UNIVERSITY OF TARTU

School of Economics and Business Administration

Arina Zaviyalova, Irina Kalinina

THE COMPARATIVE STUDY OF TRADITIONAL AND AI EYE-TRACKING: PRACTICAL UTILITY OF AI RECOMMENDATIONS AND VISUAL ATTENTION PREDICTIONS IN DIGITAL BANNER ADVERTISING

Master's thesis

Supervisors:

Kristian Pentus, Junior Lecturer

Kerli Ploom, Junior Lecturer

Tartu 2025

We have compiled the paper independently. All opinions of other authors used in the preparation of the paper, as well as data from literary sources and elsewhere, are cited. The authors contributed equally to this work.

Contents

Contents........3
Introduction........6
1. Literature Review........8
  1.1. Traditional Neuromarketing Methods For Visual Attention Measurement........8
    1.1.1. Visual Attention in Digital Banner Advertising........8
    1.1.2. The Advantages and Limitations of Eye-Tracking Method in Marketing Research........11
  1.2. Artificial Intelligence in Visual Attention Prediction........13
    1.2.1. Machine Learning, Deep Learning and Convolutional Neural Networks........13
    1.2.2. AI Eye-Tracking Prediction Softwares........14
    1.2.3. Advantages and Limitations of Eye-Tracking AI Prediction Softwares........16
2. Materials and Methods........19
  2.1. Research Design........19
  2.2. Test Samples........22
  2.3. The AI Visual Attention Predictions and Recommendations........23
  2.3. The Digital A/B Test Campaign Methodology: Pilot and Final Studies........25
  2.4. The Eye-Tracking Study Methodology........27
3. Results........31
  3.1. The International Digital Campaign........31
  3.2. The Eye-Tracking Study........40
4. Dicussion........52
  4.1. Key Findings and Their Implications........52
  4.3. Limitations and Future Research Directions........53
  4.5. Overall Conclusion........54
Author Contributions........54
Funding........55
References........56
Abbreviations........66
Appendices........67
  Appendix 1. Full Overview of the Research Process........67
  Appendix 2. Non-enhanced (A) and Enhanced (B) Banners........68
    Appendix 2.1. Non-enhanced (A) Banners........68
    Appendix 2.2. Enhanced (B) Banners........68
  Appendix 3. AI-generated Visual Attention Predictions........69
  Appendix 4. Heatmaps Gathered From the Eye-tracking Study........70
  Appendix 5. Interview Questions of the Eye-tracking Study........71
    Appendix 5.1. Estonian Questions........71
    Appendix 5.2. English Questions........71
    Appendix 5.3. Russian Questions........72
  Appendix 6. Socio-demographic Data of the Interview Participants........74

Appendix 7. Digital Campaign Guidelines........ 75
Appendix 7.1. Pilot Digital Campaign Guidelines........ 75
Appendix 7.2. Final Digital Campaign Guidelines........ 75
Appendix 8. The Results of the Final International Digital Campaign........ 76
Appendix 9. The Descriptive Statistics of the Digital campaign........ 78
Appendix 9.1. Clarity Score by Country........ 78
Appendix 9.2. Focus Score by Country........ 78
Appendix 9.3. Click-through-rate by Country........ 78
Appendix 10. ANOVA Single-Factor Test........ 79
Appendix 10.1. ANOVA Test for Clarity Score Means By Country........ 79
Appendix 10.2. ANOVA Test for Focus Score Means By Country........ 79
Appendix 10.3. ANOVA Test for Click-through-rate Means By Country........ 80
Appendix 11. Regression Analysis between AI metrics and CTR by Country........ 81
Appendix 11.1. Regression Analysis between AI Metrics and CTR for All Countries... 81
Appendix 11.2. Regression Analysis between AI Metrics and CTR for Estonia........ 83
Appendix 11.3. Regression Analysis between AI Metrics and CTR for France........ 84
Appendix 11.4. Regression Analysis between AI Metrics and CTR for Latvia........ 86
Appendix 11.5. Regression Analysis between AI Metrics and CTR for Spain........ 87
Appendix 11.6. Regression Analysis between AI Metrics and CTR for Lithuania..... 88
Appendix 12. Mean TFD Values for PenosilB Banners Compared to the Overall Noise Environment........ 91
Appendix 13. Descriptive Statistics of Eye-Tracking Data........ 92
Appendix 13.1. Descriptive Statistics of A (original) Banners........ 92
Appendix 13.2. Descriptive Statistics of the B Banners........ 94
Appendix 14. Tests of normality........ 97
Appendix 15. Mann–Whitney U test........ 97
Appendix 16. Regression........ 98
Appendix 17. Correlations........ 99
Appendix 18. Traditional Eye-Tracking and AI Eye-Tracking Differences........ 101
Summary........ 102

Abstract: The objective of this study is to clarify the practical utility of AI eye-tracking softwares in predicting visual attention and improving the performance of digital banner advertisements. In order to achieve this objective, the AI eye-tracking software was utilised to predict attention and obtain suggestions for improving the digital banner performance. To clarify the impact of these suggestions, an international digital campaign was executed, employing both non-optimised and optimised banners (A/B testing) . In order to clarify the accuracy of attention prediction, an eye-tracking study was conducted, complemented by semi-structured interviews. The findings of the digital campaign analysis demonstrate that AI design quality metrics are generally not strong predictors of click-through rate, though a few strong correlations were revealed for some countries. The findings of the eye-tracking study indicate that AI predictions exhibited modest alignment with human gaze, yet demonstrated limitations in modelling top-down attention.

Keywords: eye-tracking, AI, neuromarketing, visual attention, digital campaigns
CERCS codes: S191, S186

Introduction

The ability to capture and sustain visual attention is of crucial importance in the realm of marketing with 83% of information processing relying on visual cues as evidence (McMains & Kastner, 2009). Moreover, increased attention to an advertisement has shown to increase a likelihood of developing a positive attitude towards the brand, thus enhancing the purchase intention (Krajbich et al., 2010; Lee & and Ahn, 2012). Consequently, the accurate measurement of visual attention is essential for the evaluation of advertising performance, and a valid prediction of such phenomena would serve as an important advantage for making more data-driven decisions in terms of design and marketing strategy (Kohavi et al., 2020).

One of the most commonly used methods in marketing studies for measuring visual attention has been eye-tracking (González-Mena et al., 2022; Pentus et al., 2020). However, conventional eye-tracking systems are often associated by businesses with high costs and complexity (Winterhalter et al., 2024), thereby creating a gap between academic neuromarketing research and corporate application (Baños et al., 2020; Brenninkmeijer et al., 2019; Lin et al., 2018). Meanwhile, the development of machine learning paved the way for the widespread adoption of AI-based solutions in neuromarketing, including for visual attention prediction (Šola et al., 2024). The adoption of machine and deep learning techniques combined with big data derived from eye-tracking (Mizrahi, 2021, p. 14-15), has led to the high accuracy of 97-99% (Šola et al., 2024), offering cost-effective alternative to expensive traditional neuromarketing tools (Ahmed et al, 2022; Juárez-Varón et al., 2024; Kusá & Beličková, 2023; Šola et al., 2024). While there are still not many comparative studies conducted on this topic, AI eye-tracking softwares have demonstrated results comparable to traditional eye-tracking data (Juárez-Varón et al., 2024; Kusá & Beličková, 2023). However, as all the previous studies on the topic of AI eye-tracking softwares were limited to a single cultural or regional context (Juárez-Varón et al., 2024; Keresteš et al., 2024; Šola et al., 2025), further research is needed on the performance of AI in real international campaigns (Alam et al., 2021).

The objective of this study is to clarify the practical utility of AI eye-tracking software in predicting visual attention and improving the performance of digital banner advertisements. The present study employs two distinctive methodologies: a digital campaign experiment (A/B testing) and an eye-tracking study (comparative study of AI and traditional eye-tracking). In order to achieve the objective, the AI eye-tracking software, Attention Insight, was utilised to predict attention and obtain suggestions for improving the banner designs. To clarify the impact of AI recommendations, the international Google Ads campaigns were executed, employing both non-optimised and optimised banners. The digital campaigns have operated in the following countries: Estonia, Latvia, Lithuania, Spain and France. To clarify the accuracy of visual attention prediction, an eye-tracking study was conducted with the same stimuli, complemented by semi-structured interviews. The following tasks were assigned to achieve the objective:

- Provide an overview of factors influencing visual attention of digital banners.
- Provide an overview of advantages and limitations of traditional eye-tracking.
- Provide an overview of neuromarketing AI eye-tracking software technology.
- Provide an overview of the research gaps and limitations of neuromarketing AI eye-tracking softwares from previous studies.
- Work out a methodology for clarifying the impact of AI design improvement suggestions on the design performance.
- Work out a methodology for clarifying the validity of AI visual attention predictions.
- Clarify the validity of AI design improvement suggestions for improving the click-through-rate of digital banner designs.
- Clarify the correlation between AI design quality metrics and click-through-rates of digital banner designs.
- Describe the level of similarity between the AI attention heatmaps and the lab eye-tracking data.
- Clarify the correlation between eye-tracking results and click-through-rates of digital banner designs.

The manuscript is written in the article format according to the MDPI Behavioral Sciences (ISSN 2076-328X) peer-reviewed journal instructions (MDPI, 2025). The paper is divided into two sections: literature review and empirical study. In theoretical research, we provide an overview of traditional eye-tracking, factors influencing visual attention, AI applications in marketing research and the technology of neuromarketing AI eye-tracking softwares. The empirical study is divided into two parts: international digital campaign and eye-tracking study. This research contributes to the field of neuromarketing by providing valuable insights into the advantages and limitations of AI eye-tracking for visual attention prediction and design A/B testing. It will also offer practical guidance for international companies seeking to optimise their design processes. Furthermore, this study can serve as a stepping stone for future research on the development and application of AI in understanding and influencing human behavior.

1. Literature Review

1.1. Traditional Neuromarketing Methods For Visual Attention Measurement

1.1.1. Visual Attention in Digital Banner Advertising

In digital advertising, capturing visual attention is critical for both advertisers and consumers. For businesses, attention determines whether a banner ad effectively communicates its message and drives user engagement, influencing key performance indicators such as click-through rate (CTR) and conversion (Wedel & Pieters, 2008). For consumers, attention shapes perception, guides memory, and affects brand attitudes—even during short or unintentional exposures (Dreze & Hussherr, 2003; Briggs & Hollis, 1997).

Digital banners remain one of the most commonly used advertising formats due to their flexible integration into diverse web environments (Peker et al., 2021). While banners can increase brand awareness and purchase intention, their limited information content means that attracting initial attention is especially important (Pharr, 2004; ComScore Networks, 2008). CTR is commonly used as a proxy for ad effectiveness and is directly linked to user interest and cost efficiency in campaign planning (Chandon et al., 2003; Iankovets, 2023).

Because attention precedes clicking, understanding where users look has become essential for optimizing design. Traditional eye-tracking methods allow researchers to measure gaze behavior in detail, while recent advances in AI-based visual attention prediction provide scalable and cost-effective alternatives (Wedel & Pieters, 2008; Grewal et al., 2021). The following sections provide an overview of these two approaches and their relevance for digital banner design.

Visual attention plays a central role in evaluating the effectiveness of banner advertising. Studies show that the more attention consumers pay to an advertisement, the more likely they are to develop a positive attitude toward both the ad and the brand, increasing their purchase intention (Krajbich et al., 2010; Lee & and Ahn, 2012). Therefore, accurately measuring visual attention is essential for assessing the true impact of advertising.

Researchers have found that social media ads increase visual attention from the target audience (Muñoz-Leiva et al., 2019), especially when generate public engagement or when celebrities are used as brand representatives (Abell & Biswas, 2023; Wang et al., 2020; Zahmati et al., 2023). Furthermore, empirical evidence indicates that the inclusion of human faces in banner advertisements elicits increased user visual attention and enhances banner efficacy (Flores et al., 2014). Several studies have identified that the effect is stronger when the brand is familiar (Flores et al., 2022; Mandolfo, Di Dalmazi & Lamberti, 2022; Oliveira & Giraldi, 2019).

Based on consumer visual behavior, purchase decisions can depend on data – that is, when making a choice, participants in the study paid more attention to verbal information (Sielicka-Różyńska et al., 2021), while in other situations, graphical information received more attention than text (Pawar et al., 2023; Wu & Li, 2021; Zhang et al., 2023). Numerous studies have demonstrated that banner advertisements incorporating images attract greater consumer attention compared to text-only advertisements (Goodrich, 2011; Pieters & Wedel, 2004). However, it was also found that brand names elicit significantly greater visual attention than text or pictures (Wedel & Pieters, 2000). In a subsequent study, Pieters et al. (2002) found that attention to various ad elements had a positive effect on brand memory, an effect that was strongest for brand-related elements within an ad. Pieters and Wedel (2004) identified three primary components of banner advertisements that significantly impact visual attention: brand elements, pictorial elements, and text elements. Thus, logos, sizes, objects, and graphic elements should be centrally located and not placed in the background (García-Madariaga et al., 2019; Jiang, 2019); and slogans must be clear and promote communication (Lourenção et al., 2020; Şik & Soba, 2021). Variations in the size of advertising elements influence their salience and informative value, thereby impacting the attention directed at them (Peschel & Orquin, 2013).

Another factor influencing attention is the visual complexity of banners. Previous research on visual complexity has revealed negative correlation between visual complexity and attention, attitude and recognition (Orth & Crouch, 2014; Tuch et al., 2009). In addition, it was found that while complexity does not influence overall attention to the ad, it does negatively impact attention to brand information (Pieters et al., 2010). The Pieters et al., 2010 theory of visual complexity consists of two distinct groups of principles that help to improve the designs of advertisements: the design complexity and the brand identification difficulties The design complexity factors include: 1.Quantity of objects, 2. Irregularity of objects, 3. Dissimilarity of objects, 4. Detail of objects, 5. Asymmetry of object arrangement, and 6. Irregularity of object arrangement (Ibid.). The brand identification difficulty factors include: 1. Low brand contrast, 2. Small relative brand size, 3. Brand masking, and 4. Heterogeneity of brand background (Ibid.). The above-mentioned methodology established a foundation for a separate metric for defining the quality of banner designs, which has already been used in several studies (Bočaj & Ahtik, 2023; Liu et al., 2018).These findings reinforce the theoretical framework by Pieters et al. (2010), demonstrating that reducing visual complexity—by simplifying object quantity, enhancing brand visibility, and minimizing layout irregularity—can significantly enhance visual attention to brand-related elements in banner advertisements.

Regarding the main predictors of consumer behavior identified in studies, significant factors include ethnic and cultural differences (Ploom et al., 2020), age differences (Mičík & Kunešová, 2021), gender differences (Hamelin et al., 2022; Luo et al., 2022; Sargezeh et al., 2019; Yarosh, Kalkova & Reutov, 2021), brand loyalty and familiarity (Garczarek-Bąk et al., 2021; Kim & Lee, 2021; Peker et al., 2021). Similarly, Drèze and Hussherr (2003) found that while older and younger individuals view a similar number of areas, older adults tend to fixate on more of these areas, resulting in longer overall fixation durations (Drèze & Hussherr, 2003). It is important for researchers using quantitative models to control for variables that may affect the results, such as gender, age, ethnicity, socioeconomic, and personality traits.

Building on these findings, the present study adopts specific theoretical components to guide both AOI construction and the interpretation of AI-generated attention predictions. Based on the framework by Pieters and Wedel (2004), three key elements - brand, pictorial, and text components - have been shown to significantly shape visual attention. These elements also serve as primary drivers of brand recall (Wedel & Pieters, 2000; Pieters et al., 2002). Therefore, the following components were defined as Areas of Interest (AOIs) in this research: logo, product name, product image, and slogan.

To further evaluate visual attention patterns, the visual complexity theory developed by Pieters et al. (2010) was used to interpret the AI's design recommendations. This theory distinguishes between design complexity (e.g., number and irregularity of objects) and brand identification difficulties (e.g., low contrast, small brand size), offering a structured lens for understanding how cluttered or minimalistic layouts influence user attention.

In addition, attention allocation may vary based on individual and demographic characteristics, such as age (Mičík & Kunešová, 2021), ethnic background (Ploom et al., 2020), and especially brand familiarity, which has been shown to enhance attention and recall (Garczarek-Bąk et al., 2021; Kim & Lee, 2021). These factors are taken into account in the empirical part of this thesis, where cross-cultural and brand-related influences are examined alongside visual attention data.

1.1.2. The Advantages and Limitations of Eye-Tracking Method in Marketing Research

In recent years, the integration of neuroscience has significantly increased within management and business domains, particularly in marketing (Modi & Singh, 2023; Zhang et al., 2023). This context underscores the emergence of neuromarketing as a new and evolving field that combines psychology, economics, technology, and neuroscience to study consumer perception and behavior (Kajla et al., 2023, Mo, Yang, & Hu, 2023) and understand how psychological, physiological, and emotional phenomena influence purchasing decisions (Pawar et al., 2023; Yu, Droulers, & Lacoste-Badie, 2022). Ultimately, neuromarketing aims to provide insights into consumer decision-making by analyzing attention, emotional responses, and memory, thereby helping businesses optimize advertising and product design for better engagement and long-term effectiveness (Hamalin & Harcar, 2020; Mashrur et al., 2022; Rodrigues et al., 2022; Tüfekci & Akbiyik, 2023). One key advantage of using neuroscience techniques in marketing is the ability to measure unconscious responses to marketing stimuli, which traditional self-report measures may not capture effectively (Plassmann et al., 2012).

