pages with Copilot-generated SEO content are perceived as more relevant by Google's algorithm, leading to improved search visibility.

H2 – User Engagement (Human Level)
Product pages with Copilot-generated SEO content are perceived as more relevant and engaging by users, resulting in increased interaction.

H3 – Commercial Impact (Business Level)
Product pages with Copilot-generated SEO content lead to improved commercial outcomes by positively influencing purchase behaviour.

To empirically test these hypotheses, the study employed a controlled online field experiment in collaboration with a German SME from the beauty and healthcare industry. In this experimental setting, SEO-optimized product page content was generated using Microsoft Copilot for a selected treatment group of 10 product pages, while 10 pages served as a control group. The impact of the intervention was measured over a period of 8 weeks before and after the introduction of the treatment, using a set of clearly defined KPIs across algorithmic, user, and business levels. The underlying data was collected from Google Search Console, Google Analytics (GA4), and the company's internal sales systems. While the sample size was limited, the experimental setup still allowed for systematic comparison and provided valuable insights into the practical effectiveness of Copilot-generated SEO content under real-world conditions.

1.3 RESEARCH STRUCTURE AND DESIGN

The structure of this thesis is designed to systematically explore the application of Microsoft Copilot as a generative AI tool for SEO optimization in SMEs. It begins with a comprehensive theoretical foundation to outline the relevance of SEO in digital marketing, the specific challenges SMEs face in SEO adoption, and the emerging role of LLMs in addressing these challenges. Following this, in the research methods section, the context and design of the empirical study are outlined. This includes a detailed description of the controlled online field experiment, the applied treatment, and the rationale behind it. Furthermore, the data set used for the analysis is presented alongside relevant limitations. Lastly, the empirical strategy section details the analytical approach employed to ensure interpretability of the data despite existing constraints. In the results section, the empirical findings of the experiment are presented in a structured manner, including the findings from descriptive statistics, visualizations, and Difference-in-Difference regression analyses.

In the discussion section, the findings of the empirical analysis are critically reflected upon and placed in the context of the data set limitations. After each hypothesis is assessed separately, an overall synopsis for the main research question is drawn.

The conclusion summarizes the key findings of the study, showing how Microsoft Copilot can support SMEs in generating effective SEO content and improving product page performance across algorithmic, user-, and commercial-related levels. It also reflects on the study's limitations, such as sample size and timeframe, which help to contextualize the results. Finally, practical recommendations and suggestions for future research are provided, especially regarding long-term use and broader application of LLM tools like Copilot in SEO workflows.

2 THEORETICAL FOUNDATION

2.1 The Role of Search Engine Optimization in Digital Marketing

Search engine optimisation, a sub-discipline of Search Engine Marketing (SEM), traditionally refers to measures that aim to boost the ranking of a web page in the editorial section of a search engine results page, often referred to as SERP (Lammenett, 2025). In contrast to search engine advertising (SEA), where every click incurs costs, SEO aims to increase the organic page traffic, meaning to exploit natural ways to boost the website's visibility without a pay-for-click advertising model (Shenoy & Prabhu, 2016). This involves changing the content and structure of a website in such a way that it is ranked as highly as possible for certain search queries, so-called keywords (Erlhofer, 2020). Mastering how content is found and ranked by search engines is considered to be a core need for today's online businesses, with a fundamental role in a business's online success. In doing so, SEO practitioners can fulfil vital strategic goals for any business relying on e-commerce, including increasing the online presence and recognition of a brand (Ahmad et al., 2022; Bhandari & Bansal, 2018; Chen & Sénéchal, 2023; Erlhofer, 2020), altering website traffic by driving more visitors to a page (Erlhofer, 2020; Sharabati et al., 2024; Usmany et al., 2024) and boosting return on investment by generating significant returns from marketing efforts (Enge et al., 2023; Kritzinger & Weideman, 2017; Sharabati et al., 2024). Given SEO's strategic role in enhancing online visibility, increasing web traffic, and ultimately driving sales, its successful implementation is particularly vital for businesses operating in e-commerce environments. In order to perform SEO effectively, it is crucial to understand the technological and competitive landscape in which these optimisation efforts take place.

2.2 The Centrality of Google: Search Engine Market Dynamics and Technical Foundations

In the early days of the Internet, there were many different search engines, for different purposes, such as internationally positioned brands like Lycos or offerings specialised on local markets like Web.de in Germany (Lammenett, 2025). While, in theory, SEO can be applied across all available search engines, the actual environment is shaped by a highly concentrated market dominated most notably by one big player, Google. Today, Google has a global market share of over 90 per cent, with over 8. 3 billion Google searches are being made every day as of 2024, counting down to over 96,000 searches per second (Cardillo, 2025). This supremacy inevitably leads to the fact that most search engine optimisations are being tailored almost exclusively to Google's search algorithm.

Google's search engine function can be, in simple terms, described as a three-step process involving crawling, indexing, and retrieval (Lewandowski, 2021). Crawling describes a continuous process, where automated bots, repeatedly traverse the web by following hyperlinks to discover new or updated page content. These crawlers fetch page data such as HTML and scripts, which is then passed on for further analysis (Lewandowski, 2021). The next stage is indexing, where the task of the indexer is to break down and prepare the documents delivered by the crawler in such a way that they can be processed efficiently in the search. Google's indexing system identifies key elements within the crawled page, including keywords, metadata, page structure, and semantic context. This information is stored in the Google Index, a massive, optimized database that allows for fast and accurate retrieval (Lewandowski, 2021). When an end-user submits a search query, the searcher component is activated. It scans the index to find the most relevant results, using complex ranking algorithms that consider hundreds of factors, such as keyword relevance, page authority, user intent, and context. The end user receives the output via the search engine results page, which may include standard links, featured snippets, and other rich content formats (Lewandowski, 2021). The aim of SEO in this context, as part of online marketing, is to offer searchers the right solution for their specific problem through a search engine optimised website. This turns searchers into visitors of the website and ultimately into customers.

Since its foundation in 1998, Google has been continuously improving its search algorithm through various updates. For example, the Panda update in 2011 was particularly targeted to eliminate content farms and websites with poor content quality. By this time, around twelve per cent of search queries were reportedly affected by the update (Erlhofer, 2020; Lammenett, 2025). In 2012 and 2013, the so-called Penguin Update aimed to radically curb web spam in the search results (Erlhofer, 2020; Lammenett, 2025). Until today, Google has significantly increased its endeavours to achieve good quality results on the SERP. While in the year of 2009, four updates were introduced, Google now deploys almost 600 smaller updates and several so-called core updates per year (Lammenett, 2025). These constant updates make it increasingly difficult to determine how different SEO measures affect the algorithm's ranking choice regarding a page in relation to a certain search query.

2.3 THE CRITICAL ROLE OF RANKING DETERMINANTS AND USER CLICK BEHAVIOUR

A key question in search engine optimisation is which criteria Google uses to evaluate and weight websites for the results list, as the ranking position of a website massively determines the traffic on the impressed webpage. Click-through rate (CTR), defined as the percentage of users who click on a link after seeing it in search results, is a critical metric in evaluating this dependency (Erlhofer, 2020; Lammenett, 2025; Sistrix, 2020). A study by Sistrix (2020), based on 80 million keywords and billions of results, found that the average CTR for the first-ranked result on Google is 28.5%, meaning that, on average 28.5% of people click on the first Google search result, after appearing at the top of their SERP users (Sistrix, 2020). For the consecutive ranking positions, the CTR then drops sharply. The second position was only clicked by 15.7%, the third position by 11.0% and the 10 position by 2.5% of the users (Sistrix, 2020). Other sources provide similar values ranging from a CTR of 33.99% for the first place (Advanced Web Ranking, 2025) to up to 39.8% (FirstPageSage, 2023). A comparative analysis of web search behaviors in Switzerland and Germany from 2023 even found that in over 50% of the cases, users clicked on the first search result, and over 97% of all clicks were made on the first page of search outputs (Urman & Mykola Makhortykh, 2023).

These statistical differences can be partially explained on a perceptual basis. An eye-tracking study from 2009 found that the top 3 ranked URLs are perceived by far the most often (Usability.de, 2009). A placement in the top 5 is important because the results can be seen at standard screen resolutions without scrolling (Erlhofer, 2020). The top 10 form the first page of Google search results and consequently do not require an additional click to the next search result page (Erlhofer, 2020; Lammenett, 2025; Urman & Mykola Makhortykh, 2023). Consequently, search results on ranks 11 to 20 are even less frequently visited (Erlhofer, 2020; FirstPageSage, 2023; Lammenett, 2025; Sistrix, 2020). However, with high numbers of impressions per page, low CTRs can also still lead to decent traffic in the top 20, but rankings outside the range are almost certainly irrelevant from a business perspective (Erlhofer, 2020; Lammenett, 2025). Therefore, the aim of search engine optimisation should be to achieve rankings at least within the top 20, at best top 10 results of a search engine. As every entry in a search engine represents a potential competitor for one of the top positions, the level of competition can be extremely high depending on the search term. For some search terms, this can be several million entries, illustrating the dilemma of search engine optimisation: There are millions of entries, but only the first 10-20 are relevant from a business perspective, and everyone wants their entry to appear there (Lammenett, 2025).

This intense competition for top rankings naturally leads to the question of how Google determines which pages deserve those coveted spots, and ultimately how they can be influenced. Official statements from Google concerning concrete technical details regarding the ranking algorithm are rare and inaccurate. In an official guide to the Google Search ranking system, the company states that "Google uses automated ranking systems that look at many factors and signals about hundreds of billions of web pages and other content in our Search index to present the most relevant, useful results, all in a fraction of a second." (Google, 2025). In the info section of Google Search itself, Google states that it considers numerous factors and signals, with the weight applied to each factor varying depending on the nature of the query (Google, 2023). More specifically, five concrete key search signals are enumerated being the meaning of the query, the relevance of the content, the quality of the content, the usability of the content, and context and setting (Google, 2023).
The first factor, meaning of the query, describes the aim to understand the true intent behind a user's search query by inferring context from keywords, recognizing synonyms, and despite spelling mistakes. Relevance of the content refers to analysing the content to see how well it matches the user's needs. Beyond simple keyword matching, it assesses if a page contains truly related information, rather than just the word repeated. The system looks for quantifiable signals of relevance and does not analyse subjective concepts like political leaning or viewpoint. With quality of content on the other side, Google prioritizes content that demonstrates expertise, authoritativeness, and trustworthiness (E-A-T). Signals, like other prominent websites linking or referring to the content, are seen as indicators of trustworthiness. Usability of Content describes the accessibility and user experience of a page. Content that is mobile-friendly and loads quickly may perform better, especially for mobile users, when other signals are equal. Lastly, within context and settings, Google personalizes results based on factors like your location, past search history, and search settings to determine what is most relevant to the user at that moment (Google, 2023). Those indications are giving first hints of how the algorithm works, but still lack concrete decision-making parameters.
The professional community mostly agrees that Google uses over 200 (Evans, 2007; Luh et al., 2016; Shenoy & Prabhu, 2016; Su et al., 2010), some even speak of up to 400 distinct ranking factors (Lammenett, 2025). Matt Cutts, former head of Google's spam-fighting team, once stated that Google even uses up to 50 different variations for each of the over 200 ranking

factors (Schwartz, 2010), adding up to over 10,000 different ranking factor variations. The precise nature, weighting, and implementation of Google's ranking factors remain a closely guarded secret by the company. Nevertheless, over the years, extensive ongoing research, empirical studies, experiments, testing, and debates have contributed to a progressively refined and widely accepted common understanding of practices that can positively influence a website's ranking.

2.4 Overview of Search Engine Optimization Approaches

To better understand which elements influence a website's position in Googles search results, researchers and practitioners have employed a variety of investigative methods, including large-scale correlation analyses, e.g. by Conductor (Conductor, 2023a; Conductor, 2023b), expert surveys such as those conducted by Moz (2015), comparative studies between optimised and non-optimised pages (Evans, 2022), and even algorithmic reverse-engineering attempts aimed at predicting rankings based on defined criteria (Luh et al., 2016; Su et al., 2023). The different optimization techniques found to improve the visibility and traffic of a certain page can be split into three main categories: On-page, On-Site, and Off-Page optimization approaches (Shenoy & Prabhu, 2016). On-page optimization refers to all measures that are carried out to optimize elements directly on the individual web page itself, typically involving the optimization of the content, structure, or programming code on the specific page (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025). On-site optimization takes a broader view by discussing factors that apply to the entire website's structure and configuration, across all sub-pages of the domain (Enge et al., 2023; Lammenett, 2025; Shenoy & Prabhu, 2016). While on-page and on-site optimisations take all factors into account that lie within the own website, off-page optimisation encompasses activities performed outside of a website that aim to improve its search engine ranking, mainly by generating inbound links from other websites or channels (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025).

2.4.1 On-page SEO

2.4.1.1 Keyword research and choice

When it comes to optimizations on the individual page level, one aspect of elementary importance is the selection of relevant keywords for which a page should be optimized (Conductor, 2023a; Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025). On the one hand, optimisation must be carried out for search terms that are in high demand in the search engines,

as no optimisation will have any business relevance if the terms for which the website has been optimised are not searched for in the search engines (Lammenett, 2025). On the other hand, the search term or combination of search terms must be selected in such a way that they are highly relevant to the webpage in question (Enge et al., 2023; Lammenett, 2025). Finally, success should be realistically achievable without excessive effort or disproportionate costs, as some keywords may be too competitive to pursue efficiently (Erlhofer, 2020; Lammenett, 2025).
This importance of keyword relevance and demand is further supported by recent empirical studies. Usmany et al. (2024) identified keyword optimisation as one of the key variables significantly influencing organic rankings and website traffic, conducting a meta-analysis of ten SEO studies. Similarly, Daoud et al. (2024) found that keyword relevance has a pronounced impact on a website's ability to rank on the first page of search results.
To identify suitable keywords, various approaches can be employed. A good starting point can be brainstorming or an analysis of which keywords direct competitors rank for (Enge et al., 2023; Erlhofer, 2020). Further techniques include analysing web server logs, keyword databases, or using the suggestions of Google Autocompletion directly from Google's search bar to identify common phrase variants and what users are looking for (Erlhofer, 2020; Lammenett, 2025). Additionally, free tools, especially those offered by Google, such as Google Keyword Planner, are widely recommended due to their reliability and Google's dominant market share (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025). Since 2022/2023, there have also been publicly available AI tools like Keywordinsights.ai, which can assist in keyword research. However, caution is advised, and manual review for quality is essential (Lammenett, 2025). Regardless of the method used, the underlying core principle should always be the attempt to understand how potential visitors search and what their needs are when searching for the page under optimization (Enge et al., 2023). For a single web page, the prevailing advice is to focus on one specific topic or a small set of closely related keywords. Pages attempting to optimise for too many keywords may perform poorly, as search engine algorithms favour content that thoroughly addresses a clearly defined topic (Erlhofer, 2020; Lammenett, 2025).

*2.4.1.2 Keyword density and WDF*IDF*

Historically, connected to the chosen keywords, keyword density was considered a key ranking factor, being the frequency of a specific keyword within a webpage's text (Erlhofer, 2020; Lammenett, 2025). SEO experts debated the optimal density, with suggestions ranging from two to twelve percent, varying by industry (Lammenett, 2025). This emphasis gave rise to black-hat practices such as "keyword stuffing," where keyword terms were massively overused

to manipulate rankings (Enge et al., 2023; Zuze & Weideman, 2013). Although early search engines tolerated such techniques, evolving algorithms have become far more effective at detecting and punishing manipulated content (Enge et al., 2023). As a result, the simple practice of stuffing texts with keywords to achieve a perfect keyword density is no longer effective (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025).
As a response to the limitations of basic keyword density, the WDFIDF approach (Within Document Frequency × Inverse Document Frequency) gained traction around 2012 as a more semantic and context-aware method (Lammenett, 2025). WDF measures the frequency of a term within a document, similar to basic keyword density, while IDF measures its weighting within a group of similar documents, particularly those ranking high in search results. This method aims to identify and cover relevant subtopics more comprehensively, improving a page's topical depth and semantic relevance (Erlhofer, 2020; Lammenett, 2025). While especially keyword density is no longer as important as before, it still does not become completely meaningless, as achieving a balance between keyword usage and text readability is still considered an important factor when optimizing a web page according to SEO criteria (Erlhofer, 2020; Lammenett, 2025). Tools applying WDFIDF analysis are still used in SEO to guide content optimisation, as their evaluation is related to a higher-ranking position (Conductor, 2023).

2.4.1.3 The Use of SEO-specific HTML Elements

Another crucial component of on-page optimisation involves the strategic use of SEO-specific HTML elements, such as title tags, meta descriptions, and heading structures (h1, h2, h3, etc), which influence both search engine rankings and user behaviour.

Title tags, defined by the HTML <title> element, serve as the formal titles of web pages and are commonly displayed to the user on browser tabs and within the SERP. Despite ongoing changes in Google's algorithms, its positive influence on rankings and organic traffic remains well-documented (Conductor, 2023; Enge et al., 2023; Erlhofer, 2020; Shenoy & Prabhu, 2016; Su et al., 2010). To maximize effectiveness, a title tag should accurately reflect the page's content and include core keywords without exceeding established pixel-length limits of approximately 580 pixels for desktop and 920 for mobile. Exceeding these limits can lead to the text not being able to be fully displayed in the SERP (Erlhofer, 2020; Lammenett, 2025; Sistrix, 2021). Furthermore, titles that clearly describe the page content, rather than generic phrases such as "Welcome to our homepage," are seen as more effective in both ranking and user appeal (Erlhofer, 2020; Lammenett, 2025). A well-crafted title not only aids search engine comprehension but also improves user trust and engagement by clearly signalling relevance (Enge et al., 2023).

Similarly, the meta description is an HTML element offering a brief summary of a webpage's content that can be displayed to the user in the SERP. In doing so, it plays a dual role in search engine optimisation. First, although not considered a primary ranking factor (Erlhofer, 2020; Luh et al., 2016), some studies suggest it does contribute modestly to a page's visibility as the ranking algorithm also draws a conclusion regarding a page's relevance from it. Su et al. (2010) found that the presence of keywords in the meta description positively influences the ranking position. Similarly, Conductor (2023) reports that keyword placement within elements like the meta description likely correlates with better rankings in the SERP. Second, meta descriptions have a proven behavioural impact, as they influence click-through rates (CTR) by shaping how users perceive a page when being exposed to it in the SERP. A well-crafted description, being the first touchpoint of the user with a page within the organic search results, can act as a persuasive snippet, increasing the likelihood that users choose one result over another. According to Enge et al. (2023), Erlhofer (2020), and Lammenett (2025), meta descriptions that clearly and accurately summarise content can significantly improve CTR by engaging the user. An Eye-tracking study from 2018 confirms that while users tend to fixate on higher-ranked results, their click behaviour is ultimately guided by the perceived relevance of the snippet, highlighting the importance of a compelling meta description for driving traffic to lower-ranked results (Schultheiß et al., 2018).
Similar to title tags, to be effective, meta descriptions should stay within the recommended display limits with around 990 pixels for desktop and 1300 for mobile to avoid truncation (Lammenett, 2025; Sistrix, 2021). Correspondingly, writing clear and compelling meta descriptions, potentially incorporating calls-to-action, special characters, or persuasive language, is important for evoking the user's interest (Erlhofer, 2020; Lammenett, 2025; Sistrix, 2021). Tools such as snippet generators and AI-based plugins can support the efficient creation and testing of optimized meta descriptions that follow these best practices (Lammenett, 2025; Sistrix, 2021).

Following the optimisation of title tags and meta descriptions, another important set of HTML elements for on-page SEO are heading tags, which define the structural hierarchy of content on a webpage. These tags range from <h1> for main titles to <h2>, <h3>, and lower-level tags for subheadings. Their primary function is to organise content in a logical and accessible way, improving both readability for users and semantic interpretation by search engines (Erlhofer, 2020; Lammenett, 2025). Historically, placing keywords in these headings was considered critical for SEO, but more recent discourse suggests that Google's semantic understanding has reduced the need for explicit keyword placement in headings. Su et al. (2010) include keywords in heading tags (h1–h5) within their identified group of significant ranking features, noting that while headings may not be among the most dominant factors, they do carry measurable weight in Google's ranking algorithm (Su et al., 2010). Also practitioners continue to emphasize their importance for a search engine optimized page based on personal experience (Lammenett, 2025). Furthermore, this practice remains common among those who view clear keyword integration as beneficial for reinforcing topical relevance (Erlhofer, 2020; Lammenett, 2025). Beyond keyword optimization, headings contribute to the semantic structuring of a page, thereby aiding search engines to understand the context of the content on a page (Enge et al., 2023). This function aligns with broader content strategies such as topic clustering and internal linking models, which aim to build thematic coherence across interconnected pages (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025).

2.4.1.4 Content Quality, Relevance, and User Engagement in SEO

Following the structural and technical aspects of on-page optimisation, the quality and engagement level of the page's actual content plays a central role in SEO performance. Engaging content is considered a crucial and even increasingly important technique when it comes to SEO initiatives. According to Enge et al. (2023), high-quality and engaging content is characterized by its originality, relevance, and user-centred value (Enge et al., 2023). Rather than merely rephrasing existing material, effective content offers a unique perspective, new data, or draws upon distinct expertise. It addresses user needs by solving problems or evoking emotional and psychological responses such as humour, controversy, or practical utility, which increases its potential to be shared and linked (Enge et al., 2023). To maximize impact, the content should align closely with the core activities of the business while minimizing overt commercial intent, as heavily promotional or ad-saturated pages tend to deter user engagement (Enge et al., 2023; Lammenett, 2025; Zuze & Weideman, 2013). High-quality content can take diverse forms, including blog articles, videos, infographics, podcasts, or interactive tools, with the chosen format tailored to the preferences and expectations of the target audience (Enge et al., 2023).

As early as 2009, Google was already capable of evaluating text quality based on style, grammar, and spelling. A patent of the company described a method commonly referred to as kind of a “gibberish score”, which is detecting low-quality or gibberish content using features like unnatural syntax, grammar issues, and keyword stuffing. This allowed Google to reliably identify poor-quality texts, such as automatically translated or artificially inflated content, and rank them lower in search results (Erlhofer, 2020). With increasingly sophisticated updates, it has been shown that Google is paying more and more attention to the quality of content (Enge et al., 2023; Erlhofer, 2020).

The importance of content-focused SEO is also supported by recent empirical evidence. Usmany et al. (2024) identified content quality as one of the most influential variables contributing to improvements in organic search rankings and traffic within a meta-analysis of ten SEO studies (Usmany et al., 2024). Another study by Zhang and Cabage (2017) observed that a website using only high-quality content-based SEO still showed measurable growth in traffic, domain authority, and SERP position, without the implementation of other SEO initiatives (Zhang & Cabage, 2016). Additionally, the correlation analysis by Conductor (2023) states that factors closely tied to content, such as page traffic, authorship, and freshness, are directly correlated with higher ranking performance (Conductor, 2023a).

Knowing this, search engines are increasingly looking at engagement signals that can help them identify the best content. Metrics like engagement time or bounce rate on a page can be used by the search engine to get an estimate of how engaging a certain page appeals to the user (Conductor, 2023a; Enge et al., 2023). The golden rule "Content is King" remains a guiding principle in on-page SEO, meaning that good, relevant content is what Google wants to present to its users (Cutler, 2024; Lammenett, 2025; Shenoy & Prabhu, 2016).

In addition, there are many smaller, more general content-related points when writing SEO texts, one can pay attention to. Following the logic for keyword definition, there is also the guiding principle for page content to promote only one specific topic or set of closely related topics per page. Optimisation approaches that aim to get a page with lots of different content topics into the top 20 SERP are rarely successful (Lammenett, 2025). Another important aspect is to choose a suitable writing style that is interesting to read and doesn't tire quickly. Short sentences are, in this case, mostly better than long ones (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025). Furthermore, it is advised not to copy texts that already appear elsewhere, as search engines prefer unique content. Otherwise, there is a risk of the page being punished for the duplicate content with lower ranking positions (Erlhofer, 2020). When outlining various points, displaying them in a list format can be beneficial for the user’s engagement by creating a better overview and visual reduction of continuous text (Erlhofer, 2020). It is also recommended to illustrate the page texts with appropriate multimedia-based material like images and videos (Enge et al., 2023; Lammenett, 2025). How their integration within the webpage can be optimized for search engines is discussed in the next chapter.

2.4.1.5 Image and visual media optimization

Images and other interactive media, like videos, are components that can also be found by the Google search engine. Especially when displayed on the search engine result pages, they can be very useful in enhancing user engagement and evoking emotional responses, factors that, as previously discussed, contribute to improved SEO performance (Enge et al., 2023; Lammenett, 2025). Optimising these media elements involves several techniques. One of the most basic is the use of descriptive file names and metadata (e.g., EXIF data such as time, date, aperture, and exposure settings). By assigning filenames that clearly describe what the image or video depicts, rather than using generic names (e.g., "IMG_1234"), the content becomes more interpretable to search engines. This can help establish topical relevance between the visual element and the surrounding page content (Enge et al., 2023; Erlhofer, 2020). Also the text surrounding an image can be crucial for Google's understanding of its context. Keywords placed before the image tag (<img>) can be beneficial for the ranking position (Lammenett, 2025). However, empirical evidence suggests that the direct impact of image tag keywords on search ranking is limited. Su et al. (2010) included this factor in their evaluation of ranking features but noted that it holds relatively low influence compared to other on-page elements.

2.4.1.6 Internal Linking and Thematic Clustering

Another fundamental component of contemporary SEO practices is the integration of internal links, which structure a website's information architecture and help search engines understand content relationships. Especially for websites covering multiple thematic areas, like online shops, it is advisable to link pages within the same topic cluster more densely, while maintaining limited cross-links to unrelated topics (Erlhofer, 2020; Lammenett, 2025; Shenoy & Prabhu, 2016). This practice not only improves user navigation but also signals thematic coherence to search engines. Since 2019, this strategic approach has been increasingly discussed under the concept of "Pillar Pages" (Lammenett, 2025). In this model, a central page (the pillar) offers a comprehensive overview of a broad topic and links to various subordinate pages (cluster content) that explore specific subtopics in detail. These subpages are, where appropriate, also interlinked, thereby forming a tightly connected thematic cluster centred on the pillar topic (Lammenett, 2025). This content architecture aligns well with how modern search engines interpret content, moving away from isolated keyword matches toward broader semantic understanding and user intent (Enge et al., 2023; Lammenett, 2025).

The importance of internal linking is also reflected in empirical data. A survey of 37 SEO experts conducted by Moz (2007) ranked internal link popularity as the fifth most influential ranking factor, indicating the strong consensus among practitioners on its strategic value (Moz, 2015). Similarly, Conductor (2023) identifies site architecture and internal linking as factors likely to have a close relationship with ranking outcomes, further validating their role in effective on-page SEO (Conductor, 2023a). However, internal linking must be approached strategically, as overuse or misapplication can harm SEO performance. One common issue is keyword cannibalisation, where multiple pages on the same site target identical keywords, which can confuse search engines and dilute their individual ranking potential (Enge et al., 2023). Additionally, excessive cross-linking between unrelated domains, or within artificial link networks and link farms, may be interpreted as manipulative. In such cases, search engines may choose to ignore or even penalize the site for unnatural linking practices (Enge et al., 2023).

2.4.2 Complementary SEO Techniques: On-Site and Off-Page SEO Dimensions

Beyond the direct optimization of individual web pages, SEO can also be addressed through on-site and off-page strategies. On-site SEO encompasses technical and structural adjustments applied across the entire website, rather than just individual pages. Typical measures include establishing a logical and semantically consistent URL structure across all subpages, implementing site maps to assist with crawlability, building domain trust through secure protocols and transparent site ownership, ensuring localization for regional targeting, optimizing for mobile responsiveness, and minimizing page-load time to improve user experience and satisfy algorithmic performance thresholds (Conductor, 2023a, Erlhofer, 2020; Lammenett, 2025; Shenoy & Prabhu, 2016). These efforts are aimed at enhancing the overall usability, accessibility, and technical integrity of a website, which helps search engine crawlers to efficiently index and interpret content.

Off-page SEO, in contrast, refers to activities carried out externally to the website to increase its authority and relevance in the eyes of search engines. The most prominent measure is the acquisition of inbound links (backlinks) from reputable third-party websites, which serve as signals of credibility and trustworthiness (Conductor, 2023a; Erlhofer, 2020; Lammenett, 2025;

Lewandowski, 2021). Additional strategies include active blogging on external platforms, social media engagement, and local citation management (Erlhofer, 2020; Lammenett, 2025). While both on-site and off-page SEO can also play a role within a comprehensive optimization strategy, they fall outside the scope of this thesis, which focuses on optimization approaches on a product page level.

2.5 Measuring SEO Performance

To assess the impact and success of certain search engine optimization initiatives, it is essential to rely on a balanced set of well-defined performance indicators. As already touched upon, visibility metrics, such as ranking and traffic, are widely recognized as core KPIs, particularly because they reflect how well a page performs in terms of reach within search engine results. However, in the context of e-commerce, success cannot be reduced to simply securing a top position on Google (Erlhofer, 2020).

Instead, websites must fulfil broader marketing functions such as strengthening customer relationships, enhancing brand perception, and supporting pre-sales activities by engaging the user with compelling content (Enge et al., 2023; Erlhofer, 2020). A top-ranking position on a SERP holds little value if users quickly leave the page shortly after landing, due to a lack of perceived relevance or engagement. This highlights the importance of a second perspective, not just how relevant a page appears to be according to a search engine algorithm, but how relevant and engaging it actually is from the user's point of view. Metrics such as click-through rates or engagement time help reveal whether the content resonates with users once it appears in their SERP results (Enge et al., 2023; Erlhofer, 2020; Lammenett, 2025).

Yet even this perspective remains incomplete without considering the commercial purpose for which an online shop is originally designed for. As Lammenett (2025) and Erlhofer (2020) argue, visibility and user engagement alone are not sufficient, but what ultimately matters is, whether these efforts lead to a desired behaviour, such as the product purchase. Thus, the third critical question becomes: Does search engine optimization not only increase visibility and user engagement, but also drive sales and revenue of the targeted products? (Erlhofer, 2020; Lammenett, 2025)

Building on this logic, the present thesis proposes a multidimensional evaluation framework that captures the impact of SEO initiatives across three interrelated levels: (1) Algorithmic Relevance (Search Engine Level), (2) User Engagement (Human Level), and (3) Commercial Impact (Business Level). This three-level perspective enables a more holistic assessment of

SEO effectiveness, with each dimension evaluated through specific, purpose-driven key performance indicators.

2.5.1 Evaluation of Algorithmic Relevance (Search Engine Level)

As already discussed, one very important KPI for the evaluation of the perceived relevance from an algorithmic perspective is the ranking position within a SERP. Position can be measured in different ways, depending on the analytical focus. One approach is to calculate an average ranking position, which aggregates the ranking a specific page achieves across a broad set of search queries (Erlhofer, 2020; Google, 2019). Alternatively, ranking can be assessed on a keyword-specific basis, evaluating how prominently a page appears in response to a particular search term (Erlhofer, 2020; Google, 2019). While average ranking offers a more general indication of a page's overall visibility within search engine results, keyword-specific ranking enables more granular analysis of SEO performance in relation to individual search intents or strategically relevant keywords. As already mentioned, positionings within the top 20 are generally considered relevant from a business point of view. However, the top 10, representing the first results page, are especially critical, as user attention and click-through rates drop sharply beyond this threshold.