One of the most commonly used methods in neuromarketing research is eye-tracking (González-Mena et al., 2022; Pentus et al., 2020), which enables the analysis of visual attention patterns by identifying which elements capture the viewer's attention the most and lead to a choice or preference (Peschel, Orquin, & Loose, 2019; Zuschke, 2020). Eye-tracking is a neuromarketing method that records and tracks eye movements to understand what individuals focus on when interacting with various static or dynamic banners (Duchowski, 2017; Holmqvist et al., 2011). Eye-tracking is a method used to monitor eye movements in order to analyze respondents' visual attention to both static and dynamic stimuli. This can be conducted using either eye-tracking glasses mounted on the respondent's head or stationary systems positioned under a monitor (Duchowski, 2017; Holmqvist et al., 2011; Kondak, 2023). Eye-tracking provides a direct insight into how advertising messages are processed, reflecting where and how a person directs their gaze (González-Mena et al., 2022; Liu et al., 2023). The technology is adaptable to various settings, from controlled laboratory environments to more naturalistic conditions (Kondak, 2023). Key eye-tracking metrics include fixations, which represent the momentary pauses of the gaze, and saccades, which are rapid movements between fixations.

Eye-tracking enables both qualitative analysis, through visualizations such as heatmaps, gaze paths, and fixation maps, and quantitative analysis, including metrics like time to first fixation and total fixation duration (Kondak, 2023; Smith & Johnson, 2020). Crucially, these data reflect actual user behavior rather than relying solely on self-reported responses (Chen & Liu, 2021). This objectivity makes the method particularly valuable in marketing, where understanding unconscious consumer behavior is essential.

The digital sector, including e-commerce, online advertising, social networks, and websites, is one of the primary areas where eye-tracking has been studied. These studies highlight the potential of eye-tracking to analyze consumer attention in online environments (Bigne et al., 2021; Jiang, 2019; Küçün & Güler, 2021; Mañas-Viniegra et al., 2019; Muñoz-Leiva et al., 2019; Monica et al., 2019; Peker et al., 2021; Rúa-Hidalgo et al., 2021). In marketing communication, eye-tracking supports a range of applications that all aim to enhance the effectiveness of visual content. It allows marketers to analyze consumer attention by revealing which parts of ads, websites, or packaging attract the most gaze (Kondak, 2023). These insights are also instrumental in optimizing the layout and design of visual materials, helping to create more intuitive and appealing user experiences (Patel & Gupta, 2019; Lee & Park, 2018). Additionally, eye-tracking enables testing of campaign effectiveness by showing which visual elements succeed in capturing attention and which may require refinement (Chen & Liu, 2021; Kondak, 2023). Studies has shown that eye-tracking and its application as a marketing research tool has helped researchers, advertisers, and other industry professionals identify consumer behavior and improve customer experience when viewing products and services, making their campaigns' imagery more attractive and closer to the interests of the consumer (Neves Pereira et al., 2024).

Eye-tracking offers several advantages that make it a valuable tool in marketing research. It enables the objective and real-time measurement of visual attention, providing insight into which design elements or advertising components attract the viewer's gaze (Wedel & Pieters, 2008; Lohse, 1997; Orquin & Mueller Loose, 2013). Eye-tracking captures where and for how long viewers focus on specific areas, using quantitative metrics such as fixation count, duration, and sequence (Holmqvist et al., 2011). Additionally, visual outputs such as heatmaps and gaze plots aid in the intuitive interpretation of attention patterns (González-Mena et al., 2022). The method is adaptable to both controlled laboratory settings and naturalistic environments, and its application spans across online platforms, e-commerce, packaging design, and televised advertising (Chae & Lee, 2013; Bojko, 2013).

However, traditional eye-tracking also presents several limitations. First, the required hardware and software can be expensive and demand specialized expertise. Second, gaze data alone does not explain the cognitive or emotional reasons behind attention patterns, making it essential to combine eye-tracking with complementary qualitative methods (Holmqvist et al., 2011). Third, conducting large-scale studies may be time-consuming and logistically challenging due to the need for individual data collection (Zhao, Liu, & Yu, 2020). Moreover, the artificial nature of lab settings can limit the ecological validity of findings and may not fully reflect natural consumer behavior (Küçün & Güler, 2021).

1.2. Artificial Intelligence in Visual Attention Prediction

1.2.1. Machine Learning, Deep Learning and Convolutional Neural Networks

The term "artificial intelligence" (AI) refers to a broad field of computer science focusing on the development of intelligent models/agents for automating processes that typically require human cognitive abilities (Bellman, 1978; Carrozza et al., 2019; Russell et al., 2010). These processes include learning, problem solving, environment perception, and decision making (Juárez-Varón et al., 2024). AI-powered solutions are widely spread in the the field of consumer behaviour prediction, customer segmentation, and personalised marketing (Davahli et al., 2020; Jupalle et al., 2022). Complex AI algorithms utilising machine learning and deep learning techniques for training have gained popularity due to their ability to analyse extensive datasets and identify patterns (Ibid.). The present study concentrates on one particular branch of AI: Convolutional Neural Networks (CNNs or ConvNets). These are deep learning models (DL) that are frequently employed for image recognition and visual attention prediction (Juárez-Varón et al., 2024).

By combining gaze data with artificial intelligence, researchers have explored how to predict people's reactions to marketing materials (Gonzalez Viejo et al., 2018; Matsumoto et al., 2011; Pappas et al., 2020). This approach offers great potential in advertising and consumer behavior analysis (Abdessalem et al., 2019; Lallé et al., n.d.; Lu & Jia, 2015; Pappas et al., 2020). Human behavior analysis is also a major application area, with research covering computer use, decision-making processes, and moral choices. Eye movement data has provided valuable insights into everyday activities, social interactions, and even the assessment of surgical skills (Gonzalez Viejo et al., 2018; Matsumoto et al., 2011; Pappas et al., 2020). Such studies offer multifaceted information that can be applied in practice, such as improving user experience or workplace environments.

The main and most promising branch of AI is machine learning (ML) (Juarez et al., 2020). Machine learning has enabled AI models to analyse and generalise problem data, as well as to make data-driven decisions based on the training dataset or examples (Ibid.;

(Kalayci et al., 2019; Kohavi and Provost, 1998). The first one to coin the term "machine learning" and to build one of the earliest ML models was Arthur Samuel (1959) with his self-learning checkers-playing program. Later there emerged a more advanced technique for training AI models - deep learning (DL). DL models have a multi-layered architecture enabling them to capture the patterns within extensive and intricate datasets (Pham et al., 2023; Awad & Khanna, 2015). The advent of the World Wide Web in the 1990s followed by the increased availability of large datasets in the 2000s have led to the increased popularity of DL and ML models in various fields that require forecasting and data analysis (Russell et al., 2010). Deep learning is widely used for such complex tasks as image, sound and text recognition, and natural language processing (Tran et al., 2019).

Significant advances in image recognition have been achieved through the deployment of one specific type of DL models, CNNs (Juárez-Varón et al., 2024). Inspired by Hubel and Wiesel's early biological work on the hierarchical information processing structure of the cat's visual cortex, LeCun et al. (2015) proposed and developed the first significant CNN model, LeNet (LeCun et al., 2015). LeCun et al. (2015) used the backpropagation algorithm and stochastic gradient descent, thus laying the foundation for CNNs in image recognition (Jiang, 2024). At present, CNN is the most commonly used DL architecture in image classification and image processing, including EEG tasks (Craik et al., 2019). These models have gained popularity due to their ability to handle raw data, enhance the speed and accuracy while demanding reduced computational resources compared to other deep neural networks (Géron, 2019; Jiang, 2024; Juárez-Varón et al., 2024; Schirrmeister et al., 2017). Despite the evident potential of AI eye-tracking in other domains, its application in the field of advertising remains comparatively restricted, particularly in the domains of banners and broader marketing materials.

1.2.2. AI Eye-Tracking Prediction Softwares

The emergence of neuromarketing in 2002 (Fisher et al., 2010; Iloka & Onyeke, 2020) coincided with the rise in popularity of machine learning (ML) models. The convergence of these two fields paved the way for the widespread adoption of AI-based solutions in neuromarketing, first for EEG classification (Craik et al., 2019), customer behaviour prediction (Alimardani & Kaba, 2021), and then for visual attention prediction (Šola et al., 2024). The convergence of neuromarketing and AI, in particular the employment of advanced eye-tracking prediction softwares, offers valuable insights into human behaviour across a variety of scientific and applied domains. First eye-tracking AI prediction softwares were based on a theoretical understanding of how the brain works, while modern solutions combine big data derived from eye-tracking with machine and deep learning techniques (Mizrahi, 2021, p. 14-15). It led to the significant improvements in the accuracy, thus offering cost-effective alternatives to expensive traditional neuromarketing tools (Ahmed et al, 2022; Juárez-Varón et al., 2024; Kusá & Beličková, 2023; Šola et al., 2024). Traditional eye-tracking systems can be expensive and complex (Winterhalter et al., 2024), while AI-powered eye-tracking predictions can achieve 97-99% accuracy (Šola et al., 2024; UAB Attention Insight, 2025b) and provide real-time personalised assistance (K. Sharma et al., 2020). AI-driven solutions are undergoing rapid evolution, providing sophisticated tools for the optimisation of design and the analysis of consumer engagement. Thus, AI eye-tracking has a significant potential for enhancing the performance of marketing materials.

One of the pioneers in the market of neuromarketing eye-tracking AI prediction softwares was Neurons Inc., a Danish startup founded in 2013 by neuroscientist Thomas Z. Ramsøy and Majken Z. Ramsøy (Neurons Inc., 2025-a). Šola et al. (2025) study has shown promising results with their ML-powered software Predict (now Neurons AI) when combined with traditional eye-tracking. In recent years the AI eye-tracking market has expanded with new CNN-powered softwares such as Attention Insight, Expoze (alpha.one), Cluefy, Emotiva and others. In our work we will focus on most mentioned AI eye-tracking models - Neurons AI and Attention Insight. In previous studies, Attention Insight has shown to reliably predict areas of user focus (Keresteš et al., 2024). The above-mentioned softwares enable users to upload the marketing materials and receive the AOIs, heat maps, fog maps, as well as additional AI metrics , such as Focus Score or Clarity Score (Neurons Inc., 2025-d; UAB Attention Insight, 2025a). In addition, Neurons AI and Attention Insight offer AI recommendations powered by large language models (LLM) for enhancing the performance of uploaded marketing materials (Neurons Inc., 2025-b; UAB Attention Insight., 2025a). While there is scientific potential for the application of the technology in the fields of advertising and visual attention prediction, the research into digital campaigns conducted in real-world scenarios remains limited.

Supervised deep learning is used to train such AI models using large datasets of eye-tracking data to generate "ground truth” attention heat maps (Neurons Inc., 2025-c; UAB Attention Insight, 2025b), while relying heavily on saliency maps to simulate human attention (Cornia et al., 2018). During the training process, AI models learn to extract relevant visual features from input images, generating individual heat maps that are then normalised to produce a unified saliency map that predicts areas of high visual interest (Tatler et al., 2011; Kümmerer et al., 2017). The model's predictions are then validated against empirical eye-tracking data to assess and refine its accuracy (Kümmerer et al., 2017). MIT/Tübingen saliency benchmark is often used for evaluating the performance of such AI models through submission of their predictions for the standardized image datasets such as MIT300 and CAT2000 (Keresteš et al., 2024; MIT/Tuebingen Saliency Benchmark, n.d.-a; Kümmerer et al., 2018). The standardized datasets are used for objective performance assessment because its ground truth data are not publicly available (Ibid.) The primary metrics used by companies to calculate accuracy are AUC (area under the curve) and sAUC, which refer to the model's ability to discriminate between fixated and non-fixated areas, with the perfect metric being 1.00 (Kümmerer et al., 2018; Kadner et al., 2023). However, the validity of the current validation methods, while robust for model development, may not fully capture the complexities of consumer behaviour in dynamic, unconstrained environments.

1.2.3. Advantages and Limitations of Eye-Tracking AI Prediction Softwares

AI models are gaining popularity among neuromarketing researchers, primarily because they can produce results comparable to traditional eye-tracking, especially in generating heat maps (Kusá & Beličková, 2023). While there are still not many comparative studies conducted on this topic, previous studies provide insights into the accuracy of AI models and their potential in the neuromarketing field. AI models demonstrate significant potential, producing results analogous to those of traditional eye-tracking methods in terms of heat map generation. However, further comparative research is necessary to fully ascertain their accuracy and extend their practical applications.

Juárez-Varón et al. (2024) in their comparative study aimed to evaluate the accuracy of the CNN-based AI eye-tracking software Decoditive Spark v.1.0 in predicting visual attention for Instagram banner compared to the eye-tracking data collected with Gazepoint GP3HD (n = 28). The results of the study have shown that predicted AOIs and their percentages are mostly similar to the eye-tracking data, except for the protagonist's face area and the first line of the commercial message (Ibid.). The difference between the viewers viewing AOIs and AI predicted percentages was minimal for the logo (0,30% absolute difference), higher for the protagonist's face (25,00% absolute difference), and between 3.77% and 10.61% for the text lines (Ibid.). Thus, Decoditive Spark v.1.0 as AI eye-tracking model demonstrates comparable accuracy to traditional eye-tracking for predicting visual attention on Instagram banners, with minor differences in specific areas like faces and text, suggesting its utility in initial visual assessment.

The results of the Keresteš et al. (2024) comparative study have shown higher accuracy with another CNN-based AI model, Attention Insight. In their study Keresteš et al. (2024) have evaluated the accuracy of AI visual attention prediction in UI/UX design for banner and native ads compared to traditional eye-tracking data collected with Gazepoint device (n = 17). The results of the study suggest that AI-generated heat maps and AOI percentages align with the time spent viewing the AOIs derived from eye-tracking data and can achieve acceptable to excellent accuracy in predicting areas of user focus (Ibid.). While AI models can achieve high accuracy in predicting exogenous (bottom up; automatic) attention, there is a variance in the areas where endogenous (top down; task-oriented) attention occurs (Ibid.; Corbetta & Shulman, 2002), suggesting that AI predictions could be complemented by traditional eye-tracking data in the areas where task-specific behaviour occurs, which is justified by the Šola et al. (2025) study. While Šola et al. (2025) have not conducted a comparative study, it is the precedent of a combined neuromarketing approach using AI-driven (Neurons AI, formerly Predict) and web-cam based (Tobii Sticky) eye-tracking for gaining insights about user engagement. Thus, Attention Insight shows strong accuracy in predicting visual attention for UI/UX design, particularly for exogenous attention. However, for endogenous, task-oriented attention, a hybrid approach combining AI with traditional eye-tracking may offer more comprehensive insights into user engagement.

Concluding the results of previous studies, neuromarketing AI eye-tracking softwares offer a cost-effective alternative to expensive traditional eye-tracking tools for small businesses (Ahmed et al, 2022; Kusá & Beličková, 2023; Winterhalter et al., 2024), while providing rapid results comparable to traditional eye-tracking (Keresteš et al., 2024; Šola et al., 2025). In addition, instead of focusing on static data, AI models provide real-time insights into the performance of marketing materials and suggestions on the design improvement, thus enhancing A/B testing and enabling data-driven design decisions (Borji & Itti, 2012; Kohavi et al., 2020; Lazar et al., 2015). Although, the combination of AI and traditional neuromarketing methods, could help to get a full overview of consumer attention patterns (Keresteš et al., 2024; Šola et al., 2025).

Previous studies have also highlighted several AI limitations, such as the difficulty in identifying task-specific contexts (Keresteš et al., 2024), and cultural nuances, as well as the high cost of some AI softwares (Kusá & Beličková, 2023). Moreover, all the previous studies were limited to a single cultural or regional context (Juárez-Varón et al., 2024; Keresteš et al., 2024; Šola et al., 2025), thus further research on AI performance in real international campaigns is needed (Alam et al., 2021). The need is justified by other studies confirming the differences in cognitive processes between Western and Asian cultures (Nisbett & Miyamoto, 2005), as well as specific cross-country differences in perception and engagement when proposed to various visual stimuli (Marchesi et al., 2023; Mas et al., 2023; Stöckl & Diaz, 2025).

According to the five key contributions of neuroscience to the field of marketing laid out by Plassmann et al. (2015), further research into neuromarketing AI software could contribute to the fifth principle - making predictions about human behaviour. Past neuromarketing research has mostly contributed to the first four principles by investigating cognitive mechanisms, measuring implicit responses, distinguishing psychological processes, and understanding consumers' behaviour (Alimardani, et al, 2021). Finally, literature review reveals a significant gap between academic neuromarketing research and corporate application (Baños et al., 2020; Brenninkmeijer et al., 2019; Lin et al., 2018) that could be addressed through further research of AI models as a cost-effective and less complex alternative for businesses, thus fostering adaptation of neuromarketing tools to corporate needs. Further research into AI neuromarketing software has the potential to advance the field by enabling stronger predictions of human behaviour, building upon previous contributions in understanding cognitive processes and consumer behaviour. It is crucial to explore the potential of AI as a cost-effective and user-friendly solution, with the aim of bridging the existing gap between academic neuromarketing research and its practical corporate application.

The utilisation of AI models has undergone a significant transformation within the domain of neuromarketing, thereby providing results that are comparable to those obtained through traditional eye-tracking methodologies in generating heat maps. While Decoditive Spark v.1.0 and Attention Insight demonstrate strong accuracy, a fundamental theoretical finding is that for endogenous (task-oriented) attention and cross-cultural contexts, a hybrid approach combining AI with traditional eye-tracking is more comprehensive. This positions AI as a cost-effective and rapid tool for initial assessments, though a combined strategy provides a holistic understanding of consumer attention. This underscores the need for further research in diverse international settings. In conclusion, the enhancement of AI in neuromarketing has the potential to make a substantial contribution to the prediction of human behaviour and provide an accessible solution for businesses, thereby bridging the gap between academic research and corporate application.

2. Materials and Methods

2.1. Research Design

The empirical study employed distinct methodologies. Firstly, the AI eye-tracking software, Attention Insight, was employed to predict visual attention and gather banner design improvement suggestions. Secondly, the traditional eye-tracking was employed, utilising the Tobii X2-30.

In the absence of any previous studies on the utilisation of AI design improvement recommendations for digital banners, a pilot study was conducted in October-November 2024 with the objective of mapping out the process for the final A/B test campaign. Following the pilot study, the experiment design was revised (see the Chapter 2.3. for further details). The pilot study included only the digital A/B test campaign, and in the final study both digital A/B test campaign and eye-tracking studies were executed. The final research process included the following phases: 1. Preparation, 2. Execution, 3. Analysis. The simplified overview of research design is demonstrated in Figure 1 and the full version is available in Appendix 1.

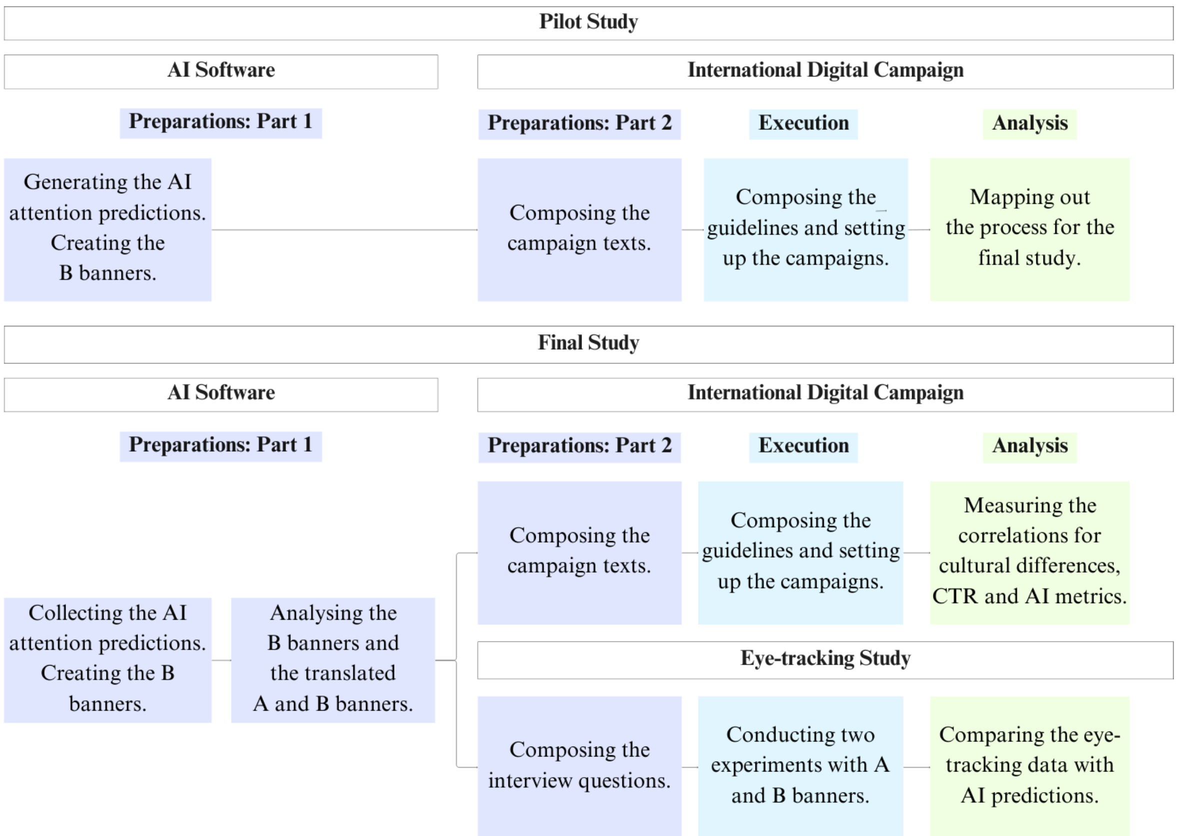


Figure 1. A simplified overview of the research process.

Source: Compiled by the authors.

In December 2024, during the Preparation phase, the AI visual attention predictions and recommendations for the non-improved banners (hereafter referred to as 'A banners') were generated. In accordance with the AI improvement suggestions, the improved banners (hereafter referred to as 'B banners') were created by the designer. Following the receipt of the B banners, the visual attention predictions were generated for them separately (Appendix 3). Then the texts for the international digital campaigns were composed, as well as the questions for the interviews (Appendix 5) that form part of the eye-tracking study. From January to March 2025, during the execution phase, the guidelines for the international digital campaigns were composed (Appendix 7), and the campaigns were set up accordingly. Simultaneously, the eye-tracking experiments were conducted for A and B banners separately. The culmination of the experiment was marked by the initiation of the final stage in late March 2025. This stage entailed a statistical analysis of the results and the execution of supplementary eye-tracking experiments, which were conducted with the objective of achieving a balanced sample. Table 1 provides a concise overview of the sources that serve as the foundation for the research design and metrics.

Table 1

The sources used for different tasks of the research.

Task	***Source***
Defining areas of interest (AOIs) and elements for which to request AI-based improvement suggestions	Brand elements, pictorial elements, and text elements have been identified as three primary components of banner advertisements that significantly impact visual attention (Pieters & Wedel, 2004). Thus it was decided to set the following banner elements as Areas of Interest (AOIs): the logo, the product name, the product image, and the slogan.
Methodology for interpreting and filtering AI design improvement suggestions	In their research, Pieters et al. (2010) provided a methodology for defining visual complexity and improving advertisement design. The methodology is divided into two distinct groups of principles, namely design complexity and brand identification difficulties (see Chapter 1.1.1.).

Metric for measuring the performance of digital banners	Click-through rate (CTR) is widely recognised as a measure of advertising effectiveness (Chandon et al., 2003; Lohita et al., 2003) and is also indicative of the quality of visuals used in advertisement (Iankovets, T., 2023).
Methodology for comparing traditional and AI eye tracking	In their study, Juárez-Varón et al. (2024) compared traditional and AI eye-tracking by measuring visual attention using metrics from each area of interest (AOI). These metrics included AOI fixation share (%), AOI average time to first fixation (sec), and AOI average fixation duration (sec).
The age as independent variable (see interview question 2 in Appendix 5)	Mičík & Kunešová (2021) have demonstrated that age differences are among the main predictors of consumer behaviour (see Chapter 1.1.1.).
The primary language of communication and country of origin as independent variables (see interview questions 4-5 in Appendix 5)	As stated by Ploom et al. (2020), ethnic and cultural differences are recognised as significant predictors of consumer behaviour (see Chapter 1.1.1).
The brand familiarity as independent variables (see interview questions 6-7 in Appendix 5)	Brand familiarity has been identified as one of the main predictors of consumer behaviour (see Garczarek-Bąk et al., 2021; Kim & Lee, 2021; Peker et al., 2021). Furthermore, individual measures of product involvement, motivation, and brand familiarity have been demonstrated to influence attention capture (Pieters et al., 2002; Pieters & Wedel, 2004).

Source: Compiled by the authors.

Together, these theoretical foundations and methodological choices provided a structured basis for conducting the empirical research. The next section describes in detail how the AI visual attention predictions and design recommendations were generated using Attention Insight software.

2.2. Test Samples

As a visual stimuli for our study, the current research employs four different designs of the square 1200 x 1200 px static digital banners (Appendix 2) received from the Penosil brand. The following brand was chosen due to several factors. Firstly, authors were able to control the majority of the experiment processes due to the semi-centralised company structure. Secondly, daughter companies in different countries agreed to participate in the study making it international, thus enabling us to analyse the performance of AI recommendations on an international level.

All the banners and texts used in the experiment were not only translated but also adapted to the local campaign product due to the geographical segmentation of the product range and marketing strategy, except for Lithuania. Although the conditions were the same for each country in order to make the results comparable in terms of A/B testing, they still differed in terms of brand awareness in the market, result prices (CPM, CPC, etc.), campaign budget and website functionalities such as e-shop. The countries with an e-shop were France, Spain, Estonia and Latvia; the Lithuanian and Romanian markets did not have an e-shop at the time of the experiment. The non-enhanced A banners received from the Penosil brand were improved according to the AI recommendations. Thus, additional four banners (B banners) were included in the study. Below there is an example of the comparison of two designs with AI insights (Figure 2). The remaining heat maps are available in Appendix 3.

Figure 2. An example of heat maps generated with Attention Insight software for the design 4 A and B banners.

Source: Compiled by the authors with Attention Insight software.

2.3. The AI Visual Attention Predictions and Recommendations

Attention Insight, a CNN-based AI eye-tracking software, was employed to predict visual attention and obtain the design improvement suggestions. The current model has been trained with 5,5+ million fixations and 550+ million gaze points collected from eye-tracking data of USA and Europe-based samples (UAB Attention Insight, 2025b). As stated by the MIT/ Tuebingen saliency benchmark, Attention Insight software possesses the heat map prediction accuracy of 92,5% for general images and 96% for all types of designs. Furthermore, Keresteš et al. (2024) have determined that it is a reliable tool for generating visual attention predictions.

The process of generating AI visual attention comprised the following steps: 1. The selection of the analysis type - marketing materials, 2. Upload of the test samples, 3. The definition of the areas of interest (AOIs). As a result, Attention Insight provided a range of visual representation and quantitative insights for the materials uploaded, including the heat map, fog map (also known as focus map), contrast map and AI metrics. The primary metrics for evaluating the quality of the design are AOI percentages and AI metrics such as Clarity and Focus scores (UAB Attention Insight, 2025a). The Clarity score, ranging from 1 to 100, is a metric used to indicate the level of design cleanliness, clearness and conflict between the design elements (Ibid.). The Focus score, ranging from 1 to 100, is a metric used to measure the level of attention density and thus concentration (Ibid.). Figure 3 shows the screenshot made from the Attention Insight dashboard after uploading the image.

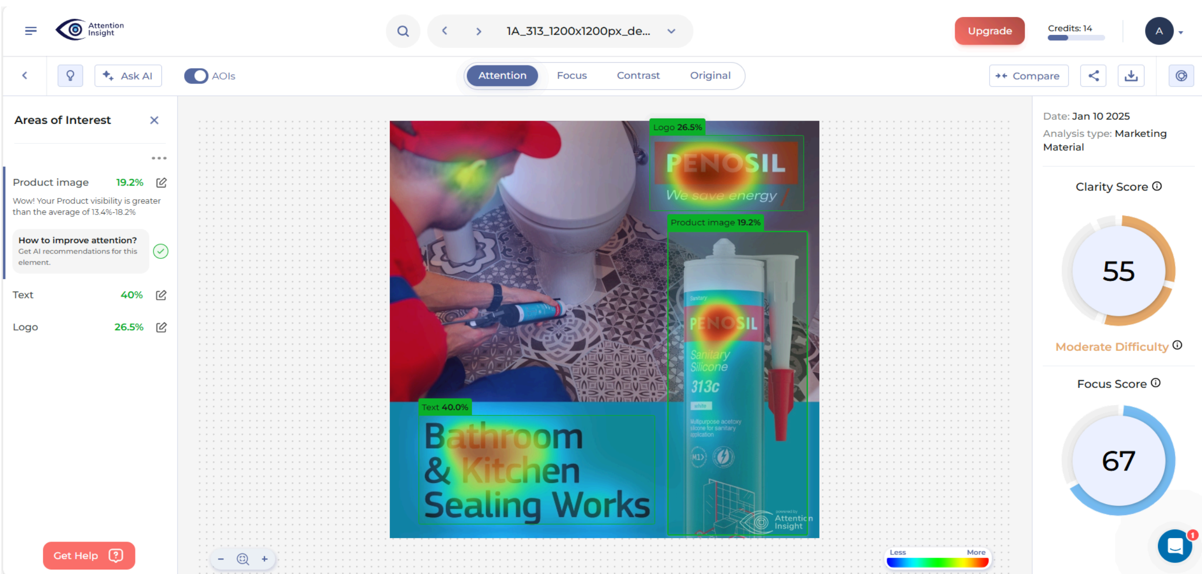


Figure 3. A screenshot of the Attention Insight dashboard after uploading the image.
Source: Compiled by the authors with Attention Insight software.

Following the generation of insights for each A banner, the LLM-based AI recommendations for further design enhancement were generated (Ibid.). The recommendations can be asked either for each AOI separately or for the design in general to optimise it for conversions, get more attention, validate design valency or to improve colours and contrast. In the pilot study, the generation of all types of recommendations was carried out in order to clarify the most applicable ones. As per the theoretical review, the primary AOI's for which the recommendations were generated included: the brand name, product image, and product name. Furthermore, the recommendation for getting attention for the general design was generated. The design enhancement process was as follows:

1. Generating the recommendations for brand name, product image, product name,

2. Generating the recommendations to increase overall attention,

3. Conducting a critical analysis and filtration of recommendations according to visual complexity principles of Pieters et al. (2010),

4. The enhanced B banners are created by the designer,

5. The banners are analysed again to ascertain if further enhancement is needed.

Below is available an example of the comparison of A and B banners gathered from Attention Insight (Figure 4). As seen from the next example, not always employment of AI recommendations led to the improvement of the design's performance (Figure 5).

COMPARISON TABLE

	CONTROL	
AOI	1 1A_313_12...	2 1B_313_12...
Clarity	55	61 +6%
Product image	19%	22% +2%
Text	40%	42% +2%
Logo	27%	34% +8%

Figure 4. An example of the comparison report generated with Attention Insight software for the design 1 A (non- improved) and B banners (improved with AI suggestions).
Source: Compiled by the authors with Attention Insight software.

COMPARISON TABLE

	CONTROL	
AOI	1 4A_313_12...	2 4B_313_12...
Clarity	64	61 -3%
Text	12%	12%
Product name	32%	24% -8%
Logo	33%	36% +3%
Product image	21%	26% +6%

Figure 5. An example of the comparison report generated with Attention Insight software for the design 4 A and B banners.
Source: Compiled by the authors with Attention Insight software.

This section outlined the use of Attention Insight, a CNN-based AI tool trained on extensive eye-tracking data, to predict visual attention and generate design improvement suggestions. Through a structured process involving AOI definition and metric evaluation (e.g., Clarity and Focus scores), AI-based recommendations were generated for key banner elements such as the brand name and product image. These insights were critically reviewed and applied to create B-version banners, though results showed that AI input did not always guarantee improved design performance.

2.3. The Digital A/B Test Campaign Methodology: Pilot and Final Studies

For the digital campaign experiment, the Google Demand Gen campaign experiment was utilised due to the format enabling using just one banner format without losing in campaign quality. Demand Gen campaigns allow advertisers to deliver visually appealing, multi-format ads across major Google platforms like YouTube (including Shorts), Discover, Gmail, and the Google Display Network (Google, 2025a). This format is particularly suited for advertisers aiming to drive engagement and action (Ibid.). The evaluation criteria for the format comprised the following: 1. The capacity for conducting A/B testing, 2. Minimal materials required without the forfeiture of Google quality points (Google, 2025c). With regard to the test samples, it should be noted that only a square banner format was available. This choice is therefore justified, despite the higher prices of the aforementioned PPC campaign in comparison to other similar campaigns. As an evaluation metric for measuring the campaign performance, the click-through-rate (CTR) was chosen. As stated in previous studies, CTR is a reliable metric for measuring the campaign performance (Chandon et al., 2003; Lohita et al., 2003) and specifically for evaluating the appealingness of advertising creatives (Iankovets, 2023).

The pilot digital campaign involved France, Spain, Portugal and Romania. The pilot study process for the digital campaign consisted of three distinct stages: 1. Preparation, 2. Execution, 3. Analysis. In the Analysis stage, the visual attention predictions and design improvement suggestions for A banners were generated, and the shortened instructions were sent to the brand manager, who filtered it according to the CVI rules and forwarded to the designer. Following the receipt of the B banners, the texts for the international Google Demand Gen experiment campaign were composed with the aim to ensure the same campaign content in all the countries. Although the risk of human factors remained, as the translation was undertaken by marketing specialists in each country. Following the completion of translations, the verification and correction process was undertaken to ensure the translation alignment with the original English text.