Another important KPI to be considered in this regard is the number of times a link to a distinct page is impressed towards a user. According to Google (2019), "An impression means that a user has seen (or potentially seen) a link to a page" (Google, 2019). More specifically, "an impression is counted whenever an item appears in the current page of results, whether or not the item is scrolled into view, as long as the user does not need to click to see more results" (Google, 2019). From a SEO perspective, impressions are important because they signal that a page has been deemed relevant enough by the search algorithm to be presented to the user (Erlhofer, 2020; Lammenett, 2025). But as Google emphasizes: "You should aim not simply for more impressions, but meaningful impressions." (Google, 2019), underscoring the need for both algorithmic relevance and content quality, as visibility alone is not sufficient if it does not reach and engage a qualified audience.

A further indicator commonly used to assess a website's search performance are synthetically constructed visibility indices, which provide a quantitative estimate of how visible a website is within Google's SERP. While it is not a single, standardized metric, various SEO tools have developed different visibility indices based on their own methodologies. Among the most well-known is the SISTRIX Visibility Index, which is widely used in German-speaking SEO practice (Erlhofer, 2020). It is calculated based on a domain's ranking for a large, representative keyword set, weighted by expected click-through rates and search volume, also incorporating competitive benchmarking and long-term trend analysis dating back to 2008 (Erlhofer, 2020; Sistrix, 2024). Other tools, such as Searchmetrics and XOVI, offer similar indices, each with slightly different calculation models. Despite their popularity and ease of use, visibility indices have certain limitations. One key critique is their lack of transparency, as the exact parameters and weightings used in the calculation are often proprietary and not fully publicly disclosed. As Erlhofer (2020) notes, this makes it difficult to trace which keywords are responsible for changes in the index or what precise actions are needed to influence it. Moreover, since each tool uses different keyword sets, data sources, and weighting algorithms, visibility scores for the same domain can differ widely across platforms. This introduces an element of subjectivity and limited comparability. Other downsides include a potential lack of representation for niche topics or low-volume keywords, no full accounting for seasonal fluctuations, and an exclusion of short-term trending search terms, local search variations, or non-Google visibility (Erlhofer, 2020; Sistrix, 2024).

2.5.2 Evaluation of User Engagement (Human Level)

While algorithmic improvements might increase visibility, they do not guarantee user interaction. As Lammenett (2024) and Enge et al. (2023) note, content must also appeal to users by being relevant, engaging, and trustworthy. This leads to the second dimension, measuring the extent to which SEO-optimized content resonates with users.

In doing so, the first fundamental step is the measurement of traffic on a page, as it reflects whether a user perceives a webpage to be relevant enough to warrant a visit. As discussed earlier, traffic is influenced by a page's position in SERP, with higher-ranking pages generally receiving significantly more attention (Conductor, 2023a; Erlhofer, 2020; Sistrix, 2020). However, as already discussed, ranking alone is not sufficient to guarantee user clicks. Whether a user decides to visit a page also depends on how well the page appears to meet their search intent. This decision starts with the SERP snippet, especially the title tag and meta description, as these elements are critical to converting impressions into actual traffic by pitching the user the website's content when being displayed in the SERP (Schultheiß et al., 2018).

To assess this level of SEO performance, several traffic-related metrics are commonly used.

Views or pageviews a metrics counting each time a webpage is loaded or reloaded in a user's browser, reflecting on how often a particular page is accessed (Cutler, 2024; Google Analytics, 2023). While pageviews can provide a general indication of a page's popularity, they also have notable limitations. The same user can generate multiple pageviews by reloading or revisiting the page, which means the metric does not reflect the number of unique users. More critically for SEO evaluation, pageviews are not limited to visits originating from organic search via a search engine but also include traffic coming from other sources such as email campaigns, paid ads, or social media. As a result, while pageviews contribute to understanding total web activity, they offer only limited insight into SEO-specific performance (Cutler, 2024; Google Analytics, 2023). A more sophisticated way is to account for active users, which represents the number of unique users who engaged with the website in a specified date range (Google Analytics, 2024). This metric helps to identify the reach and effectiveness of efforts to attract new audiences, especially in campaign-driven contexts. However, like pageviews, unique visitor counts are also influenced by non-organic channels and do not isolate search engine-driven traffic. Furthermore, due to cross-device usage and privacy-related tracking limitations, the precision of this metric can vary. In contrast, the most direct and SEO-relevant traffic metric is the number of clicks, as recorded by Google Search Console. A click is registered when a user selects a link to a website directly from the Google search results. Furthermore, according to Google: "Clicking a search result in the Google search page to an outside page count as a click. Returning and clicking the same link again does not count as a second click." (Google, 2019). This definition ensures that clicks reflect unique user interaction with a specific search result, thereby offering a more precise measure of how well a page attracts traffic from organic search alone. Unlike pageviews or unique visitors, clicks directly link a user's behavior to Google's SERP, making this the most meaningful absolute metric for evaluating the effectiveness of SEO efforts in driving relevant user engagement.

While the absolute number of clicks provides a concrete measure of user interaction with search results, it does not account for how effectively a page converts its visibility into actual clicks. This is where the already mentioned Click-Through Rate becomes a critical KPI. Expressed by the ratio of clicks to impressions, CTR indicates what percentage of users who see a page in the SERP actually choose to visit it. This makes CTR one of the most telling relative performance metrics, as it reflects not just how visible a page is, but sets it into relation to how persuasive it is in attracting users' attention (Cutler, 2024; Enge et al., 2023; Erlhofer, 2020; FirstPageSage, 2023; Sistrix, 2020). In doing this, the metrics is building the bridge between the evaluation of the level of Algorithmic Relevance and User Engagement. In this way, CTR also ties into broader conversion-oriented thinking: Similar to how businesses measure the success of leads or transactions, CTR reflects how efficiently a listing converts search interest into user engagement (Erlhofer, 2020).

Beyond the initial decision to click on a search result, it becomes crucial to understand how users interact with the website once they arrive. Click- and Click-through metrics alone cannot capture the depth of engagement or indicate whether users find the content valuable. To assess the quality and depth of user interaction beyond the click, it is essential to examine engagement-based metrics, such as engagement time and bounce rate.
Average engagement time, also referred to as time on page or dwell time, measures how long a user keeps a webpage in active focus during a session (Google Analytics, 2023). This data is recorded when a session ends due to user actions like navigating away, closing the browser, or switching tabs. From an SEO perspective, the underlying assumption of this metric is that the more time a user spends on a website, the more satisfied they are with the information offered there (Erlhofer, 2020). High engagement times often suggest that users found the content relevant or useful, and are a strong indicator of positive user experience, which is also considered a contributing factor to search engine rankings (Lammenett, 2024). In parallel, also the other perspective can be of relevance, namely the proportion of visitors who leave the website after viewing only one page, without initiating further interaction. This value is known as the bounce rate (Cutler, 2024; Erlhofer, 2020). A high bounce rate may signal a disconnection between user expectations and page content, technical issues, or weak information architecture. Conversely, a low bounce rate implies that users are exploring further, which in turn can serve as an indicator of content quality and navigational clarity (Erlhofer, 2020; Lammenett, 2025). Google defines an engaged session as one that lasts over 10 seconds, includes multiple page views, or triggers a key event (action of particular importance for the business) (Google Analytics, 2023). Based on this, engagement rate and bounce rate form two sides of the same behavioural spectrum, allowing not just an assessment of whether a user visits, but whether they stay and interact, offering a more comprehensive picture of how well a page performs from the user's perspective.

2.5.3 Evaluation of Commercial Impact (Business Level)

Ultimately, especially for e-commerce platforms such as online shops, the true success of SEO lies in whether visitors perform the desired actions, which primarily is making a purchase (Erlhofer, 2020; Lammenett, 2025). Evaluating SEO at the commercial level closes the loop between visibility, engagement, and economic value, ensuring that optimization efforts do not merely increase reach, but also contribute directly to the organization's bottom line (Cutler, 2024; Erlhofer, 2020; Lammenett, 2025; Usmany et al., 2024). Therefore, the third dimension evaluates the economic contribution of SEO, using KPIs such as sales volume and revenue per product page. However, it is essential to attribute these figures correctly by excluding sales that result from non-organic channels, such as paid advertisements, newsletters, or social media promotions, as they would bias the attributed effect to SEO initiatives. Only transactions that are traceably linked to organic search traffic provide an accurate picture of SEO's business-level impact.

2.6 Search Engine Optimization within the Segment of SMEs

In recent years, also a growing number of small and medium-sized enterprises have recognized the strategic necessity of enhancing their online visibility, driven by shifting consumer behaviours, declining relevance of traditional marketing channels, and increasing competitive pressure in the digital space (Erlhofer, 2020; Mou et al., 2022; Sharabati et al., 2024). In parallel, a growing body of literature is highlighting the potential benefits that SMEs can derive from successfully performing SEO practices. Mou et al. (2022) state that especially for SMEs, SEO has become a critical enabler for digital visibility, by enhancing their search engine rankings, attracting high-quality organic traffic, and increasing conversion rates - outcomes which are of particular importance for businesses with limited resources. When combined with web analytics, the authors stress that SEO not only improves marketing efficiency but also supports SMEs' strategic innovation in areas such as digital branding, business model development, and long-term planning (Mou et al., 2022). Congruently, Sharabati et al. (2024) underline the importance of digital marketing, including SEO, for SME effectiveness, being an engine of digital transformation and leading to stronger economic results and an enlarged market presence (Sharabati et al., 2024).

Despite the increasing relevance and high potential of search engine optimization, the implementation of concrete SEO strategies in the segment of SMEs remains considerably underdeveloped. A survey conducted by the business news platform The Manifest, surveying owners and managers at 500 small and medium-sized companies in the U.S., found that just 30% have a SEO strategy in place to improve their website's organic visibility. However, these low figures cannot simply be attributed to a general lack of awareness regarding the topic. A study conducted by Hochschule Offenburg in 2016 surveyed 162 companies and analysed 2,138 websites in order to investigate online marketing practices among SMEs in Germany. On the one hand, the survey clearly shows that the participants are well aware that search engine optimisation is important, as 54 % of the participating companies stated that good results (e.g., higher sales, customer acquisition) were achieved, and 73 % stated that customers are better reached through search engine optimisation. However, according to the study, there is a significant mismatch between SMEs' self-perception of their SEO efforts and the objective reality. While 72% of the SMEs surveyed fully or somewhat agree that their company website is search engine optimised, and 51% stated that they use search engine optimisation regularly or frequently, the analysis of the websites shows that 98.4% of all websites examined had some significant weaknesses in terms of optimisation. For example, the title or meta description was missing on 46.7% of the websites analysed (Zerres & Israel, 2016).

Even though most research in the domain of digital marketing is still centred around large firms (Awa et al., 2015; Sharabati et al., 2024), there are already some indications from the literature shedding light on the reasons for this low adoption rate within the SME segment. The survey by Zerres and Israel (2016) already offers initial insights into the perceived barriers for SMEs. 58% of the respondents report to expect a highly time- and resource-demanding implementation. Furthermore, 56% reported lacking sufficient experience with SEO as a significant hurdle, and 50% expressed financial concerns, estimating SEO as too expensive (Zerres & Israel, 2016). A study by Sharabati et al. (2024), investigating how various digital marketing strategies, including SEO, influence SME success, mentions congruent aspects (Sharabati et al., 2024). According to the authors, SMEs often face great financial budget-related constraints. Concerns regarding the return on investment for digital projects can make SMEs sceptical about investing. Furthermore, limited human resources and a lack of skills were reported, as many may not have the in-house expertise required for digitization. Another factor mentioned is the fast-changing digital landscape, making it tough for small businesses to keep up with the latest developments and implementation of new tools. Finally, since implementing new digital technologies and strategies, including those for SEO, often involves large changes to the way of working. Another barrier mentioned was a culturally based resistance to change from within an organization (Sharabati et al., 2024). Also the paper by Mou et al. (2022) acknowledges similar hurdles, such as limited financial resources, uncertainty over return on investment, skills shortages, vastly growing technological complexity, and a lack of strategic digital vision (Mou et al., 2022). In addition to this, the authors state that SMEs in developing economies are particularly disadvantaged due to infrastructural limitations and lower access to digital literacy training programs, as well as high technology costs. Lastly, also growing concerns regarding cybersecurity threats and regulatory compliance, particularly related to data privacy laws like GDPR, are mentioned, which are especially hard to deal with by small companies with no specialized competence department (Mou et al., 2022).

After all, these barriers help to explain the slower SEO uptake in SMEs. However, overcoming them is essential not only to unlock the significant benefits SEO offers but also to remain competitive against larger enterprises, which increasingly makes SEO a matter of long-term survival in the digital marketplace. This makes it imperative for SMEs to systematically embed SEO into their organisational structures and core business processes. For larger companies, there are many approaches to implement tools, structures, and processes for qualified search engine optimisation. A common practice, if e-commerce plays a significant role in a company's core business, is to establish an internal SEO function within the marketing department, staffed by dedicated employees responsible for search engine optimisation and marketing. Beyond their departmental responsibilities, these employees also help to carry a SEO focused perspective into the company structures and processes. Furthermore, even companies with their own in-house SEO experts often use freelancers or agencies for strategic advice (SEO consulting) and a professional exchange (sparring). Also access to expensive tools and the efficient optimisation of entire work processes is a common way of enhancing SEO capabilities (Erlhofer, 2020).

For start-ups and other smaller companies, those approaches often do not provide viable options. In most cases, it only makes sense for companies above a certain size to set up their own online marketing department, and so self-taught people often take care of optimisation as a side job (Erlhofer, 2020). Some SMEs turn to SEO agencies for one-time consulting, especially during the initial development of their website (Erlhofer, 2020; Süß, 2022). However, since effective SEO requires continuous adjustment to evolving algorithms and competitive dynamics, sustained results are unlikely without ongoing SEO effort, something most SMEs cannot afford due to limited budget-bound resources (Mou et al., 2022; Sharabati et al., 2024; Zerres & Israel, 2016). There are also some free basic versions of SEO specific tools that can support unskilled employees with foundational tasks. But if SEO is carried out properly and on a regular basis, monthly tool expenses with costs up to several hundred euros per month are often inevitable in order to gain access to advanced data and analysis features (Erlhofer, 2020).

Taken together, these structural, resource-based, and technical hurdles have long made sustained and effective SEO seem out of reach for many SMEs, despite their awareness of its importance and potential. Limited budgets, a lack of in-house expertise, and the need for constant adaptation to evolving algorithms have created a playing field where mostly larger companies could fully commit to ongoing optimisation efforts. However, as the pressure on SMEs to stay digitally visible and competitive continues to mount, a new radical innovation has recently begun to shake these barriers at their core. With the rapid rise of generative AI, a new class of tools is emerging, potentially capable of revolutionizing how SEO has been done before.

2.7 ARTIFICIAL INTELLIGENCE

2.7.1 Early-stage integration of AI in Google's Search Engine Technology

Even before the skyrocketing rise of ChatGPT and other LLMs in 2022, the use of artificial intelligence within the field of search engines was not new. Already some years before, Google Search introduced AI and machine learning based models, each with specialized roles to deliver helpful results. As early as 2015, the first AI-based deep learning system was deployed in Google Search, called RankBrain. It was groundbreaking because it helped Google understand how words relate to concepts, aiding in returning relevant content by understanding the intent of a query without exact keyword matching (Google Search Central, 2025; Nayak, 2022), making methods like Keyword Density and WDF*IDF less important (Lammenett, 2025). By this, RankBrain helps find information that was previously inaccessible by broadly understanding how words in a search query relate to real-world concepts. Despite being the first deep learning model, it remains one of the major AI systems powering Google Search today (Google Search Central, 2025; Nayak, 2022). This marked only the initial breakthrough, followed by numerous other AI-driven search enhancements such as the neural matching algorithm. Introduced in 2018, this system utilizes neural networks to better understand how queries relate to pages by analyzing an entire query or page rather than just individual keywords (Google Search Central, 2025; Nayak, 2022). Another important step was the launch of BERT (Bidirectional Encoder Representations from Transformers) in 2019, representing a significant advancement in natural language understanding. The LLM helps Google to grasp the context of a sentence or query by connecting the dots between how combinations of words express different meanings and intents in order to then generate a compelling and accurate response (Cutler, 2024). It plays a critical role in almost every English query, as its complex language understanding enables it to quickly rank documents for relevance (Google Search Central, 2025; Nayak, 2022). One of the most recent milestones in AI search technology resembles the so-called Multitask Unified Model (MUM), introduced in 2021. MUM is said to be 1000 times more powerful than BERT and is designed to both understand and generate language, having been trained across 75 languages, and to perform a wide range of tasks simultaneously. Additionally, MUM is multimodal, meaning it can process and relate information across different formats such as text and images (Google Search Central, 2025; Nayak, 2022). After AI began to transform the way search engines work, the technology continued to empower users and SEO practitioners to actively harness AI for a wide range of SEO tasks.

2.7.2 Use of generative AI-based tools for SEO purposes

Following the launch of ChatGPT and the resulting global hype surrounding LLMs, the SEO industry also began exploring their potential applications. Since then, an increasing number of AI-powered SEO tools have become available to support various search engine optimization tasks, including keyword research, content generation, text structuring, and the automated creation of SEO-optimized title tags and meta descriptions (Lammenett, 2025; Lembke & Meil, 2022). While these tools can offer valuable support in automating isolated SEO tasks, they often remain limited to specific sub-disciplines, such as keyword research or content formatting, and typically come with separate subscription costs (Lammenett, 2025; Lembke & Meil, 2022). Especially for SMEs, this fragmented, cost-intensive tool landscape reinforces several of the barriers previously identified, including limited budgets and the absence of integrated workflows. Furthermore, most of these tools are based on freely available large language models, often using the ChatGPT algorithm. By this, they primarily function as customized and fine-tuned interfaces rather than offering fundamentally new capabilities (Lammenett, 2025). Given that many of the currently available SEO tools are both subscription-based and built on freely accessible large language models, the question arises whether direct use of these foundational models might offer a more cost-effective and flexible alternative. Rather than relying on fragmented, limited-scope applications, SMEs could potentially benefit from engaging directly with LLMs to address a broader range of SEO tasks in a more integrated and scalable manner.

2.7.3 Potential benefits and barriers of free-to-use LLMs for a holistic SEO approach

Looking at the landscape of generative AI models in the first half of 2025, there are already a couple of LLMs on the market, such as ChatGPT, Gemini, or Claude, that are, in their base

version, free to use. Almost every month new models or improvements of existing versions are launched, drastically outperforming their predecessors.
In parallel, the academic literature in this field is also rapidly growing. Though still in an early stage, recent literature in the crossover section of the usage of LLMs for SEO already identifies different fields of applications for this technology, especially in on-page SEO. One paper by Krol et al. (2024) confirms that different LLM-based AIs, including Microsoft Copilot, can be potentially useful for SEO optimization, particularly in generating text for website components like meta tags and headings. The authors also outline that, while most LLMs also excel at content optimization, which can assist firms to improve existing website content for search engines, the AI is not capable of conducting SEO audits in real-time and lacks direct access to analytical data and website quality metrics. But according to the paper, AI can be helpful for interpreting reports and data from third-party SEO tools, such as Google Search Console, and with that maybe simplify the analysis process, especially for SMEs. Still, the authors conclude that LLMs cannot replace human creativity and critical thinking yet completely, but rather optimize existing texts and or website components, and can be helpful in understanding SEO data reports (Krol et al., 2024).
Another paper by Kauppinen (2024) also comes to the conclusion that generating SEO content with generative AI tools can be a time-efficient way for firms to create new SEO content aimed at improving search engine rankings but points out that human editing and writing remain necessary to ensure accuracy, tone of voice, and to incorporate and emphasize the most important insights. One important finding was that the quality of AI-generated content can be improved by training the generative AI with existing branded content (Kauppinen, 2024).
Those findings are also supported by the paper of Guelailia & Bouziane (2024). The authors list different possible fields where AI has the potential to significantly enhance SEO practices, including keyword research, SEO strategy development, content creation and optimization, and enhancing accuracy in online listings by leveraging Machine Learning (ML) and Natural Language Processing (NLP) (Guelailia & Bouziane, 2024). Another study by Wohllebe and Lagodka (2024) elaborates on an experiment that compared the performance of SEO texts generated by ChatGPT with those written by human authors over an eight-week period on a German e-commerce website. The experiment analysed the Google ranking changes of 60 product category pages, with 30 containing ChatGPT-generated texts and 30 containing human-authored texts, for specific keywords. Neither type of text existed on the pages at the start of the experiment. Their findings revealed that both AI- and human-generated content had a statistically significant positive impact on search rankings compared to no SEO text

optimization before. But they could not find a statistically significant difference in effectiveness between the human writers and AI. Furthermore, the use of ChatGPT-enabled content creation resulted in substantial time and cost savings, as AI-generated texts required only a few seconds to produce (Wohllebe & Lagodka, 2024).

While in theory, this growing body of literature clearly highlights the benefits of using LLMs in SEO, in practice, especially SMEs often encounter significant hurdles when it comes to the implementation. In 2024, only around 11% of small enterprises in the EU reported to use use AI technologies, compared to 41% among large enterprises (Eurostat, 2025). Typically, models like ChatGPT operate on external servers, meaning that any internal or sensitive information input into the system leaves the direct control of the user. This poses substantial data security and compliance risks for organizations, especially in regulated industries or where customer data is involved. As current literature shows, while secure, closed-system integrations of such AI models are theoretically possible, they are currently associated with high implementation costs and demand advanced technical capabilities that many SMEs lack. By performing a systematic literature review, Oldemeyer et al. (2024) investigated the state of AI adoption in small and medium-sized enterprises. They identified 27 distinct implementation barriers, most notably limited knowledge, high costs, and inadequate infrastructure (Oldemeyer et al., 2024). The authors also found concerns around data privacy and security, stating that while the issue of data protection is currently underrepresented in the SME-focused literature due to the historically low data volumes and usage in smaller firms (Empl & Pernul, 2021; Oldemeyer et al., 2024), some scholars suggest this is set to change. As AI adoption grows and SMEs begin to collect and process larger datasets, data security and privacy concerns are expected to rise significantly in importance (von Joerg & Carlos, 2022; Oldemeyer et al., 2024). Similarly, Govori and Sejdija (2023) explore the barriers and prospects of AI adoption in SMEs within a review-based meta-analysis. The authors confirm that limited financial and technological resources continue to constrain SMEs in leveraging AI effectively. However, the study also emphasizes the growing potential of AI in this sector, noting that declining costs and ongoing innovation may gradually lower these entry barriers (Govori & Sejdija, 2023). Congruently, the paper by Iyelolu et al. (2024) on AI adoption in SMEs confirms that while AI holds strong potential, its implementation is hindered by persistent barriers like limited financial resources, a lack of technical expertise, and growing concerns over data security and privacy, but also employee resistance to change (Iyelolu et al., 2024).

In summary, current literature consistently identifies limited financial resources, lack of technical expertise, and inadequate infrastructure as key barriers to AI adoption in SMEs. Furthermore, while data privacy concerns have so far received limited attention due to historically low data usage, they are expected to grow in significance as AI adoption increases.

Acknowledging these persistent hurdles, this thesis proposes to explore an alternative solution by examining the potential of Microsoft Copilot. With Copilot, Microsoft, the world's biggest player in the office software supply segment (Office Software, 2024), provides an already accessible, integrated, and free-to-use AI model for their clients. It operates on the basis of Chat GPT 4, applying LLM technology within a prompt and answer interface, including the option to upload and analyse files. It is furthermore integrated into all office applications like PowerPoint, Word, and Excel. Moreover, as it is tailored for the usage within a firm's office software environment, Copilot offers robust enterprise data protection, ensuring a safe environment for organizational data and protecting it from leaking outside the organization (DHB-MSFT, 2024). With these prerequisites, Copilot could represent a great opportunity, especially for SMEs with limited resources, to integrate AI into their everyday business practices in a meaningful and simple way. If Copilot proves to be capable to support SMEs in carrying out of effective SEO, then a large number of firms already using Microsoft Office could directly benefit from its seamless integration by bypassing many of the prevailing barriers to both AI and SEO adoption, such as high implementation costs, limited in-house expertise, fragmented tool landscapes, and data security concerns.

So far, the performance of Copilot as an LLM interface has only been in other areas such as healthcare (Alhur, 2024; Du et al., 2025; Tepe & Emekli, 2024), education (Adetayo et al., 2024; Harun et al., 2024; Kateryna Osadcha et al., 2024), or in communications (Brantner et al., 2025; Kuznetsova et al., 2024), all with mixed results. Up to this day, there is no academic literature investigating the concrete performance of Copilot for on-page SEO appliances. This study is tackling this research gap by performing an online controlled experiment together with a German SME company, in order to shed light on the research question:

To what extent can the performance of product pages be improved by using Microsoft Copilot to generate SEO-optimized content?

To address this question more concretely, and following the insights from the literature review

(Chapter 2.5), the research question is examined by three hypotheses related to the three levels of SEO performance impact, being algorithmic relevance, user engagement, and commercial impact:

H1 – Algorithmic Relevance (Search Engine Level)

Product pages with Copilot-generated SEO content are perceived as more relevant by Google's algorithm, resulting in improved search visibility.

H2 – User Engagement (Human Level)

Product pages with Copilot-generated SEO content are perceived as more relevant and engaging by users, resulting in increased interaction.

H3 – Commercial Impact (Business Level)

Product pages with Copilot-generated SEO content lead to improved commercial outcomes by positively influencing purchase behaviour.

3 RESEARCH DESIGN & METHODOLOGY

3.1 EMPIRICAL CONTEXT

In order to tackle the identified research question and its corresponding hypotheses, a controlled online field experiment (Kohavi et al., 2020) was conducted for this thesis together with a German retail company. The company in this experiment can be classified as a German SME within the beauty and healthcare industry sector, operating in over 50 countries with 280 employees and a corporate net revenue of about 50 million € in 2023, of which about 60% was generated in the German market (Unternehmensregister, 2025). By the time of the experiment, the employees of the company had already started working with different publicly available AI tools, including ChatGPT for generalized tasks, but, as in most other companies, were not allowed to use them for any tasks involving company, product, or customer-specific information because of data privacy restrictions. Although the use of this type of sensitive data was permitted when using MS Copilot, most employees did not have access to the AI by this time, as the company-wide roll-out of MS Office 365, and thus the access to the secure environment of the AI, had only just taken place.

3.2 THE CONTROLLED ONLINE FIELD EXPERIMENT

Within the field experiment, 20 products from three different product lines were randomly but equally divided into a treatment (2) and a control group (1), where each product has one corresponding product page in the firm's B2C online shop. The decision on which product lines were integrated in the sample was influenced by the degree of prior SEO optimization attempts. Some product pages within the firm's portfolio were already (in parts) subject to SEO initiatives, either through company internals or through the implementation of advice received by an external SEO agency. But due to time and resource limitations, those changes were not deployed to the full extent across all products. For the experiment, only product lines were included that had not been touched by explicit SEO initiatives before in order to capture the unbiased effects of the treatment. Following the guidelines for online controlled experiments by Kohavi et al. (2020), the assignment to treatment and control group was randomized but equal within every product line (Kohavi et al., 2020).

For all product pages of the treatment group, MS Copilot was prompted to create new, search engine optimized product texts, replacing the preexisting human-written texts. The prompt used for receiving the search engine optimized product texts was the same for every product, except product specific information such as product name and product specific keywords, ensuring comparability across the treatment group. The prompt contained a rough structure regarding the product pages' textual layout, following company internal guidelines, instructions regarding company specific tone, style, and content restrictions, as well as parts dedicated to SEO techniques such as the use of product specific focus and a secondary keyword with corresponding proof keywords or the generation of a SERP snippet. In order to leverage the full potential of the AI, the prompt was optimized according to the latest discussed prompt engineering techniques. The full prompt, the applied prompt engineering techniques, and a description of the purpose of each prompt section are outlined in detail in chapter 3.3. To ensure that Copilot has accurate information about the products, the AI was provided with product specific internal information material in each prompt, which is normally used by human content creators as well. These internal documents always consisted of the product- and efficacy descriptions approved by the legal department but could also include all kinds of other sources such as the instruction leaflet, product catalogues, descriptions of active ingredients, advertising materials, press releases, training documents for internal sales staff or application instructions for the end consumer or B2B customers such as cosmetics institutes. The use of Microsoft Copilot, connected with the use of internal company data, as well as the prompt engineering aspect, are crucial aspects considering the tackled research gap. The prompt, as well as all generated outputs, were formulated in German as this is the operating language of the company. After a review process through the Legal and Scientific Affairs department of the company, the final product texts, as well as the new SERP titles and meta descriptions generated by the AI were deployed for the treatment group on all German product pages, while the product pages of the control group remained untouched. After a period of 8 weeks, the collected data was analysed for common patterns using various statistical methods as outlined in section 3.6. below.

3.3 TREATMENT DESIGN AND IMPLEMENTATION

The treatment in this experiment was the exchange of a product text, formerly written by an employee from the marketing department of the company, for a product text that was fully generated by MS Copilot. In order to ensure comparability throughout the whole experiment, it is important to use the same prompt for the generation of all product texts in the treatment group, as it was done, for example, by Wohllebe & Lagodka (2024). In this case, the only differences in the prompt across the treatment group were product related differences, such as the name, the product line, product-specific keywords, and other product specific information to be included by the AI.

Wohllebe and Lagodka (2024), for example, used one rather simple and generic prompt in their experiment:

Generate an SEO text about [keyword].

- 275 words in German
- Use of three to five keywords with a search volume of ≥ 500
- An H2 title with one keyword
- Structure of SEO text into sections
- One H3 heading with one keyword per section
- Three to five internal links
- Relevant keywords in bold
- No mention of online stores, customer service or delivery
- Focus on the added value of the product
- Inform the reader about factors relevant to the purchase decision
- Address the reader in an informal manner using “Du” (German for “you”)

Figure 3. Prompt Example from Wohllebe and Lagodka (Wohllebe & Lagodka, 2024)

But in practice, the complexity is often embedded within the finer details. For example, as in this case, the company had precise guidelines anchored in its corporate identity regarding style and tone when it comes to the communication towards customers, ranging from certain forms of addressing the customer (e.g. formal vs informal) over certain writing styles (informational, scientific, casual, entertaining, etc.) to a consistent page structure throughout all product pages. Also, when it comes to the generated claims concerning a product and its efficacy, AI should not simply reproduce all available information unregulated. Most product descriptions must adhere at least to a certain extent to legally and scientifically backed up product promises. How close the AI needs to adhere to the wording of certain characteristics of the product is often a tricky act of balance and must be precisely outlined in the prompt in order to be understood by the AI. Including all those requirements in a clear and unmistakable way automatically leads to the prompt being more elaborate.

Another rationale for using a more elaborate prompt lies in the fact that the quality of the output generated by a LLM is highly influenced by the quality of the input provided by the user when communicating the task at hand (Chen et al., 2023; Knoth et al., 2024; Schorcht et al., 2024; White et al., 2023). Therefore, to accurately evaluate the capabilities of an LLM in performing a complex task such as generating search engine optimized content, it is essential to apply prompt engineering techniques. These techniques involve designing prompts in a way that clearly clarifies the intent, provides relevant context, and gives the model a structured framework to follow. This enables the LLM to interpret the prompt more effectively and produce responses that are better aligned with the task's objectives (Chen et al., 2023; White et al., 2023). Without this, the assessment would risk underestimating the AI's true capabilities due to poorly formulated inputs.