In the execution step, the A/B test campaign guidelines (Appendix 7) were composed with the objective to ensure the same setup and minimal risk of external factors. The guidelines were inspired by the previous setups utilised by the brand, campaign objective and best practices (Google, 2025b; Google, 2025c). The Google experiment campaign setup was conducted by countries, and then reviewed and corrected by the present authors. Subsequent to the initiation of the campaigns, diligent monitoring was implemented to ensure the minimisation of potential errors.

In the analysis step, the data was collected from Google Ads accounts to ascertain the types of data that could be received and what statistical analysis could be implemented. Finally, all the risks have been identified and the updated experiment design has been mapped out according to the online A/B test best practices (Fabijan et al., 2019).

The final digital campaign was conducted in Estonia, Latvia, Lithuania, Spain, and France. In accordance with the insights obtained from the pilot study, modifications have been made to the research design. Specifically, in the preparation step the design and brand complexity principles inspired by the study of (Pieters et al., 2010) were chosen as a framework for a critical analysis, interpretation and filtering of the AI design improvement suggestions. The framework choice is justified by the previous studies showing a negative correlation between visual complexity and attention, attitude, and recognition (Orth & Crouch, 2014; Tuch et al., 2009), thus supporting the objective of banner enhancements. In addition, we have followed other principles mentioned in the theoretical review (see the Chapter 1.1.3).

Furthermore, during the final study's preparation, it was decided to conduct an additional analysis of B banners to ascertain if additional enhancement is needed in case of points left for the additional analysis. In addition, the updated guidelines (Appendix 7.2) were composed according to the new product advertised in the final digital campaign. Finally, in the analysis stage of the final study, it was decided to undertake an additional generation of AI insights for each country following the campaign.

2.4. The Eye-Tracking Study Methodology

This study employed two separate eye-tracking experiments to investigate how digital advertising banners influence visual attention. The primary aim was to determine which design elements (defined as Areas of Interest, AOIs) attract attention most effectively and how quickly, and to compare original versus AI-modified banner versions.

To ensure diversity, a combination of convenience, snowball, and quota sampling methods was used. A total of 66 male volunteers (n = 33 in each experiment), aged 21 to 64, participated. Particular emphasis was placed on including individuals aged 18–35, given their greater responsiveness to emotionally and visually engaging advertising (Lee & Ahn, 2012; Peker et al., 2021). Including older participants allowed for the examination of attentional variation across age groups. The all-male sample reflected the target audience for the products advertised, which were construction chemicals—a sector predominantly composed of men. Prior research supports aligning the participant profile with the advertisement's intended demographic, especially in neuromarketing and eye-tracking research (Juárez-Varón et al., 2024). As the digital campaign was part of a broader international initiative distributed across multiple markets, cultural diversity was intentionally incorporated into the study design. Participants of various linguistic and ethnic backgrounds were included to mirror the real-world advertising environment and to permit, if necessary, the interpretation of cultural trends. However, due to the limited scope of this thesis, culture-specific analyses were not performed.

Before the experiments, participants completed structured interviews to collect demographic information (age, ethnicity, native language, vision quality) and self-reported familiarity with the advertised brands and products, using the availability heuristic approach (Pieters et al., 2010). Recruitment took place in four Estonian cities: Tallinn, Tartu, Viljandi, and Pärnu, through personal contacts and social media, including groups targeting international communities (Neves Pereira et al., 2024). The socio-demographic data are presented in Appendix 6.

The experiments took place individually using a Tobii X2-30 stationary eye tracker and Tobii Studio software. Participants were seated approximately 60–70 cm from a 14-inch monitor. After receiving an overview of the study's purpose and providing informed consent, participants were positioned so their entire face remained visible to the integrated camera. A standard 9-point calibration procedure was used to ensure accurate gaze tracking, repeated as necessary. Although emotional responses were also recorded using FaceReader software, emotion-based data were excluded from this thesis due to their time-intensive nature.

Participants were first shown a "noise" slide simulating a cluttered digital environment. Each slide consisted of 15 square, static digital banners (1200 × 1200 px), arranged in a grid. Four of these were Penosil banners, either in their original or AI-modified forms, depending on the assigned experiment. To control for placement effects, two versions of the noise slide were created with banners positioned differently. One version was shown to 17 participants, the other to 16. Each participant saw only one version. This approach allowed focus on the design's influence on visibility, minimizing location bias. Participants in Experiment A saw noise slides with original Penosil banners, while those in Experiment B viewed noise slides with AI-enhanced Penosil banners. Following the noise slide, participants viewed a series of 15 static banners (also 1200 × 1200 px) on a neutral gray background. Each banner was displayed for 5 seconds, followed by a 1-second gray screen to reset the gaze and reduce carryover from previous stimuli. Banner presentation order was randomized to prevent sequence effects.

After the banner viewing session, brief post-interviews were conducted to supplement the eye movement data regarding participants' awareness, memory, and preferences. The interviews employed both open-ended and directed questions to assess the recognition and impact of the displayed advertisements in the participants' native languages (Estonian, Russian, English). The questions from both the pre- and post-interviews are provided in Appendix 5. This combined methodology (eye-tracking and post-interview) is significant in advertising and brand research, as eye movement data alone may not fully reflect conscious information processing (Lee & Ahn, 2012; Zito et al., 2021; Pieters et al., 2010; Neves Pereira et al., 2024). The post-interviews enabled the evaluation of ad memorability, preferences, and the differentiation of brand elements, considering both the salience of areas of interest (AOIs). Specifically, it facilitated the analysis of the recognition and comparison of Penosil banners (original vs. AI-optimized versions), which was a central aspect of the study.Throughout both phases, the Tobii device recorded continuous eye-tracking data, including fixations, saccades, and pupil movement at a sampling rate of 30 Hz. Predefined

AOIs in Tobii Studio marked elements such as logos, product name, product visuals, and text feature. To ensure methodological consistency, particular attention was given to the definition and application of Areas of Interest (AOIs) in both the AI-based and lab-based analyses. The AI software restricts AOI shapes to simple rectangles, which limits the ability to precisely outline irregular or complex visual elements. To maintain comparability between the two data sources, the manually drawn AOIs in the lab were aligned as closely as possible with the rectangular structure of the AI-based AOIs - both in terms of content and spatial positioning. This alignment helped ensure the validity of the comparative analysis between AI-predicted attention patterns and actual eye-tracking data.

While the AI system cannot interpret precise object contours or semantic relevance, the manual AOI mapping allowed the researchers to trace visually meaningful shapes more accurately (e.g., curved logos or irregular product outlines). Nevertheless, the final AOIs in the lab environment were intentionally aligned with the general rectangular structure of the AI-generated AOIs to maintain analytical equivalence. An illustrative comparison between an AI-generated AOI and the corresponding manually drawn AOIs for the same banner is presented in Figure 6.

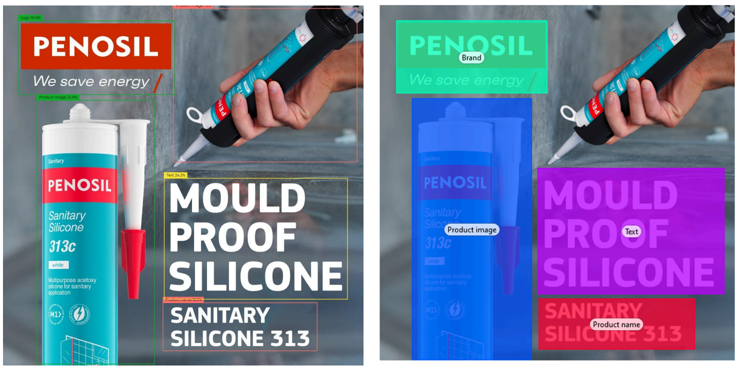


Figure 6. Illustrative Comparison Between AI-Predicted (left) and Manually Drawn AOIs (right) on Penosil Banner B3.

Source: Compiled by the authors with Attention Insight software and Tobii Studio.

This deliberate alignment between AI and human-defined AOIs allows for a valid comparison of attention metrics across systems and supports the integrity of the mixed-method research design.

Post-experiment analysis was conducted using Microsoft Excel and SPSS. Two primary eye-tracking metrics were employed to assess visual attention:

- **Time to First Fixation (TFF):** The time elapsed before a participant first fixated on a designated Area of Interest (AOI), serving as an indicator of visual salience.
- **Total Fixation Duration (TFD):** The total amount of time spent fixating on an AOI, reflecting the depth or volume of attention.

These metrics were selected based on prior research demonstrating their relevance in evaluating advertising effectiveness (Lee & Ahn, 2012; Neves Pereira et al., 2024). In addition to numerical data, heatmaps were generated to provide a visual representation of gaze distribution. While heatmaps offer valuable insights into general viewing patterns, they were used primarily for illustration purposes and not as the basis for statistical inference - consistent with literature warning against overinterpretation of visualizations alone (Keresteš et al., 2024). The data collection approach in this study - combining eye-tracking metrics, heatmap analysis, and structured interviews - provided a comprehensive basis for evaluating both the design effectiveness and the resulting viewer attention. This mixed-method design allowed for an in-depth exploration of how specific design elements performed across different viewer groups and experimental conditions.

To guide the analysis, the empirical eye-tracking study was structured into three analytical stages, each addressing a distinct research focus and set of attention-related questions. This structured approach enabled both quantitative measurement and qualitative interpretation, creating a cohesive framework for evaluating the impact of AI-based design changes.

The first stage (see Chapter 3.2.1) examined the visibility of Penosil advertisements within a competitive, cluttered visual environment. AI-enhanced Penosil banners were tested alongside other commercial banners to assess whether they could effectively stand out. The focus here was on TFD values, used descriptively to evaluate overall visibility compared to competing ads.

The second stage (see Chapter 3.2.2) focused on a comparative AOI-level analysis between the original (A-version) and AI-optimised (B-version) banners. TFF, TFD, and Viewers % were used to evaluate the performance of specific visual elements such as the logo, product image, and text. This allowed a detailed examination of how AI recommendations influenced user gaze patterns and attention distribution.

The third stage (see Chapter 3.2.3) involved a direct comparison between real eye-tracking data and AI-generated visual attention predictions. The analysis included correlation testing between predicted and actual attention values across AOIs, supported by heatmap interpretation. This final stage offered valuable insights into the practical reliability and limitations of AI models in predicting human visual behavior.

3. Results

3.1. The International Digital Campaign

The final digital campaign was conducted from 03.02 to 28.02.2025 with Estonia, France, Latvia, Spain and Lithuania participating, and an average budget of 274,87€. The minimum budget was 200,82€ in Lithuania, the mean budget of 271,08€ was in Spain and the maximum budget of 400,61€ in France. The A/B testing results were gathered from the Google Ads platform and an additional analysis was made with Microsoft Excel software.

In order to analyse the practical utility of AI design improvement suggestions, an A/B testing methodology was utilised with the objective of comparing the non-optimised (A banners) and optimised banners (B banners). The descriptive statistics in Table 2 demonstrate that the mean CTR for B banners (1,62%) is slightly higher than that for A banners (1,59%). However, the median CTR for B banners (1,46%) is slightly lower than for A banners (1,56%), suggesting some variability in the data. The standard deviations are quite similar, indicating comparable spread in the data for both banner types.

Table 2

Central tendency and spread of click-through-rates across A and B banners.

	Mean	**Median**	**Variance**	**Standard deviation**
A banners	1,59%	1,56%	0,01%	0,85%
B banners	1,62%	1,46%	0,01%	0,90%

Source: Compiled by the authors.

To assess whether the observed differences in CTRs between A and B banners were statistically significant, we used the paired samples t-Test (see Table 3). Since the p-value (0,4382) is greater than our significance level (0,05), we fail to reject the null hypothesis. Consequently, there is no statistically significant difference in the mean CTR between the A and B banners.

Table 3

The results of t-Test for CTR means of A and B banner groups.

	CTR - A banners	***CTR - B banners***
Mean	0,015895	0,01619
Variance	0,000072959	0,000080284
Observations	20	20
Pearson Correlation	0,39492423	
Hypothesized Mean Difference	0	
df	19	
t Stat	-0,1369554	
P(T<=t) one-tail	0,44625345	
t Critical one-tail	1,72913281	
P(T<=t) two-tail	0,8925069	
t Critical two-tail	2,09302405	

Source: Compiled by the authors.

In addition, we have reviewed the Google Demand Gen experiment results (Appendix 8) in the Google Ads accounts of each country to control the winning sets. According to Table 4, B banners have shown to bring greater results in comparison to A banners, in the case of the Baltic countries. In France, however, A and B banners have demonstrated a similar performance. In the context of Spain, the A banners became a winning set.

Table 4

The results of the international Demand Gen experiment campaign.

Country	Estonia	France	Latvia	Spain	Lithuania
A Banners	0,60%	1,58%	0,83%	1,93%	2,08%
B Banners	0,71%	1,45%	0,96%	1,77%	2,75%
Winning arm	B banners	Similar performance	B banners	A banners	B banners
Absolute difference	+0,11%	-0,13%	+0.13%	-0,17%	+0,67%

Relative difference	+19%	-8,3%	+15.4%	-8,6%	+32,3%
Confidence interval	95%	95%	95%	95%	95%
P-value	0,04	0,06	0,00	0,09	<0,01
Average CTR in 2024	3,16%	1,77%	N/A	4,97%	2,09%

Source: Compiled by the authors based on the results assessed through Google Ads platform.

As the results of A/B testing were controversial due to the insignificant difference in CTR means and the median CTR to show the opposite result than the mean, it was decided to additionally investigate the correlation between CTR and AI design quality metrics.

The mean Clarity Score of the translated banners ranged from 51,88 in Spain to 56,63 in Latvia, with an overall mean of 54,90 (Appendix 9.1). The median Clarity Score varied from 51,00 in Spain to 61,00 in Latvia, with an overall median of 57,00. The lowest variance of Clarity Score was shown in Estonia (16,00) and the highest in Latvia (134,84). The results demonstrate that there occurred some differences in Clarity Scores after translation of initial banners in each country (Figure 7). The frequencies of Clarity Scores in the combined data demonstrate a slight right skew, with a higher frequency in the middle to upper ranges and relatively few low scores and relatively normal distribution across countries. Latvia and Lithuania though show a right-skewed distribution, indicating a concentration of higher Clarity Scores. In order to address the differences across countries, the ANOVA test was conducted. The ANOVA test indicates that, despite some differences in the mean Clarity Scores across countries (Estonia: 55,5; France: 54,375; Latvia: 56,625; Spain: 51,875; Lithuania: 56,125), these differences are not statistically significant (Appendix 10.1).

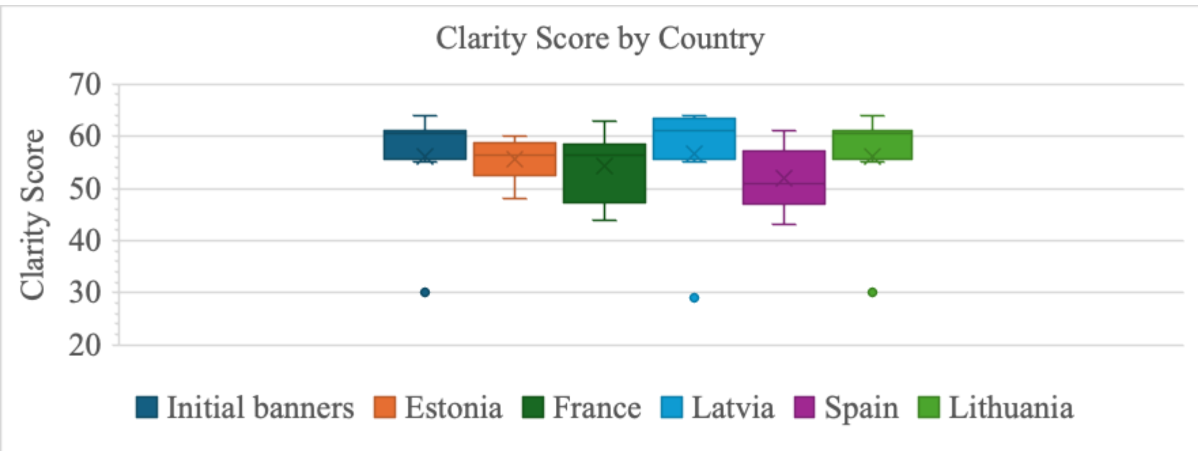


Figure 7. Clarity Score distributions by country and for initial English banners.

Source: Compiled by the authors.

The mean Focus Score of the translated banners ranged from 60,63 in Spain to 69,00 in France, with an overall mean of 60,63 (Appendix 9.2). The median Focus Score varied from 60,00 in Spain to 67,50 in Estonia, France and Latvia, with an overall median of 67,00. The lowest variance of Focus Score was shown in Spain (31,41) and highest in France (50,00). The results demonstrate that the differences after translation occured in Focus Scores as well, although the scores are generally high across all countries (Figure 8). Similar to Clarity Score, the ANOVA test for Focus Score suggests that the observed differences in average Focus Scores (Estonia: 67,5; France: 69; Latvia: 65,375; Spain: 60,625; Lithuania: 65,625) are not statistically significant (Appendix 10.2).

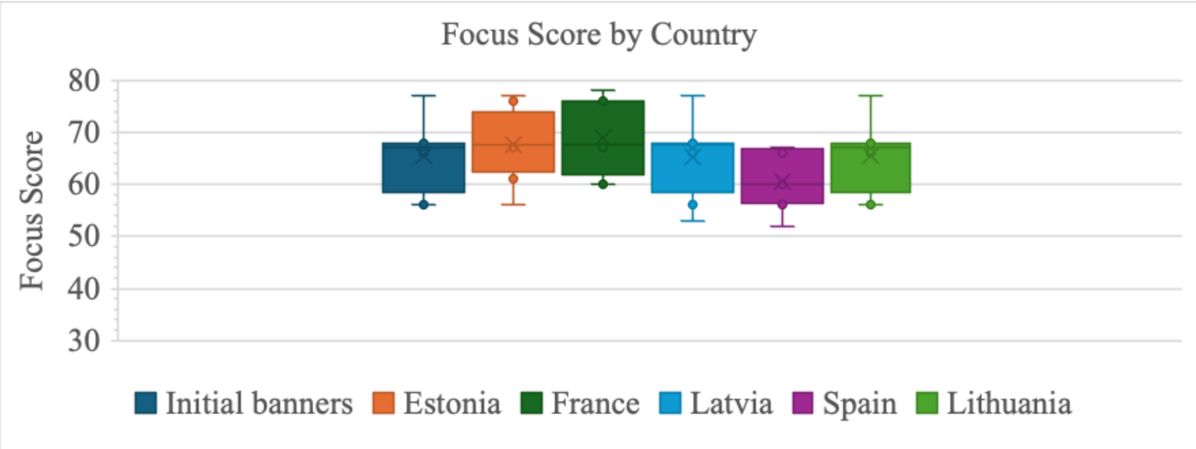


Figure 8. Focus Score distributions by country and for initial English banners.
Source: Compiled by the authors.

The mean click-through-rates (CTR) ranged from 0,94% in Latvia to 2,35% in Lithuania, with an overall mean of 1,60% (Appendix 9.3). The median CTR varied from 0,88% in Spain to 2,44% in Lithuania, with an overall median of 1,47%. The lowest variance of CTR was shown in France (0,0002%) and highest in Lithuania (0,0123%). The results demonstrate that CTR significantly varies across countries, with highest results demonstrated in Lithuania and lowest in Estonia and Latvia (Figure 9). The frequencies of Focus Scores in the combined data demonstrate a slight left skew, with a higher frequency in between 60 and 70 and a normal distribution of Focus Scores, although Spain distribution is slightly shifted towards lower Focus Scores. The frequencies of CTRs in the combined data and frequencies across countries demonstrate a strong left skew, with a higher frequency from 0,35% to 2,35%, indicating that most banners have lower CTRs, with a few high-performing designs. Lithuania shows a wider spread of CTR results. In contrast with Clarity and Focus score results, the ANOVA test for CTR (Appendix 10.3) has demonstrated that there are statistical differences in average CTR across the countries (Estonia: 1,11%; France: 1,55%; Latvia: 0,94%; Spain: 2,07%; Lithuania: 2,35%). Thus, although the differences between Clarity and Focus scores don't differ significantly, the performance of campaigns is still significantly different across countries.

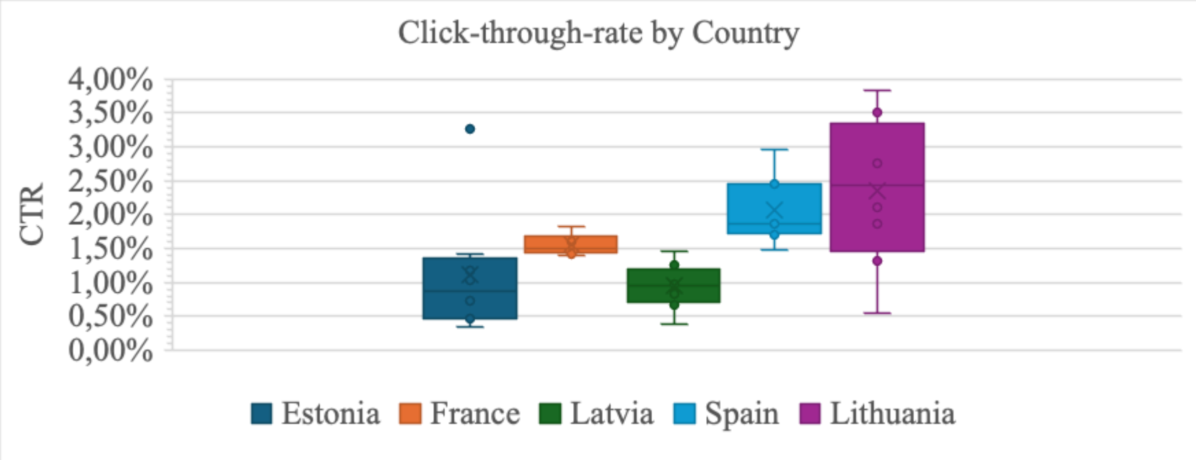


Figure 9. Click-through-rate distributions by country.
Source: Compiled by the authors.

Following the evaluation of data distributions, the correlation analysis was conducted. Correlation between Clarity Score and CTR in the combined data appears to be low with a R-squared value of 0,0121 (Figure 10). In most countries (Estonia, Latvia, Spain, Lithuania), the linear correlation between Clarity Score and CTR appears also to be weak, with data points scattered and low R-squared values (ranging from 0 to 0,26).

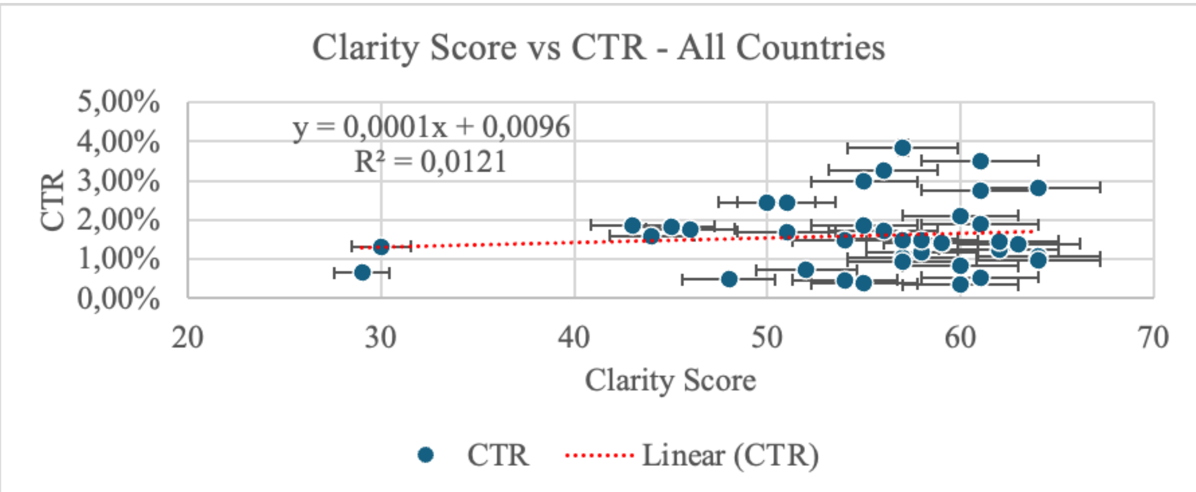


Figure 10. Scatter plot of Clarity Score versus click-through-rate in all countries combined.
Source: Compiled by the authors.

However the overall correlation is weak, additional analysis revealed that there are a few strong correlations in some countries. For example, France shows a slightly stronger positive correlation with an R-squared value of 0,5106 (Figure 11).

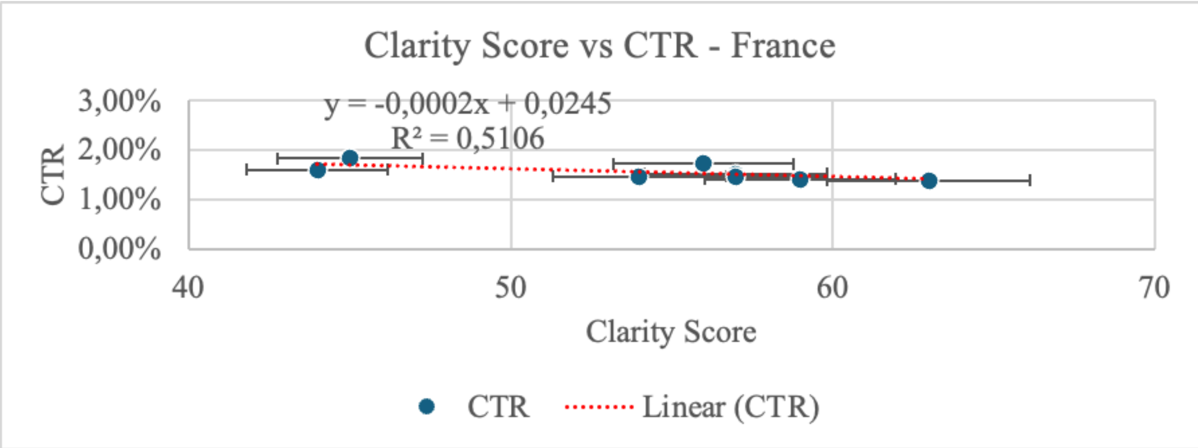


Figure 11. Scatter plot of Clarity Score versus click-through-rate in France.
Source: Compiled by the authors.

Correlation between Focus Score and CTR in the combined data appears to also be low with a R-squared value of 0,0112 (Figure 12), with very weak correlations with R-squared values close to 0 in most of the countries.

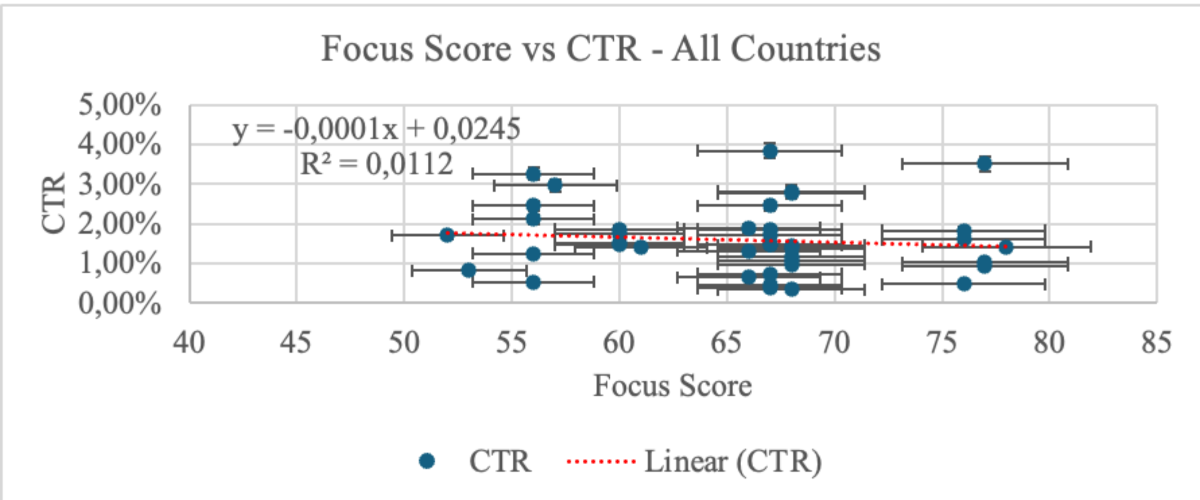


Figure 12. Scatter plot of Focus Score versus click-through-rate in all countries combined.
Source: Compiled by the authors.

However, the additional analysis again revealed the strong correlations in some countries. For example, Estonia shows moderate negative correlation between Focus Score and CTR, with an R-squared value of 0,5039 (Figure 13).

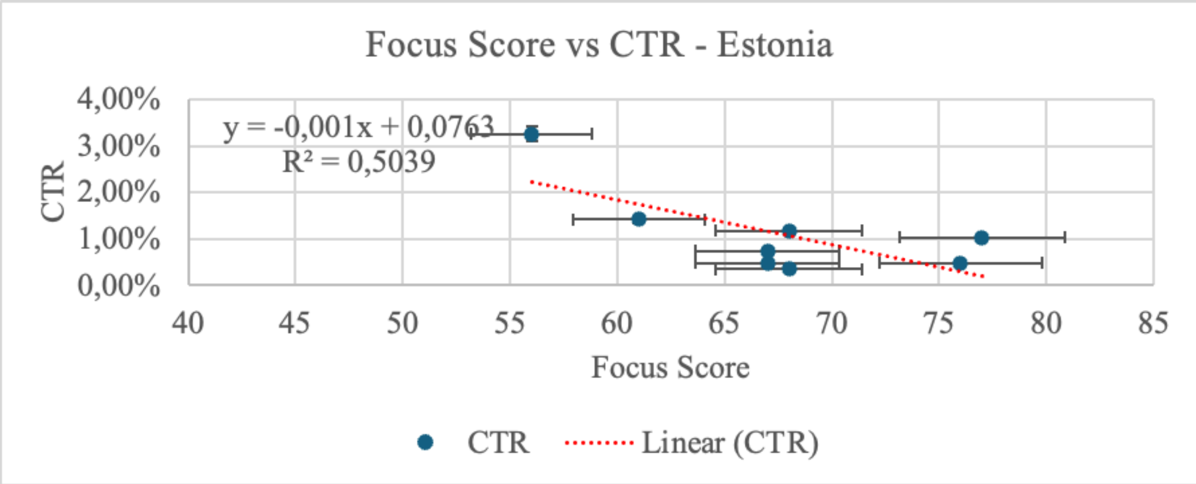


Figure 13. Scatter plot of Focus Score versus click-through-rate in Estonia.

Source: Compiled by the authors.

In contrast, Lithuania shows a moderate positive correlation between Focus Score and CTR, with an R-squared value of 0,4575 (Figure 14).

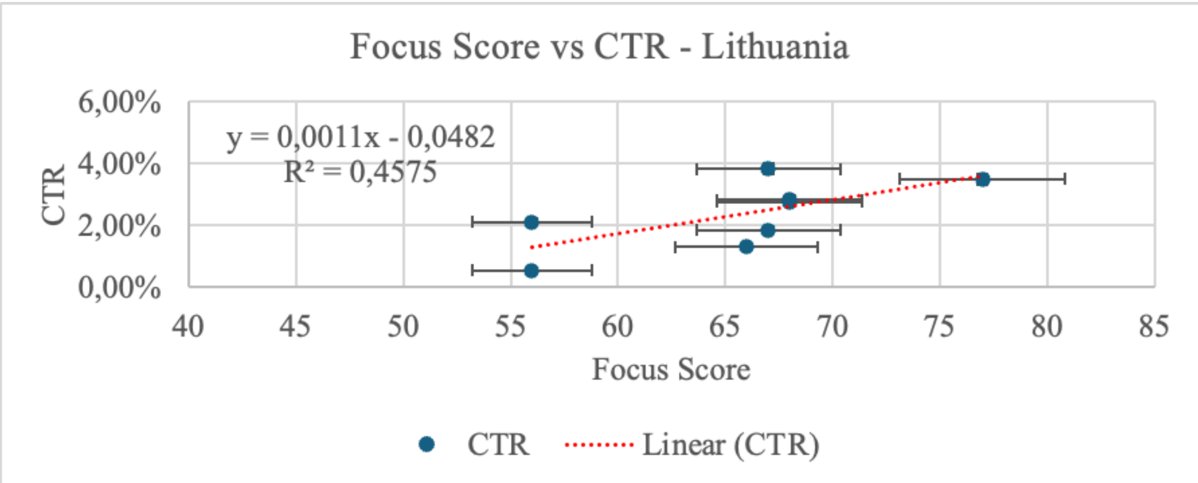


Figure 14. Scatter plot of Focus Score versus click-through-rate in Lithuania.

Source: Compiled by the authors.

In addition to scatter plots, the correlation coefficients were calculated. Table 5 demonstrates that correlation between Clarity Score and CTR varies significantly across countries.

Table 5

The correlation analysis between Clarity Score, Focus Score and click-through-rate.

Country	Variables	Clarity Score	Focus Score	CTR
Estonia	Clarity Score	1.00	-0.37	0.22
	Focus Score	-0.37	1.00	-0.71
	CTR	0.22	-0.71	1.00
France	Clarity Score	1.00	-0.47	-0.71
	Focus Score	-0.47	1.00	0.28
	CTR	-0.71	0.28	1.00
Latvia	Clarity Score	1.00	-0.09	0.51
	Focus Score	-0.09	1.00	-0.05
	CTR	0.51	-0.05	1.00
Spain	Clarity Score	1.00	0.37	0.02
	Focus Score	0.37	1.00	-0.20
	CTR	0.02	-0.20	1.00
Lithuania	Clarity Score	1.00	-0.02	0.36
	Focus Score	-0.02	1.00	0.68
	CTR	0.36	0.68	1.00
All Countries	Clarity Score	1.00	-0.04	0.11
	Focus Score	-0.04	1.00	-0.11
	CTR	0.11	-0.11	1.00

Source: Compiled by the authors.