Considering these aspects, there are still, in fact, endless possibilities for designing the prompt for this certain purpose. It is therefore of highest importance to clarify in advance which criteria the output must fulfil and assess the output according to these criteria in order to come up with a prompt that is leveraging upon the true potential of the AI.

3.3.1 Prompt design & testing

When designing the prompt for the experiment, the goal was to ensure consistently high-quality output that was not only optimized for search engines but also aligned with brand guidelines and compliant with legal requirements across repeated uses of the same prompt. To achieve this, the prompt development followed a structured, iterative process based on principles from current prompt engineering literature and was designed around three sequential quality gates. Each quality gate introduced a specific layer of evaluation and refinement, involving distinct stakeholder groups.

The process began with a self-evaluation phase to ensure baseline output quality through internal testing and refinement (Quality Gate 1). The AI-generated outputs were evaluated for common errors, including basic issues with adherence to the given tone, structure, basic factual accuracy, and basic SEO techniques such as proper keyword integration. Already in this stage, the prompt was refined using established prompt engineering techniques until a sophisticated level of quality was reached throughout most outputs.

Once the first version of the prompt passed quality gate 1, the prompt was again refined within structured feedback loops with experts from digital marketing and e-commerce (Quality Gate 2), who evaluated the generated outputs for alignment with brand identity, presumed SEO effectiveness, and stylistic and structural consistency. This step ensured alignment with operational marketing standards and completed the second quality gate.

In the final stage, the AI-generated texts were reviewed by the company's Legal and Scientific Affairs department (Quality Gate 3). The focus was on ensuring factual and legal accuracy, particularly regarding claims related to product efficacy and active ingredients. Feedback from this department was again iteratively implemented through targeted prompt refinements and re-testing. This stage concluded the third and final quality gate, leading to the final prompt version.

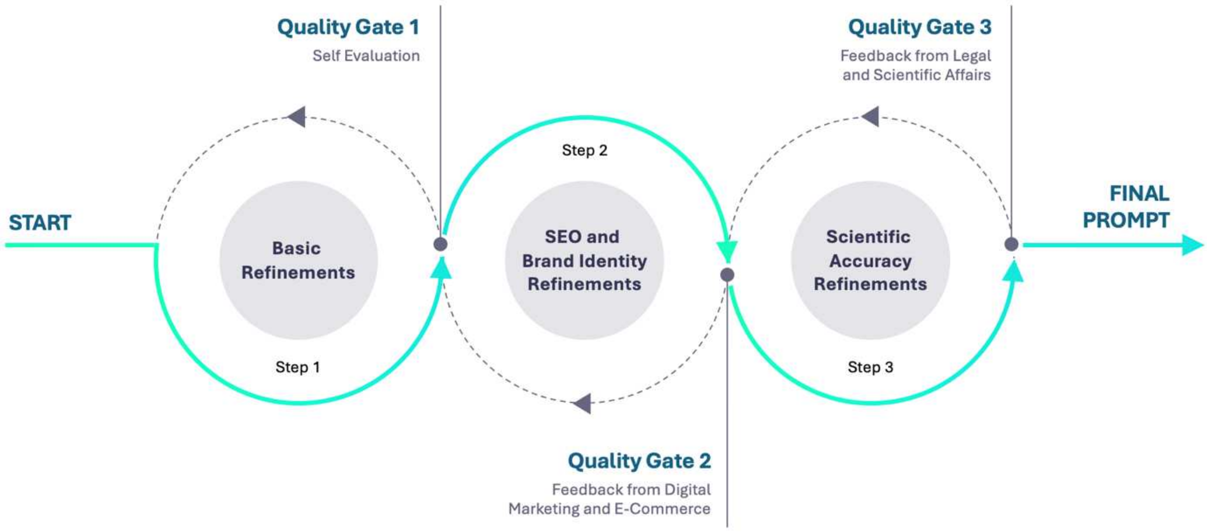


Figure 4. Iterative prompt design and testing approach

This iterative prompt engineering approach made it possible to systematically align the prompt with technical, commercial, and regulatory requirements. In total, 110 different test outputs along all three testing steps were created and evaluated. The tests were conducted with all products from the treatment group and their corresponding information material. Also it is important to mention that for every single prompt test, as well as for the creation of the final outputs, a new chat window was opened in Copilot instead of continuing an ongoing dialogue within the same chat window. This decision was made to ensure that every output was generated based solely on the prompt and product input, without any influence from prior interactions or contextual memory within the chat. While it is possible that refining the output through a multi-step conversation might have improved the quality of individual texts, it would have also introduced inconsistencies in the input-output process, making systematic comparison across versions less reliable. By resetting the chat for each test, the experiment maintained a controlled environment, enabling a more objective evaluation of prompt effectiveness across all outputs.

Et the end of the iterative design & testing phase, the prompt that was continuously delivering the highest quality output across all quality gates was a lot more sophisticated compared to the one used in the experiment by Wohllebe and Lagodka (2024), due to the inclusion of corporate tone, style and content restrictions as well as several prompt engineering techniques, as outlined in detail in the next section.

3.3.2 The final prompt

The original coherent prompt in German, together with the English translation, can be found in the Appendix (7.1.1. & 7.1.2.). It was structured into four clearly defined sections: Role, goal, working procedure, and the final work instruction, with each section being clearly separated by three hash marks in front of the section's header (e.g., "### Role"). This inclusion of explicit and consistent formatting (e.g., hash marks or section numbers) is a prompt engineering technique, especially for encapsulating large and more complex prompts, clearly structuring the input for the AI (Chen et al., 2023).

1) *### Role*

The prompt begins by assigning the AI a role, in this case professional identity of a SEO writing expert specialized on cosmetic products. This use of Role Prompting, also referred to as the Persona Pattern, guides the model's behaviour and output style. It sets an expert-level perspective that primes the LLM to focus on relevant vocabulary, structure, and tone from the outset (Chen et al., 2023; White et al., 2023).

```
### Role
You are an SEO writing expert who specialises in creating SEO-optimised texts for
cosmetic products. Your task is to create texts that are based on the provided internal
documents, and which fulfil SEO criteria in order to improve the ranking on Google and
other search engines.
```

Figure 5. Prompt head including Role Prompting

2) *### Goal*

This section clarifies the objective of the task: Generating SEO-optimized product page content based exclusively on internal company documentation. It sensitises the AI for given boundaries (e.g., no unfounded product claims), specifies the audience (end consumers), and how the content should address this audience (provide added value, convince to buy). By doing so, it builds upon the Persona Pattern, explaining the context for the AI. It applies the Context Manager Pattern, ensuring irrelevant content is systematically excluded from the output (Chen et al., 2023; White et al., 2023).

Goal
Create an SEO-optimised text for the product page of the product '*insert product name*', from the product line '*insert product line*', of the brand [redacted], which strictly adheres to the information from the attached internal documents. It is aimed to target end consumers for home use and not at beauticians and cosmetic institutes.
Detailed product descriptions offer good potential to offer potential buyers the best possible experience. Use the data provided as a basis to personalise the customer journey for specific target groups. The text on the product detail page should offer potential customers added value, provide directional advice and convince them to buy the product.

Important:
The text must not contain any product-related claims that are not explicitly given within the documents. This is particularly important for ingredients, application and promises of effectiveness

Figure 6. Definition of prompt objectives

3) *### Working procedure*

The Working Procedure section forms the core of the prompt and is divided into seven logically sequenced, numbered instruction steps that are expected to be followed consecutively by the AI. Each step includes a critical part for the content creation process, including (1) The scanning of the provided documents, (2) A reflection on SEO related practises, (3) The structural framework for the content creation, (4) Content specifications for the product texts, (5) Tonality guidelines, (6) Fact checking and self-control and (7) Instructions for the creation of a meta description. This stepwise instruction approach applies the Recipe Pattern, enabling the model to execute this complex, multicomponent task by configuring specific "ingredients" into a clearly defined content format (White et al., 2023).

STEP 1: Review Internal Documents

At first, the AI is instructed to read through the provided info material and to focus only on product-specific information from attached internal sources, emphasizing constraints around claims and ingredient usage. In this case, it is advised to stick closely to the claims of what is called "Produktauslobung lang", which can be translated to "long product claim". The long product claim is a fixed product description, defined uniquely by the Legal and Scientific Affairs Department for every product, and contains the precise claims regarding a product's effectiveness, which are legally backed up by product-specific efficacy studies. This part of the prompt prevents hallucinations, reinforces the usage of specific factual input, and ensures compliance with legal guidelines, functioning as a Context Manager Pattern.

Working procedure

1 **View internal documents**:

- Carefully read the attached internal documents. Concentrate exclusively on information about the product '*insert product name*', from the product line '*insert product line*', of the brand [redacted]
- Pay particular attention to the information on ingredients, application and promises of efficacy, which can be found in the „Produktauslobung lang"

Figure 7. Working procedure step 1: Review of internal documents

STEP 2: A reflection on SEO related practices

In this step, the AI is prompted to "think out loud" about key SEO techniques before starting the actual content generation. This instruction applies a combination of two prompt engineering techniques: Generated Knowledge and the Golden Chain-of-Thought Pattern. The Generated Knowledge technique encourages the model to brainstorm potentially relevant SEO related strategies and best practices, such as appropriate heading structure, keyword placement, readability, or internal linking, even if these are not explicitly mentioned elsewhere in the prompt. By doing so, the model is nudged to self-generate a richer contextual foundation to draw from in the writing phase (Chen et al., 2023). Simultaneously, the Golden Chain-of-Thought Pattern prompts the model to reason through these elements in a predefined, structured, and sequential way, ensuring that its reflections are not only broad but also logically organized (Chen et al., 2023; White et al., 2023).

2. **Consideration of SEO criteria**:

- Think out loud about which measures are important to optimise the texts on the product page according to SEO criteria. The aim for the product page is to rank as high as possible in search engines such as Google.
- Consider the following points, among others:
    - Use of H1, H2, H3, etc. headings
    - Use of secondary and proof keywords
    - Structure and readability of the text
    - Keyword density/stuffing
    - Internal links
    - Content optimisation incl. product description, application notes, customer experience, customer benefits/special problems that the product solves
    - Call to action
    - Other important text-related SEO techniques

Figure 8. Working procedure step 2: Consideration of SEO criteria

STEP 3: The structural framework for the content creation

In step 3, the AI receives a detailed content template for the product detail page, including fixed sections such as product description, usage instructions, target group fit, and beauty routine integration. Each of these sections was labelled and formatted consistently, using H1 and H2 tags where appropriate. This implements the Template Pattern, which is designed to ensure that the generated text adheres to a certain structure, improving output comparability and usability for web integration (White et al., 2023). The corresponding extract from the prompt is shown in the Appendix (7.1.3).

STEP 4: Content specifications

This step specifies different aspects regarding the content to be generated. One important part is the integration of focus and secondary keywords, alongside with their related proof keywords, including rules for their usage and placement (e.g., include focus or secondary keyword in H2 headings, maintaining natural language flow, avoiding keyword stuffing, etc.). As already outlined in the literature review, the definition of the relevant keywords is a complex process, highly intertwined with the strategic positioning of the firm and its products within their competitive environment. In this case the focus- and secondary keywords for which the site was optimised were not determined by the AI itself but were provided by the company's internal SEO expert on the basis of keyword research (via Google Keyword Planner), in view of existing average rankings (via Google Search Console) and taking into account factors such as search volume, keyword type (informational/transactional) & opportunities (competition high/medium/low). This was done because the integration of the designated proof keywords is an essential SEO technique, and Copilot did not have access to the internet by the time of the experiment, and thus had no ability to access them on its own. Maybe at a later development stage, the machine might be able to perform this task on its own. In a second step, a WDF*IDF tool was used to generate a list of relevant proof keywords for each of these keywords as input for the AI. Furthermore, the content specifications in this section include the repeated order for close adherence to the provided information, as well as content aspects that may not be addressed by the AI, like claims regarding availability or delivery (Content Manager Pattern) (Chen et al., 2023; White et al., 2023). The corresponding extract from the prompt is shown in the Appendix (7.1.3).

STEP 5: Tonality

In this step, the model was instructed with the tone of use: a fact-based but sales-oriented style, short sentences, neutral forms of address, etc. This section ensures that the final output consistently aligns with the brand-owned tonality, modern standards for politically correct addressing forms, and user readability. By setting these stylistic parameters, the prompt limits undesirable model behaviours and creates a consistent tone across all generated outputs (Chen et al., 2023; White et al., 2023).

```
5. **Tonality**:
  -  Use sales-promoting and at the same time factually correct formulations
  -  Make sure to work in rather short sentences (avoid nested sentences)
  -  When addressing the customer, use a gender and persona neutral form or the
     formal form of “you” instead of the informal form.
  -  Language: German
```

Figure 9. Working procedure step 5: Tonality

STEP 6: Fact checking and self-control

In step 6, the AI is prompted to generate a set of post-task checklists to perform self-control on key factual elements included in the text, such as product claims, ingredients, used proof keywords, and recommended internal products. On the one hand, this can serve as a measure of self-control regarding aspects of high importance like factual claims. This technique reflects the Fact Checklist Pattern, which is designed to mitigate the risk of hallucinated or unverifiable content through the AI itself (White et al., 2023). On the other hand, it also adds transparency and facilitates internal review by, for example, the legal department, and with this also supports operational efficiency.

```
6. **Fact Checking**
  -  Create a list of the most important facts and ingredients contained in the text and
     check them for accuracy.
  -  Create a list of the proof keywords you have included
  -  Create a list of suggested products for internal linking
```

Figure 10. Working procedure step 6: Fact checking & self-control

STEP 7: Instructions for the creation of a meta description

In the final step, the AI was asked to generate a title and meta description for the Google SERP, effectively summarizing the page’s content and activating the user to visit the site.

7. **Title and meta description for the SERP snippet**
- Create a title and meta description for the product page
- These should reflect the content of the page in a short, correct and succinct manner and encourage the user to click on the search engine hit with a call to action.
- The ending of the title should always look like this: ‘| [redacted] ’. The end of the description should NOT include this ending with the company name
- Stick to the optimum number of characters (including spaces):
    - Title: maximum 50 characters
    - Meta description: maximum 120 characters

Figure 11. Working procedure step 7: SERP Snippet generation

4) ### Final work instruction

The final line of the prompt instructs the AI to begin processing the task step by step without skipping any part. This reinforces the Recipe Pattern, emphasizing a procedural and controlled execution of the task.

Start with the editing. Work step by step. No point may be skipped.

Figure 12: Final work instruction

3.3.3 Content generation & deployment

After designing the final prompt, it was used to generate the final output for the product page. This was done by generating two separate output versions per product in the treatment group. In both times, identical input was used in separate chat windows, including the same set of internal information material and focus-, secondary- and respective proof key words. After the creation of the two outputs per product, selected experts from the company’s digital marketing and e-commerce teams reviewed both generated product texts and selected the superior version. This choice was based on a combination of objective criteria such as factual accuracy, inclusion of proof keywords, and logical consistency. Following their assessment, they had to choose one of the two texts as a whole to be the final text. Mixing passages or editing the outputs was not permitted. This approach was based on the understanding that large language models remain probabilistic systems, producing different results even when prompted with identical input. As such, there is no fixed one-to-one relationship between input and output. To account for this inherent variability and better leverage the model’s full potential, the process included this selection step of the best version. By doing so, the experiment aimed to simulate a realistic workflow while still isolating and measuring the AI’s best possible performance within a controlled selection framework.

After the final version of the generated product text was chosen by the experts, it was passed to the company’s Legal and Scientific Affairs department for review, a regular and mandatory requirement for every text published in the name of the company. All passages containing wrong or misleading claims from a scientific or legal perspective were respectively eliminated from the text. After this, no further changes were applied to the AI-generated texts. Following this last review step, the generated product texts and the corresponding SERP snippet were set to fully replace the previous text on the product pages for all products from the treatment group.

3.4 DATA COLLECTION AND VARIABLE OPERATIONALIZATION

After the go-live of the AI-generated texts in May 2025, the data for the experiment was tracked using Google Search Console, Google Analytics (GA4), and internal sales reports of the company. The logs included panel data for every page on a daily basis, recording the development of different SEO related KPIs that were in a later stage used to test the formulated hypotheses. The evaluation of the aggregated panel data set was performed using R and RStudio and is available upon request.

For H1 (algorithmic relevance), the ranking of the page (*rank*) was tracked, being a core metric for assessing visibility within search engine results. To account for the non-linear meaning of changes in ranking (changes from position 10 to 9 are not as “valuable” as from position 2 to 1), dummy variables were introduced, capturing the number of pages ranking in the top 20 (*top20*), top 10 (*top10*), top 5 (*top5*) and top 3 (*top3*) of the Google SERP. In the case of *rank*, a value of zero was never assigned, as this would misrepresent the absence of a ranking. Instead, if a page was not ranked on a specific day, the value was coded as ‘NA’ to accurately reflect the lack of a ranking position. In addition to *rank*, impression counts were analysed to reflect whether Google’s algorithm considers a page relevant enough to display it in response to a search query (*impressions*). The variable was coded zero if the page was not shown to a user on a specific day. Together, these variables form a robust basis for testing H1 while aligning with both the theoretical foundation and practical insights outlined in the literature review (chapter 2.5).

For H2 (User engagement), the number of clicks (*clicks_engagement*) was chosen, as they exactly determine how many users clicked on the displayed page in their SERP, capturing in how far the page was actually perceived as relevant by the users. The variable was coded as zero only when a page was impressed but not clicked, and as NA when the page was not impressed, since a click was not possible in that case. This distinction prevents artificial deflation of click rates and more accurately reflects true user engagement opportunities. Furthermore, CTR (*ctr*) complements this by offering a relative measure of success, reflecting how well a page converts impressions into actual clicks. It links the dimension of visibility with user interest and thus captures how persuasive a page's snippet is in the SERP, bridging the algorithmic and user engagement perspectives. It was set to zero when impressions occurred without clicks, and coded as NA when there were no impressions, since division by zero is undefined. This coding aligns with the logic applied to the clicks variable and ensures accurate representation of user engagement. Finally, average engagement time per active user (*eng_time_pau*) allows for deeper insight into the engagement of the user. It measures the duration (in seconds) for which a user actively interacts with a page after arriving on the site. This metric includes all users, regardless of whether they come from organic search, paid advertising, or any other channel. (Google Analytics, 2023). In this context, measuring the engagement of users from all traffic sources is not a limitation, as it focuses solely on user behaviour after already landing on the page. Importantly, the variable never reaches a value of zero, as it is only recorded when a user actually visits the page, ensuring it always reflects active engagement. Similar to rank and clicks, if there is no exposure of the page to a user, the value is set to NA.

Lastly, to evaluate H3 (the commercial impact), the quantity of products sold (*sales*) was selected as the driving KPI as it provides a direct measure of the core objective in e-commerce: increasing the number of transactions attributable to SEO activities. Using the quantity sold rather than revenue mitigates potential biases introduced by price variability across products, ensuring that observed effects reflect changes in sales volume rather than fluctuations in pricing. To provide a tangible sense of commercial impact, changes in sales are contextualized using the average product price, translating sales effects into estimated revenue changes while controlling for price heterogeneity. Following Cutler (2024) and Usmany et al. (2024), only those sales traceably linked to organic search traffic were considered, in order to isolate the impact of SEO from other marketing channels.

Within the dataset, each variable was analysed for outliers, showing that all variables contained values beyond the 1st and 99th percentiles. After thorough consideration, it was decided to deliberately retain these outliers for two main reasons.

First, there was no evidence to suggest that these extreme values resulted from measurement or transcription errors or other sources for outliers common in this field (e.g., bots or spam behaviour) (Dobson & Barnett, 2018; Kohavi et al., 2020). On the contrary, they reflect genuine real-world observations that are plausible within the empirical context of this study and thus are essential for accurately representing the actual range and variability in the real product page performance.

Second, robustness checks were performed by winsorizing the variables at the 1st and 99th percentiles and comparing the new page means in each group and period, overall group and period means, and their difference-in-differences with those from the unaltered data (Fox, 2016). The comparison between capped and uncapped data showed only negligible changes, with no meaningful impact on the overall descriptive patterns and substantive conclusions derived from it (see Appendix 7.1.4). This observed stability is likely due to the limited number and moderate strength of the present outliers, combined with the sufficient number of daily observations per mean, ensuring that averages are not unduly influenced by occasional extremes. Additionally, residual diagnostics for the difference-in-differences models showed that for each variable of the uncapped data, fewer than 5% of standardized residuals exceeded a value of ±2 (see Appendix 7.1.4). Exceeding this threshold may indicate poor fit due to influential outliers that could bias the DiD regression estimates, which was not the case for any of the variables of interest (Fox, 2016). These findings collectively justify the decision to retain all observed values in the final analyses, ensuring a valid and comprehensive representation of the underlying data.

3.5 Methodological Constraints and Statistical Sensitivity of the Experimental Design

Due to constraints in time and scope inherent to this master's thesis, the empirical analysis was limited to an 8-week post-treatment observation period, resulting in a total analysis window of 16 weeks (comprising 8 weeks pre-treatment and 8 weeks post-treatment). Furthermore, limitations imposed by the company side restricted the sample to a total of 20 product pages, with 10 pages assigned to the treatment group and 10 pages to the control group. This sampling frame yielded a balanced panel data set, comprising 2,280 daily observations per outcome variable and 112 daily observations per variable for each individual product page.

Although the dataset contains daily observations for each product page, these repeated measurements cannot be considered statistically independent. This is because the intervention

was implemented at the page level, and the content of each page remained unchanged throughout the post-treatment observation period. Consequently, both the unit of randomization and the appropriate unit of statistical inference are the product pages themselves, rather than the individual daily observation. Treating daily measurements as independent data points would violate the independence assumption underlying standard statistical tests, resulting in inflated estimates of statistical power and an increased risk of Type I error. Therefore, the effective sample size for conventional statistical analyses is restricted to the number of unique product pages, being 10 in the treatment group and 10 in the control group.
To estimate the sensitivity of the given experimental design, Cohen's d and the minimum detectable effect size were calculated for each outcome variable using the variability observed in the pre-treatment period. Given a sample size of 10 pages per group, 80% power, and a 95% confidence level (Kohavi et al., 2020), the minimum detectable standardized effect size (Cohen's d) was calculated to be 1.32 (see Appendix 7.1.5). According to conventional benchmarks (Cohen, 1988), this corresponds to a large effect size, indicating that the study is only sufficiently powered to reliably detect large differences between groups with statistical significance. In the context of SEO interventions, particularly when implementing one-time changes, the expected impact on KPIs is often rather small (Cutler, 2024; Erlhofer, 2020; Lammenett, 2025). As such, a large amount of data is typically required to detect these subtle shifts with adequate statistical significance (Kohavi et al., 2020). Given the mentioned sample size limitations, it is acknowledged that the statistical power to detect small effects and associated formal hypothesis testing may be limited. However, since this study can be seen as a pilot study in its field with no prior experiments having applied the same treatment under comparable conditions, in the environment of Copilot before, the KPI shifts known from existing literature may not be perfectly predictive, but serve only as approximations. In light of these considerations, a structured and methodologically sound empirical strategy was employed to extract as much informative value as possible from the available data, given the mentioned limitations. Moreover, the value of this analysis extends beyond the initial observation window, as all product pages will remain unchanged beyond the scope of this thesis, enabling continued data collection and longer-term evaluation. As such, this pilot serves as a foundation for a more robust, longitudinal future research, increasing the potential statistical validity of the results over time.

3.6 EMPIRICAL EVALUATION STRATEGY

As outlined in the limitations section, the experimental design is underpowered due to the small sample size and the panel structure. Consequently, formal hypothesis tests such as t-tests were not performed, as these would not yield meaningful statistical inference. Instead, to identify possible indications of treatment effects, a multi-level analytic approach was employed for each variable.

Initially, the distribution of each variable was examined in its raw form in order to assess the presence of non-normality, missing data, or outliers and to better understand the underlying structure and variability of the data. This initial exploratory was performed for identifying potential data quality issues and for selecting suitable subsequent analytical techniques given the observed characteristics of each variable.

Building on this, the second step included the aggregation of each variable to page-level means of the experimental periods to account for intra-page temporal correlation and to ensure the independence of analytical units. On these page-level aggregates, extensive exploratory data analysis was conducted, utilizing a range of descriptive statistics and graphical visualizations to reveal patterns or trends suggestive of treatment effects. Differences-in-differences were examined descriptively by comparing changes in page-level means across groups and periods, allowing for the identification of group-specific trends that might not be visible in simple before-and-after comparisons. Histograms and boxplots were used to inspect the distribution and spread of key variables, facilitating the detection of skewness or outliers. Line plots were employed to visualize trends and temporal dynamics in mean values across the observation window. Accumulation plots illustrated the build-up of outcomes over time, highlighting differences in the timing and magnitude of effects. Dumbbell plots provided a clear depiction of changes in key metrics from pre- to post-intervention, providing deeper insights on the direction and magnitude of changes on a page level within groups. Each of these visualization techniques contributed complementary perspectives on the data, providing a compelling strategy for the interpretability of the descriptive analysis.

In the final step of the analysis, a difference-in-differences regression was performed, leveraging the strengths of the panel data structure, with the key parameter of interest being the coefficient of the interaction term between treatment assignment and the post-intervention period. The analytical model incorporated a fixed effects specification in order to account for common temporal shocks, ensuring that estimated effects are not driven by persistent differences between pages. Robust and clustered standard errors at the page level were employed to correct for potential within-page correlation and heteroscedasticity, thereby providing more reliable inference given the repeated measures design. This regression-based approach added a layer of statistical rigor and causal interpretability, complementing the descriptive analyses by formally testing for differential changes over time attributable to the intervention. It should be noted that for some variables, this approach may not represent the optimal modelling strategy. For example, changes in average ranking position are difficult to interpret due to the ordinal nature of the rank variable, and variables such as clicks may exhibit high zero inflation, which can challenge the assumptions of standard regression models. These limitations in interpretability are recognized for the DiD regression analysis. The application of more specialized methods, such as ordinal regression for rank data or beta regression for proportions, was deemed beyond the methodological scope of this thesis but could be considered in future research.

4 RESULTS

4.1 H1 - Algorithmic Relevance (Search Engine Level)

4.1.1 Rank

The rank variable, representing the average daily search engine position of each product page, has a mean rank of 10.96 (SD = 10.98), with a median of 8, indicating that half of the daily observed ranks are 8 or better and half are under. The observed values ranged from 1 to 99 (see Appendix 7.2.1.1.). Appendix figure 7.2.1.2. displays the density of daily rank values, highlighting a right-skewed distribution with the majority of values concentrated between ranks 1 and 10. Further stratification by group and period (Appendix figure 7.2.1.3.) indicates a broadly similar distributional pattern in both the control and treatment groups, before and after the intervention.

Figure 13 displays the weekly mean rank by group, with standard error bars, over the study period. Both groups show relatively stable mean ranks prior to the intervention. Directly after the intervention, mean ranks of the treatment group decreased sharply from about 13 to 9, indicating better search positions, while the control group stayed stable. This led to an assimilation of the treatment's average ranking towards the previously higher control group ranking, which stayed stable until the last week of the experiment. In the last week mean ranks of the treatment increased while the control decreased, resulting in diverging trend lines.

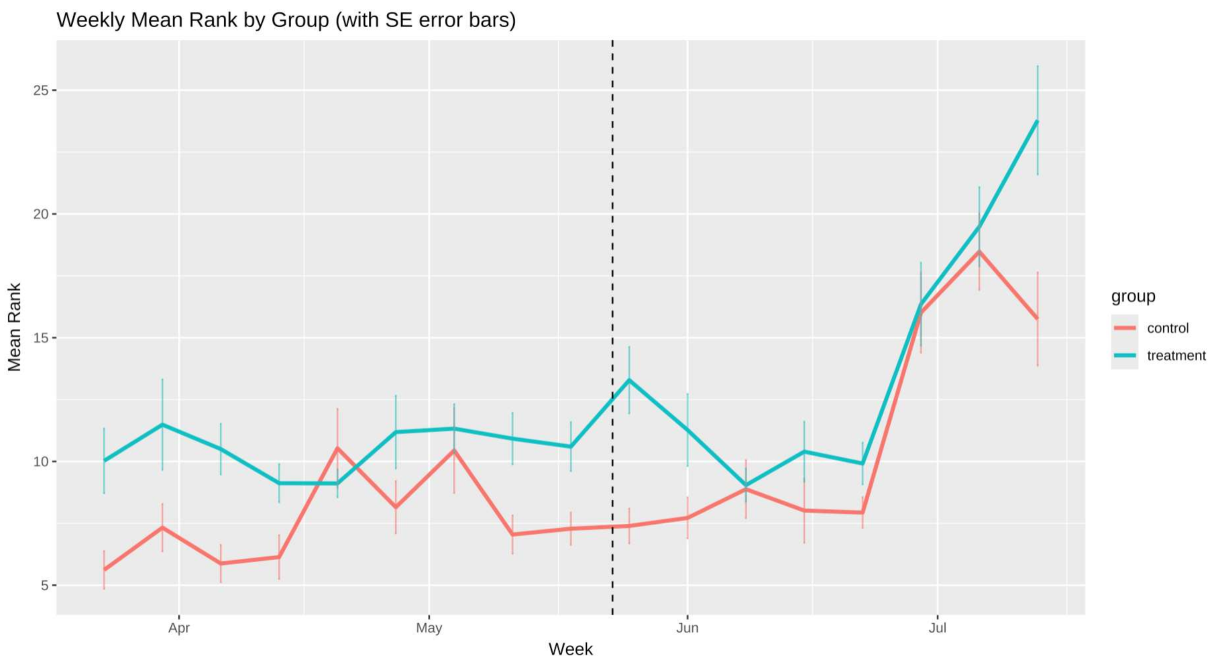


Figure 13. Mean weekly rank by group

To account for temporal correlation in daily observations, subsequent analyses aggregate rank at the page level, comparing mean ranks per page across groups and periods. The descriptive difference-in-differences (DiD) table (Appendix 7.2.1.4.) and accompanying plot (figure 14) show that mean daily rank per page increased from pre- to post-intervention in both groups. The control group's mean rank rose from 8.10 to 11.67 (Δ = 3.58), while the treatment group increased from 10.40 to 13.86 (Δ = 3.46), resulting in a difference-in-difference of -0.12. The 95% confidence intervals shown in figure 14 indicate substantial overlap between groups and periods, suggesting that the observed differences in mean rank are not statistically significant. The pattern observed in the overall difference-in-differences analysis is also mirrored at the individual page level in the dumbbell plot (figure 15), which shows that nearly all pages in both groups notably worsened in terms of mean daily rank from pre- to post-intervention, with similar shifts visible within both control and treatment groups. This common shift can also be seen in the faceted histograms (figure 16). In the box plot, the control group shows consistently lower medians then the treatment group, comparing pre and post period groups. After the intervention, both groups display a noticeable upward shift in their medians and inflation of their Interquartile Range (IQR), indicating that heterogeneity in page-level ranks increased in the post period, especially for the treatment group (figure 17).