In France, there's a moderate negative correlation (-0,71), suggesting that higher Clarity Scores might be associated with lower CTR. Latvia (0,51) and Lithuania (0,36) show moderate positive correlations, indicating that higher Clarity Scores tend to correlate with higher CTR. Estonia (0,22) and Spain (0,02) show weak positive correlations. When calculated for all countries combined, the correlation between Clarity Score and CTR is weak (0,11), thus it can be misleading not to consider cross-country differences. The correlation between Focus Score and CTR also differs across countries. Estonia shows a strong negative correlation (-0,71), suggesting that higher Focus Scores are associated with lower CTR. Lithuania (0,68) demonstrates a strong positive correlation, indicating that higher Focus Scores correlate to higher CTR. France (0,28) shows a weak positive correlation. While Spain (-0,20) and Latvia (-0,05) demonstrate a weak negative correlation. In all countries combined, the correlation between Focus Score and CTR is also weak but in contrast to Clarity Score also negative (-0,11).

These correlation coefficients are consistent with the patterns observed in the scatter plots, which also showed:

- A lack of strong linear relationships between the variables in many cases.
- Variability in the relationships across different countries.

The analysis reveals that the relationships between Clarity Score, Focus Score, and CTR are complex and vary by country. However, in many cases, these scores are still not strong predictors of CTR, suggesting that other factors may play a more significant role in influencing ad performance.

Finally, the regression analysis was conducted to model and predict CTR based on Clarity Score and Focus Score in each country with CTR as dependent variable. The regression analysis conducted for combined data (Appendix 11.1) shows a weak positive relationship (Multiple R: 0,1496) between the independent variables (Clarity Score and Focus Score) and the dependent variable (CTR). The R Square value of 0,0224 indicates that only 2.24% of the variation in CTR is explained by Clarity Score and Focus Score. The adjusted R-squared is negative (-0,0305), indicating that adding more variables (Clarity and Focus Scores) decreases the explanatory power of the model. The significance F value of 0,6578 also indicates that the Clarity and Focus scores do not significantly predict CTR. According to the coefficients table, for each one-unit increase in Clarity Score while Focus Score constant, CTR is estimated to increase by 0,001%, though this effect is also not statistically significant (P-value = 0,5200). For each one-unit increase in Focus Score while Clarity Score constant, CTR is estimated to decrease by 0,01%, though this effect is also not statistically significant (P-value = 0,5359). In conclusion, based on this regression analysis, Clarity Score and Focus Score are not good predictors of CTR in this dataset, when all countries are observed in a combined dataset.

Across all five countries, none of the regression models (Appendices 11.2-11.6) are statistically significant at the 0,05 alpha level. The explanatory power of the models (R-squared) varies, with Lithuania's model explaining the most variance (59,46%) and Spain's the least (11,99%). In most countries, both Clarity Score and Focus Score have statistically insignificant relationships with CTR. The direction of the relationships varies (some positive, some negative), but the effects are generally small. Adjusted R-squared values are often substantially lower than R-squared, or even negative, indicating that the inclusion of additional Clarity and Focus Scores data doesn't improve the model's ability to predict CTR. In conclusion, while there are varying degrees of correlation between Clarity/Focus Scores and CTR in each country, none of the regression models provide statistically significant evidence that these scores are reliable predictors of CTR. Given that the targeting, ad copy and platform choice were consistent across all countries, the results suggest that there are factors such as cultural, linguistic, and market-specific factors not included in the model.

3.2. The Eye-Tracking Study

3.2.1 Visibility of Penosil Advertisements in a Competitive Visual Environment

The first phase of the eye-tracking study focused on the visibility of Penosil banners within a context of so-called visual clutter, where advertisements were presented in a dense, multi-brand environment. The objective was to assess whether Penosil advertisements could visually stand out under such conditions. The analysis was limited to two B-version banners with different layouts, each featuring four AI-modified Penosil designs. One of the test scenarios (Noise 1) is illustrated in Figure 15.

Figure 15. Example of “Noise” slide from the second experiment.

Source: Author’s own material from the eye-tracking experiment

Although the experiment was not originally designed for inter-brand comparisons, the collected TFD dataset allowed for preliminary observations regarding the visual impact of Penosil advertisements. The average TFD values within Penosil AOIs ranged from 0,33 to 0,61 seconds, which was generally comparable to or higher than those of competing brands (e.g., Soudal4 = 0,24 s; GESealants1 = 0,32 s). The highest average was recorded for Penosil 3B (M = 0,61 s), followed by Pencil 4B (M = 0,57 s), suggesting that these particular designs were more visually attention-grabbing. Appendix 12 presents the mean TFD values for the Penosil banners in comparison to the average across the entire noise environment.

Given that the experimental environment was static, consistently designed, and all banners were displayed in identical size, it can be concluded that the AI-enhanced Penosil banners stood out in the competitive context and received more visual attention than several other brands. The majority of participants fixated on at least one Penosil banner, and the TFD results indicated above-average visibility. While no statistical comparisons with other brands were conducted, the descriptive analysis supports the assumption that Penosil advertisements were competitive within the test setting.

3.2.2 AOI-Level Comparison: Visual Performance of A- vs B-Version Banners

In the second phase of the analysis, a quantitative comparison was conducted at the AOI (Area of Interest) level. Two key eye-tracking metrics -Time to First Fixation (TFF) and Total Fixation Duration (TFD) - were compared between the A- and B-version banners. In addition, the proportion of participants who fixated on specific banner elements -such as the logo, product image, product name, and accompanying text -was examined. This enabled an assessment of whether the AI-based design modifications enhanced the visual prominence of these elements or instead dispersed attention.

The heatmaps of the banners reveal which areas received the most visual attention. Although the visual interpretation of heatmaps can be speculative and should be approached with caution, they are used illustratively in this analysis, based on conventions from previous research. The primary focus remains on the statistical comparison of TFF and TFD metrics. All corresponding values for the A- and B-version Penosil banners are provided in Appendices 13.1 and 13.2.

Figure 16 presents heatmaps that visualize gaze distribution across four Penosil banners (A – original design; B – AI-modified design).

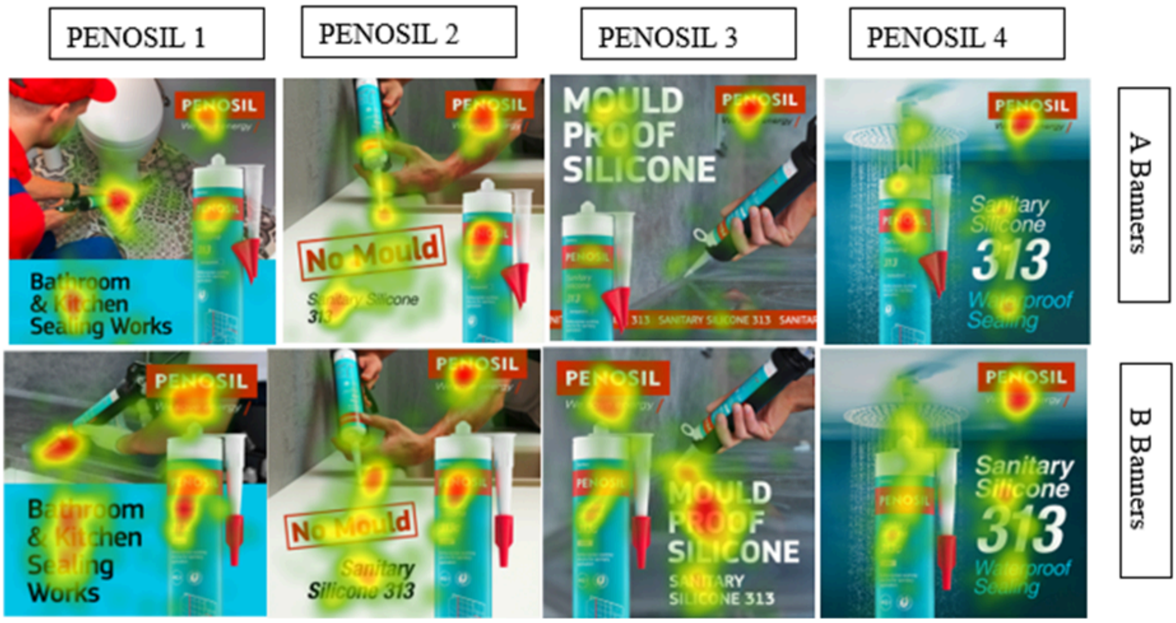


Figure 16. Heatmaps Gathered From the Eye-tracking Study.
Source: Author's own material from the eye-tracking experiment

The elements that received the most visual attention were the brand logo, product image, and large-format text. These elements were typically positioned near the center or top of the screen, employed strong visual contrasts, and were graphically emphasized—factors that help explain their high attention-capturing potential. Certain differences emerged between the A- and B-versions, particularly in the cases of Penosil 1 and Penosil 3. For instance, in the A-version of Penosil 1, participants' gaze was strongly drawn to the worker's face and hands, which distracted from key advertising elements. In the B-version, the visual prominence of the worker was reduced, redirecting attention more effectively toward the logo and product image. In the case of Penosil 3, participants' attention in the A-version was primarily focused on the text "Mould Proof Silicone," whereas in the B-version, attention was more evenly distributed across the product and logo areas. This suggests that the AI-based design modifications helped direct visual attention more effectively to content-relevant elements.

The results of the analysis showed that the visibility of AOIs (Areas of Interest) was generally higher in the AI-modified B-version banners compared to the original A-versions. The highest levels of attention in the B-versions were recorded for the "Product Image" and "Text" AOIs in the Penosil 1 banner, each noticed by 97,0% of participants (n = 32). A noticeable increase also occurred in the "Product Name" AOI of the Penosil 2 banner, where visibility rose from 81,8% in the A-version to 90,9% in the B-version. An exception was the

“Brand” AOI of Penosil 2, which showed slightly lower visibility in the B-version (72,7%) compared to the A-version (75,8%). Overall, the findings indicate that the AI-based design modifications enhanced the visual prominence of key elements, thereby supporting the more effective delivery of the advertising message. A comparison of AOI-level visibility between the A- and B-version banners is presented in Table 6.

Table 6

The results of an AOI-level visibility comparison between the A- and B-version banners.

		A Banner		B Banner	
Banner	AOI	Count (A)	% (A)	Count (B)	% (B)
PENOSIL 1	Brand	26	79	31	94
	Product Name	-	-	-	-
	Product Image	29	88	32	97
	Text	25	76	24	73
PENOSIL 2	Brand	27	82	30	91
	Product Name	29	88	32	97
	Product Image	33	100	32	97
	Text	25	76	28	85
PENOSIL 3	Brand	19	58	20	61
	Product Name	28	85	30	91
	Product Image	33	100	32	97
	Text	25	76	25	76
PENOSIL 4	Brand	28	85	32	97
	Product Name	33	100	33	100
	Product Image	22	67	20	61
	Text	26	79	31	94

Source: Eye-tracking data exported from Tobii software; compiled and formatted by the authors.

The comparison between the A- and B-versions revealed several meaningful differences in both Total Fixation Duration (TFD) and Time to First Fixation (TFF). For several AOIs, the average TFD was higher in the B-version banners, indicating increased visual engagement. For example, the TFD for the “Text” AOI in Penosil 1 increased from 0,93 seconds to 1,31 seconds, and for the “Product Image” AOI in Penosil 4, from 1.16 seconds to 1,31 seconds.

The longest average fixation time was observed for the “Text” AOI in the Penosil 3 B-version (1,49 s), suggesting strong visual prominence of the text. However, the TFF for this AOI was longer in the B-version (1,19 s) than in the A-version (0,22 s), even though one might intuitively expect faster detection. This may suggest that despite the text’s strong visual presence, its placement was not optimal for immediate attention. A similar pattern emerged in Penosil 2, where the “Brand” AOI was noticed more slowly in the B-version (TFF: 2,82 s) than in the A-version (2,18 s), although the TFD remained nearly the same.

The visual analysis of the heatmaps supports these findings. In the AI-modified banners, gaze distribution was more frequently concentrated on the “Product Image” and “Text” AOIs, while peripheral elements (e.g., background or tools) received less attention. This suggests that the AI-driven modifications improved visual guidance and helped direct attention toward intended content elements.

In summary, the AI-modified designs were able to enhance visual engagement in several cases and influenced which advertising elements attracted attention. However, the results also showed that TFD and TFF do not always correlate—an element may receive prolonged attention without being among the first to be noticed. An overview of the differences in TFF and TFD metrics between the A and B banners is presented in Table 7.

Table 7

Summary of AOI-Level TFF and TFD Differences Between A- and B-Versions

		A Banner		B Banner	
Banner	AOI	TTF (s)	TFD (s)	TTF (s)	TFD (s)
PENOSIL 1	Brand	1,53	0,56	1,88	0,6
	Product Name	-	-	-	-
	Product Image	1,83	0,72	1,31	1,02
	Text	1,34	0,93	1,14	1,31

PENOSIL 2	Brand	2,18	0,62	2,82	0,67
	Product Name	2,51	0,59	2,06	0,7
	Product Image	2,17	0,69	1,96	0,84
	Text	0,89	0,8	0,78	0,75
PENOSIL 3	Brand	1,94	0,75	1,61	0,78
	Product Name	3,15	0,45	3,11	0,49
	Product Image	1,97	0,7	0,41	0,83
	Text	0,22	1,2	1,19	1,49
PENOSIL 4	Brand	1,71	0,73	2,18	0,91
	Product Name	1,33	0,79	1,19	0,89
	Product Image	0,73	1,16	0,52	1,31
	Text	2,66	0,57	2,61	0,46

Source: Eye-tracking data exported from Tobii software; compiled and formatted by the authors.

Descriptive statistics alone (e.g., percentages and means) do not permit conclusions about whether the differences between A- and B-versions are statistically significant. Therefore, the next step involved hypothesis testing using SPSS to assess whether the AI-based design modifications (B-versions) led to significant changes in visual attention distribution compared to the original designs (A-versions).

The first objective was to determine whether there was a statistically significant difference in Total Fixation Duration (TFD) between the two independent groups—A and B. Initially, a t-test was considered for comparing the two groups. However, the Shapiro–Wilk test for normality (see Appendix 14) indicated that the assumption of normal distribution was not met in either group (A: $p = .006$; B: $p = .033$). As a result, the Mann–Whitney U test was selected as the appropriate non-parametric alternative for comparing TFD values between the two independent samples.

Since each AOI–banner combination had to be tested individually, the analysis became time-consuming and difficult to generalize. Therefore, the analysis was supplemented with a linear regression model (see Appendix 16) to evaluate the effect of the AI-modified

B-versions on visual attention across the entire dataset. The dependent variable in the regression model was TFD (Total Fixation Duration in seconds). The independent variables included:

- **Test_Code** (1 = A-version, 2 = B-version),
- **AOI_Code** (identifying the Area of Interest),
- **Banner** (specific banner number),
- **Familiarity** (participant's self-rated domain knowledge),
- **Language_Code** (participant's language group),
- **Age** (participant's age).

The explanatory power of the regression model was modest ($R^2 = 0{,}038$), yet the overall model was statistically significant: $F(6, 838) = 5.47$, $p < .001$. This indicates that at least some of the independent variables—particularly the experimental condition (A or B) — had a significant effect on the duration of visual attention.Although the coefficient of determination was low, this is common in behavioral and perception research. In data related to human behavior, high R^2 values are not typically expected; the key value lies in whether the independent variables reliably predict the dependent variable. The effects of the independent variables in the regression analysis are illustrated in Figure 17.

Independent Variable	B Coefficient	p-value	Statistical Significance	Interpretation
Katse_kood (1=A, 2=B)	0,083	0,08	Marginally significant (p < .10)	B-version banners increase TFD slightly, borderline significance
AOI_kood	0,066	0,001	Highly significant (p < .001)	AOI type strongly influences TFD duration
Familiarity	0,068	0,016	Significant (p < .05)	Greater familiarity increases TFD
Banner	0,029	0,131	Not significant	Banner number has no significant impact on TFD
Language_kood	0,036	0,152	Not significant	Language does not significantly affect TFD
Age	-0,006	0,041	Significant (p < .05)	Older participants tend to fixate shorter

Figure 17. Linear regression results predicting Total Fixation Duration (TFD) from selected independent variables.

Source: Data exported from SPSS software; compiled and formatted by the authors.

The result obtained from the regression model suggests that the AI-modified B-version banners may attract slightly longer visual attention than the original A-versions (*B* = 0,083), but this effect was not statistically significant at the 5% level (*p* = 0,080). Therefore, no definitive conclusion can be drawn from this model regarding the advantage of AI-modified designs in capturing visual attention.In addition, the following effects were observed:

Familiarity with the topic was a positive predictor (B = 0,068, p = .016), indicating that more at capturing viewer attention. Age had a negative effect on TFD (B = –0,006, p = .041), meaning that older participants spent less time fixating on AOIs—consistent with findings from previous studies.

The banner number did not have a statistically significant effect (p = .132), suggesting that the specific advertisement layout was not a determining factor—rather, the design version (A vs. B) played a more critical role.

The language background variable (*Language_kood*) included in the model did not have a statistically significant effect on TFD (*B* = 0,036, *p* = 0.152). Although the direction of the coefficient suggests that non-Estonian-speaking participants may have tended to view the advertisements slightly longer, the available data do not allow for reliable conclusions regarding the impact of language on visual attention.