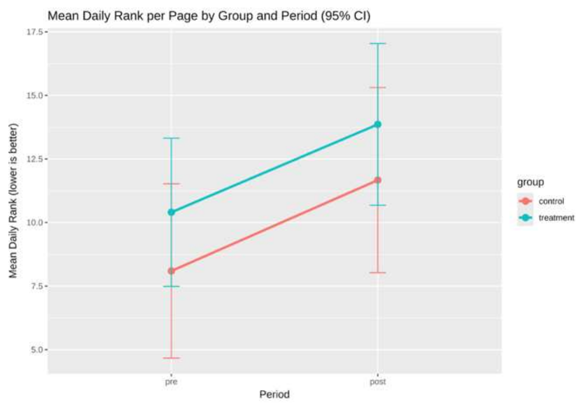


Figure 14. Mean rank, faceted by group and period

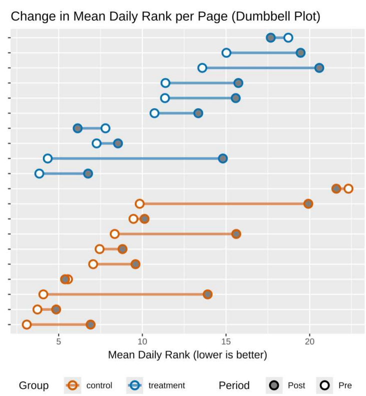


Figure 15. Mean page level rank, by group and period

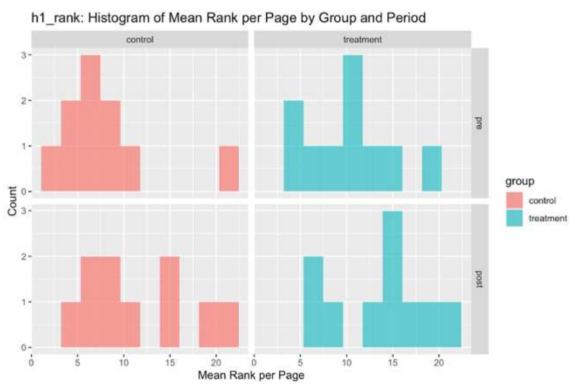


Figure 16. Faceted histogram of mean rank

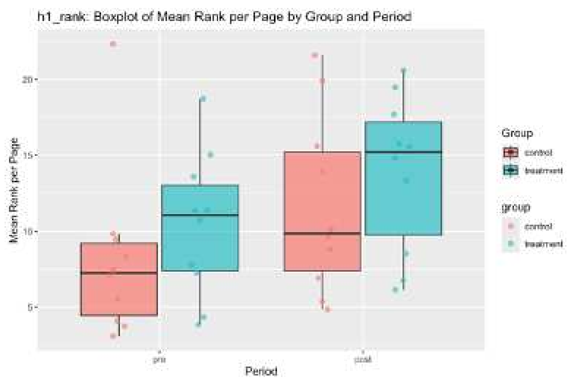


Figure 17. Boxplot of mean rank by group and period

To conclude, a difference-in-differences (DiD) analysis was performed using a fixed effects model with robust, page-clustered standard errors to account for unobserved heterogeneity across pages and the correlation of repeated measures within each page over time. The parallel trends assumption, as shown in the earlier weekly line plot, appears reasonable, with both groups displaying similar rank trajectories prior to the intervention. While the distribution of mean ranks is somewhat right-skewed, log-transforming rank is not appropriate given its ordinal nature. The regression yields a non-significant treatment effect (treat_post = -0.20, SE = 1.60), indicating no measurable difference in rank development between treatment and control groups post-intervention. The explained variance by the model is rather low ($R^2 = 0.18$, adj. $R^2 = 0.12$), indicating that most rank variation is due to page and time factors rather than treatment (see figure 18).

	Dependent variable:
	rank
treat_post	-0.197
	(1.599)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Observations	2,029
R2	0.179
Adjusted R2	0.121
F Statistic	3.617*** (df = 114; 1895)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.

Figure 18. DiD regression table of rank

4.1.2 Rank Dummies

Because rank is an ordinal variable and differences in rank are not equally meaningful across its scale, using dummy variables for rank "buckets" (e.g., Top 20, Top 10, Top 5, Top 3) allows for a more interpretable assessment of SEO performance. This approach makes it easier to evaluate the effect of the treatment on pages in practically relevant positions in the search results. The summary statistics for the binary ranking dummies (top20, top10, top5, top3) reflect the relative frequency with which pages occupied these search positions during the observation period. The mean for each dummy can be interpreted as the proportion of days a page was ranked in the respective bucket, while the median indicates whether the majority of observations fell in or out of the bucket. Results show that pages were in the top 20 on 86.9% of days (mean

= 0.869, median = 1), in the top 10 on 62.7% of days (mean = 0.627, median = 1), in the top 5 on 30.1% of days (mean = 0.301, median = 0), and in the top 3 on 15.9% of days (mean = 0.159, median = 0) (see Appendix figure 7.2.1.5.).

Figures 19 and 20 examines the changes in average share of days each page appears within specific rank “buckets” (Top 20, Top 10, Top 5, Top 3), comparing pre- and post-intervention periods for both groups.

Within the top 20 bucket both groups started from similar high proportions with 93% of the pages of the control group and 91% of pages from the treatment group having an average ranking of 20 or higher. Both also declined by a similar margin from pre to post (control: -10 percentage points (p.p.), treatment: -11 p.p.), following a parallel trend.

In the top 10 bucket initial proportions were higher in the control group (80%) compared to treatment (61%). Similar to the top 20 bucket both groups experienced a parallel notable drop post-intervention (control: -15 p.p., treatment: -0.13 p.p.), maintaining a roughly constant gap and parallel changes.

While in the top 5 bucket the two groups also started with a notably difference (control: 46%, treatment: 25%), the proportion of pages with ranking on average in the top 5 dropped again substantially in the control group (-18 p.p.), while the treatment group remained somewhat stable against this trend (-1 p.p.). While being on different proportions in the pre period, both groups had much more similar proportions in the post period (control: 28%, treatment: 24%)

Finally, this trend of showing resistance against declining can also be observed in the top 3 bucket. Both groups had diverging proportions of days in the Top 3 pre-intervention (control: 28%, treatment: 11%). Again, the control group declined visibly (-12 p.p.), while the treatment group even saw a minor increase (+1 p.p.) and with this reducing the difference between groups.

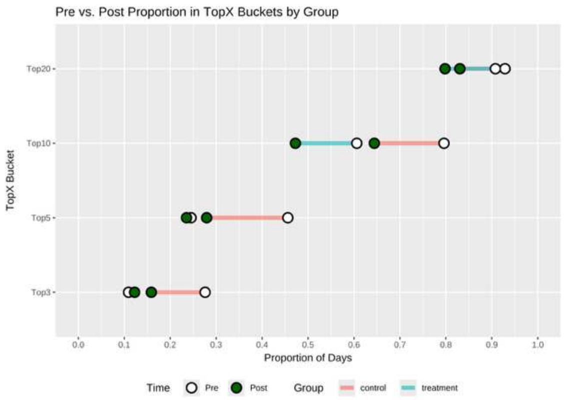


Figure 19. Proportion in TopX bucket by group & period I

Table: Proportion in TopX Buckets (per Group and Period)

dummy	group	pre	post	delta
Top20	control	0.929	0.830	-0.098
Top20	treatment	0.908	0.798	-0.110
Top20	Treat - Control	-0.021	-0.032	-0.011
Top10	control	0.796	0.644	-0.152
Top10	treatment	0.606	0.472	-0.134
Top10	Treat - Control	-0.190	-0.172	0.018
Top5	control	0.456	0.279	-0.177
Top5	treatment	0.245	0.236	-0.010
Top5	Treat - Control	-0.211	-0.044	0.167
Top3	control	0.276	0.159	-0.117
Top3	treatment	0.109	0.123	0.013
Top3	Treat - Control	-0.167	-0.036	0.130

Figure 20. Proportion in TopX bucket by group & period II

Top 20

	Declined	Improved	Stable
control	6	1	3
treatment	8	1	1

Top 10

	Declined	Improved	Stable
control	10	0	0
treatment	8	1	1

Top 5

	Declined	Improved	Stable
control	9	1	0
treatment	6	3	1

Top 3

	Declined	Improved	Stable
control	5	3	2
treatment	3	4	3

Figure 21. Proportion in TopX bucket by group & period III

This trend also becomes visible when looking at the changes on an individual page level (figure 21). A page-level comparison of changes in ranking buckets shows that, for the Top 10, Top 5, and Top 3 categories, the treatment group consistently had fewer pages declining, more pages remaining stable and more pages ascending into these higher-ranking buckets from pre to post period, compared to the control group, suggesting a more favourable development for treatment pages in achieving or maintaining higher search positions after the application of the treatment. In addition, the difference-in-differences regression with page and date fixed effects and robust, page-clustered standard errors was conducted for each ranking bucket. The common trend assumption found to be supported through examination of the pre-treatment development of both groups (see Appendix 7.2.1.6. – 7.2.1.9.). Results show no significant treatment effect for the Top 20, Top 10, or Top 3 buckets. However, for the Top 5 bucket, there is a positive and statistically significant effect (coef. = 0.16, SE = 0.09, $p < 0.1$), indicating a 16% higher likelihood for a treatment page to be ranked in the Top 5 in comparison to the control group after the deployment of the Copilot-generated text (figure 22).

	Dependent variable:			
	top20	top10	top5	top3
	Top20	Top10	Top5	Top3
	(1)	(2)	(3)	(4)
treat_post	-0.018	0.016	0.164*	0.120
	(0.056)	(0.048)	(0.094)	(0.081)
Date FE	Yes	Yes	Yes	Yes
Page FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Observations	2,029	2,029	2,029	2,029
R2	0.197	0.177	0.110	0.074
Adjusted R2	0.141	0.119	0.048	0.009
F Statistic (df = 114; 1895)	4.077***	3.569***	2.065***	1.331**
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.			

Figure 22. DiD regression tables of topX rank dummies

4.1.3 Impressions

The impressions variable, representing the daily number of Google search impressions per product page, has a mean of 7.54 (SD = 9.41) and a median of 5, indicating that on average a page receives 7.5 clicks a day but half of the observed daily impressions are 5 or fewer (see Appendix 7.2.1.10). Observed impressions range from 0 to 93 per page per day, with a right-skewed distribution concentrated at lower values, as shown in Appendix figure 7.2.1.12. When faceted by group and period (Appendix figure 7.2.1.13.), the distributional pattern remains right-skewed for both control and treatment groups across periods. Notably, the treatment group already shows higher mean impressions than the control group in the pre-intervention period (pre: 8.51 vs. 4.80; post: 11.77 vs. 5.09), highlighting pre-existing group differences in baseline visibility (see Appendix 7.2.1.11.).

Following this, also the weekly mean impressions (figure 23) were already from the beginning consistently higher in the treatment group over the whole study period. While both groups mostly moved in the same direction over time, the treatment group exhibited change of greater amplitude, especially after the intervention, while the control group remained comparatively stable throughout. The cumulative impressions plot shows that the treatment and control groups began to diverge early in the observation period, with the treatment group accumulating more impressions from the outset (figure 24). After the intervention, this divergence seemed to accelerate even more, with the gap between the two groups widening more rapidly than before.

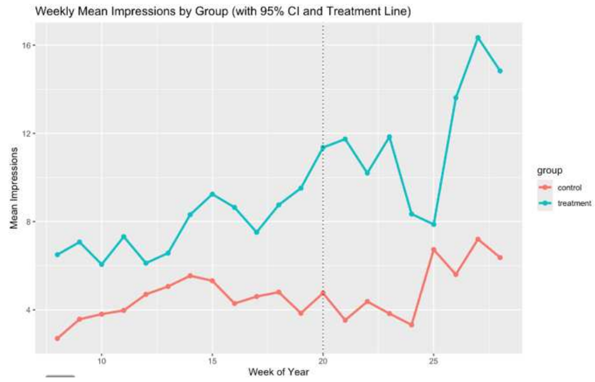


Figure 23. Mean weekly impressions by group

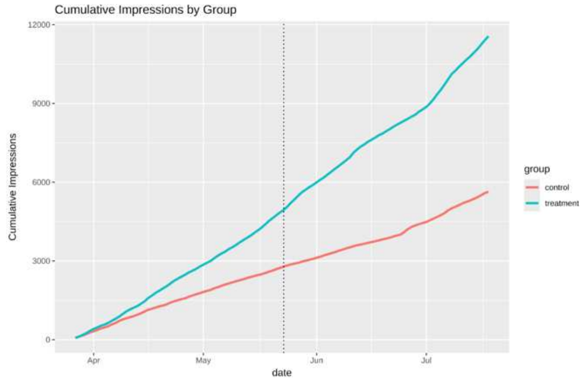


Figure 24. Accumulated impressions by group

The aggregated mean impressions at the page level are supporting this assumption. The difference-in-differences plot in figure 25 (corresponding values in Appendix 7.2.1.14.) show that mean daily impressions per page increased from pre- to post-intervention in both groups. The control group's mean impressions rose slightly from 4.80 to 5.09 (Δ = 0.29), while the treatment group increased more substantially from 8.51 to 11.77 (Δ = 3.26), resulting in a difference-in-differences of 2.98 impressions. Although 95% confidence intervals are relatively wide, especially for the treatment group, there is little overlap post-intervention, hinting towards a possible group-specific difference in impression gains. The dumbbell plot (figure 26) shows that this overall difference-in-differences effect was largely driven by a substantial increase in impressions on three individual treatment pages, each rising by approximately 10 impressions on average from pre- to post-intervention. For the remaining pages in the treatment group, changes from pre to post were relatively stable, with some pages gaining and others losing minor counts of impressions. In contrast, these minor variations are recorded for all pages in the control group with no major changes on page level between periods. The faceted histogram per page by group and period (figure 27) supports this by displaying how the distribution for the control group remained tightly clustered at lower impression counts both before and after the intervention, indicating little change within this group over time. In contrast, the treatment group's distribution shifted upward post-intervention, with several pages moving into much higher impression ranges. This resulted in increased variability and higher overall mean impressions in the treatment group compared to the control.

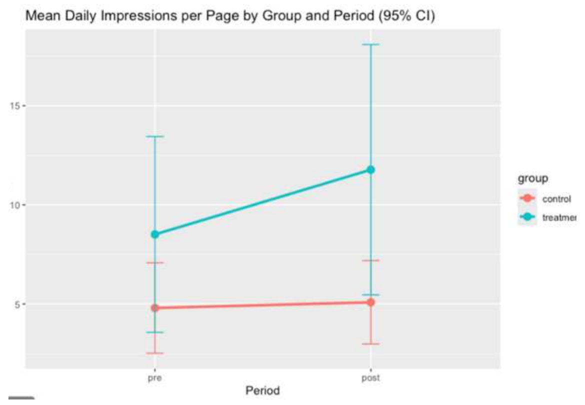


Figure 25. Mean impressions by group and period

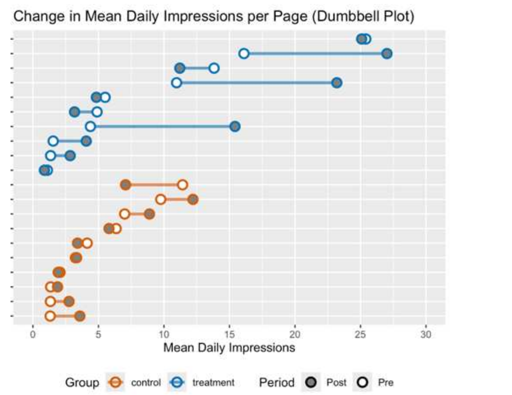


Figure 26. Mean page level impression

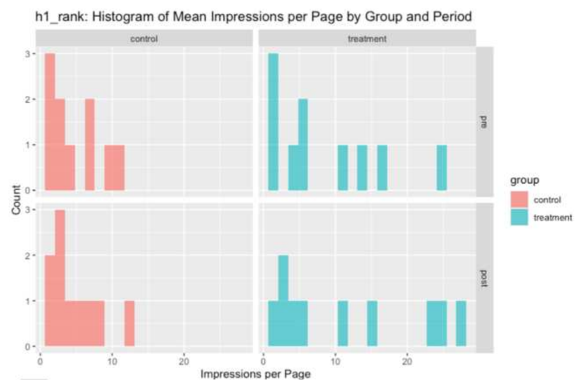


Figure 27. Faceted histogram of mean impressions

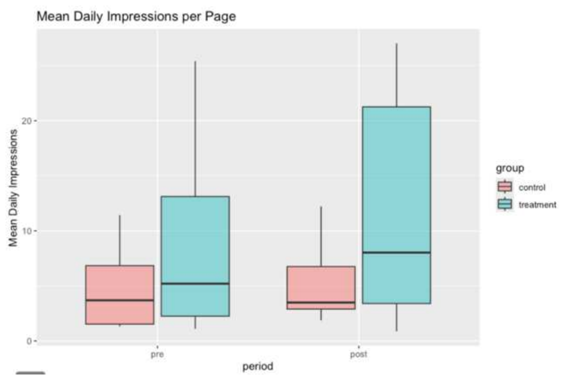


Figure 28. Faceted boxplot of mean impressions

The boxplots (Figure 28) provide additional detail by showing not only the shift in central tendency but also the change in variability across groups and periods. While the control group remained somewhat stable in both median and spread, the treatment group showed both a higher median and a substantial increase in IQR and overall spread after the intervention. Reflecting that the higher averages are stemming from the greater variability through the heterogeneous changes in the treatment group.

To assess whether the observed changes in impressions for the treatment group were statistically significant, a DiD regression was performed. Because the impressions variable was highly right-skewed (Appendix 7.2.1.12.), the analysis was conducted on the log-transformed variable to better approximate normality and stabilize variance. The common trend assumption was supported by the parallel patterns observed in the pre-intervention period (see Figure 23). The DiD regression (figure 29) estimated a treatment effect of 0.22 (SE = 0.18) on log-impressions, which was not statistically significant at conventional thresholds.

	Dependent variable:
	log_impressions
treat_post	0.224
	(0.179)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Observations	2,280
R2	0.155
Adjusted R2	0.102
F Statistic	3.442*** (df = 114; 2146)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.

Figure 29. DiD regression table of impressions

4.2 H2 - User Engagement (Human Level)

4.2.1 Clicks

The clicks variable, which measures the number of times a page was clicked when impressed for a user in the Google SERP, contained a substantial number of missing values (n_NA = 251 out of 2,280) (see Appendix 7.2.2.1.), reflecting the days where a certain page was not impressed to a certain user, and thus could not possibly been clicked. There is no indication for any missing values beyond this cause such as measurement errors or technical incidents. Among non-missing values, the distribution was highly zero-inflated and right-skewed, with a median of 0 and an IQR of 0, indicating that at least 75% of all observed values were also zero. Further investigation of how many days per page had no clicks compared by group/period confirmed the equal distribution of zero-inflation between treatment and control, with both groups showing a slight increase in the proportion of zeros from pre- to post-intervention (treatment: 91% to 93%; control: 88% to 91%) (Appendix 7.2.2.2.). This proportion of zero values per group is underscoring the extreme sparsity of clicks in the dataset, regardless of group or intervention period. The density plot further highlights this heavy concentration at zero (Appendix 7.2.2.3.). Values above zero were ranging from 1 to a maximum of 4 clicks resulting in an average of 0.11 clicks over all groups and periods per day. Analysis of the proportion of pages with at least one click showed that, in the pre-intervention period, 80% of pages in both groups recorded at least one click (Appendix 7.2.2.4.). After the intervention, this proportion increased to 90% in the control group but decreased to 70% in the treatment group, meaning that after the treatment 1 out of 10 pages in the control and 3 out of 10 in the treatment group did not receive any clicks.

Figure 30. Faceted histogram of mean clicks

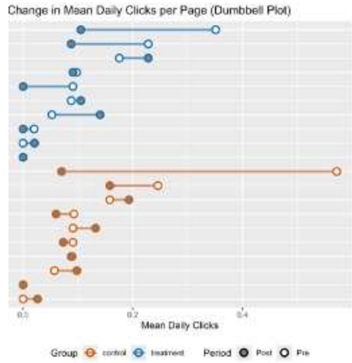


Figure 31 Mean page level clicks

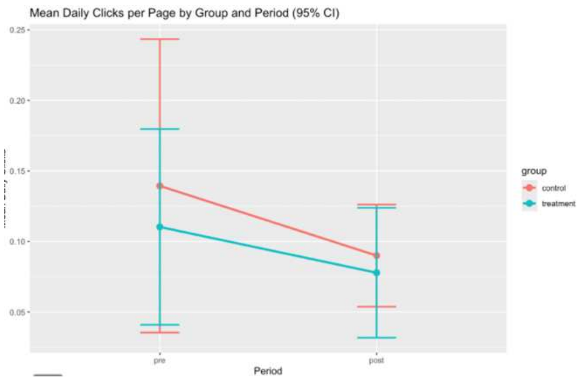


Figure 32. Mean clicks by group and period

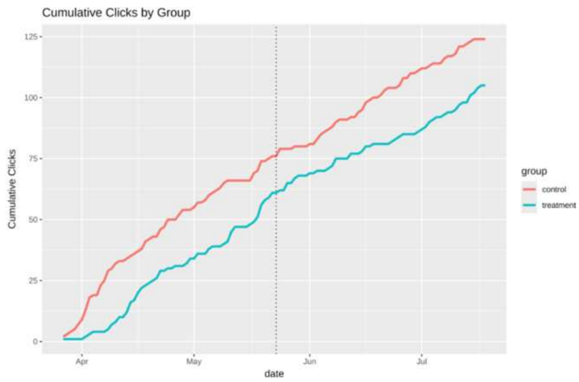


Figure 33. Accumulated clicks by group

Aggregating clicks at the page level shifted the distribution away from absolute zeros, with most mean daily clicks per page now falling between 0 and 0.2 for both groups and periods, as shown by the histogram (Figure 30). Also both groups had some higher click means (0.2 - 0.5) in the pre-period, means in the post period are more aggregated towards the lower end of the distribution. The histogram gives first tentative explanations for this common shift as both groups lose the highest values within the range of the pre-period distribution and shift towards lower average clicks. This is also visible in the Dumbbell Plot where most pages showed very little movement except the pages with the higher click rates in the pre period declining strongly towards lower averages in the post period (figure 31). While in the control group one sharp drop can be observed of by far best clicked page in the pre period, several mid-range declines can be observed in the treatment group, also effecting the previously best clicked pages. Besides that, both groups show a couple of small changes in mixed directions. In the plot of mean daily clicks per page by group and period with 95% confidence intervals this averages into a communal downward trend in both groups after the intervention, with the control group declining from 0.16 to 0.10 and the treatment group from 0.12 to 0.08 (figure 32; Appendix 7.2.2.5.). The

difference-in-differences between groups was minimal (0.02), and the confidence intervals were very wide, indicating a high degree of uncertainty and no statistically meaningful difference between groups. This parallel tendency can also be observed looking at the cumulative clicks with no major divergence trend appearing after the treatment (figure 33).
Finally, the difference-in-differences regression was performed, under acknowledgment of the limitations given by the distribution of the variable. Despite a logarithmic transformation, the distribution remained heavily zero-inflated and right-skewed (Appendix 7.2.2.6.). Because the visual inspection of pre-treatment trends for log-transformed daily clicks suggested that the parallel trends assumption may not be fully met (Appendix 7.2.2.7.), page-level pre-period trend was included as a covariate to account for differences in parallel trends. The coefficient of the interaction term was indicated with 0.012 (SE = 0.039), but with no statistical significance (Figure 34). The model explained little variance ($R^2 = 0.069$, adjusted $R^2 = 0.043$), reflecting the substantial heterogeneity and sparsity observed in the underlying data.

	Dependent variable:
	log_clicks_engagement
treat_post	0.012
	(0.039)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Pre-Trend Slope	Yes
Observations	2,029
R2	0.069
Adjusted R2	0.003
F Statistic	1.227* (df = 114; 1895)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Pre-period trend (slope) included as covariate. Cluster-robust SEs by page.

Figure 34. DiD regression table of clicks

4.2.2 CTR

The CTR defined as clicks divided by impressions quantifies the probability for a page of being clicked under the condition of being displayed to a user by the search algorithm. Following this logic, like clicks, it is set to NA whenever there were no impressions for a page (resulting in the same number of NAs) and is zero whenever the page was impressed but not clicked (thus, matching the zero-inflation observed in clicks). In parallel, there is no indication for any missing values beyond this cause such as measurement errors or technical incidents, and the zero inflation is evenly distributed between control and treatment group. The variable displayed a mean of 0.017 (SD = 0.077), implying an average click through rate of 1.7% (see Appendix

7.2.2.8.). The median of 0, and an IQR of 0 show a similarly strong zero-inflation to clicks, with nearly all observed CTR values being zero.

The development of the average daily CTR per page showed that the control group remained on average stable from pre- to post-intervention (0.0216 to 0.0214), while the treatment group declined from 0.0166 to 0.0079 (Appendix 7.2.2.9.). The resulting difference-in-differences was -0.0085 (0.85 p.p.), indicating a small decrease in CTR for the treatment group in comparison to the control. The plot with 95% confidence intervals illustrates this trend, showing a decline for the treatment group and stable values for the control group, with overlapping and wide CIs suggesting that the differences were too small to be statistically meaningful (Figure 35). The faceted histogram of mean CTR per page by group and period shows that, while the control group's distribution remained fairly stable between pre- and post-periods, the treatment group experienced a visible downward shift post-intervention, especially concerning the pages with the highest CTR (Figure 36), which is consistent with the observations in the click variable. Congruently, the boxplot shows that while the control group's median and IQR for mean CTR remained largely unchanged, the treatment group's median and spread both decreased after the intervention, with 50% of the values centred around the median collectively moving towards smaller CTRs (Figure 37). For further investigation of this distribution, the dumbbell plot shows that most pages in the control group actually exhibited substantial but mixed changes in CTR, with an even share of pages increasing and decreasing in similar magnitude, leading to no visible changes in the average mean (Figure 38). In contrast, the treatment group's overall decline in mean CTR was primarily driven by three pages that started with the highest CTR values of the group and saw sharp drops after the intervention, while the remaining treatment pages showed little change.

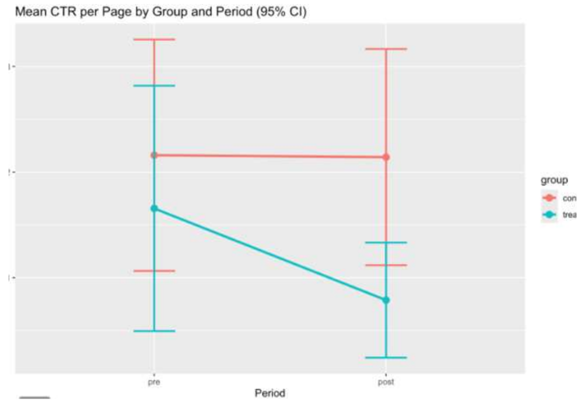


Figure 35. Mean CTR by group and period

Figure 36. Faceted histogram of mean CTR

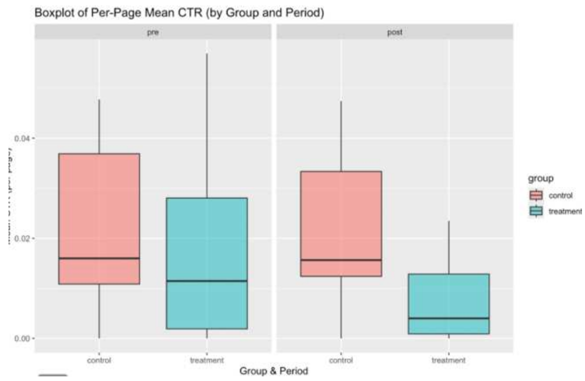


Figure 37. Faceted boxplot of mean CTR

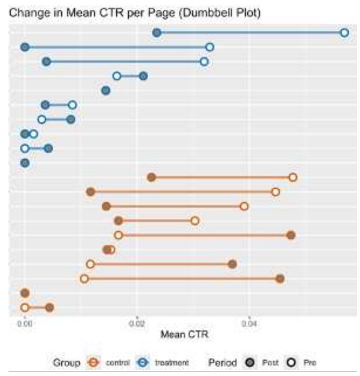


Figure 38 Mean page level CTR

The difference-in-differences regression was performed on the untransformed CTR variable to preserve interpretability, as logging did not improve the distribution due to the high concentration of values near zero (see Appendix 7.2.2.10.). Parallel trends were visually confirmed before the intervention (see Appendix 7.2.2.11.). The results indicated a treatment effect estimate of -0.009 (SE = 0.010), which was not statistically significant (Figure 39). The model explained little variance ($R^2 = 0.075$, adjusted $R^2 = 0.010$), reflecting the limited explanatory power and lack of a measurable effect of the intervention on click-through rates.

	Dependent variable:
	ctr
treat_post	-0.009
	(0.010)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Observations	2,029
R2	0.075
Adjusted R2	0.010
F Statistic	1.352*** (df = 114; 1895)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.

Figure 39. DiD regression table of CTR

4.2.3 Engagement time per active user

The last variable for the assessment of hypothesis two is the engagement time per active user (in seconds). Throughout the 16-week observation period, it showed an overall mean of 40.0 seconds (SD = 25.5) per user, indicating that on average a user would spend 40 seconds on a page (see Appendix 7.2.2.12.). The median of 36.4 seconds indicates that half of user sessions lasted less than 36 seconds and half longer. The minimum observed engagement time was 0.5 seconds, while the maximum reached 360 seconds, with an interquartile range of 23.6 seconds, reflecting moderate spread within the central 50% of values. There were 2,280 valid observations, with 105 missing values, representing the days where no visitor was on the respective page, who could have engaged with it. Again, there is no indication of any missing values beyond this cause. The density plot (Appendix 7.2.2.13.) highlights a right-skewed distribution, with most engagement times clustering around 35 - 40 seconds and a long tail of higher values, suggesting a minority of sessions with substantially higher engagement. Unlike for the two other variables for Hypothesis 2, a rather lower proportion of zeros in this variable is reflected through the inclusion of all active users on a page, regardless of their traffic source, as already outlined in chapter 2.5.2.

The comparison of mean daily engagement time per active user, aggregated at the page level, showed that on average, the treatment group started slightly higher than the control group (40.25 vs. 38.45 seconds pre-intervention) (Appendix 7.2.2.14.). In the post-intervention period, mean engagement time in the treatment group increased to 42.43 seconds, while the control group remained stable at 38.33 seconds. This resulted in a pre-post change of +2.18 seconds for the treatment group and -0.12 seconds for the control group, yielding a difference-in-differences of

2.29 seconds. However, the 95% confidence intervals displayed in the corresponding plot indicate substantial overlap across all means and periods, suggesting that these differences were not statistically significant (Figure 40). The dumbbell plot of mean engagement time at the page level showed that changes from pre- to post-intervention showed mixed patterns in both groups (Figure 41). In the control group, most pages experienced small changes, relatively balanced in both directions, resulting in a stable group average over time. In contrast, while the majority of pages in the treatment group showed a clear tendency for larger increases in engagement time from pre to post, except for one notable outlier, which showed a substantial decline. This individual's decrease dampened the overall average increase for the treatment group, but aside from this outlier, the general tendency within the treatment group was towards higher engagement times following the intervention.