3.2.3 Comparison of AI-Based Predictions and Eye-Tracking Experiment Results

The AI-based data were generated using a model that predicts eye movement trajectories and fixation density based on the visual characteristics of advertisements. The results were visualized as heatmaps and compared with real eye-tracking data from a laboratory experiment, in which participants’ gaze behavior and fixations were recorded while viewing the same banners.

Although AI does not attempt to capture the full extent of individual variability in human behavior, it can be used to assess whether the primary attention-attracting areas (e.g., logo, product image) sufficiently overlap with actual fixations. This AOI-based approach allows for meaningful conclusions about the applicability of AI in advertising analysis.

To assess how well the AI-predicted attention patterns aligned with actual gaze behavior, a correlation analysis was conducted between the relative TFD percentage (Total Fixation Duration, transformed into a proportion per AOI) measured in the laboratory data and the AI-predicted attention percentages (Appendix 18).

Pearson correlation analysis showed a moderate positive relationship between the two variables ($r = 0{,}536$, $p = 0{,}002$), which is statistically significant at the 1% level. As normal distribution could not be fully assumed and some values were ordinal, a non-parametric Spearman's rho test was also conducted. This test confirmed a moderate positive correlation ($\rho = 0{,}458$, $p = 0{,}011$), which is statistically significant at the 5% level.

These results indicate that AI-based visual attention prediction partially aligns with real eye-tracking data, but not entirely. While there is a general trend—higher AI-predicted attention corresponds to higher observed fixation time on AOIs — the relationship is not strong enough to consider AI a full substitute for laboratory testing.

The discrepancies may be due to the fact that AI models rely on assumptions of logical visual hierarchy, whereas actual human attention is influenced not only by design, but also by context, cultural differences, and individual perception patterns.

The second phase of the analysis aimed to assess how reliably an AI-based model can predict human visual attention. For this purpose, eye-tracking heatmaps recorded during the lab experiment were compared with AI-generated heatmaps predicting attention distribution across advertising banners. The heatmaps are provided in Appendices 3 and 4.

The results showed that, with regard to Total Fixation Duration (TFD), the AI model demonstrated substantial alignment with actual eye-tracking data. In particular, it accurately predicted concentrated attention in the areas of the logo and product image. This finding supports the conclusion by Juárez-Varón et al. (2024), who noted that AI can reliably predict which AOIs attract longer viewing times (e.g., average time viewed), although larger discrepancies were observed in Time to First Fixation (TFF), likely due to AI's assumptions about the sequential order of gaze movements.

Penosil 1 A-Banner Example:

AI-Based Attention Prediction (Figure 18, left):

The AI model predicted that visual attention would primarily concentrate on the human figure (application context), specifically the face and posture, as well as on the product logo and, to a lesser extent, the product image. The highest predicted attention was directed toward the craftsman's hand and the application area (glue line). Text elements were predicted to attract very little visual interest.

Eye-Tracking Experiment Results (Figure 18, right):

According to the eye-tracking data, most actual attention was directed at the product logo and product image. A notable portion of participants also fixated on the text segment *"Bathroom & Kitchen Sealing Works."* Gaze directed toward the human figure (builder) was present as well, but considerably less intense than predicted by the AI model.

Figure 18. Visual Attention on Banner 1A: Eye-Tracking (right) Versus AI Prediction.

Source: Heatmaps generated using Tobii Pro Lab and Attention Insight; compiled by the authors.

The results indicate that the AI model underestimated the importance of the verbal message and overestimated the visual impact of the activity context (application scene). The eye-tracking data revealed that branding and product information—logo, product image, and text—attracted more real attention than the AI model had predicted.

This example demonstrates that, while AI is capable of identifying certain visual attention focal points (e.g., the logo or product handling) with reasonable accuracy, it cannot fully replicate human responses to the content and context of advertisements. The limitations of AI are particularly evident in modeling attention to textual and contextually meaningful elements, where significant content may be undervalued. Therefore, AI-based prediction models require further refinement to better account for the communicative role of textual and contextual elements in advertising visuals.

The qualitative comparison of heatmaps revealed that the main visual anchors—logo, product image, and large-format text—overlapped between the AI-generated and eye-tracking-based data in most cases. Differences primarily emerged in the direction of the initial gaze and the timing of fixations, rather than in overall AOI-level visibility. Thus, it can be concluded that the attention patterns predicted by AI were sufficiently similar to the laboratory results to support the hypothesis that visual attention can, to a certain extent, be reliably modeled algorithmically.

The attention percentages predicted by AI for various AOIs (e.g., "Logo Attention", "Product Image") indicated that the model expected the highest visual focus on the logo and product image areas. The highest predicted logo attention was found in the Penosil 3B (35,8%) and 4B (35,5%) designs, while the largest predicted share of attention to the product image was in Penosil 4B (26,2%).

Comparison of these predictions with the TFD values from the eye-tracking experiment highlighted a number of meaningful overlaps:

- In **Penosil 4B**, AI predicted high attention to both the product image (26,2%) and the logo (35,5%), which aligned with the eye-tracking results (TFD: product image = 1,31 s; logo = 0,91 s).
- In **Penosil 3B**, AI predictions were also consistent with the laboratory data - logo (35,8%) and product image (22,8%) received high predicted attention, which matched observed TFD values (logo = 0,78 s; product image = 0,83 s).

However, some notable exceptions also emerged:

In **Penosil 4A**, AI predicted the highest attention to the text element (31,9%), but in the actual eye-tracking data, the total fixation time on text was lower (TFD = 0,57 s) compared to the product image (TFD = 1,16 s) and the logo (TFD = 0,91 s). This suggests that AI overestimated the visual appeal of the text, while human viewers preferred to focus on the visual brand elements (figure 4A).

Figure 19. Visual Attention on Banner Penosil 4A: Eye-Tracking (left) Versus AI Prediction. Source: Heatmaps generated using Tobii Pro Lab and Attention Insight; compiled by the authors.

These findings are consistent with the results of Juárez-Varón et al. (2024), who concluded that while AI can reliably predict general zones of attention (e.g., logo, product), it may be inaccurate in estimating gaze sequence and precise fixation timing.

A key observation was that the overlap between AI and eye-tracking results was especially strong in the B-version designs. This suggests that AI-supported visual optimization (e.g., clearer, simplified, visually focused layouts) enhances predictability and results in a more coordinated distribution of visual attention. Compositional simplicity appears to improve both actual and AI-modeled visual guidance.

In addition to the earlier analysis comparing Total Fixation Duration (TFD%) with AI-predicted attention percentages, a second analysis focused on the Viewers % metric. This represents the proportion of participants who noticed a specific AOI at all, allowing an assessment of which elements actually attracted viewers' attention. Unlike TFD%, which measures how long an area was viewed, Viewers % captures whether it was noticed in the first place.

While the TFD%-based analysis focused on the temporal dimension and revealed a strong positive correlation with AI predictions—i.e., higher AI-predicted attention corresponded to longer actual viewing—the Viewers % metric measures the binary aspect of attention (noticed or not), which may relate to AI predictions differently.

A correlation analysis between AI-predicted attention percentages and the Viewers % measured in the eye-tracking data revealed a weak negative association (Spearman's $\rho = -0{,}345$, $p = 0{,}062$), which fell just short of conventional statistical significance ($p < 0{,}05$) but trended toward marginal significance. A similar pattern was found using Pearson's correlation ($r = -0{,}329$, $p = 0{,}076$). While these results do not allow for firm conclusions, they indicate a potential trend worth further investigation.

The negative direction of the correlation does not reflect a methodological flaw but instead highlights the fundamental difference between the two metrics. AI models estimate attention based on visual salience, contrast, and layout—assuming that visually prominent elements will receive more attention. However, in real-world conditions, viewer behavior is also shaped by context, cultural background, and individual perception patterns. This may result in situations where an element deemed visually dominant by AI is not frequently noticed in practice.

For example, AI might predict high attention to a small but high-contrast logo that only a few participants actually noticed—though those who did looked at it for a relatively long time. Conversely, a large text block may have been noticed by many, yet received only brief attention—resulting in lower AI-predicted attention. Thus, Viewers % and AI-predicted attention % measure different phenomena: the former reflects reach (noticed or not), while the latter captures depth (how long it was viewed).

Consequently, AI-based predictions may not reliably reflect actual noticing behavior. The findings underscore that AI cannot replace eye-tracking—not only in terms of accuracy, but also in terms of what is being measured. AI models and laboratory experiments provide insight into different aspects of visual attention. Therefore, it is recommended that these approaches be used complementarily rather than interchangeably.

4. Dicussion

This research contributes to better understanding of the validity of AI eye-tracking predictions and suggestions in improving the performance of digital banner advertising and accuracy of AI-predicted visual attention predictions. This understanding is of crucial importance for determining the scope of tasks for which it should be used, as well as for the further investigation and optimisation of modern AI models.

4.1. Key Findings and Their Implications

The empirical campaign analysis conducted as part of this study indicated that the suggestions for improving AI design proposed by Attention Insight did not result in a statistically significant increase in the click-through rate (CTR). While the implementation of AI suggestions in B banners resulted in a minor increase in the mean click-through rate (1,62%) in comparison to A banners (1,59%), this variation was not statistically significant ($p > 0,05$). This finding suggests that AI-generated recommendations may not consistently enhance click-through rate (CTR), with observed outcomes more likely due to random chance than a systematic AI effect. This finding aligns with a broader academic caution against evaluating digital banner advertising solely on CTR (Bergkvist & Melander, 2000; Dreze & Hussherr, 2003; Briggs & Hollis, 1997). Consequently, while AI eye-tracking has the capacity to provide feedback on designs for improving visual attention, this study's exclusive emphasis on CTR did not encompass such potential benefits.

Furthermore, the correlation analysis between the AI design quality metrics (Clarity Score, Focus Score) and CTRs of digital banner designs revealed mixed insights. While the overall analysis of the campaign data did not demonstrate a strong correlation, a number of significant correlations were identified in specific countries. Specifically, the Clarity Score and the CTR in the combined data demonstrated a weak positive relationship (R-squared value = 0,0121). However, France exhibited a moderately higher positive correlation (R-squared value = 0,5106). The correlation between Focus Score and CTR in the combined data was also found to be insignificant, with an R-squared value of 0,0112. However, Estonia exhibited a moderate negative correlation (R-squared value = 0,5039) between Focus Score and CTR, while Lithuania demonstrated a moderate positive correlation (R-squared value = 0,4575). The observed correlations demonstrate significant variation across countries, suggesting that the relationships between Clarity Score, Focus Score, and CTR are intricate and country-specific. According to regression analysis, AI design quality metrics are not strong predictors of CTR ($p > 0,05$), suggesting that other factors may play a more significant role in influencing the performance of digital banner advertisements.

The eye-tracking study offered key insights into how AI recommendations affect visual attention in digital banners. Notably, B banners attracted more visual attention, with higher Total Fixation Duration (TFD) and improved visibility of logos, product images, and slogans. These results align with Pieters and Wedel's (2004) emphasis on branded, pictorial, and textual elements, and support earlier findings that strong visual hierarchy and reduced design complexity enhance ad effectiveness (Pieters et al., 2010; Orth & Crouch, 2014). A key finding is the moderate positive correlation between AI-predicted and actual eye-tracking data ($r = 0,536$, $p = .002$), confirmed by Pearson and Spearman tests. This suggests AI models can approximate bottom-up attention patterns. However, discrepancies, like delayed TFF, highlight AI’s limitations in predicting top-down attention, which is shaped by personal intentions and cultural context (Corbetta & Shulman, 2002). Therefore, tools like Attention Insight show promise but should be applied with context-awareness, especially in diverse campaigns. Finally, in a “noise” test, AI-enhanced banners performed as well as or better than competitors in terms of TFD. These descriptive results suggest AI-guided improvements can increase a banner’s visual salience in cluttered digital environments. In summary, the eye-tracking study supports the partial validity and practical utility of AI-generated attention predictions in guiding banner design. The regression model showed that AI can increase user engagement at a visual attention level ($p = 0,048$). Furthermore, AI-enhanced banners performed as well as or better than competitors in a "noise" test, indicating improved visual salience. This underscores the value of combining AI with traditional eye-tracking to capture both stimulus-driven and goal-directed attention.

4.3. Limitations and Future Research Directions

It is important to consider several limitations when interpreting the results of this study. The present study investigates solely the Attention Insight AI eye-tracking model; consequently, the results may not be applicable to other AI models. The execution of digital campaigns in five European countries solely in Google Ads, in addition to eye-tracking experiments conducted exclusively in Estonia, suggests that the conclusions may not be universally applicable to other platforms or diverse regions. Finally, the samples of digital banner advertising and eye-tracking data were limited to male subjects due to the target demographic for the products under discussion, namely construction chemicals.

In consideration of the study's findings and limitations, future research could explore several promising research directions. In order to enhance the applicability of findings regarding the practical utility of AI eye-tracking, it is recommended that future studies include a more extensive and varied sample of countries, industries, brands and AI models. A systematic examination of the role of cultural and linguistic factors in digital advertising performance, and their interaction with AI design recommendations, could offer valuable insights. In order to achieve a more holistic understanding of the practical utility of AI eye-tracking, it would be beneficial to incorporate a wider array of campaign performance metrics beyond click-through rate (CTR), including session duration or conversion rates. The impact of an iterative design process on banner design performance, in which AI suggestions are applied, tested and further refined, could be studied.

4.5. Overall Conclusion

Findings of the study demonstrate that AI suggestions on banner design improvement did not lead to statistically significant improvements in CTR. However, moderate correlations between AI design quality metrics, namely Clarity and Focus scores, and CTR emerged in specific countries, suggesting that external factors may play a significant role in campaign performance. In the eye-tracking experiment, AI predictions moderately aligned with human gaze behaviour in terms of TFD but showed clear differences in TFF and scanpath order, reflecting the limits of AI in modelling top-down attention. In conclusion, AI tools can be used in design decisions, especially in the early phases, but should be used as complementary solutions in addition to other scalable methods. Their practical utility and accuracy varies by context, and human interpretation remains essential. Future studies should explore emotional metrics and deeper cultural variables for a fuller understanding of user visual attention.

Author Contributions

The original research idea and the concept of comparing AI-based and traditional eye-tracking methods in international advertising were proposed by A.Z., who began gathering theoretical material on the topic already in 2023. However, most of the relevant scientific studies in this field emerged in 2024, which meant that although the research concept was well-formed early on, there was initially limited literature available to support it empirically. A.Z. curated the digital campaign study and collected relevant literature in the AI attention prediction field.

I.K. joined the project as a collaborator with experience and expertise in lab-based eye-tracking methodology. She conducted experimental studies in Tartu, Pärnu, and Viljandi and performed in-depth visual attention analysis using eye-tracking software. A.Z. conducted complementary eye-tracking experiments in Tallinn and was responsible for the digital campaign implementation and AI-related statistical analysis.

The theoretical framework was developed collaboratively, with A.Z. focusing on AI-driven visual attention modelling and I.K. on traditional neuromarketing techniques. Both authors jointly defined the methodology, each writing and describing their respective part (AI vs. lab-based approach).

Data analysis was shared: A.Z. performed statistical analysis using external tools (e.g., correlation analysis), and I.K. performed the statistical analysis performed in SPSS .

Both authors actively collaborated in interpreting the findings, developing tables and figures, responding to supervisor feedback, and finalising the manuscript. Equal contribution is acknowledged.

Funding

It is important to note that the present research has not been funded by any corporation or private entity. However, subscription fees for the Attention Insight platform and designs were paid by Penosil, a brand owned by Wolf Group OÜ. Moreover, the digital campaigns were funded by the daughter companies of the aforementioned brand.

Informed Consent Statement

All participants voluntarily took part in the eye-tracking experiment after being informed about the purpose and procedures of the study.

Acknowledgments

The authors would like to thank Junior Lecturer Kristian Pentus and Kerli Ploom, Junior Lecturer for supervision and critical feedback, and all study participants for their time and contribution during the eye-tracking sessions. The authors also thank Reena Naelapää for assistance with the eye-tracking study, and the Penosil team for providing banners and access to their Attention Insight account.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Abdessalem, H. B., Chaouachi, M., Boukadida, M., & Frasson, C. (2019). Toward Real-Time System Adaptation Using Excitement Detection from Eye Tracking. In *Lecture notes in computer science* (pp. 214–223). https://doi.org/10.1007/978-3-030-22244-4_26

2. Alam, M. Z., Moudud-Ul-Huq, S., Sadekin, M. N., Hassan, M. G., & Rahman, M. M. (2021). Influence of social distancing behavior and Cross-Cultural motivation on consumers' attitude to using M-Payment services. *Sustainability*, *13*(19), 10676. https://doi.org/10.3390/su131910676

3. Ali Taban, Et Al., R. (2018). Eye Tracking Based Directional Control System using Mobile Applications. *International Journal of Computing and Digital Systems*, *7*(6), 365–374. https://doi.org/10.12785/ijcds/070605

4. Alimardani, M., & Kaba, M. (2021). Deep learning for neuromarketing; Classification of user preference using EEG signals. In *AH2021: 12th Augmented Human International Conference* (pp. 1–7). https://doi.org/10.1145/3460881.3460930

5. Awad, M., & Khanna, R. (2015). Chapter 9: Deep Learning. In *Efficient learning machines: Theories, Concepts, and Applications for Engineers and System Designers* (pp. 167–184). Apress.