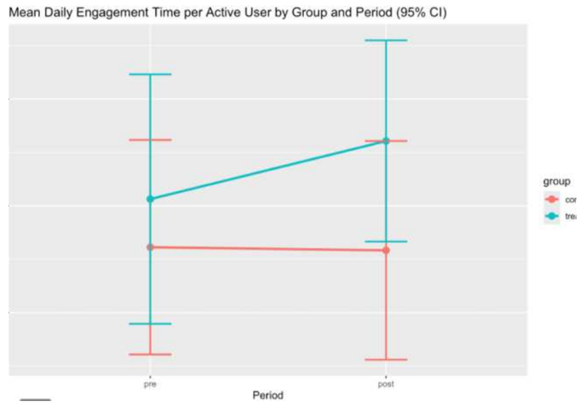


Figure 40. Mean engagement time by group and period

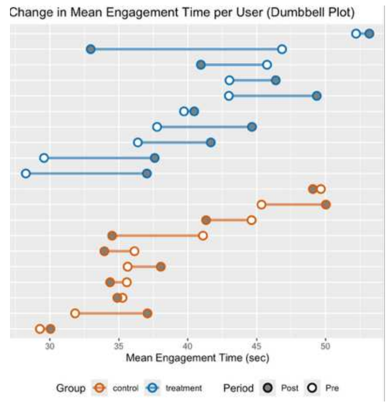


Figure 41 Mean page level user engagement

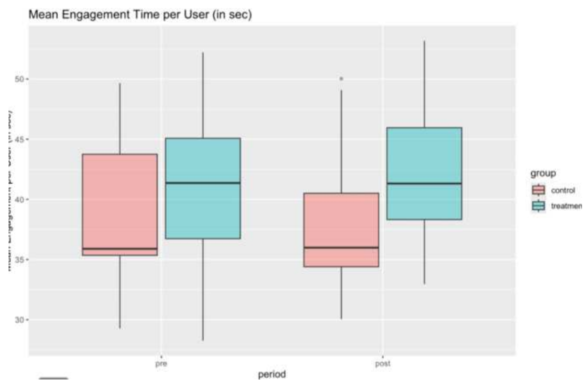


Figure 42. Faceted boxplot of mean user engagement

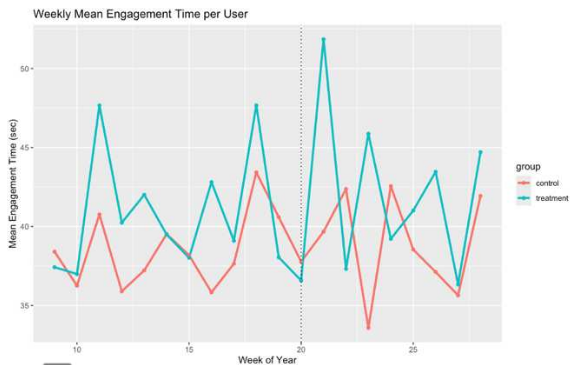


Figure 43. Mean weekly user engagement by group

The box plots provide an additional perspective on the distributional changes observed in the control and treatment groups (Figure 42). For the control group, the box plot shows a marked reduction in interquartile range from pre- to post-periods, indicating that values became more compressed around the median, while the median itself remained largely unchanged. This finding is congruent with the dumbbell plot, which illustrates that pages in the control group moved from higher and lower engagement times toward the centre, resulting in less dispersion but stable central tendency. On the other side, the box plots for the treatment group displayed a stable median and interquartile range in engagement time from pre- to post-intervention, suggesting that the centred 50% of user engagement values remained largely unchanged across most pages. At first glance, this observation appears to contradict the dumbbell and the DiD line plot, which showed that the majority of treatment pages experienced increases in average engagement time from pre to post. This apparent contradiction can be further examined by considering the trend in weekly mean engagement time per group (see Figure 43). The line plot illustrates that both groups followed an approximately parallel trend before the intervention, with the treatment group consistently displaying slightly higher averages than the control group. Notably, immediately after the introduction of the treatment, there was a pronounced but short-lived spike in the average engagement time for the treatment group, while the control group remained stable. This temporary jump likely affected the overall post-period average of the treatment pages, but not the median and interquartile range shown in the box plot, as they are more robust to isolated outliers.

To conclude the analysis, the DiD regression analysis was performed with the logarithmic transformation of the variable in order to reduce its pronounced right skewness (Appendix 7.2.2.15.). Graphical inspection of the weekly mean engagement time by group indicated broadly parallel pre-intervention trends between treatment and control groups, thereby supporting the plausibility of the common trend assumption (Figure 43). The interaction term showed no statistically significant treatment effect (coef. = 0.009, SE = 0.060), indicating no measurable impact of the intervention on user engagement time (see figure 44).

	Dependent variable:
	log_eng_time_pau
treat_post	0.009
	(0.060)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Observations	2,175
R2	0.070
Adjusted R2	0.009
F Statistic	1.345** (df = 114; 2041)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.

Figure 44. DiD regression table of engagement time

4.3 H3 - COMMERCIAL IMPACT (BUSINESS LEVEL)

To evaluate the commercial impact of the treatment, the number of sales generated through organic page traffic was tracked. Descriptive statistics show that overall daily sales per product page were highly right-skewed, as depicted in the density plot (Appendix 7.2.3.1.). The average number of sales per day was 0.21 (SD = 0.53), while both the median and interquartile range were zero, reflecting that the majority of page-days had no sales at all (Appendix 7.2.3.2.). Daily sales counts ranged from 0 to a maximum of 6 per day across all observations, with only 16.6% of page-days recording at least one sale and only a small number of instances with more than one sale in a single day (Appendix 7.2.3.3.). The average purchase price per product was €25.86 (Appendix 7.2.3.4.).

Looking from a page-level perspective, all product pages generated at least one sale during the pre-period. This proportion remained unchanged in the treatment group during the post period, while it declined slightly to 9/10 pages in the control group, where one page did not generate any sales after the intervention (Appendix 7.2.3.5.). The descriptive difference-in-differences analysis of total sales per product page, aggregated by period and group, is displayed in Figure 45 (Appendix 7.2.3.6). According to this, in the control group, one page generated on average 13.2 sales during the pre-period and 13.3 sales during the post period (Δ = 0.1), indicating almost no change over time. For the treatment group, mean total sales per page increased from 9.0 in the pre-period to 11.4 in the post-period (Δ = 2.4), yielding a descriptive DiD of 2.3 additional total sales per page in the treatment compared to the control group. Translating these figures into revenue, and assuming the average product price of 25.86€, the DiD effect corresponds to an average increase of approximately 59.48€ in sales per page and 594.8€ in total in the treatment group, relative to the control. However, as illustrated in Figure 45, the 95% confidence intervals around group means are wide and show substantial overlap across groups and periods, suggesting no statistically different means. The histograms showed that the overall distribution of total sales numbers per page remained broadly similar from pre to post in both groups (Figure 46). In the control group, the two most extreme values shifted towards more central values in the post period, while in the treatment group, a new, more extreme sales number emerged. Aside from these changes at the distribution tails, the general spread and pattern of sales totals across pages remained largely unchanged.

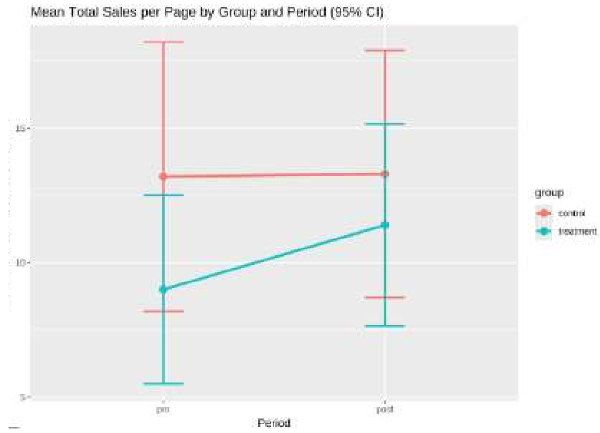


Figure 45. Mean accumulated sales by group and period

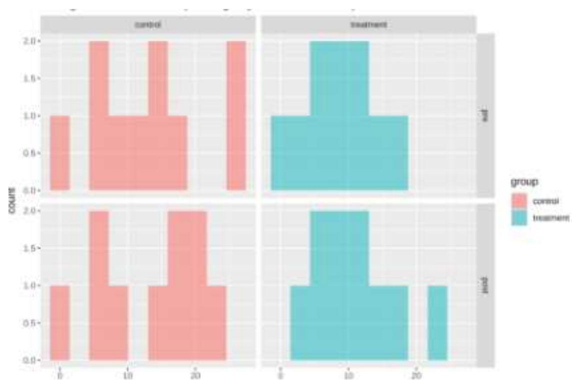


Figure 46. Faceted histogram of mean accumulated sales

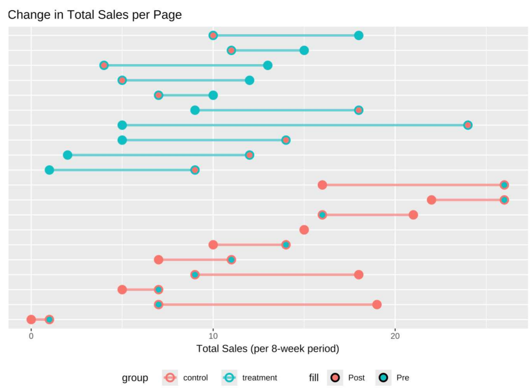


Figure 47 Mean accumulated page level sales

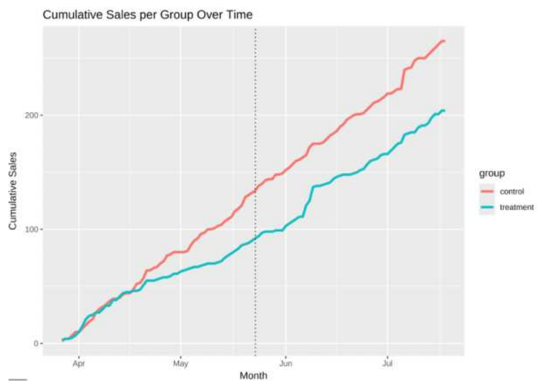


Figure 48. Accumulated sales by group

While the overall distribution of total sales per page remained broadly similar across periods, the dumbbell plot provides additional insight into within-group dynamics, especially within the treatment group (Figure 47). In this group, the pages that had achieved the highest total sales during the pre-period experienced large declines, such that they became the pages with the lowest total sales in the post-period. Conversely, those pages that had accumulated the lowest sales before the intervention showed the large, substantial gains and became the highest-selling pages after the intervention. Thus, there was a clear reversal in rank order among treatment pages. Overall, this turnover then contributed to the modest but overall increase in total sales among treatment pages after the intervention. By contrast, the control group showed a more mixed pattern, with both increases and decreases among individual pages but no consistent pattern and overall little change in the group mean. This subtle increase in total sales for the treatment group is also reflected in the cumulative sales plot (Figure 48). Before the intervention, the treatment group consistently accumulated sales at a slower rate than the control group, resulting in a widening gap between the two lines. After the intervention, however, the rate of accumulation in the treatment group appears to increase, as indicated by the lines for both groups running more parallel. This suggests that the difference in cumulative sales between the groups did not widen further after the intervention, indicating a relative improvement in sales dynamics for the treatment group.

For completeness, a difference-in-difference analysis was also conducted, however, as previously noted, the sales variable is highly zero-inflated, which poses challenges for standard linear modelling. Attempts to address this skewness using a logarithmic transformation (log(sales + 1)) did not meaningfully alter the distribution, as the zero-inflation persisted and the transformed variable remained heavily concentrated at zero (see Appendix 7.2.3.1.). Graphical analysis supports the validity of the common trend assumption with no evidence of systematic divergence prior to treatment (Appendix 7.2.3.7.). The DiD results, presented in Figure 49, indicate a small and statistically insignificant treatment effect (coefficient = 0.040, SE = 0.064). Model fit remains low ($R^2 = 0.098$), reflecting the challenge of explaining variance in a dataset dominated by zeros.

	Dependent variable:
	sales
treat_post	0.040
	(0.064)
Date FE	Yes
Page FE	Yes
Date FE	Yes
Observations	2,280
R2	0.098
Adjusted R2	0.042
F Statistic	2.051*** (df = 114; 2146)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Cluster-robust SEs by page. Page and date fixed effects included but not shown.

Figure 49. DiD regression table of sales

5 DISCUSSION

5.1 H1 - ALGORITHMIC RELEVANCE (SEARCH ENGINE LEVEL)

Product pages with Copilot-generated SEO content are perceived as more relevant by Google's algorithm, leading to improved search visibility.

The results of the data analysis of the rank variable first provided mixed signals when it came to possible treatment effects. On the one side, there was no significant treatment effect measured by the DiD regression. Also, the descriptive difference-in-differences between group and period means, along with the shifts observable in the histograms and average page ranks in the Dumbbell plot painted a coherent picture of the same common downward shift in rankings for both the treatment and control groups during the observation period. As such, as it seemingly affected both groups equally, this parallel decline may be attributed to different natural exogenous factors, such as increased competitor activity or seasonal fluctuations in demand, rather than treatment-related effects. However, a nuanced insight emerges when observing the weekly line plot over time. Immediately following the implementation of the AI-generated content, the treatment group's rankings exhibited a sharp and distinct improvement compared to the control group, decreasing the initial and somewhat stable difference between control and treatment, and with that bringing both groups' mean ranks notably closer together. This initial improvement persisted for the majority of the subsequent weeks, maybe implying a short-term positive impact from the intervention. However, in the very last week of observation, the rankings of both groups diverge strongly again. This brief yet strong worsening may partially explain why the final averaged difference in difference, and thus treatment effect measured throughout the whole 8 weeks, was negligible, as the initial positive deviation and later negative divergence effectively neutralized each other. Still, this nuance could be interpreted as a first indication that the treatment had at least a short-term positive effect on the rankings of the treatment group. Due to the limited observation window of eight weeks, it cannot be ascertained whether this initial advantage would be sustainable or ruled out over a more extended period, as indicated by the divergence in the last week. Future studies extending beyond this timeframe would be necessary to confirm whether this short-lived boost translates into long-term ranking improvements.

Furthermore, the granular analysis of the ranking dummies provides additional important insights. In the broadest buckets (Top 20 and Top 10), both treatment and control pages declined in parallel, with no hint of a positive treatment effect. However, when focusing on the most competitive positions (Top 5 and Top 3), the treatment group appears somewhat more resilient to decline. Specifically, while the control group experienced notable losses in these high-ranking buckets, the treatment group remained stable or even improved slightly in comparison to the control group. This divergence, although modest and based on small numbers, suggests that AI-generated SEO text may help protect or even enhance a page's presence in the most commercially relevant search positions. Most notably, the difference-in-differences regression for the Top 5 bucket yielded a positive and marginally significant effect, indicating a 16% higher likelihood for a treated page to achieve a Top 5 ranking compared to the control. Although this result is only weakly significant (90% significance level) and should be interpreted with caution, especially given the limited sample size and high sample variation, it does offer a further tentative signal that AI-generated content could have a beneficial effect at the upper end of the ranking spectrum. From a SEO perspective, these results are particularly insightful because high-ranking positions (Top 3 and Top 5) are most strongly correlated with increased visibility and clicks (Erlhofer, 2020; FirstPageSage, 2023; Lammenett, 2025; Sistrix, 2020; Urman & Mykola Makhortykh, 2023; Usability.de, 2009). Thus, even though broader ranking averages did not significantly improve, the observed resistance against decline in top-ranking positions represents an important hint towards the beneficial potential of AI-generated texts for maintaining competitive visibility.

The data on impression counts introduces another perspective on the question of an existing treatment effect on search engine visibility, reflecting actual exposure of product pages to users in the organic search results due to algorithmic choice. The longitudinal analysis shows that the treatment pages already entered the observation window with a notably higher impression average and preserved this throughout. Weekly means for impressions moved somewhat in parallel across groups, but the magnitude of the swings was consistently larger for the treatment pages. The cumulative plot slightly points towards the tendency of the treatment accelerating this divergence, with an even faster accumulation of impressions in the treatment group compared to the control group. Also the descriptive difference-in-differences analysis indicates the same pattern. The resulting DiD of 2.98 impressions per page per day suggests that, on average, a page in the treatment group gained approximately three more daily impressions than pages in the control group after the intervention. Despite wide confidence intervals, particularly notable within the treatment group, the minimal overlap post-intervention hints at a potential group-specific improvement in visibility attributable to the treatment, but also towards higher volatility in the post treatment period. This is also confirmed by the box plot and histogram, indicating an isolated heterogeneous increase of pages towards higher impression counts. A page-level view on the individual average changes with the help of the dumbbell plot confirms this by showing that the rise in average impressions per page is not general across the treatment group but rather driven by three pages that exhibit substantial jumps from pre- to post-intervention. For the remaining treated pages, changes were, as in the control group, small and inconsistent in direction, thus somewhat stable.

This asymmetry presents an interpretative dilemma. On one hand, the absence of any substantial increases among control page averages could indicate that the large increases seen in those three treated pages might indeed relate to the intervention. This nearly 3-impression difference per page per day in favour of the treatment group could be practically meaningful in high-volume e-commerce contexts, especially in connection with stabilized high ranking positions, which would then possibly convert to high click-through rates (Erlhofer, 2020; FirstPageSage, 2023; Lammenett, 2025; Sistrix, 2020), if attributable to the intervention. However, on the other hand, the lack of a broader response across the remaining treatment pages tempers this interpretation. If the AI-generated content had a uniform and systematic positive impact, one may expect more than three pages to benefit substantially, perhaps indicating that the treatment was effective, but only under certain conditions or for certain types of pages. Also the formal difference-in-differences regression analysis, conducted on the log-transformed impressions variable, did not find overall statistically significant treatment effects. Although statistical significance is not expected to be the primary benchmark in this exploratory analysis, given the discussed limitations, the absence of significance further suggests that the observed shifts are primarily driven by a few pages that do not convincingly represent a consistent overall treatment effect.

Taken together, the impression data offered tentative and localized signs of treatment efficacy. While the overall treatment effect was not generally visible, the strong response observed in a small subset of pages, combined with the absence of such responses in the control group, may suggest a possible signal worth investigating further. Thus, it could be speculated that the Copilot-generated content may have improved search visibility, but under specific content, product, or keyword conditions that were met only by a portion of the treatment group.

Taken as a whole, both ranking-level metrics and impression data offer tentative but directionally consistent indications that the Copilot-generated product texts may have positively influenced the perceived relevance of a page from a search engine perspective. While broad-based statistically significant treatment effects were mostly absent (partially explained by the limited and variable sample), several small findings consistently point in a direction in favour of a treatment effect, by stabilizing competitive search positions and supporting exposure. Importantly, at no point did the introduction of high-quality AI-generated content appear to be penalized by the algorithm. Although the positive effects were often localized or modest, the absence of any negative impact alongside several promising shifts is also a finding in itself, furthermore supporting the practical argument for a possible adoption of AI-generated content in SEO practice. These results, while exploratory, justify further research on a larger scale to more definitively assess the long-term and widespread efficacy of such interventions in real-world e-commerce settings.

5.2 H2 - User Engagement (Human Level)

Product pages with Copilot-generated SEO content are perceived as more relevant and engaging by users, resulting in increased interaction.

The analysis of user clicks as a proxy for perceived relevance and user engagement presents a highly inconclusive picture, primarily due to the extreme sparsity of the data. With over 90% of daily observations registering zero clicks and most non-zero values falling between 1 and 4, the dataset provides only a limited empirical basis to detect meaningful effects. Importantly, this structural limitation affects both control and treatment groups equally, thereby precluding any systematic bias.

From a descriptive standpoint, there is no clear evidence of a positive treatment effect. In fact, the proportion of treatment pages that recorded at least one click decreased after the intervention, while the control group saw a modest increase. This mild asymmetry suggests that the treatment, if anything, did not safeguard against post-intervention declines and may have coincided with slightly worsened click behaviour. Cumulative and mean click trajectories, as well as confidence interval plots, underline no visible divergence between groups after the intervention. Both groups followed a similar downward trend in clicks, and the very small descriptive difference-in-differences estimate, paired with wide confidence intervals, indicates that any observed differences are likely due to noise rather than treatment effects. This common downward trend is further reinforced by the dumbbell plot, which shows that the pages with the highest average clicks in the pre-period experienced the strongest declines in the post-period, simultaneously across both groups, which may be suggesting a general loss of user relevance likely driven by external factors, with no indication that the treatment mitigated this downward trend. The results of the regression analysis support this interpretation. Despite controlling for pre-period page-level trends, the model fails to find a significant treatment effect, and the

overall explained variance remains extremely low. Given the highly skewed and zero-inflated nature of the data, this lack of significance is not surprising but nonetheless confirms the fragility of any potential conclusion regarding treatment efficacy.
In conclusion, the analysis of click behaviour offers no compelling evidence that the AI-generated content led to improved user engagement at the point of search result interaction. Given that at this point of the user journey, users base their click decisions primarily on the search snippet displayed to them in the organic search results, this may suggest that the treatment did not meaningfully enhance the appeal or perceived relevance of these previews compared to the existing, non-optimized snippets. However, due to the extremely sparse and zero-inflated nature of the data, any such interpretation must be approached with considerable caution and remains highly tentative.

The analysis of CTR as a ratio of user clicks to impressions offers a refined lens on user engagement by isolating the effectiveness of a page's search snippet at prompting clicks when it is actually seen. However, like the click data itself, the CTR variable suffers from extreme zero-inflation and limited variance, severely constraining the ability to detect treatment effects with confidence.
Following the extracted data, the treatment group experienced a small but visible decline in average CTR post-intervention, whereas the average CTR of the control group remained nearly unchanged. While the absolute difference-in-differences is minimal, the directionality of this shift does point toward a negative treatment effect. Visual analyses reinforce this impression as the treatment group shows a downward shift in the distribution of mean CTRs across pages, especially for initially higher CTRs per page, while the control seems to stay somewhat stable. This further suggests that the treatment did not enhance but may even have diminished the effectiveness of search snippets in attracting user clicks. The dumbbell plot provides further nuances on this shift. In the treatment group, the observed decline in CTR was not uniform but largely driven by a small subset of pages that initially had the highest click-through rates. After the intervention, these high-performing pages experienced pronounced drops, while the remaining pages stayed relatively stable at already low levels. The control group, by contrast, showed a more balanced pattern of moderate increases and decreases, averaging out to no meaningful change overall. Regression analysis confirms the descriptive patterns. The treatment coefficient, while slightly negative, is statistically insignificant, and the model itself explains very little of the variance in CTR.

In summary, at first glance, the observed data suggest a disproportionate decline in CTR for the treatment group, even indicating a negative effect of the intervention on user click behaviour relative to visibility. When attempting to relate these findings to those of the CTRs component variables (clicks and impressions), an apparent contradiction emerges. Solely based on the previously observed patterns in clicks and impressions, one would reasonably expect the CTR to remain stable or decline in the control group (with impressions largely constant and clicks declining), and a potential rise or stability in CTR in the treatment group (where impressions increased while clicks also declined). However, the just presented results show the opposite trend: CTR remained stable in the control group but seemingly decreased in the treatment group.

This apparently counterintuitive outcome can likely be explained by the mathematical sensitivity of CTR under conditions of zero-inflated click data. By definition, CTR is set to zero whenever clicks equal zero, regardless of how many impressions were recorded ($\frac{0\ clicks}{n(impressions)} = 0$). Thus, in a dataset like the present one, characterized by very frequent zero-click days, any increase in impressions will not influence CTR unless accompanied by at least one click. As such, even if a page receives more impressions over time, a simultaneous drop from one click to zero will cause the CTR to drop to zero, effectively masking any visibility gains in the denominator. This tendency is supposed to be especially influential in the underlying data, as there are mostly days with 0 or 1 clicks per day and an observed downward trend from pre to post period (more days with 0 instead of 1 click). Therefore, besides being already fairly influential due to the general high zero inflation of clicks, the effect of this bias is likely to be even stronger in the treatment period. Moreover, this effect may appear even more pronounced in the treatment group, as this group showed consistently lower click counts and a higher proportion of days with zero clicks compared to the control group. In larger datasets or in settings where web pages receive higher and more consistent levels of traffic, this distortion would likely be less severe. CTR would then reflect a more continuous and meaningful ratio, including both the development of impressions and clicks. However, under the constraints of this study, the high zero-inflation of the click variable and particularly in the treatment group, CTR is highly vulnerable, potentially leading to severe interpretative bias when comparing group averages.

Summing up, in this scenario, the CTR becomes highly sensitive to minimal changes, being disproportionately biased by infrequent binary changes (whether or not a single click occurred), rather than offering a stable indication of performance. Given these limitations, the CTR findings should be drawn into conclusion with extreme caution, if at all, with respect to making assumptions regarding a given treatment effect for Hypothesis 2.

On the contrary, the analysis of engagement time per active user offers a comparatively more stable perspective on user behaviour than the preceding metrics, due to its lower susceptibility to zero-inflation and its broader inclusion of all traffic sources. As such, it allows for a more meaningful interpretation of how users interacted with product pages once they arrived, providing insight into the perceived usefulness or interest level of the content beyond the search interface. The descriptive analysis of engagement time per active user indicated a modest overall increase for the treatment group relative to the control group. While the control groups' overall mean user engagement time on page remained largely stable, the treatment group exhibited a slight upward shift in mean engagement time post-intervention, even though small in size (2 seconds on average). Notably, also on a page level, this group-level improvement in the treatment condition was visible for most pages, increasing from pre to post period, only dampened by a single outlier page that showed a substantial decline. This overall trend of increased mean user engagement time was not mirrored in the control group, suggesting a potential treatment-related effect across the majority of treated pages.
However, when placing these results in the context of the weekly trend line and box plot, a more nuanced interpretation emerges. The time series shows a sharp but brief spike in engagement for the treatment group immediately after the intervention, which then quickly normalizes and follows a similar trend as the control group. This transient peak raises questions about the plausibility of the assumed treatment effects. Given that most users are likely to be new to the page and not returning visitors, a rather long-lasting change in how content is perceived by the user would be expected to manifest more consistently over time, not just briefly following the content update. The short-lived nature of this increase challenges the assumption that the AI-generated content alone led to sustained improvements in user interaction.
In summary, despite the generally positive trend across most treatment pages, the short-lived spike in engagement time raises questions about the plausibility of a sustained treatment effect. Given the abrupt nature of this increase and its immediate normalization, it is equally possible that the observed pattern reflects random variation rather than a meaningful behavioural change. This interpretation is further supported by the inherent variability in engagement data and the limited sample size. Future research should investigate this dynamic more closely, ideally incorporating additional, more robust user engagement metrics (e.g., scroll depth, interaction

events, or bounce rates), to better understand whether AI-generated content can systematically enhance on-page user behaviour. At the same time, it is notable that no hints for a decline in user engagement were observed following the intervention. Seemingly, users on AI-treated pages did not disengage more quickly than those on control pages, suggesting that the machine-generated content was not perceived negatively. Even if the treatment did not substantially increase engagement, the absence of detrimental effects may support the view that the Copilot-generated texts perform at least on par with traditional content in terms of user experience. This functional equivalence in itself could already be interpreted as a positive takeaway for practitioners, indicating that AI-assisted SEO optimization can be deployed without risking user alienation.

In conclusion, the analysis assessment of hypothesis 2, using clicks, CTR and engagement time per active user yielded no robust evidence supporting a positive treatment effect of AI-generated content on perceived user relevance or engagement. Both clicks and CTR data were heavily constrained by extreme zero-inflation and low variance, severely limiting interpretability. While clicks did not show any indications for a treatment effect, CTR showed a disproportionate post-treatment decline in the treatment group, but likely attributable to mathematical distortion arising from sparse click data, rather than a true behavioural shift. In contrast, in engagement time per active user, most treatment pages showed increased engagement time post-intervention, but this trend was overshadowed by a short-lived spike immediately after the treatment, raising questions about its sustainability and suggesting potential random variation. Importantly, no indications of disproportional decreased click rates or engagement time were observed, pointing towards the assumption that AI-generated content did not adversely affect user interaction. While these findings do not explicitly support Hypothesis 2, they do suggest that AI-generated texts may perform comparably to traditional content in terms of perceived user experience, which is an encouraging indication for practical application. Nevertheless, further research with richer and more robust engagement metrics is needed to reliably assess the long-term impact of AI-generated content on user behaviour and to distinguish genuine treatment effects from random variation.

5.3 H3 - Commercial Impact (Business Level)

Product pages with Copilot-generated SEO content lead to improved commercial outcomes by positively influencing purchase behaviour.

The analysis of sales data as an indicator of commercial impact presents a constrained foundation for inference, as only around 17% of the observations count at least one sale per day. The difference-in-differences regression reinforces the fragility of these descriptive insights. Despite testing the treatment effect with a log-transformed specification to address skewness, the model yielded a statistically insignificant estimate and explained little variance in the data. This confirms that, under the current conditions and given the prevalence of zero-sales days, reliable conclusions about treatment impact are difficult to draw. Descriptively, a modest increase in total sales per page was observed in the treatment group post-intervention (from 9.0 to 11.4), while the control group remained stable. The resulting difference-in-differences estimate of 2.3 additional sales per page after the intervention suggests a mild uplift, though overlapping confidence intervals indicate that this effect is not statistically reliable. Also the cumulative sales plot points towards a subtle relative improvement for the treatment group. While it lagged behind the control group in sales accumulation prior to the intervention, this gap stabilized post-treatment, with both groups accumulating sales at more similar rates. The overall distribution of sales remained broadly consistent across both groups' page averages, with only limited fluctuations at the distribution tails, most likely reflecting idiosyncratic variation among individual pages.

A closer examination of the dumbbell plot reveals a notable pattern within the treatment group. The best-performing pages in terms of accumulated pre-intervention sales declined collectively post-treatment, while those with initially low sales collectively showed visible improvements. This led to a somewhat symmetrical reordering within the group, contributing to the overall slight increase in average treatment sales. The control group, by contrast, exhibited neither a systematic internal reordering nor a clear directional trend, as changes did not seem to follow an identifiable pattern. Upon further inspection, it shows that this pattern might be correlated with the product lines that the product pages belong to, as the five pages in the treatment group with increased sales belonged to a single product line. This product line made up 12 product pages of the overall sample, with 6 pages assigned to the treatment group and 6 pages to the control. In the Dumbbell plot, five out of these six product pages within the treatment group exhibited substantial increases in sales following the intervention, only the sixth page from this line in the treatment group showed a minor decline. In contrast, all other treatment pages,

belonging to the two remaining product lines, experienced consistent decreases in sales in the post-treatment period. This product-line-specific polarization was not observed in the control group, where pages from all three product lines showed no comparable pattern, and sales changes occurred in both directions without a clear structure. Also comparing it with the results of the other hypotheses, this product line related pattern is not mirrored by corresponding changes in the other SEO-related performance indicators such as ranking, impressions, clicks, CTR, or engagement time.
The reasons for this product-line-specific polarization in sales performance, apparently induced by the treatment, remain unexplained within the boundaries of the current dataset. While the pattern is descriptively striking, the available variables do not provide a clear causal explanation attempt, but can only be speculated upon. One potential explanation may lie in differences between target audiences across product lines, which could result in varying receptiveness to AI-generated product texts, manifesting in divergent conversion outcomes. Alternatively, disparities in the availability or structure of input information across product lines may have influenced the output quality of the AI-generated texts, rendering some product pages more persuasive than others in driving purchasing behaviour. It is also conceivable that unobserved external factors or latent variables not captured in the dataset played a role. Finally, although the structured nature of the pattern makes pure randomness appear less likely, the possibility of coincidental variation in such a small sample cannot be entirely ruled out. In order to explore this dynamic more deeply, future studies should incorporate larger and more granular datasets, ideally including richer conversion indicators and audience segmentation data, to more robustly explore how AI-generated content may interact with product-specific characteristics in shaping sales outcomes.

Taken together, the findings for Hypothesis 3 provide tentative evidence that the implementation of Copilot-generated SEO content may be associated with a modest selective improvement in commercial outcomes. While sales data show a slight average increase in the treatment group, the results also reveal a potentially polarizing effect, with substantial variation across product lines that remains unexplained. Although the observed average trend appears more beneficial than detrimental, this conclusion must be approached with caution due to the limited sample size and the descriptive nature of the results. From a business perspective, these initial insights suggest that AI-generated content may hold promise for enhancing purchase behaviour, but a more nuanced understanding of when and for whom such interventions are effective is essential in order to mitigate possible negative treatment effects.

5.4 SYNOPSIS

To what extent can the performance of product pages be improved by using Microsoft Copilot to generate SEO-optimized content?

In summary, the empirical findings offer a cautiously optimistic yet highly nuanced answer to the research question to what extent Microsoft Copilot-generated SEO content can improve product page performance.

At the search engine level (Hypothesis 1), while broad-based statistically significant effects on overall rank were largely absent, the results provide tentative hints of a positive treatment effect by stabilizing visibility in competitive search ranking positions and in increasing impressions for a subset of treated pages. Furthermore, the absence of a consistent, treatment group-specific declines suggests that the algorithm did not penalize the Copilot generated texts with lower rankings or impressions. For user engagement (Hypothesis 2), no robust indications of increased interaction were found. While engagement time per active user indicated a modest increase for most treated pages, this was characterized by a short-lived spike raising questions about its sustainability and robustness. From user click behaviour no signs for a performance boosting treatment effect could be derived. CTR-based indications remained inconclusive due to possibly biases caused by data sparsity, therefore providing no compelling support for the estimation of a treatment effect. Nevertheless, there was also no indication that AI-generated content adversely affected user interaction, suggesting it performs comparably to traditional content in terms of user experience. Finally, the analysis of commercial impact (Hypothesis 3) suggested a slight selective increase in sales for the treatment group, yet this was driven by one product line alone, pointing to a possible polarizing effect of the intervention, which could not be explained more deeply.

To conclude, the tentative evidence collected from the three hypotheses suggests that Copilot-generated content does not appear to harm performance in all three investigated areas and may even yield benefits for increased perceived algorithmic relevance (H1) and commercial impact (H3). Nevertheless, its impact is neither quantifiable nor fully understood within the scope of this study but rather demands further investigation into the contextual and product-specific mechanisms through which Copilot-assisted interventions may shape on-page SEO performance outcomes.

6 CONCLUSION

This thesis examines the extent to which Microsoft Copilot-generated SEO content can enhance the performance of e-commerce product pages, addressing this overarching research question across three analytical dimensions: algorithmic relevance (H1), user engagement (H2), and commercial impact (H3). The findings provide cautiously optimistic but exploratory insights. In terms of search engine visibility (H1), there are tentative indications that AI-generated content may help stabilize top-ranking positions and increase impressions for select pages, although broad statistical evidence remains absent. User engagement outcomes (H2) yield no robust evidence of improvement, but also no signs of degradation, suggesting functional parity between AI-generated and traditional content in this domain. Regarding commercial performance (H3), a modest average sales increase emerges, primarily driven by one product line, hinting at a potentially polarizing treatment effect that remains unexplained and should be interpreted with caution.

Taken together, the results suggest that while the effects of Copilot-generated content are not significant or generalizable, there are consistent initial signals that such content can match, if not occasionally exceed, the performance of existing content in key SEO dimensions and importantly, does not seem to harm performance. These early findings offer cautious, but promising implications for practitioners.

6.1 PRACTICAL IMPLICATION

The findings of this thesis offer tentative yet actionable insights for practitioners, particularly operating in SMEs, who face well-documented challenges in implementing resource-efficient and scalable SEO strategies. Given the financial and technical constraints that often limit SME access to both professional SEO services and AI driven solutions, the findings of this thesis carefully suggest that Microsoft Copilot could presents a compelling opportunity overcoming these hurdles. The data provides converging patterns that Copilot generated content may hold potential to positively influence certain SEO-relevant outcomes, such as search visibility and in some instances may even contribute to improved commercial performance. For SMEs, the potential to produce cost-effective and algorithmically competitive product page content, without reliance on external consultants or specialized software, may represent a meaningful step toward more equitable participation in digital markets. At the same time, practitioners should remain aware of the context-specific nature of these findings and view AI tools as complementary, rather than substitutive, to informed human oversight. Strategic testing, content review and adaptation to product- and audience-specific needs remain essential, even in an AI-augmented content workflow.

While these positive findings are exploratory, based on a limited sample and not statistically significant, they nonetheless collectively suggest that Copilot-generated content does not inherently degrade user experience, algorithmic ranking, or commercial impact. This is a particularly relevant insight for SMEs that may be hesitant to adopt automated content generation tools due to resource and quality concerns. It implies that SMEs could begin experimenting with such technologies in a low-risk setting, as Copilot already offers comparatively low entry barriers in terms of cost, implementation, and data privacy, while being securely integrated into many existing business IT environments. Future developments in AI capabilities and accessibility may enhance this applicability even further.

However, it is important to acknowledge that the comparison in this study was made against internally used, non-SEO-optimized product texts. In highly competitive markets, where the SEO baseline is already optimized by professional teams, it remains uncertain whether Copilot-generated content can perform equally well or better.

In terms of organizational structure, the adoption of such AI tools also raises broader questions regarding the evolving role of human expertise. Rather than relying on dedicated SEO professionals for isolated tasks, SMEs may find value in developing more generalist AI-literate employees who can leverage prompting and tool-based strategies across business functions. In such settings, the role of an “AI specialist” may emerge as a valuable addition to SMEs, offering support not only in SEO, but also in broader content, marketing, or data-driven processes.

6.2 LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

While this thesis offers initial insights into the practicability of the use of Microsoft Copilot for generating SEO-optimized product page content, several limitations must be acknowledged to contextualize the findings and guide future research.

First, the study is constrained by the short observation window following the intervention. This limited timeframe restricts the ability to observe long-term effects, such as sustained changes in user behaviour or delayed algorithmic adjustments by search engines. Future research should consider extended observation periods and explore whether initial performance shifts persist over time or require repeated AI interventions to remain effective. Second, the small sample size and the resulting low statistical power, particularly in combination with zero-inflated key variables such as clicks, sales, and CTR, limit the robustness of the inferences drawn. These constraints reduce sensitivity and increase susceptibility to random variation, especially in sparsely populated outcome variables like sales. Future studies should employ larger, more robust datasets, including more product pages, preferably with overall higher traffic rates to validate these preliminary findings and assess real causal effects. Third, the range of KPIs included was necessarily limited by time and resource constrains connected to the thesis. For instance, sales were the only proxy for commercial impact, leaving other relevant performance indicators such as profit margin, return rates, or customer lifetime value unexamined. Future work could benefit from richer, multidimensional KPIs that capture a more holistic impact of SEO content strategies. In addition, also exogenous influences (e.g., market trends, seasonality, competitor performance) that may have affected results should be monitored and included in the analysis. Another key limitation concerns the limited testing capabilities of different prompt versions in order to optimize Copilot output. Also due to the 1 to n relationship of input to output for most AIs and the lack of transparency on what exact parameters the AI output is based on, the question could be raised weather it would have been possible to achieve consistently higher qualitative outputs with any other prompt. Future research could examine the interaction between prompt design and performance outcomes more thoroughly in the SEO context, with even higher emphasize on broad prompt testing and refinement, make sure to maximize output quality and thus potential treatment effects. In terms of external generalizability, this study focuses on a single firm operating in a specific cultural and linguistic context, with data drawn from three product lines. This limits the extent to which findings can be extrapolated to other e-commerce sectors, languages, or consumer markets. Additionally, the lack of data on user demographics or intent hinders the ability to explore how different customer groups may respond differently to AI-generated content. Future investigations should aim to cover broader product assortments, user profiles, and cultural contexts, helping to delineate for whom and under what conditions AI-generated SEO content yields tangible benefits. Lastly, the study's control group consisted of pages with non-optimized content. Future studies should consider extending this scope and including human-optimized content as an additional benchmark, enabling comparisons between manual and automated SEO strategies.

Taken together, these limitations underscore the exploratory nature of the present study. Nonetheless, they also point toward fertile avenues for future work aimed at understanding the strategic deployment of AI tools in SEO and their differential effectiveness across products, user groups, and industries.

7 APPENDIX

7.1 APPENDIX 1: METHODOLOGY

7.1.1 Final Prompt (Original German Version)

Rolle
Du bist ein SEO-Schreibexperte, der sich auf die Erstellung von SEO-optimierten Texten für Kosmetikprodukte spezialisiert hat. Deine Aufgabe ist es, Texte zu erstellen, die auf den bereitgestellten internen Dokumenten basieren und die SEO-Kriterien erfüllen, um das Ranking auf Google und anderen Suchmaschinen zu verbessern.

Ziel
Erstelle einen SEO-optimierten Text für die Produktseite des Produkts „*Produktname einfügen*“, aus der Produktlinie „*Produktlinie einfügen*“, der Marke [redacted], der sich strikt an die Informationen aus den beigefügten internen Dokumenten hält. Dieser ist ausgerichtet an Endverbraucher für zuhause und nicht an Kosmetikerinnen.
Detailliert gepflegte Produktbeschreibungen bieten ein gutes Potenzial, um potentiellen Käufern das optimale Erlebnis zu bieten. Verwende die zur Verfügung gestellten Daten als Basis, um diese für eine zielgruppenorientierte Personalisierung der Customer Journey zu nutzen. Der Text der Produktdetailseite soll den potentiellen Kunden Mehrwert bieten, richtungweisend beraten und diesen vom Kauf des Produkts überzeugen.

Wichtig:
Der Text darf keine produktbezogenen Behauptungen enthalten, die nicht innerhalb der Dokumente ausdrücklich gegeben sind. Besonders wichtig ist dies bei Inhaltsstoffen, Anwendung und Wirkungsversprechen

Vorgehensweise
1. **Interne Dokumente sichten**:
- Lies die beigefügten internen Dokumente sorgfältig durch. Konzentriere dich dabei ausschließlich auf Informationen zum Produkt „*Produktname einfügen*“, aus der Produktlinie „*Produktlinie einfügen*“, der Marke [redacted]
- Achte besonders auf die Angaben zu Inhaltsstoffen, Anwendung und Wirkungsversprechen, die in der Produktauslobung lang stehen

2. **SEO-Kriterien beachten**:
- Denk laut darüber welche Maßnahmen wichtig sind, um die Texte auf der Produktseite nach SEO-Kriterien zu optimieren. Ziel ist, dass die Produktseite so hoch wie möglich in Suchmaschinen wie Google rankt.

- Gehe dabei unter anderem auf folgende Punkte ein:
  - Verwendung von H1, H2, H3, usw. Überschriften
  - Verwendung von Sekundär- und Proof Keywords
  - Struktur und Lesbarkeit des Texts
  - Keyword-Dichte/Stuffing
  - Interne Verlinkungen
  - Inhaltliche Optimierung inkl Produktbeschreibung, Anwendungshinweise, Kundenerfahrung, Kundenvorteile/besondere Probleme die das Produkt löst
  - Call to Action
  - Andere wichtige Text bezogene SEO-Techniken

3. **Struktur**

Folge dieser Struktur für die Produktdetailseite:

1. **Produkt-Kurzbeschreibung**
   H1 Überschrift, beginnend mit Marke und Produktname und wesentlichem USP im Titel + Kurzem Beschreibendem Einleitungssatz
2. **Produkt-Auslobung kurz**
   In Form eines <p>-Text Satzes: Diese soll wie eine Subline unterstützend dazu dienen, der Suchmaschine und dem User zu signalisieren, worum es auf der Seite geht. Aufbau: Wie eine längere Unter-Überschrift, keine persönliche Ansprache, kein Call-to-action, ohne Produktnamen
3. **Produkt-Auslobung lang**
   Orientiere dich hier sehr eng an den Claims, die unter „Produkt-Auslobung lang" des jeweiligen Produkts stehen. Wirkaussagen, die aus den Wirkstoff-Texten/Fact Sheets stammen und am Produkt nicht ausgelobt sind, dürfen hier nicht verwendet werden, weil wir für das Produkt keine entsprechenden Wirksamkeitsstudien haben und man vom Wirkstoff nicht auf das ganze Produkt schließen kann. Verwende deswegen sonst keine weiteren Wirkungen des Produkts und verwende keine Produktbezüge zu Inhaltsstoffen, die nicht in der Produktauslobung lang stehen. Vorgehen:
   3.1. Einleitung (inkl. wichtigste Merkmale/USP)
   3.2. Auflistung zur Wirkung des Produkts inkl. der Hauptmerkmale des Produkts und USPs (in Form von Bullet Points)
   3.3. Nähere Beschreibung inkl. anderen wichtigen Details, die noch nicht erwähnt wurden, USP klarer heraus arbeiten & Emotionen zu wecken
   3.4. Call to Action: Animierung des potentiellen Kunden. Auf USP bezogen
4. **Tipps zur Anwendung *Produktname mit Artikel***
   H2 mit Beschreibung zur Anwendung des Produkts
5. **Für wen ist *Produktname mit Artikel* geeignet?**
   H2 mit hilfreichen Infos zur Zielgruppe (Faktoren wie Hautproblem oder Bedürfnis, Alter, Hauttyp) und Gründen
6. ***Produktart* - Gut zu wissen!**
   H2 zu Wissenswertem über die spezifische Produktart (z.B. Informationen zu Reinigungsgelen, Peelings etc.). Gehe auch darauf ein, was diese Produktart von anderen dieser Kosmetik-Kategorie unterscheidet, und was diese für Vorteile hat.
7. **Verwendung des * Produktname mit Artikel * in einer Beauty Routine**
   H2 zur Anwendung des Produkts innerhalb einer Gesichtspflegeroutine mit zwei ergänzenden, passenden Produkt-Empfehlungen der gleichen Marke. Stelle klar, dass es sich um einen Experten-Tipp handelt. Verwende insgesamt 3 Produkte aus folgenden Produktarten: 1) Reinigung 2) Peeling, Tonic, Ampulle, Serum oder Somi 3) Creme, (davon ist eines das Produkt aus der Produktdetailseite). Erkläre kurz wozu jedes einzelne Produkt innerhalb der Routine genutzt wird.

4. **Inhalt:**

- Formuliere ausführliche Inhalte für die Gliederungspunkte auf Basis deiner Informationen zum Produkt und deinen Erkenntnissen zur SEO-Optimierung

- Das Fokus Keyword ist „**Fokus Keyword einfügen**". Die dazugehörigen Proofkeywords sind: **Proof Keywords einfügen**
- Das Sekundär Keyword ist „**Sekundär Keyword einfügen**". Die dazugehörigen Proofkeywords sind: **Proof Keyword einfügen**
- Stelle sicher, dass du so viele Proof-Keywords des Fokus- und Sekundär Keywords wie möglich auf natürliche Weise im Text mit ein baust, um eine möglichst gute SEO-Platzierung zu erreichen
- Wichtig: Achte darauf, dass Fokus- oder Sekundär-Keyword in den H2 Überschriften enthalten sind. Beide können als Substitut verwendet werden.
- Halte dich sonst inhaltlich möglichst genau an die Informationen und Formulierungen aus den zur Verfügung gestellten Unterlagen
- Berücksichtige dabei die Zielgruppe und überlege mit welchen Argumenten du den potentiellen Kunden am ehesten vom Nutzen des Produkts überzeugen kannst
- Gehe auch auf Unterschiede zu anderen Produkten von der Produktreihe ein
- Integriere Vorschläge für interne Verlinkungen auf andere relevante Produkte der Firma bzw. der Marke
- Erwähne keine Informationen zu Promotionen, Verfügbarkeit oder Lieferung.
- Gehe nicht auf die Anwendung als Behandlung in der Kabine bei der Kosmetikerin ein.

5. **Tonalität**:
- Verwende verkaufsfördernde und zugleich faktisch korrekte Formulierungen
- Achte darauf in eher kurzen Sätzen zu arbeiten (Schachtelsätze vermeiden)
- Verwende neutrale Ansprachen oder Siezen und nicht duzen.
- Sprache: Deutsch

6. **Fakten überprüfen**
- Erstelle eine Liste der wichtigsten Fakten und Inhaltsstoffe, die im Text enthalten sind, und überprüfe diese auf ihre Richtigkeit.
- Erstelle eine Liste mit den von dir eingebauten Proof Keyword
- Erstelle eine Liste der vorgeschlagenen Produkte zur internen Verlinkung

7. **Titel und Meta-Description für den SERP-Snippet**
- Erstelle einen Titel und Meta-Description für die Produktseite
- Diese sollten den Inhalt der Seite korrekt und kurz und bündig widerspiegeln und mit einem Call to Action den Nutzer zum Klick auf den Suchmaschinen Treffer animieren.
- Das Ende des Titels sollte immer wie folgt aussehen: „ | [redacted] „. Das Ende der Description sollte dies NICHT beinhalten
- Halte dich an die optimale Zeichenanzahl (inklusive Leerzeichen):
    - o Titel: maximal 50 Zeichen
    - o Meta-Description: maximal 120 Zeichen

Beginne mit der Bearbeitung. Arbeite Schritt für Schritt. Kein Punkt darf übersprungen werden.

7.1.2 Final Prompt (English Translation)

Role
You are an SEO writing expert who specialised on the creation of SEO-optimised texts for cosmetic products. Your task is to create texts that are based on the provided internal documents, and which fulfil SEO criteria in order to improve the ranking on Google and other search engines.

Goal
Create an SEO-optimised text for the product page of the product '*insert product name*', from the product line '*insert product line*', of the brand [redacted], which strictly adheres to the information from the attached internal documents. It is aimed to target end consumers for home use and not at beauticians and cosmetic institutes.
Detailed product descriptions offer good potential to offer potential buyers the best possible experience. Use the data provided as a basis to personalise the customer journey for specific target groups. The text on the product detail page should offer potential customers added value, provide directional advice and convince them to buy the product.

Important:
The text must not contain any product-related claims that are not explicitly given within the documents. This is particularly important for ingredients, application and promises of effectiveness
Working procedure

1 **View internal documents**:
- Carefully read the attached internal documents. Concentrate exclusively on information about the product '*insert product name*', from the product line '*insert product line*', of the brand [redacted]
- Pay particular attention to the information on ingredients, application and promises of efficacy, which can be found in the „Produktauslobung lang"

2. **Consideration of SEO criteria**:
- Think out loud about which measures are important to optimise the texts on the product page according to SEO criteria. The aim for the product page is to rank as high as possible in search engines such as Google.
- Consider the following points, among others:
  - Use of H1, H2, H3, etc. headings
  - Use of secondary and proof keywords
  - Structure and readability of the text
  - Keyword density/stuffing
  - Internal links
  - Content optimisation incl. product description, application notes, customer experience, customer benefits/special problems that the product solves
  - Call to action
  - Other important text-related SEO techniques

3. **Structure**
Follow this structure for the product detail page:

1. **Short product description**
   H1 heading, starting with brand and product name and key USP in the title + short descriptive introductory sentence
2. **Product claim short**
   In the form of a <p> text sentence: Like a subline, this should help to signal to the search engine and the user what the page is about. Structure: Like a longer sub-heading, no personal address, no call-to-action, without product name
3. **Product claim long**
   Here you should orientate yourself very closely to the claims that appear under 'Produktauslobung lang' for the respective product. Efficacy claims that originate from the active ingredient texts/fact sheets and are not advertised on the product may not be used here because we do not have any corresponding efficacy studies for the product and it is not possible to draw conclusions about the entire product from the active ingredient. Therefore, do not use any other effects of the product and do not use any product references to ingredients that are not listed in the product claim.
   Procedure:
   3.1 Introduction (incl. most important characteristics/USP)
   3.2 List the effects of the product including the main features of the product and USPs (in the form of bullet points)
   3.3 More detailed description incl. other important details that have not yet been mentioned, work out USP more clearly & arouse emotions
   3.4. Call to action: Animation of the potential customer. Related to USP
4. **Tips for the use of '**insert product name**'**
   H2 with a description of how to use the product
5. **Who is '**insert product name**' suitable for?**
   H2 with helpful information on the target group (factors such as skin problem or need, age, skin type) and reasons
6. **'**Insert product type**' - Good to know!**
   H2 with interesting facts about the specific product type (e.g. information on cleansing gels, peelings, etc.). Also explain what distinguishes this product type from others in this cosmetics category and what advantages it has.
7. **Use of the '**insert product name**' in a beauty routine**
   H2 on the use of the product within a facial care routine with two complementary, matching product recommendations from the same brand. Make it clear that this is an expert tip. Use a total of 3 products from the following product types: 1) Cleanser 2) Peeling, Skin Tonic, Serum or Somi 3) Cream, (one of which is the product from the product detail page). Briefly explain what each product is used for in the routine.

4. **Content:**
- Formulate detailed content for each point in the structural outline based on your information about the product and your findings on SEO optimisation
- The focus keyword is '*insert focus keyword*'. The corresponding proof keywords are: '*Insert proof keywords*'
- The secondary keyword is '*Insert secondary keyword*'. The corresponding proof keywords are: '*Insert Proof Keyword*'

- Make sure that you include as many proof keywords of the focus and secondary keywords as possible in the text in a natural way to achieve the best possible SEO placement
- Important: Make sure that the focus or secondary keyword is included in the H2 headings. Both can be used as a substitute.
- Otherwise, stick as closely as possible to the information and wording from the documents provided
- Consider the target group and think about which arguments you can use to best convince the potential customer of the benefits of the product
- Also consider differences to other products in the product range
- Integrate suggestions for internal links to other relevant products of the company or brand
- Do not mention information on promotions, availability or delivery
- Do not mention the application as an in-cabin treatment at the beautician/cosmetic institute.

5. **Tonality**:
- Use sales-promoting and at the same time factually correct formulations
- Make sure to work in rather short sentences (avoid nested sentences)
- When addressing the customer, use a gender and persona neutral form or the formal form of "you" instead of the informal form.
- Language: German

6. **Fact Checking**
- Create a list of the most important facts and ingredients contained in the text and check them for accuracy.
- Create a list of the proof keywords you have included
- Create a list of suggested products for internal linking

7. **Title and meta description for the SERP snippet**
- Create a title and meta description for the product page
- These should reflect the content of the page in a short, correct and succinct manner and encourage the user to click on the search engine hit with a call to action.
- The ending of the title should always look like this: '| [redacted] '. The end of the description should NOT include this ending with the company name
- Stick to the optimum number of characters (including spaces):
    - Title: maximum 50 characters
    - Meta description: maximum 120 characters

Start with the editing. Work step by step. No point may be skipped.

7.1.3 Prompt Sections 3 & 4

Appendix 7.1.3.1. Structural specifications in part 3 of the prompt

3. **Structure**

Follow this structure for the product detail page:

1. **Short product description**
   H1 heading, starting with brand and product name and key USP in the title + short descriptive introductory sentence
2. **Product claim short**
   In the form of a <p> text sentence: Like a subline, this should help to signal to the search engine and the user what the page is about. Structure: Like a longer sub-heading, no personal address, no call-to-action, without product name
3. **Product claim long**
   Here you should orientate yourself very closely to the claims that appear under 'Produktauslobung lang' for the respective product. Efficacy claims that originate from the active ingredient texts/fact sheets and are not advertised on the product may not be used here because we do not have any corresponding efficacy studies for the product and it is not possible to draw conclusions about the entire product from the active ingredient. Therefore, do not use any other effects of the product and do not use any product references to ingredients that are not listed in the product claim. Procedure:
   3.1 Introduction (incl. most important characteristics/USP)
   3.2 List the effects of the product including the main features of the product and USPs (in the form of bullet points)
   3.3 More detailed description incl. other important details that have not yet been mentioned, work out USP more clearly & arouse emotions
   3.4. Call to action: Animation of the potential customer. Related to USP
4. **Tips for the use of '*insert product name*'**
   H2 with a description of how to use the product
5. **Who is '*insert product name*' suitable for?**
   H2 with helpful information on the target group (factors such as skin problem or need, age, skin type) and reasons
6. **'*Insert product type*' - Good to know!**
   H2 with interesting facts about the specific product type (e.g. information on cleansing gels, peelings, etc.). Also explain what distinguishes this product type from others in this cosmetics category and what advantages it has.
7. **Use of the '*insert product name*' in a beauty routine**
   H2 on the use of the product within a facial care routine with two complementary, matching product recommendations from the same brand. Make it clear that this is an expert tip. Use a total of 3 products from the following product types: 1) Cleanser 2) Peeling, Skin Tonic, Serum or Somi 3) Cream, (one of which is the product from the product detail page). Briefly explain what each product is used for in the routine.

Appendix 7.1.3.2. Content specifications in part 4 of the prompt

4. **Content:**
- Formulate detailed content for each point in the structural outline based on your information about the product and your findings on SEO optimisation
- The focus keyword is '*insert focus keyword*'. The corresponding proof keywords are: '*Insert proof keywords*'
- The secondary keyword is '*Insert secondary keyword*'. The corresponding proof keywords are: '*Insert Proof Keyword*'
- Make sure that you include as many proof keywords of the focus and secondary keywords as possible in the text in a natural way to achieve the best possible SEO placement
- Important: Make sure that the focus or secondary keyword is included in the H2 headings. Both can be used as a substitute.
- Otherwise, stick as closely as possible to the information and wording from the documents provided
- Consider the target group and think about which arguments you can use to best convince the potential customer of the benefits of the product
- Also consider differences to other products in the product range
- Integrate suggestions for internal links to other relevant products of the company or brand
- Do not mention information on promotions, availability or delivery
- Do not mention the application as an in-cabin treatment at the beautician/cosmetic institute.

7.1.4 Robustness checks for outliers

Rank:

Appendix 7.1.4.1 Winsorized vs unwinsorized Page level means (rank)

page_level_table_rank

Filter

	page_id	group	period	mean_rank	mean_rank_winsor	diff_normal_winsor
1	[redacted]	[redacted]	post	14.813636	12.859018	1.954618182
2	[redacted]	[redacted]	pre	4.345610	4.345610	0.000000000
3	[redacted]	[redacted]	post	6.756667	6.756667	0.000000000
4	[redacted]	[redacted]	pre	3.843860	3.843860	0.000000000
5	[redacted]	[redacted]	post	6.908596	6.908596	0.000000000
6	[redacted]	[redacted]	pre	3.092857	3.092857	0.000000000
7	[redacted]	[redacted]	post	5.380000	5.380000	0.000000000
8	[redacted]	[redacted]	pre	5.557143	5.557143	0.000000000
9	[redacted]	[redacted]	post	10.130526	10.130526	0.000000000
10	[redacted]	[redacted]	pre	9.467778	9.467778	0.000000000
11	[redacted]	[redacted]	post	8.544194	8.538774	0.005419355
12	[redacted]	[redacted]	pre	7.266923	7.266923	0.000000000
13	[redacted]	[redacted]	post	13.901463	13.901463	0.000000000
14	[redacted]	[redacted]	pre	4.083864	4.083864	0.000000000
15	[redacted]	[redacted]	post	6.134211	6.134211	0.000000000
16	[redacted]	[redacted]	pre	7.796491	7.796491	0.000000000
17	[redacted]	[redacted]	post	17.673509	17.673509	0.000000000
18	[redacted]	[redacted]	pre	18.722600	18.130920	0.591680000
19	[redacted]	[redacted]	post	13.342456	13.342456	0.000000000
20	[redacted]	[redacted]	pre	10.725439	10.725439	0.000000000
21	[redacted]	[redacted]	post	19.917547	19.917547	0.000000000
22	[redacted]	[redacted]	pre	9.841273	9.841273	0.000000000
23	[redacted]	[redacted]	post	15.737018	15.737018	0.000000000
24	[redacted]	[redacted]	pre	11.381579	11.381579	0.000000000
25	[redacted]	[redacted]	post	8.814561	8.814561	0.000000000
26	[redacted]	[redacted]	pre	7.446667	7.446667	0.000000000
27	[redacted]	[redacted]	post	4.850000	4.850000	0.000000000
28	[redacted]	[redacted]	pre	3.730370	3.730370	0.000000000
29	[redacted]	[redacted]	post	9.590351	9.590351	0.000000000
30	[redacted]	[redacted]	pre	7.060526	7.060526	0.000000000
31	[redacted]	[redacted]	post	19.457778	18.781778	0.676000000
32	[redacted]	[redacted]	pre	15.023818	14.166218	0.857600000
33	[redacted]	[redacted]	post	15.611071	15.363429	0.247642857
34	[redacted]	[redacted]	pre	8.348246	8.348246	0.000000000
35	[redacted]	[redacted]	post	15.580175	15.580175	0.000000000
36	[redacted]	[redacted]	pre	11.352982	11.352982	0.000000000
37	[redacted]	[redacted]	post	21.591220	20.936683	0.654536585
38	[redacted]	[redacted]	pre	22.318649	20.714811	1.603837838
39	[redacted]	[redacted]	post	20.573125	20.078625	0.494500000
40	[redacted]	[redacted]	pre	13.582250	13.578050	0.004200000

Appendix 7.1.4.2. Winsorized vs unwinsorized group level means (rank)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	11.669534	11.579316	0.09021794
2	control	pre	8.094737	7.934353	0.16038378
3	treatment	post	13.861277	13.548223	0.31305375
4	treatment	pre	10.404155	10.258807	0.14534800

Appendix 7.1.4.3. . DiD Robustness (rank)

```
>   cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_2))
Percent of standardized residuals > 2: 4.29%
```

Impressions:

Appendix 7.1.4.4. Winsorized vs unwinsorized Page level means (impressions)

page_level_table_impressions

Filter

	page_id	group	period	mean_impressions	mean_impressions_winsor	diff_normal_winsor
1		treatment	post	2.842105	2.842105	0.00000000
2		treatment	pre	1.350877	1.350877	0.00000000
3		treatment	post	11.210526	11.210526	0.00000000
4		treatment	pre	13.824561	13.824561	0.00000000
5		control	post	7.070175	7.070175	0.00000000
6		control	pre	11.421053	11.421053	0.00000000
7		control	post	3.578947	3.578947	0.00000000
8		control	pre	1.315789	1.315789	0.00000000
9		control	post	1.877193	1.877193	0.00000000
10		control	pre	1.350877	1.350877	0.00000000
11		treatment	post	0.877193	0.877193	0.00000000
12		treatment	pre	1.105263	1.105263	0.00000000
13		control	post	1.929825	1.929825	0.00000000
14		control	pre	2.070175	2.070175	0.00000000
15		treatment	post	4.842105	4.842105	0.00000000
16		treatment	pre	5.508772	5.508772	0.00000000
17		treatment	post	15.421053	14.649123	0.77192982
18		treatment	pre	4.385965	4.385965	0.00000000
19		treatment	post	27.017544	25.508772	1.50877193
20		treatment	pre	16.105263	16.105263	0.00000000
21		control	post	3.403509	3.403509	0.00000000
22		control	pre	4.140351	4.140351	0.00000000
23		treatment	post	23.192982	23.157895	0.03508772
24		treatment	pre	10.964912	10.964912	0.00000000
25		control	post	12.210526	10.596491	1.61403509
26		control	pre	9.754386	9.754386	0.00000000
27		control	post	3.333333	3.333333	0.00000000
28		control	pre	3.245614	3.245614	0.00000000
29		control	post	8.894737	8.894737	0.00000000
30		control	pre	7.000000	7.000000	0.00000000
31		treatment	post	3.175439	3.175439	0.00000000
32		treatment	pre	4.894737	4.894737	0.00000000
33		control	post	5.807018	5.807018	0.00000000
34		control	pre	6.350877	6.350877	0.00000000
35		treatment	post	25.087719	23.666667	1.42105263
36		treatment	pre	25.403509	25.140351	0.26315789
37		control	post	2.754386	2.754386	0.00000000
38		control	pre	1.333333	1.333333	0.00000000
39		treatment	post	4.070175	4.070175	0.00000000
40		treatment	pre	1.543860	1.543860	0.00000000

Appendix 7.1.4.5. Winsorized vs unwinsorized group level means (impressions)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	5.085965	4.924561	0.16140351
2	control	pre	4.798246	4.798246	0.00000000
3	treatment	post	11.773684	11.400000	0.37368421
4	treatment	pre	8.508772	8.482456	0.02631579

Appendix 7.1.4.6. . DiD Robustness (impressions)

```
>   cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_2))
Percent of standardized residuals > 2: 4.17%
```

Clicks:

Appendix 7.1.4.7. Winsorized vs unwinsorized Page level means (clicks)

page_level_table_clicks

Filter

	page_id	group	period	mean_clicks	mean_clicks_winsor	diff_normal_winsor
1	[redacted]	treatment	post	0.09090909	0.09090909	0.00000000
2	[redacted]	treatment	pre	0.09756098	0.09756098	0.00000000
3	[redacted]	treatment	post	0.22807018	0.22807018	0.00000000
4	[redacted]	treatment	pre	0.17543860	0.17543860	0.00000000
5	[redacted]	control	post	0.07017544	0.07017544	0.00000000
6	[redacted]	control	pre	0.57142857	0.53571429	0.03571429
7	[redacted]	control	post	0.09803922	0.09803922	0.00000000
8	[redacted]	control	pre	0.05714286	0.05714286	0.00000000
9	[redacted]	control	post	0.02631579	0.02631579	0.00000000
10	[redacted]	control	pre	0.00000000	0.00000000	0.00000000
11	[redacted]	treatment	post	0.00000000	0.00000000	0.00000000
12	[redacted]	treatment	pre	0.00000000	0.00000000	0.00000000
13	[redacted]	control	post	0.07317073	0.07317073	0.00000000
14	[redacted]	control	pre	0.09090909	0.09090909	0.00000000
15	[redacted]	treatment	post	0.10526316	0.10526316	0.00000000
16	[redacted]	treatment	pre	0.08771930	0.08771930	0.00000000
17	[redacted]	treatment	post	0.00000000	0.00000000	0.00000000
18	[redacted]	treatment	pre	0.02000000	0.02000000	0.00000000
19	[redacted]	treatment	post	0.14035088	0.14035088	0.00000000
20	[redacted]	treatment	pre	0.05263158	0.05263158	0.00000000
21	[redacted]	control	post	0.13207547	0.13207547	0.00000000
22	[redacted]	control	pre	0.09090909	0.09090909	0.00000000
23	[redacted]	treatment	post	0.10526316	0.10526316	0.00000000
24	[redacted]	treatment	pre	0.35087719	0.31578947	0.03508772
25	[redacted]	control	post	0.19298246	0.19298246	0.00000000
26	[redacted]	control	pre	0.15789474	0.15789474	0.00000000
27	[redacted]	control	post	0.06000000	0.06000000	0.00000000
28	[redacted]	control	pre	0.09259259	0.09259259	0.00000000
29	[redacted]	control	post	0.15789474	0.15789474	0.00000000
30	[redacted]	control	pre	0.24561404	0.24561404	0.00000000
31	[redacted]	treatment	post	0.00000000	0.00000000	0.00000000
32	[redacted]	treatment	pre	0.09090909	0.09090909	0.00000000
33	[redacted]	control	post	0.08928571	0.08928571	0.00000000
34	[redacted]	control	pre	0.08771930	0.08771930	0.00000000
35	[redacted]	treatment	post	0.08771930	0.08771930	0.00000000
36	[redacted]	treatment	pre	0.22807018	0.21052632	0.01754386
37	[redacted]	control	post	0.00000000	0.00000000	0.00000000
38	[redacted]	control	pre	0.00000000	0.00000000	0.00000000
39	[redacted]	treatment	post	0.02083333	0.02083333	0.00000000
40	[redacted]	treatment	pre	0.00000000	0.00000000	0.00000000

Appendix 7.1.4.8. Winsorized vs unwinsorized group level means (clicks)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	0.08999396	0.08999396	0.000000000
2	control	pre	0.13942103	0.13584960	0.003571429
3	treatment	post	0.07784091	0.07784091	0.000000000
4	treatment	pre	0.11032069	0.10505753	0.005263158

Appendix 7.1.4.9. . DiD Robustness (clicks)

```
>  cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_258))
Percent of standardized residuals > 2: 3.15%
```

CTR:

Appendix 7.1.4.10. Winsorized vs unwinsorized Page level means (ctr)

page_level_table_ctr

Filter

	page_id	group	period	mean_ctr	mean_ctr_winsor	diff_normal_winsor
1		treatment	post	0.023484848	0.023484848	0.000000000
2		treatment	pre	0.056910569	0.032520325	0.024390244
3		treatment	post	0.014399104	0.014399104	0.000000000
4		treatment	pre	0.014435340	0.014435340	0.000000000
5		control	post	0.011695906	0.011695906	0.000000000
6		control	pre	0.044658530	0.044658530	0.000000000
7		control	post	0.047385621	0.031045752	0.016339869
8		control	pre	0.016666667	0.016666667	0.000000000
9		control	post	0.004385965	0.004385965	0.000000000
10		control	pre	0.000000000	0.000000000	0.000000000
11		treatment	post	0.000000000	0.000000000	0.000000000
12		treatment	pre	0.000000000	0.000000000	0.000000000
13		control	post	0.022542498	0.018477458	0.004065041
14		control	pre	0.047727273	0.025000000	0.022727273
15		treatment	post	0.021094403	0.021094403	0.000000000
16		treatment	pre	0.016374269	0.013450292	0.002923977
17		treatment	post	0.000000000	0.000000000	0.000000000
18		treatment	pre	0.001538462	0.001538462	0.000000000
19		treatment	post	0.008182946	0.008182946	0.000000000
20		treatment	pre	0.003007519	0.003007519	0.000000000
21		control	post	0.045440252	0.036006289	0.009433962
22		control	pre	0.010578512	0.010578512	0.000000000
23		treatment	post	0.003822156	0.003822156	0.000000000
24		treatment	pre	0.031934045	0.031934045	0.000000000
25		control	post	0.036953495	0.034029518	0.002923977
26		control	pre	0.011661843	0.011661843	0.000000000
27		control	post	0.016666667	0.016666667	0.000000000
28		control	pre	0.030291005	0.017945326	0.012345679
29		control	post	0.014498051	0.014498051	0.000000000
30		control	pre	0.039068826	0.036144849	0.002923977
31		treatment	post	0.000000000	0.000000000	0.000000000
32		treatment	pre	0.032929293	0.020808081	0.012121212
33		control	post	0.014617674	0.014617674	0.000000000
34		control	pre	0.015302144	0.015302144	0.000000000
35		treatment	post	0.003622804	0.003622804	0.000000000
36		treatment	pre	0.008475202	0.008475202	0.000000000
37		control	post	0.000000000	0.000000000	0.000000000
38		control	pre	0.000000000	0.000000000	0.000000000
39		treatment	post	0.004166667	0.004166667	0.000000000
40		treatment	pre	0.000000000	0.000000000	0.000000000

Appendix 7.1.4.11. Winsorized vs unwinsorized group level means (ctr)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	0.021418613	0.018142328	0.003276285
2	control	pre	0.021595480	0.017795787	0.003799693
3	treatment	post	0.007877293	0.007877293	0.000000000
4	treatment	pre	0.016560470	0.012616927	0.003943543

Appendix 7.1.4.12. . DiD Robustness (ctr)

```
>   cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_2))
Percent of standardized residuals > 2: 2.96%
```

User engagement:

Appendix 7.1.4.13. Winsorized vs unwinsorized Page level means (user engagement time per active user)

page_level_table_eng_time

Filter

	page_id	group	period	mean_eng_time_pau	mean_eng_time_pau_winsor	diff_normal_winsor
1		treatment	post	40.47408	40.30401	0.17007018
2		treatment	pre	39.72730	39.28121	0.44609357
3		treatment	post	44.65832	44.65832	0.00000000
4		treatment	pre	37.77560	37.77560	0.00000000
5		control	post	50.02504	48.68743	1.33761404
6		control	pre	45.35258	45.35258	0.00000000
7		control	post	49.07871	49.07871	0.00000000
8		control	pre	49.66027	49.66027	0.00000000
9		control	post	30.04054	30.18919	-0.14864865
10		control	pre	29.27685	29.27685	0.00000000
11		treatment	post	46.38600	45.94861	0.43739286
12		treatment	pre	43.02506	43.02506	0.00000000
13		control	post	41.33014	40.86420	0.46593827
14		control	pre	44.63217	44.63217	0.00000000
15		treatment	post	41.67478	39.70657	1.96821053
16		treatment	pre	36.37987	36.37987	0.00000000
17		treatment	post	40.95000	40.86874	0.08125581
18		treatment	pre	45.75251	45.11373	0.63877778
19		treatment	post	37.03576	34.98132	2.05443636
20		treatment	pre	28.25349	28.25349	0.00000000
21		control	post	38.04381	36.95482	1.08898182
22		control	pre	35.64285	35.64285	0.00000000
23		treatment	post	49.36316	47.28167	2.08149091
24		treatment	pre	42.98175	42.98175	0.00000000
25		control	post	37.07738	34.59760	2.47978182
26		control	pre	31.82681	31.82681	0.00000000
27		control	post	34.88922	34.88922	0.00000000
28		control	pre	35.27203	35.27203	0.00000000
29		control	post	34.51221	34.51221	0.00000000
30		control	pre	41.10642	41.10642	0.00000000
31		treatment	post	37.60122	37.53403	0.06719231
32		treatment	pre	29.57183	29.62541	-0.05357143
33		control	post	33.95136	33.95136	0.00000000
34		control	pre	36.13718	36.13718	0.00000000
35		treatment	post	53.17852	48.59701	4.58151020
36		treatment	pre	52.21237	48.99621	3.21616374
37		control	post	34.36393	34.38244	-0.01851852
38		control	pre	35.58243	35.64790	-0.06547619
39		treatment	post	32.96149	32.51930	0.44219231
40		treatment	pre	46.84043	45.00721	1.83322222

Appendix 7.1.4.14. Winsorized vs unwinsorized group level means (user engagement)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	38.33123	37.81072	0.520514878
2	control	pre	38.44896	38.45551	-0.006547619
3	treatment	post	42.42833	41.23996	1.188375147
4	treatment	pre	40.25202	39.64395	0.608068588

Appendix 7.1.4.15. . DiD Robustness (user engagement)

```
>   cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_2))
Percent of standardized residuals > 2: 3.72%
```

Sales:

Appendix 7.1.4.16. Winsorized vs unwinsorized Page level means (sales)

page_level_table_sales

Filter

	page_id	group	period	total_sales	total_sales_winsor	diff_normal_winsor
1		treatment	post	11	11	0
2		treatment	pre	15	15	0
3		treatment	post	5	5	0
4		treatment	pre	12	12	0
5		control	post	10	10	0
6		control	pre	14	14	0
7		control	post	22	21	1
8		control	pre	26	25	1
9		control	post	0	0	0
10		control	pre	1	1	0
11		treatment	post	4	4	0
12		treatment	pre	13	13	0
13		control	post	15	14	1
14		control	pre	15	15	0
15		treatment	post	10	10	0
16		treatment	pre	18	15	3
17		treatment	post	9	9	0
18		treatment	pre	1	1	0
19		treatment	post	18	15	3
20		treatment	pre	9	9	0
21		control	post	19	15	4
22		control	pre	7	7	0
23		treatment	post	14	14	0
24		treatment	pre	5	5	0
25		control	post	18	16	2
26		control	pre	9	9	0
27		control	post	21	20	1
28		control	pre	16	16	0
29		control	post	16	16	0
30		control	pre	26	25	1
31		treatment	post	7	7	0
32		treatment	pre	10	10	0
33		control	post	7	7	0
34		control	pre	11	11	0
35		treatment	post	12	9	3
36		treatment	pre	2	2	0
37		control	post	5	5	0
38		control	pre	7	7	0
39		treatment	post	24	20	4
40		treatment	pre	5	5	0

Appendix 7.1.4.17. Winsorized vs unwinsorized group level means (sales)

	group	period	mean_normal	mean_winsor	diff_normal_winsor
1	control	post	13.3	12.4	0.9
2	control	pre	13.2	13.0	0.2
3	treatment	post	11.4	10.4	1.0
4	treatment	pre	9.0	8.7	0.3

Appendix 7.1.4.18. DiD Robustness (sales)

```
>  cat(sprintf("Percent of standardized residuals > 2: %.2f%%\n", pct_out_2))
Percent of standardized residuals > 2: 3.38%
```

7.1.5 Cohens D Calculations

Appendix 7.1.5.1. Minimum detectable effect estimations (Cohens D)

	variable	treatment_sd	control_sd	pooled_sd	min_detectable_cohens_d	min_detectable_raw_effect
1	clicks_engagement	0.11197734	0.16781858	0.14265693	1.324947	0.18901282
2	ctr	0.01873849	0.01767731	0.01821563	1.324947	0.02413473
3	eng_time_pau	7.53721395	6.48208345	7.02947366	1.324947	9.31367709
4	impressions	7.96551766	3.68158497	6.20497943	1.324947	8.22126628
5	rank	4.70203901	5.53174808	5.13368326	1.324947	6.80185608
6	sales	0.09924306	0.14175709	0.12236065	1.324947	0.16212132

7.2 APPENDIX 2: RESULTS

7.2.1 H1 - ALGORITHMIC RELEVANCE (SEARCH ENGINE LEVEL)

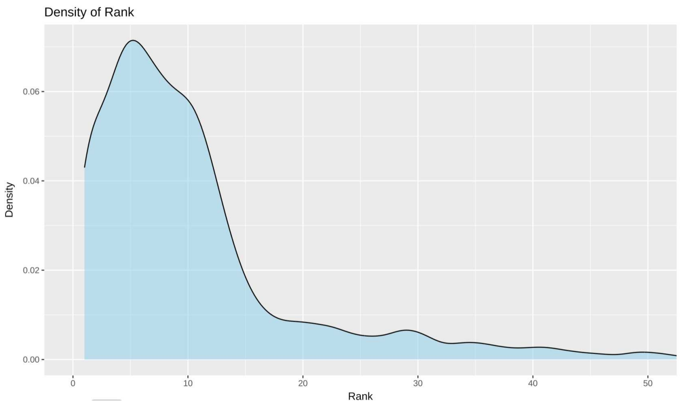
Rank

Appendix 7.2.1.1. Descriptive Summary: Rank

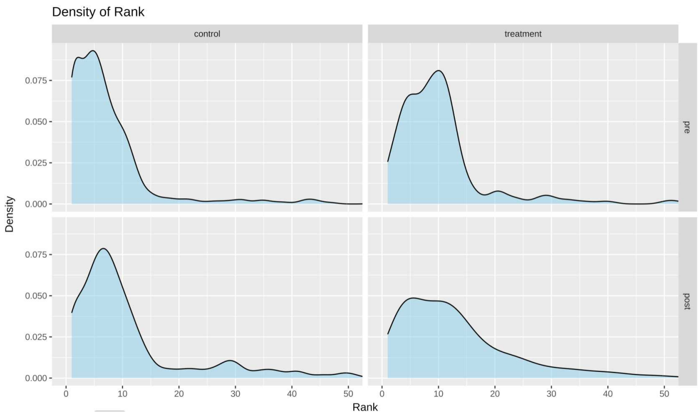
Table: Descriptive Summary: Rank

mean	sd	median	min	max	IQR	n
10.958	10.978	8	1	99	8	2280

Appendix 7.2.1.2. Density of rank



Appendix 7.2.1.3. Density of rank, faceted by group and period



Appendix 7.2.1.4. Descriptive DiD Group level means (rank)

Table: Summary Table: Mean Rank, Group Changes, and DiD

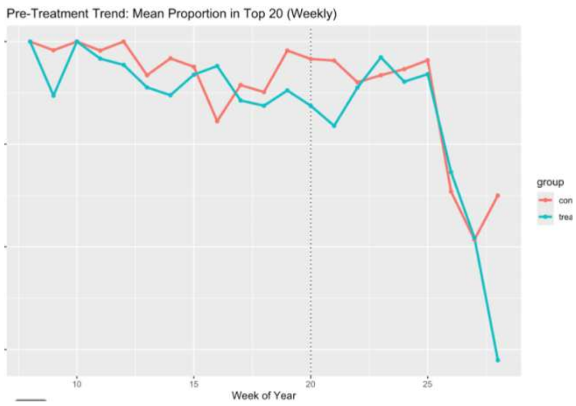
group	pre	post	delta
control	8.095	11.670	3.575
treatment	10.404	13.861	3.457
Treat - Control	2.309	2.192	-0.118

Ranking Dummies

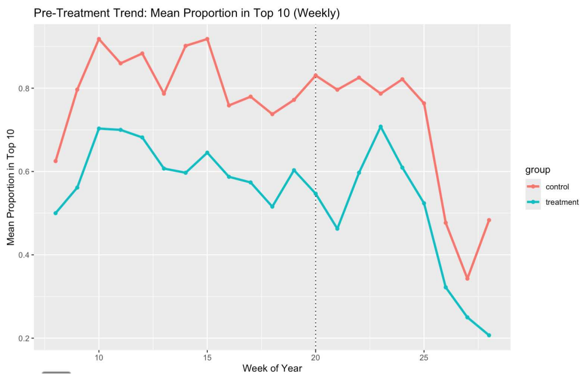
Appendix 7.2.1.5. Descriptive Summary: Rank Dummies

	dummy	mean	sd	min	max	median	IQR	n
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	top20	0.869	0.338	0	1	1	0	2280
2	top10	0.627	0.484	0	1	1	1	2280
3	top5	0.301	0.459	0	1	0	1	2280
4	top3	0.159	0.365	0	1	0	0	2280

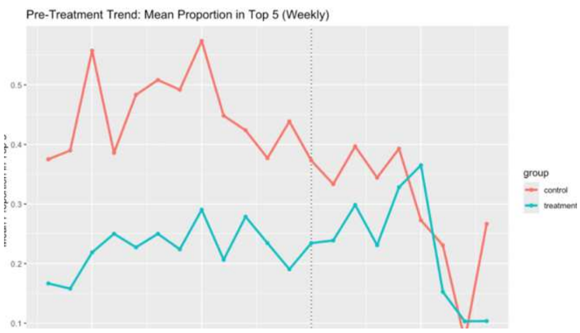
Appendix 7.2.1.6. Common Trend Top 20



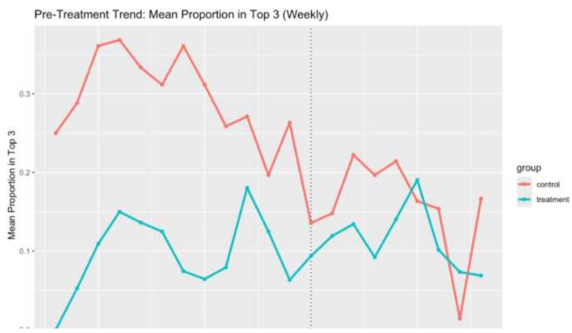
.Appendix 7.2.1.7 Common Trend Top 10.



Appendix 7.2.1.8. Common Trend Top 5



Appendix 7.2.1.9. Common Trend Top 3



Impressions

Appendix 7.2.1.10. Descriptive Summary: Rank

Table: Descriptive Summary: Impressions

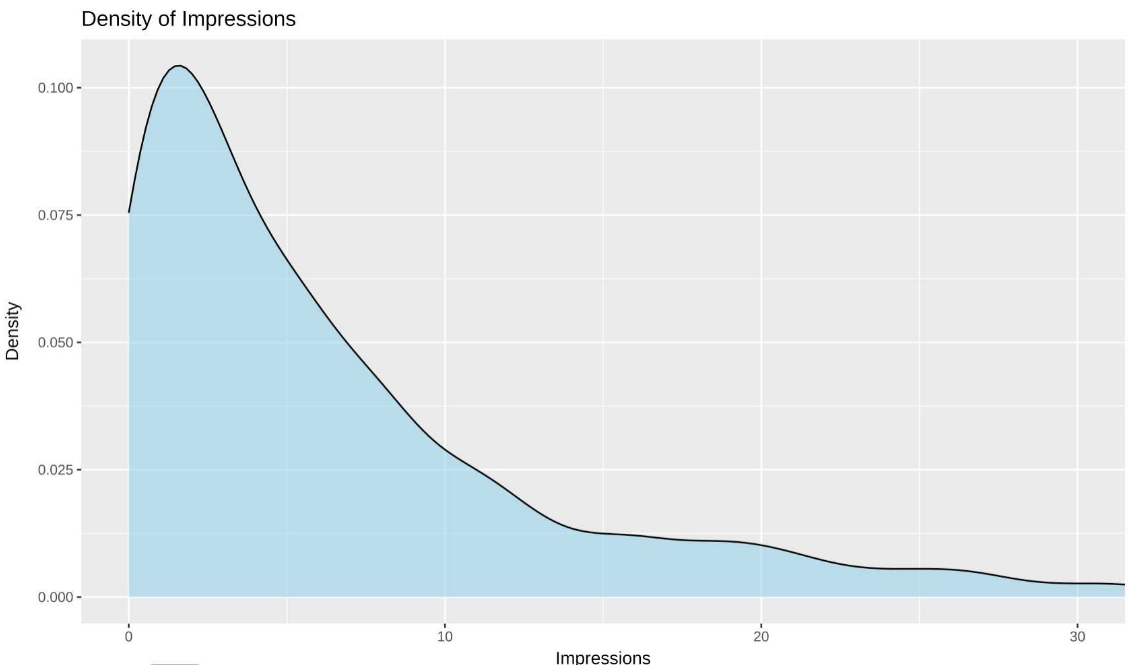
mean	sd	median	min	max	IQR	n
7.542	9.4	5	0	93	8	2280

Appendix 7.2.1.11. Descriptive Summary: Rank (by group and period)

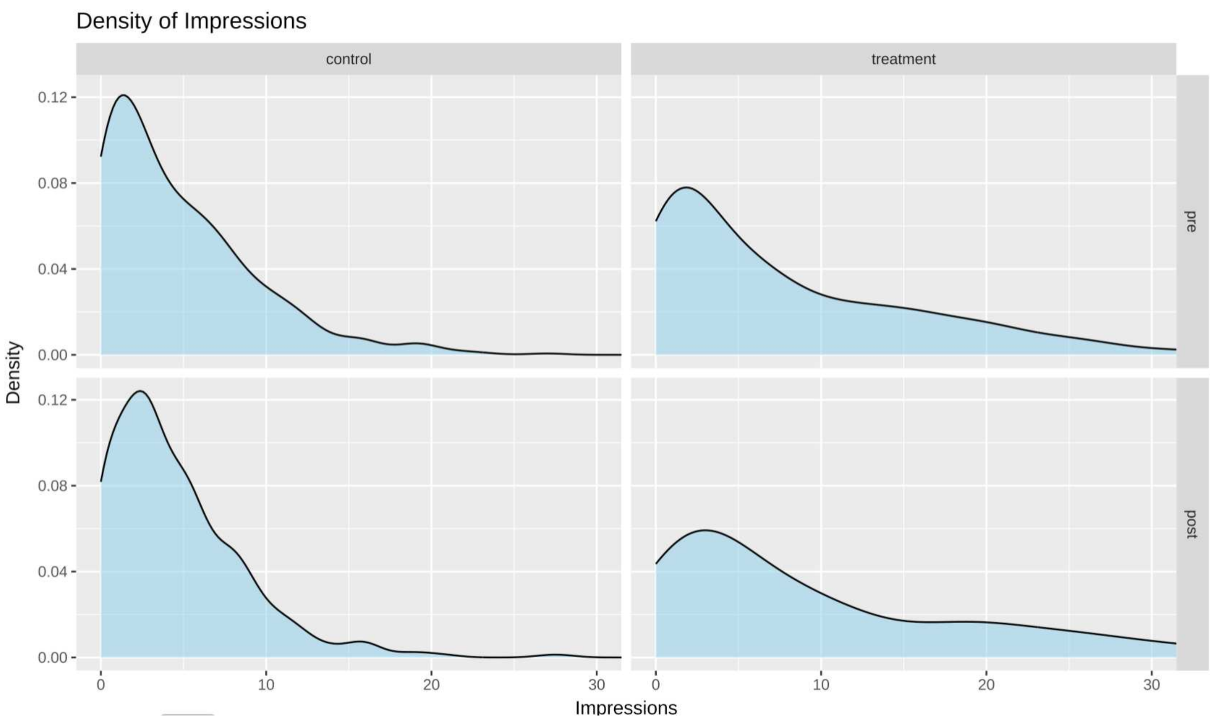
Table: Descriptive Summary: Impressions per Page by Group and Period

group	period	mean	sd	median	min	max	IQR	n
control	post	5.086	6.948	4	0	93	5	570
control	pre	4.798	4.564	4	0	27	6	570
treatment	post	11.774	13.006	7	0	83	15	570
treatment	pre	8.509	9.120	5	0	51	11	570

Appendix 7.2.1.12. Density of impressions



Appendix 7.2.1.13. Density of impressions, faceted by group and period



Appendix 7.2.1.14. Descriptive DiD Impressions

Table: DiD Table: Mean of Mean Daily Impressions per Page

group	pre	post	delta
control	4.80	5.09	0.29
treatment	8.51	11.77	3.26
Treat - Control	3.71	6.69	2.98

7.2.2 H2 - USER ENGAGEMENT (HUMAN LEVEL)

Clicks

Appendix 7.2.2.1. Descriptive Summary: Clicks

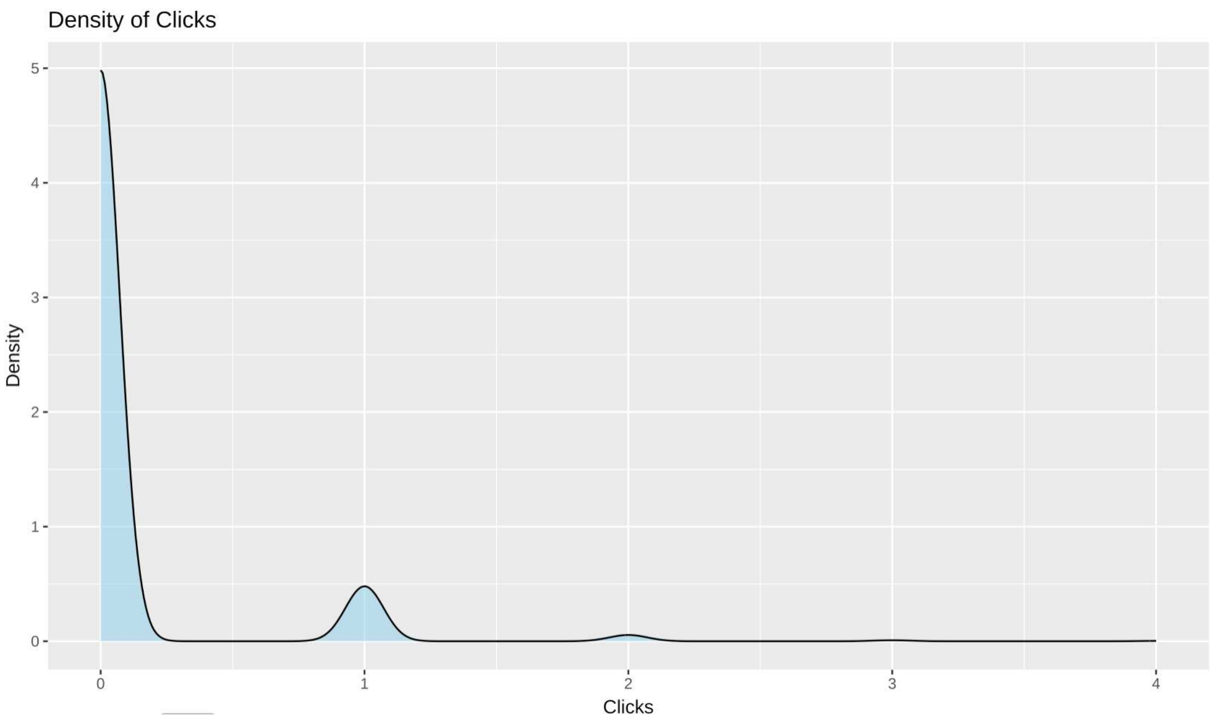
Table: Descriptive Summary: Clicks

mean	sd	median	min	max	IQR	n	n_NA
0.113	0.367	0	0	4	0	2280	251

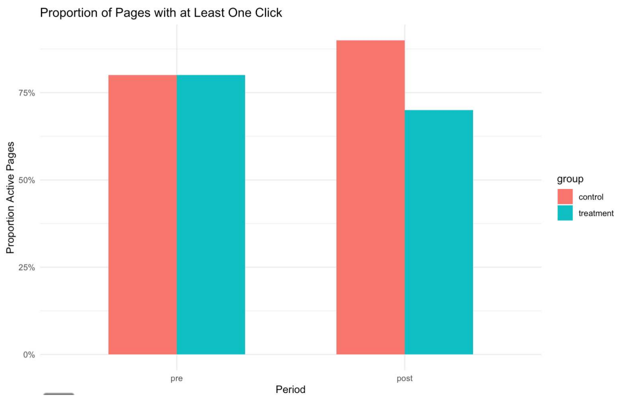
Appendix 7.2.2.2. Percentage of days with zero clicks faceted by group and period

	group	period	mean_prop_zero
	<fct>	<fct>	<dbl>
1	control	pre	0.880
2	control	post	0.914
3	treatment	pre	0.907
4	treatment	post	0.933

Appendix 7.2.2.3. Density of Clicks



Appendix 7.2.2.4 .Proportion of Pages with at least one click



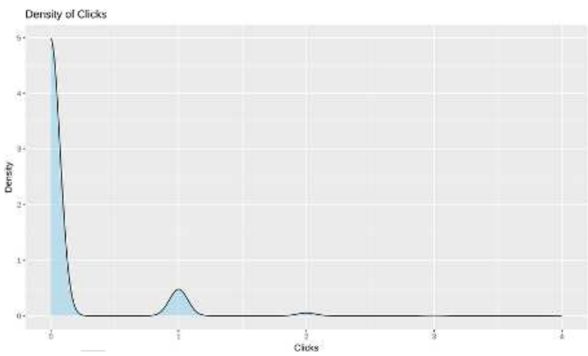
	group	period	prop_active
	<fct>	<fct>	<dbl>
1	control	pre	0.8
2	control	post	0.9
3	treatment	pre	0.8
4	treatment	post	0.7

Appendix 7.2.2.5. Descriptive DiD Clicks

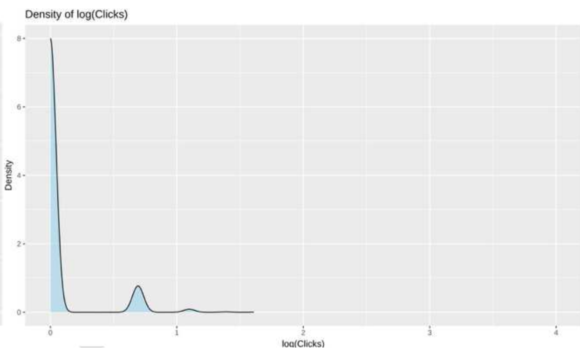
```
Table: Summary Table: Mean Clicks, Group Changes, and DiD

|group           |    pre|   post|  delta|
|:---------------|------:|------:|------:|
|control         |  0.139|  0.090| -0.049|
|treatment       |  0.110|  0.078| -0.032|
|Treat - Control | -0.029| -0.012|  0.017|
```

Appendix 7.2.2.6. Distribution before and after log-transformation (Clicks)

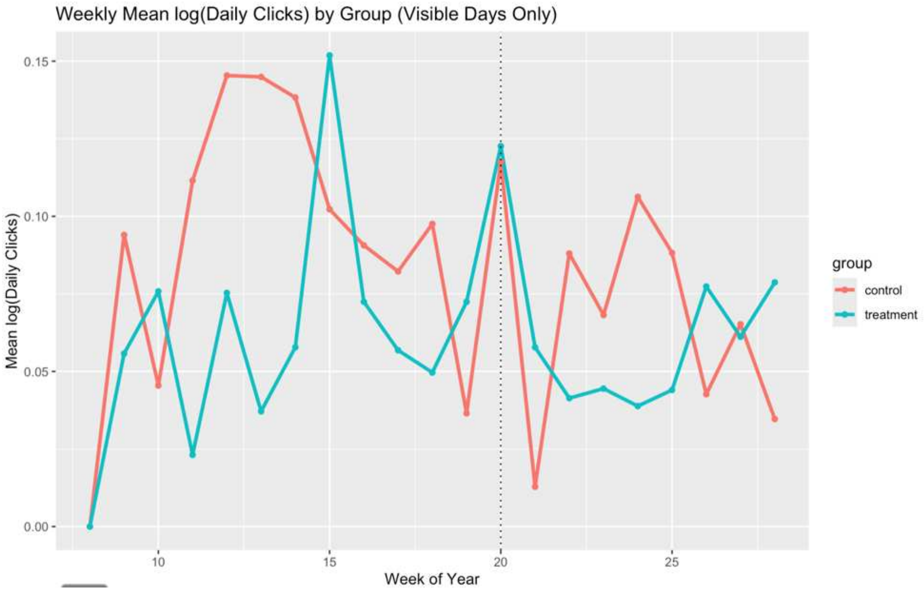


Before



After

Appendix 7.2.2.7. Common trend assumption: log (Clicks)



CTR

Appendix 7.2.2.8. Descriptive Summary: CTR

Table: Descriptive Summary: CTR

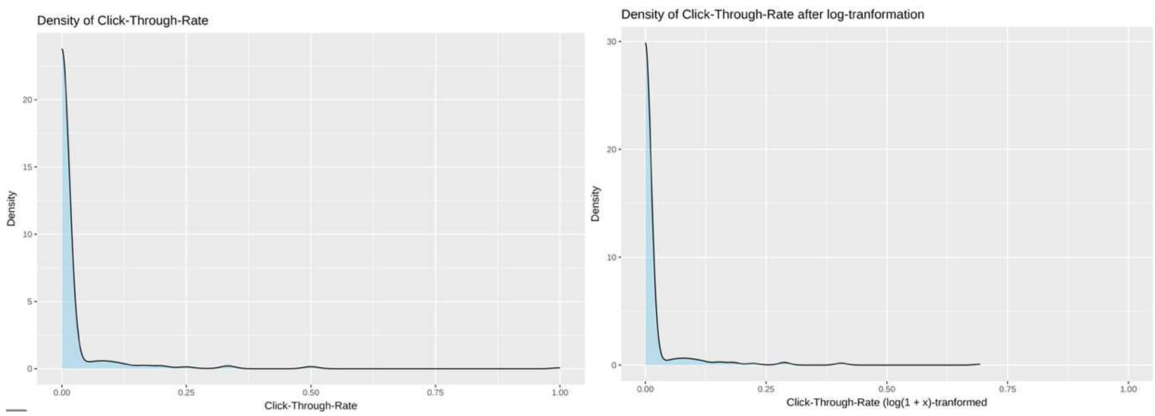
mean	sd	median	min	max	IQR	n	n_NA
0.017	0.077	0	0	1	0	2280	251

Appendix 7.2.2.9. Descriptive DiD Table (CTR)

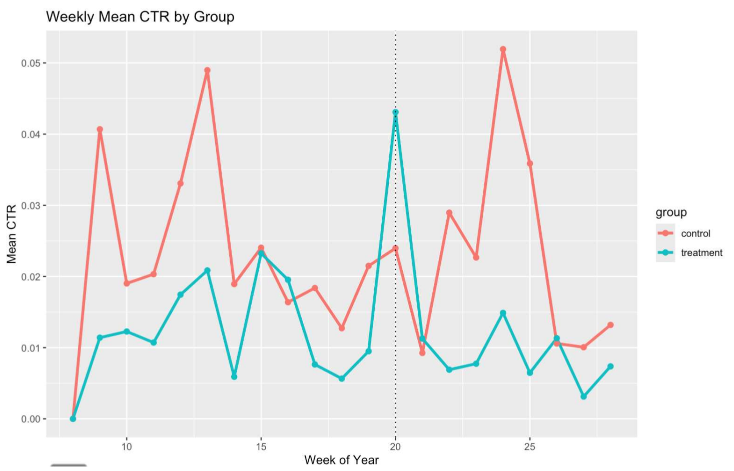
Table: Summary Table: Mean CTR, Group Changes, and DiD

group	pre	post	delta
control	0.0216	0.0214	-0.0002
treatment	0.0166	0.0079	-0.0087
Treat - Control	-0.0050	-0.0135	-0.0085

Appendix 7.2.2.10. Distribution before and after log-transformation (CTR)



Appendix 7.2.2.11. Common Trend assumption (CTR)



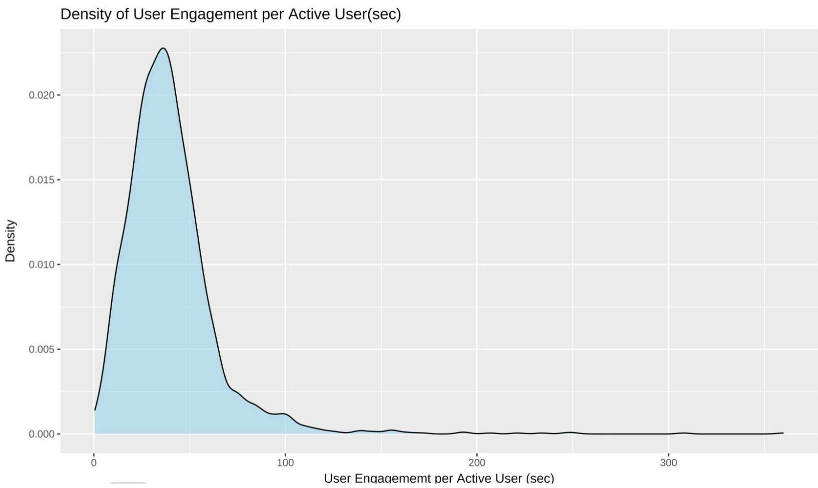
Engagement Time

Appendix 7.2.2.12. Descriptive Summary: Engagement Time

Table: Descriptive Summary: Engagement time

mean	sd	median	min	max	IQR	n	n_NA
40.027	25.479	36.444	0.5	360	23.646	2280	105

Appendix 7.2.2.13. Density of User Engagement

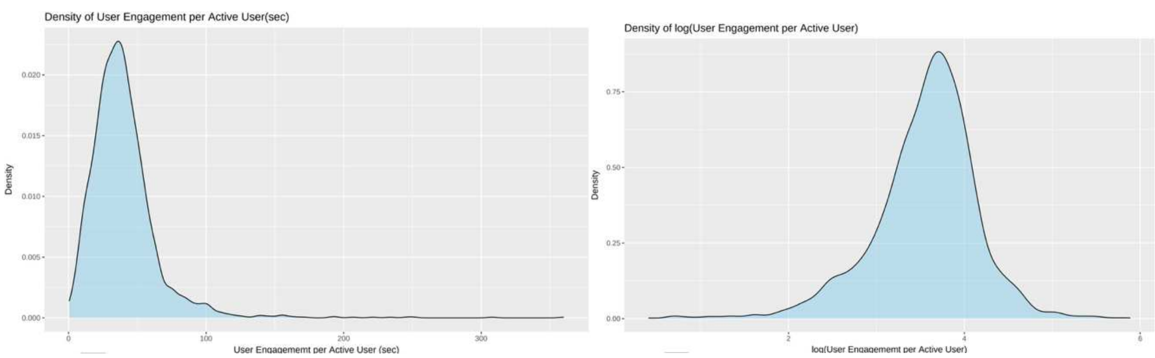


Appendix 7.2.2.14. Descriptive DiD: Engagement Time

Table: Summary Table: Mean Engagement Time per User, Differences, and DiD

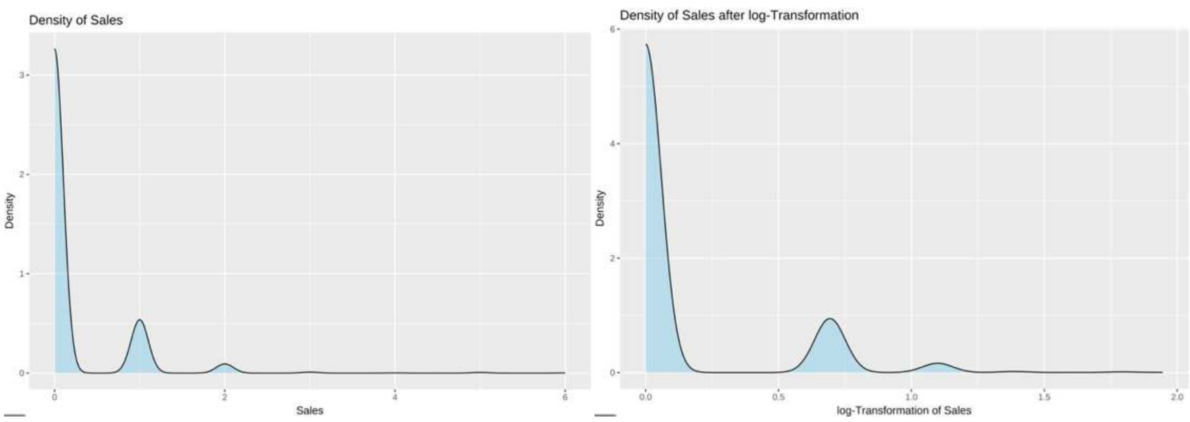
group	pre	post	delta
control	38.449	38.331	-0.118
treatment	40.252	42.428	2.176
Treat - Control	1.803	4.097	2.294

Appendix 7.2.2.15. Distribution before and after the transformation



7.2.3 H3 - COMMERCIAL IMPACT (BUSINESS LEVEL)

Appendix 7.2.3.1. Distribution before and after transformation



Before *After*

Appendix 7.2.3.2. Descriptive Summary: Sales

```
Table: Descriptive Summary: Sales

|  mean|     sd| median| min| max| IQR|    n|
|-----:|-----:|------:|---:|---:|---:|----:|
| 0.206| 0.529|      0|   0|   6|   0| 2280|
```

Appendix 7.2.3.3. Proportion of days with at least one sale per day

```
Table: Proportion of Days with at least one Sale

| n_pages| n_not_zero| prop_not_zero|
|-------:|----------:|-------------:|
|    2280|        379|         0.166|
```

Appendix 7.2.3.4. Average price per product

```
> overall_avg_price
[1] 25.87568
```

Appendix 7.2.3.5. . Proportion of days with at least one sale per day (faceted by group and period)

Table: Fraction of Pages Ever Receiving Nonzero CTR (by Group and Period)

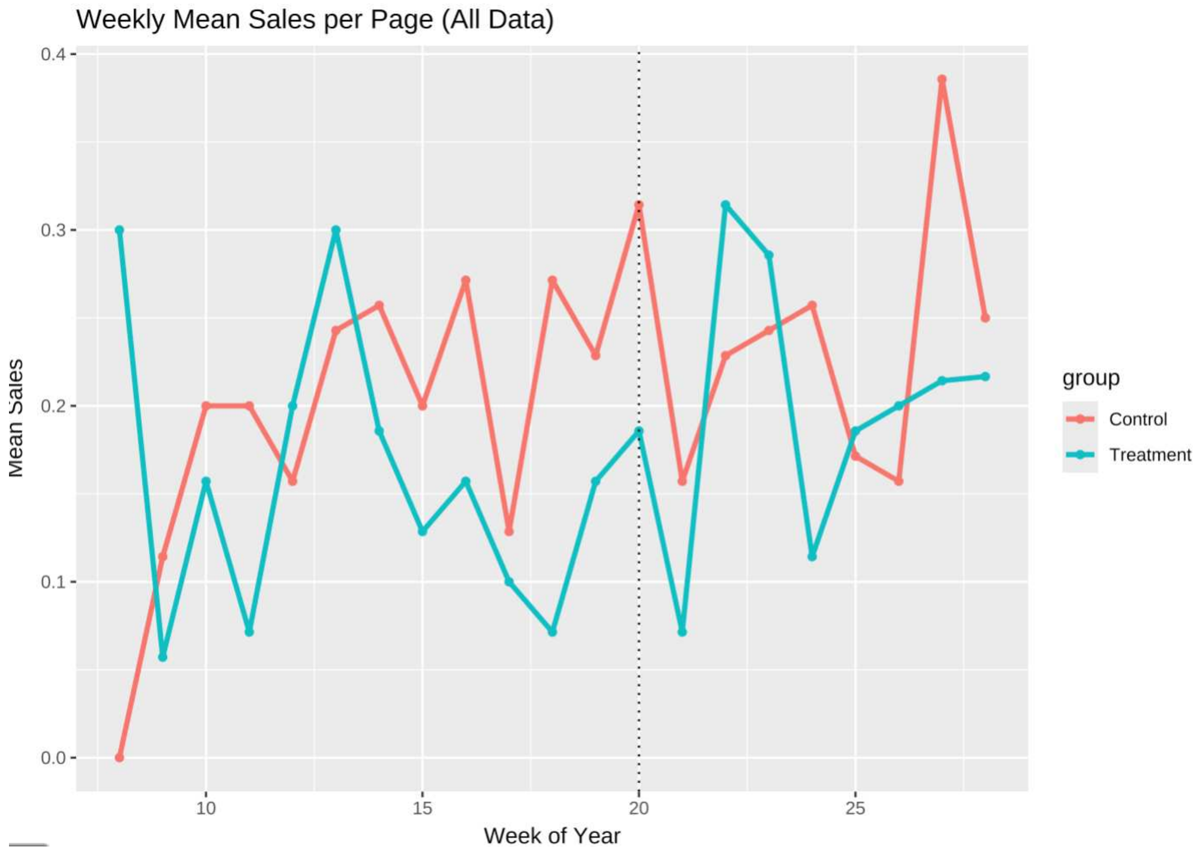
group	period	n_pages	n_ever_nonzero	prop_ever_nonzero
control	pre	10	8	0.8
control	post	10	9	0.9
treatment	pre	10	8	0.8
treatment	post	10	7	0.7

Appendix 7.2.3.6. Descriptive DiD: Sales

Table: Total Sales per Page, per Period, Group, and DiD

group	pre	post	delta
control	13.2	13.3	0.1
treatment	9.0	11.4	2.4
Treat - Control	-4.2	-1.9	2.3

Appendix 7.2.3.7. Common trend assumption



8 LIST OF REFERENCES

Adetayo, A., Aborisade, M., & Sanni, B. (2024). Microsoft Copilot and Anthropic Claude AI in education and library service. *Library Hi Tech News*. https://doi.org/10.1108/lhtn-01-2024-0002

Advanced Web Ranking. (2025, May). *Google Organic CTR History - Advanced Web Ranking*. Www.advancedwebranking.com. https://www.advancedwebranking.com/free-seo-tools/google-organic-ctr

Ahmad, U. F., Mahdee, J., & Abu Bakar, N. (2022). Search engine optimisation (SEO) strategy as determinants to enhance the online brand positioning. *F1000Research*, *11*, 714. https://doi.org/10.12688/f1000research.73382.1

Alhur, A. (2024). Redefining Healthcare With Artificial Intelligence (AI): The Contributions of ChatGPT, Gemini, and Co-pilot. *Curēus*, *16*(4). https://doi.org/10.7759/cureus.57795

Awa, H. O., Ojiabo, O. U., & Emecheta, B. C. (2015). Integrating TAM, TPB and TOE frameworks and expanding their characteristic constructs for e-commerce adoption by SMEs. *Journal of Science and Technology Policy Management*, *6*(1), 76–94.

Bhandari, R. S., & Bansal, A. (2018). Impact of Search Engine Optimization as a Marketing Tool. *Jindal Journal of Business Research*, *7*(1), 23–36. Sagepub. https://doi.org/10.1177/2278682117754016

Brantner, C., Karlsson, M., & Kuai, J. (2025). Sourcing behavior and the role of news media in AI-powered search engines in the digital media ecosystem: Comparing political news retrieval across five languages. *Telecommunications Policy*, *49*(5), 102952. https://doi.org/10.1016/j.telpol.2025.102952

Cardillo, A. (2025, June 5). *How Many Google Searches Are There Per Day? (March 2025)*. Exploding Topics. https://explodingtopics.com/blog/google-searches-per-day#how-many-google-searches-are-there-per-day

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2023, October 27). *Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review*. ArXiv.org. https://doi.org/10.48550/arXiv.2310.14735

Chen, J.-C., & Sylvain Sénéchal. (2023). The reciprocal relationship between search engine optimization (SEO) success and brand equity (BE): an analysis of SMEs. *European Business Review*, *35*(5), 860–873. https://doi.org/10.1108/ebr-02-2023-0045

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://utstat.utoronto.ca/brunner/oldclass/378f16/readings/CohenPower.pdf

Conductor. (2023a, October 9). *Ranking*. Conductor. https://www.conductor.com/de/academy/glossar/rankings/

Conductor. (2023b, October 9). *Ranking-Faktor*. Conductor. https://www.conductor.com/de/academy/glossar/ranking-faktor/

Cutler, K. (2024). *Search Marketing: A Strategic Approach to Seo and Sem*. Kogan Page.

Daoud, M. K., Sharabati, A.-A., Samarah, T., Alqurashi, D., Alfityani, A., Allahham, M., & Nasereddin, A. Y. (2024). Optimizing online visibility: A comprehensive study on effective SEO strategies and their impact on website ranking. *Journal of Infrastructure Policy and Development*, *8*(7), 4860–4860. https://doi.org/10.24294/jipd.v8i7.4860

DHB-MSFT. (2024, March 5). *Data, Privacy, and Security for Microsoft Copilot for Microsoft 365*. Learn.microsoft.com. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy

Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models*. Crc Press Taylor & Francis Group. https://reneues.wordpress.com/wp-content/uploads/2010/01/an-introduction-to-generalized-linear-models-second-edition-dobson.pdf

Du, Y., Ji, C., Xu, J., Wei, M., Ren, Y., Xia, S., & Zhou, J. (2025). Performance of ChatGPT and Microsoft Copilot in Bing in answering obstetric ultrasound questions and analyzing obstetric ultrasound reports. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-99268-2

Empl, P., & Pernul, G. (2021). A Flexible Security Analytics Service for the Industrial IoT. *Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*, 23–32. https://doi.org/10.1145/3445969.3450427

Enge, E., Spencer, S., & Stricchiola, J. C. (2023). *The Art of SEO : mastering search engine optimazion* (4th ed.). Beijing O'reilly August.

Erlhofer, S. (2020). *Suchmaschinen-Optimierung : das umfassende Handbuch* (10th ed.). Rheinwerk Verlag.

Eurostat. (2025, January). *Use of artificial intelligence in enterprises*. Ec.europa.eu. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises

Evans, M. P. (2007). Analysing Google rankings through search engine optimization data. *Internet Research*, *17*(1), 21–37. https://doi.org/10.1108/10662240710730470

FirstPageSage. (2023, September 7). *Google Click-Through Rates (CTRs) by Ranking Position in 2023*. First Page Sage. https://firstpagesage.com/reports/google-click-through-rates-ctrs-by-ranking-position/

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Sage.

Google. (2019). *What are impressions, position, and clicks? - Search Console Help*. Google Search Console Help. https://support.google.com/webmasters/answer/7042828#position

Google. (2023). *Ranking results – How Google Search works*. Google Search – Discover How Google Search Works; Google Search. https://www.google.com/intl/en_uk/search/howsearchworks/how-search-works/ranking-results/

Google. (2025, February 4). *A Guide to Google Search Ranking Systems | Google Search Central | Documentation*. Google Developers; Google. https://developers.google.com/search/docs/appearance/ranking-systems-guide

Google Analytics. (2023). *The difference between Google Ads Clicks, and Sessions, Users, Entrances, Pageviews, and Unique Pageviews in Analytics - Analytics Help*. Support.google.com. https://support.google.com/analytics/answer/1257084?hl=en#zippy=%2Cin-this-article

Google Analytics. (2024). *GA4 - Understand user metrics - Analytics Help*. Support.google.com. https://support.google.com/analytics/answer/12253918?hl=en

Google Search Central. (2025, June). *A Guide to Google Search Ranking Systems | Google Search Central | Documentation*. Google Developers. https://developers.google.com/search/docs/appearance/ranking-systems-guide

Govori, A., & Sejdija, Q. (2023). Future prospects and challenges of integrating artificial intelligence within the business practices of small and medium enterprises. *Journal of Governance and Regulation*, *12*(2), 176–183. https://doi.org/10.22495/jgrv12i2art16

Guelailia, R., & Bouziane, M. (2024). Enhancing Search Engine Optimization through Artificial Intelligence. *Beam Journal of Economic Studies*, *8*(2), 489–500. https://asjp.cerist.dz/en/article/253421

Harun, N., Syed, M., & Abdullah, F. (2024). Revolutionising costume design education: a case study on the impact of Microsoft Copilot. *Quality Assurance in Education*, *33*(2). https://doi.org/10.1108/qae-06-2024-0119

IFM Bonn. (2023). *Mittelstand im Überblick*. Www.ifm-Bonn.org; Institut für Mittelstandsforschung Bonn. https://www.ifm-bonn.org/statistiken/mittelstand-im-ueberblick/volkswirtschaftliche-bedeutung-der-kmu/deutschland

Iyelolu, V., Ebele, E., Idemudia, N. C., & Ignatius, T. (2024). Driving SME Innovation with AI solutions: Overcoming Adoption Barriers and Future Growth Opportunities. *International Journal of Science and Technology Research Archive*, *7*(1), 036–054. https://doi.org/10.53771/ijstra.2024.7.1.0055

Kateryna Osadcha, Justyna Szynkiewicz, & Chishti, M. S. (2024). Using Microsoft Copilot Chat in the Work of IT Educators: Pilot Study. *Norsk IKT-Konferanse for Forskning Og Utdanning*, *4*. https://doi.org/10.5324/nikt.6202

Kauppinen, R. T. (2024). *Exploring the role of generative artificial intelligence in digital content marketing* [Master's Thesis].

Knoth, N., Tolzin, A., Janson, A., & Jan Marco Leimeister. (2024). AI Literacy and its Implications for Prompt Engineering Strategies. *Computers and Education. Artificial Intelligence*, *6*, 100225–100225. https://doi.org/10.1016/j.caeai.2024.100225

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Cambridge University Press.

Kritzinger, W. T., & Weideman, M. (2017). Parallel search engine optimisation and pay-per-click campaigns: A comparison of cost per acquisition. *SA Journal of Information Management*, *19*(1). https://doi.org/10.4102/sajim.v19i1.820

Krol, K., Jaworska, M., & Goshchynska, D. (2024). Seo auditing using large language models as a key university marketing strategy component. *SCIENTIFIC PAPERS of SILESIAN UNIVERSITY of TECHNOLOGY ORGANIZATION and MANAGEMENT SERIES*, *207*. http://dx.doi.org/10.29119/1641-3466.2024.207.18

Kuznetsova, E., Makhortykh, M., Vziatysheva, V., Stolze, M., Baghumyan, A., & Urman, A. (2024). In generative AI we trust: can chatbots effectively verify political information? *Journal of Computational Social Science*, *8*(1). https://doi.org/10.1007/s42001-024-00338-8

Lammenett, E. (2025). Praxiswissen Online-Marketing. In *Springer eBooks* (10th ed.). Springer Nature. https://doi.org/10.1007/978-3-658-48516-0

Lembke, G., & Meil, C. (2022). *Smartes Marketing mit künstlicher Intelligenz*. Redline Verlag.

Lewandowski, D. (2021). *Suchmaschinen verstehen*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-63191-1

Luh, C.-J., Yang, S.-A., & Huang, T.-L. D. (2016). Estimating Google's search engine ranking function from a search engine optimization perspective. *Online Information Review*, *40*(2), 239–255. https://doi.org/10.1108/oir-04-2015-0112

Mou, A. J., Hossain, M. S., & Siddiqui, N. A. (2022). DIGITAL TRANSFORMATION IN MARKETING: EVALUATING THE IMPACT OF WEB ANALYTICS AND SEO ON SME GROWTH. *American Journal of Interdisciplinary Studies*, *03*(04), 61–90. https://doi.org/10.63125/8t10v729

Moz. (2015). *Ranking Factors Expert Survey*. Moz. https://moz.com/search-ranking-factors/survey-results?_gl=1

Nayak, P. (2022, February 3). *How AI powers great search results*. Google. https://blog.google/products/search/how-ai-powers-great-search-results/

Office Software. (2024). Statista. https://de.statista.com/outlook/tmo/software/produktivitaetssoftware/office-software/custom?token=Gr9PG8ZSSugoLMSI0Tzkpa2sa6hx4no3y-HS9xnCaf4qhZcd4XBfy9QbMyjfVP6iAeZJROlET1ohHckitzb5lPBUj48M7NzGVYcRh292dvwya7DnVs6a

Oldemeyer, L., Jede, A., & Teuteberg, F. (2024). Investigation of artificial intelligence in SMEs: a systematic review of the state of the art and the main implementation challenges. *Management Review Quarterly*. https://doi.org/10.1007/s11301-024-00405-4

Schorcht, S., Buchholtz, N., & Lukas Baumanns. (2024). Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1386075

Schultheiß, S., Sünkler, S., & Lewandowski, D. (2018). We Still Trust in Google, but Less than 10 Years Ago: An Eye-Tracking Study. *Information Research: An International Electronic Journal*, *23*(3). http://informationr.net/ir/23-3/paper799.html

Schwartz, B. (2010, November 11). *Key Takeaways From Google's Matt Cutts Talk At PubCon*. Search Engine Land. https://searchengineland.com/key-takeaways-from-googles-matt-cutts-talk-at-pubcon-55457

Sharabati, A.-A. A., Ali, A., Allahham, M. I., Hussein, A. A., Alheet, A. F., & Mohammad, A. S. (2024). The Impact of Digital Marketing on the Performance of SMEs: An Analytical Study in Light of Modern Digital Transformations. *Sustainability*, *16*(19), 8667–8667. MDPI. https://doi.org/10.3390/su16198667

Shenoy, A., & Prabhu, A. (2016). *Introducing SEO*. Apress. https://doi.org/10.1007/978-1-4842-1854-9

Sistrix. (2020, July 13). *Wieso (fast) alles, was du bislang über die Google CTR wusstest, nicht mehr stimmt - SISTRIX*. SISTRIX. https://www.sistrix.de/news/wieso-fast-alles-was-du-bislang-ueber-die-google-ctr-wusstest-nicht-mehr-stimmt/

Sistrix. (2021, July 28). *SERP Snippet Generator - Optimiere Title und Description*. SISTRIX. https://app.sistrix.com/de/serp-snippet-generator

Sistrix. (2024, May 8). *SISTRIX Sichtbarkeitsindex: Stärken & Grenzen*. SISTRIX. https://www.sistrix.de/sichtbarkeitsindex/staerken-grenzen/

Statista. (2024a). *Informationsquellen für neue Produkte in Deutschland 2024| Statista*. Statista. https://de.statista.com/prognosen/999856/deutschland-informationsquellen-fuer-neue-produkte

Statista. (2024b, October). *Werbung - Europa | Statista Marktprognose*. Statista. https://de.statista.com/outlook/amo/werbung/europa#digitale-werbeausgaben

Su, A.-J., Hu, Y. C., Kuzmanovic, A., & Koh, C.-K. (2010). How to Improve Your Google Ranking: Myths and Reality. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*. https://doi.org/10.1109/wi-iat.2010.195

Süß, Y. (2022). *E-Commerce Für Kleine und Mittelständische Unternehmen* (2nd ed.). Springer Vieweg.

Tepe, M., & Emekli, E. (2024). Assessing the Responses of Large Language Models (ChatGPT-4, Gemini, and Microsoft Copilot) to Frequently Asked Questions in Breast Imaging: A Study on Readability and Accuracy. *Cureus*. https://doi.org/10.7759/cureus.59960

The Economist. (2020, June 25). *The advertising business is becoming less cyclical—and more concentrated*. The Economist. https://www.economist.com/business/2020/06/25/the-advertising-business-is-becoming-less-cyclical-and-more-concentrated

Unternehmensregister. (2025). *Consolidated financial statements for the 2023 financial year of a German cosmetics company*. Unternehmensregister.de; Bundesanzeiger Verlag

GmbH. https://www.unternehmensregister.de/ureg/index.html. (The name of the company remains undisclosed for reasons of anonymisation) .

Urman, A., & Mykola Makhortykh. (2023). You are how (and where) you search? Comparative analysis of web search behavior using web tracking data. *Journal of Computational Social Science*, *6*(2), 741–756. https://doi.org/10.1007/s42001-023-00208-9

Usability.de. (2009). *Goodbye Golden Triangle*. Usability.de. https://www.usability.de/usability-user-experience/publikationen/goodbye-golden-triangle.html#download

Usmany, P., Rachmawati, R., Rembe, E., Sopacua, F., Santosa, T. A., Arifin, A. H., Fitria, A., & Suhardi, S. (2024). The Effectiveness Of Search Engine Optimization (SEO) In Marketing: A Meta-Anlysis Study. *Journal of Economic, Bussines and Accounting (COSTING)*, *7*(5), 807–811. Researchgate. https://doi.org/10.31539/costing.v7i5.11446

von Joerg, G., & Carlos, J. (2022). Design Framework for the Implementation of AI-based (Service) Business Models for Small and Medium-sized Manufacturing Enterprises. *Journal of the Knowledge Economy*, *14*. https://doi.org/10.1007/s13132-022-01003-z

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *ArXiv*. https://doi.org/10.48550/arxiv.2302.11382

Wohllebe, A., & Lagodka, C. (2024). Can ChatGPT Replace Human Content Writers in SEO Copywriting? An Empirical Comparison for Performance Measurement of Generative AI in Search Engine Optimization in E-Commerce. *Springer EBooks*, 303–313. https://doi.org/10.1007/978-3-658-45132-5_21

Zerres, C., & Israel, K. (2016). *Online-Marketing Nutzung bei klein- und mittelständischen Unternehmen*. Hochschule Offenburg. https://opus.hs-offenburg.de/frontdoor/deliver/index/docId/1268/file/Online-Marketing+-+Nutzung+bei+KMU+-+web.pdf

Zhang, S., & Cabage, N. (2016). Search Engine Optimization: Comparison of Link Building and Social Sharing. *Journal of Computer Information Systems*, *57*(2), 148–159. https://doi.org/10.1080/08874417.2016.1183447

Ziakis, C., Vlachopoulou, M., Kyrkoudis, T., & Karagkiozidou, M. (2019). Important Factors for Improving Google Search Rank. *Future Internet*, *11*(2), 32. mdpi. https://doi.org/10.3390/fi11020032

Zuze, H., & Weideman, M. (2013). Keyword stuffing and the big three search engines. *Online Information Review*, *37*(2), 268–286. https://doi.org/10.1108/oir-11-2011-0193