6. Baños-González, M., Baraybar-Fernández, A., & Rajas-Fernández, M. (2020). The application of neuromarketing techniques in the Spanish advertising industry: weaknesses and opportunities for development. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.02175

7. Bellman, R. (1978). *An introduction to artificial intelligence : can computers think?* Boyd & Fraser Pub. Co.

8. Bočaj, N., & Ahtik, J. (2023b). Effects of visual complexity of banner ads on website users' perceptions. *Applied Sciences*, *13*(24), 13317. https://doi.org/10.3390/app132413317

9. Borji, A., & Itti, L. (2012). State-of-the-Art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. https://doi.org/10.1109/tpami.2012.89

10. Brenninkmeijer, J., Schneider, T., & Woolgar, S. (2019). Witness and Silence in Neuromarketing: Managing the Gap between Science and Its Application. *Science Technology & Human Values*, *45*(1), 62–86. https://doi.org/10.1177/0162243919829222

11. Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: motivationally biased attention. *Motivation and Emotion*, *28*(3), 221–243. https://doi.org/10.1023/b:moem.0000040153.26156.ed

12. Carrozza, M. C., Oddo, C., Orvieto, S., di Minin, A., & Montemagni, G. (2019). Automation and autonomy: From a definition to the possible applications of artificial intelligence. *Biolaw Journal-Rivista Di Biodiritto* (3), 237–254. <Go to ISI>://WOS:000503075800014

13. Chandon, J., Chtourou, M. S., & Fortin, D. R. (2003). Effects of configuration and exposure levels in responses to web advertisements. *Journal of Advertising Research*, *43*(2), 217–229. https://doi.org/10.2501/jar-43-2-217-229

14. Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

15. Cornia, M., Baraldi, L., Serra, G., & Cucchiara, R. (2018). Predicting human eye fixations via an LSTM-Based Saliency Attentive model. *IEEE Transactions on Image Processing*, *27*(10), 5142–5154. https://doi.org/10.1109/tip.2018.2851672

16. Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, *16*(3), 031001. https://doi.org/10.1088/1741-2552/ab0ab5

17. Drèze, X., & Hussherr, F. (2003). Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, *17*(4), 8–23. https://doi.org/10.1002/dir.10063

18. Fabijan, A., Dmitriev, P., Olsson, H. H., Bosch, J., Vermeer, L., & Lewis, D. (2019). *Three Key Checklists and Remedies for Trustworthy Analysis of Online Controlled Experiments at Scale* (pp. 1–10). 41st ACM/IEEE International Conference on Software Engineering (ICSE). https://doi.org/10.1109/icse-seip.2019.00009

19. Fisher, C. E., Chin, L., & Klitzman, R. (2010). Defining Neuromarketing: Practices and professional challenges. *Harvard Review of Psychiatry*, *18*(4), 230–237. https://doi.org/10.3109/10673229.2010.496623

20. Flores, W., Chen, J. V., & Ross, W. H. (2013). The effect of variations in banner ad, type of product, website context, and language of advertising on Internet users' attitudes. *Computers in Human Behavior*, *31*, 37–47. https://doi.org/10.1016/j.chb.2013.10.006

21. Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly.

22. Goodrich, K. (2011). Anarchy of effects? Exploring attention to online advertising and multiple outcomes. *Psychology and Marketing*, *28*(4), 417–440. https://doi.org/10.1002/mar.20371

23. Google. (2025a). About Demand Gen campaigns - Google Ads Help. Google Ads Help. Retrieved May 21, 2025, from https://support.google.com/google-ads/answer/13695777?hl=en

24. Google. (2025b). Create A/B experiments for Demand Gen campaigns. Google Ads Help. Retrieved April 23, 2025, from https://support.google.com/google-ads/answer/13719071?hl=en&sjid=4727671776367489810-EU

25. Google. (2025c). Demand Gen campaign asset specs and best practices. Google Ads Help. Retrieved April 23, 2025, from https://support.google.com/google-ads/answer/13704860?hl=en&sjid=4727671776367489810-EU

26. Iankovets, T. (2023). Media planning of digital advertising campaigns. *Eastern-European Journal of Enterprise Technologies*, *6*(13 (126)), 42–53. https://doi.org/10.15587/1729-4061.2023.293074

27. Iloka, B. C., & Onyeke, K. J. (2020). Neuromarketing: a historical review. *Neuroscience Research Notes*, *3*(3), 27–35. https://doi.org/10.31117/neuroscirn.v3i3.54

28. Jiang, J. (2024). The eye of artificial intelligence - Convolutional Neural Networks. *Applied and Computational Engineering*, *76*(1), 273–279. https://doi.org/10.54254/2755-2721/76/20240613

29. Juárez-Varón, D., Mengual-Recuerda, A., Zuluaga, J. C. S., & Corvello, V. (2024). Application of artificial intelligence in neuromarketing to predict consumer behaviour towards brand stimuli. *International Journal of Software Science and Computational Intelligence*, *16*(1), 1–18. https://doi.org/10.4018/ijssci.347214

30. Juárez-Varón, D., Tur-Viñes, V., Rabasa-Dolado, A., & Polotskaya, K. (2020). An adaptive machine learning methodology applied to neuromarketing analysis: prediction of consumer behaviour regarding the key elements of the packaging design of an educational toy. *Social Sciences*, *9*(9), 162. https://doi.org/10.3390/socsci9090162

31. Kadner, F., Thomas, T., Hoppe, D., & Rothkopf, C. A. (2023). Improving saliency models' predictions of the next fixation with humans' intrinsic cost of gaze shifts. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2103–2113. https://doi.org/10.1109/wacv56688.2023.00214

32. Kalayci, C. B., Ertenlice, O., & Akbay, M. A. (2019). A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. *Expert Systems With Applications*, *125*, 345–368. https://doi.org/10.1016/j.eswa.2019.02.011

33. Kędras, M., & Sobecki, J. (2023). What is hidden in clear Sight and how to find It—A survey of the integration of artificial intelligence and eye tracking. *Information*, *14*(11), 624. https://doi.org/10.3390/info14110624

34. Keresteš, N. M., Golubović, G., Dedijer, S., Pavlović, Ž., & Janjić, T. (2024). *AI MODELS FOR PREDICTING VISUAL ATTENTION IN DIGITAL APPLICATIONS: A COMPARATIVE PILOT ANALYSIS WITH EYE TRACKING RESULTS*. https://doi.org/10.24867/grid-2024-p47

35. Kondak, A. (2023b). The application of eye tracking and artificial intelligence in contemporary marketing communication management. *Scientific Papers of Silesian University of Technology Organization and Management Series*, *2023*(186), 239–253. https://doi.org/10.29119/1641-3466.2023.186.18

36. Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

37. Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2018). Saliency benchmarking made easy: separating models, maps and metrics. In *Lecture notes in computer science* (pp. 798–814). https://doi.org/10.1007/978-3-030-01270-0_47

38. Kümmerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). Understanding Low- and High-Level contributions to fixation prediction. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 4799–4808). https://doi.org/10.1109/iccv.2017.513

39. Kohavi, R., & Foster, P. (1998). Glossary of Terms. *Machine Learning*, *30*(2/3), 271–274. https://doi.org/10.1023/a:1017181826899

40. Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108653985

41. Kusá, A., & Beličková, P. (2023). The role of artificial intelligence in neuromarketing research: Insights and applications. In *Conference Proceedings from the Annual International Scientific Conference „Marketing and Media Identity: AI – The Future of Today“* (pp. 269–275). https://doi.org/10.34135/mmidentity-2023-27

42. Lallé, S., Conati, C., & Carenini, G. (2016). Predicting Con-fusion in Information Visualization from Eye Tracking and Interaction Data. In *Conference Proceedings from the International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 2529–2535.

43. Lazar, J., Goldstein, D., & Taylor, A. (2015). Ensuring digital accessibility through process and policy. In *Elsevier eBooks* (1st ed.). Morgan Kaufmann. https://doi.org/10.1016/C2013-0-13367-3

44. Lee, J., & Ahn, J. (2012). Attention to banner ads and their effectiveness: An Eye-Tracking approach. *International Journal of Electronic Commerce*, *17*(1), 119–137. https://doi.org/10.2753/jec1086-4415170105

45. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

46. Lin, M., Cross, S. N., Jones, W. J., & Childers, T. L. (2018). Applying EEG in consumer neuroscience. *European Journal of Marketing*, *52*(1/2), 66–91. https://doi.org/10.1108/ejm-12-2016-0805

47. Lohtia, R., Donthu, N., & Hershberger, E. K. (2003). The impact of content and design elements on banner advertising click-through rates. *Journal of Advertising Research*, *43*(4), 410–418. https://doi.org/10.1017/s0021849903030459

48. Lu, W., & Jia, Y. (2015). Inferring User Preference in Good Abandonment from Eye Movements. In *Lecture notes in computer science* (pp. 457–460). https://doi.org/10.1007/978-3-319-21042-1_40

49. Marchesi, S., Abubshait, A., Kompatsiari, K., Wu, Y., & Wykowska, A. (2023). Cultural differences in joint attention and engagement in mutual gaze with a robot face. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-38704-7

50. Mas, J. M., Jaszus, K., Gómez, A., & Monleon, F. G. (2023). Image destination engagement from a cultural perspective—A neuromarketing approach for a Spanish-German study. *Smart Tourism*, *4*(2), 2463. https://doi.org/10.54517/st.v4i2.2463

51. McMains, S. A., & Kastner, S. (2008). Visual attention. In Binder, M. D., Hirokawa, N., & Windhorst, U. (Eds.), *Encyclopedia of Neuroscience* (pp. 4296–4302). Springer eBooks. https://doi.org/10.1007/978-3-540-29678-2_6344

52. MDPI. (2025). Behavioral Sciences: Instructions for Authors. Retrieved May 21, 2025, from https://www.mdpi.com/journal/behavsci/instructions

53. Mizrahi, R. (2021). Webcam Eye-Tracking: Use Machine Vision to Improve Customers' Attention Through Design Changes. Insights, 33, 14-15.

54. *MIT/Tuebingen Saliency Benchmark: About*. (n.d.). MIT/Tuebingen Saliency Benchmark. Retrieved May 20, 2025, from https://saliency.tuebingen.ai/

55. *MIT/Tuebingen Saliency Benchmark: Leaderboard MIT300*. (n.d.). MIT/Tuebingen Saliency Benchmark. Retrieved May 20, 2025, from https://saliency.tuebingen.ai/results.html

56. Neurons Inc. (2025-a). About us. Neurons. https://www.neuronsinc.com/about

57. Neurons Inc. (2025-b). Neurons AI insights and recommendations. Neurons Knowledge Base. https://knowledge.neuronsinc.com/ai-insights-and-recommendations

58. Neurons Inc. (2025-c). Neurons attention prediction. Neurons Knowledge Base. https://knowledge.neuronsinc.com/how-neurons-attention-prediction#eye-tracking

59. Neurons Inc. (2025-d). Running predictions with neurons AI. Neurons Knowledge Base. https://knowledge.neuronsinc.com/how-to-run-your-first-prediction

60. Orth, U. R., & Crouch, R. C. (2014). Is beauty in the aisles of the retailer? package processing in visually complex contexts. *Journal of Retailing*, *90*(4), 524–537. https://doi.org/10.1016/j.jretai.2014.05.004

61. Pan, B., Zhang, L., & Law, R. (2012). The complex matter of online hotel choice. *Cornell Hospitality Quarterly*, *54*(1), 74–83. https://doi.org/10.1177/1938965512463264

62. Pappas, I. O., Sharma, K., Mikalef, P., & Giannakos, M. N. (2020). How quickly can we predict users' ratings on aesthetic evaluations of websites? employing machine learning on Eye-Tracking data. In *Lecture notes in computer science* (pp. 429–440). https://doi.org/10.1007/978-3-030-45002-1_37

63. Pereira, M. H. N., De Melo, F. L. N. B., Soares, A. M. J., Ferreira, P. B. S., Da Silva, M. P., & Morya, E. (2024). Eye-Tracking como correlato fisiológico do comportamento do consumidor: *ReMark - Revista Brasileira De Marketing*, *23*(1), 300–365. https://doi.org/10.5585/remark.v23i1.23271

64. Peschel, A. O., & Orquin, J. L. (2013). A review of the findings and theories on surface size effects on visual attention. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00902

65. Pham, N. T., Bunruangses, M., Youplao, P., Garhwal, A., Ray, K., Roy, A., Boonkirdram, S., Yupapin, P., Jalil, M. A., Ali, J., Kaiser, S., Mahmud, M., Mallik, S., & Zhao, Z. (2023). An exploratory simulation study and prediction model on human brain behavior and activity using an integration of deep neural network and biosensor Rabi antenna. *Heliyon*, *9*(5), e15749. https://doi.org/10.1016/j.heliyon.2023.e15749

66. Pieters, R., Warlop, L., & Wedel, M. (2002). Breaking through the clutter: benefits of advertisement originality and familiarity for brand attention and memory. *Management Science*, *48*(6), 765–781. https://doi.org/10.1287/mnsc.48.6.765.192

67. Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, Pictorial, and Text-Size effects. *Journal of Marketing*, *68*(2), 36–50. https://doi.org/10.1509/jmkg.68.2.36.27794

68. Pieters, R., Wedel, M., & Batra, R. (2010). The Stopping Power of Advertising: Measures and Effects of Visual Complexity. *Journal of Marketing*, *74*(5), 48–60. https://doi.org/10.1509/jmkg.74.5.48

69. Plassmann, H., Venkatraman, V., Huettel, S., & Yoon, C. (2015). Consumer Neuroscience: applications, challenges, and possible solutions. *Journal of Marketing Research*, *52*(4), 427–435. https://doi.org/10.1509/jmr.14.0048

70. Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A Modern Approach*. Prentice Hall.

71. Rayner, K., Miller, B., & Rotello, C. M. (2007). Eye movements when looking at print advertisements: the goal of the viewer matters. *Applied Cognitive Psychology*, *22*(5), 697–707. https://doi.org/10.1002/acp.1389

72. Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology Applied*, *7*(3), 219–226. https://doi.org/10.1037/1076-898x.7.3.219

73. Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

74. Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

75. Sharma, K., Giannakos, M., & Dillenbourg, P. (2020). Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments*, *7*(1). https://doi.org/10.1186/s40561-020-00122-x

76. Šola, H. M., Qureshi, F. H., & Khawaja, S. (2024). Predicting Behaviour Patterns in Online and PDF Magazines with AI Eye-Tracking. *Behavioral Sciences*, *14*(8), 677. https://doi.org/10.3390/bs14080677

77. Šola, H. M., Qureshi, F. H., & Khawaja, S. (2025). AI and Eye Tracking Reveal Design Elements' Impact on E-Magazine Reader Engagement. *Education Sciences*, *15*(2), 203. https://doi.org/10.3390/educsci15020203

78. Stöckl, A., & Diaz, D. (2025). Cultural differences in perception and engagement of AI-generated online ads. *AHFE International*, *161*. https://doi.org/10.54941/ahfe1005916

79. Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5. https://doi.org/10.1167/11.5.5

80. Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, *67*(9), 703–715. https://doi.org/10.1016/j.ijhcs.2009.04.002

81. Tüfekci, Ö. K., & Akbiyik, F. (2023). CONCEPTUAL LEVEL STRATEGY RECOMMENDATIONS FOR UNDERSTANDING MARKETING COMMUNICATION EFFORTS IN DIGITAL CONSUMPTION CULTURE WITH NEUROSCIENCE TECHNIQUE. *Mehmet Akif Ersoy Üniversitesi İktisadi Ve İdari*

Bilimler Fakültesi Dergisi, *10*(2), 1710–1723. https://doi.org/10.30798/makuiibf.1282255

82. UAB Attention Insight. (2025a). *Features - AI heatmaps, Clarity Score, Areas of Interest | Attention insight*. Attention Insight. https://attentioninsight.com/features/

83. UAB Attention Insight. (2025b). *Predictive eye tracking technology explained | Attention insight*. Attention Insight. https://attentioninsight.com/technology/

84. Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science*, *19*(4), 297–312. https://doi.org/10.1287/mksc.19.4.297.11794

85. Winterhalter, L., Kofler, F., Ströbele, D., Othman, A., & Von See, C. (2024). AI-Assisted Diagnostics in Dentistry: An Eye-Tracking Study on User behavior. *Journal of Clinical and Experimental Dentistry*, e547–e555. https://doi.org/10.4317/jced.61491

Abbreviations

AI artificial intelligence

AOI area of interest

Attention Insight neuromarketing AI eye-tracking prediction software

AUC, sAUC metric used to measure the AI model performance in a 2 alternative forced choice task where the model has to decide which one of two locations has been fixated

Clarity Score AI design quality metric used to indicate the level of design clearness

CNN convolutional neural network

CTR click-through-rate

DL deep learning

EEG electroencephalogram

ET eye tracking

FC eye-tracking metric used to measure fixation count

fMRI functional magnetic resonance imaging

Focus Score AI design quality metric used to measure the level of attention density

MIT Massachusetts Institute of Technology

ML machine learning

TFD eye-tracking metric used to measure total fixation duration

TFF eye-tracking metric used to measure time to first fixation

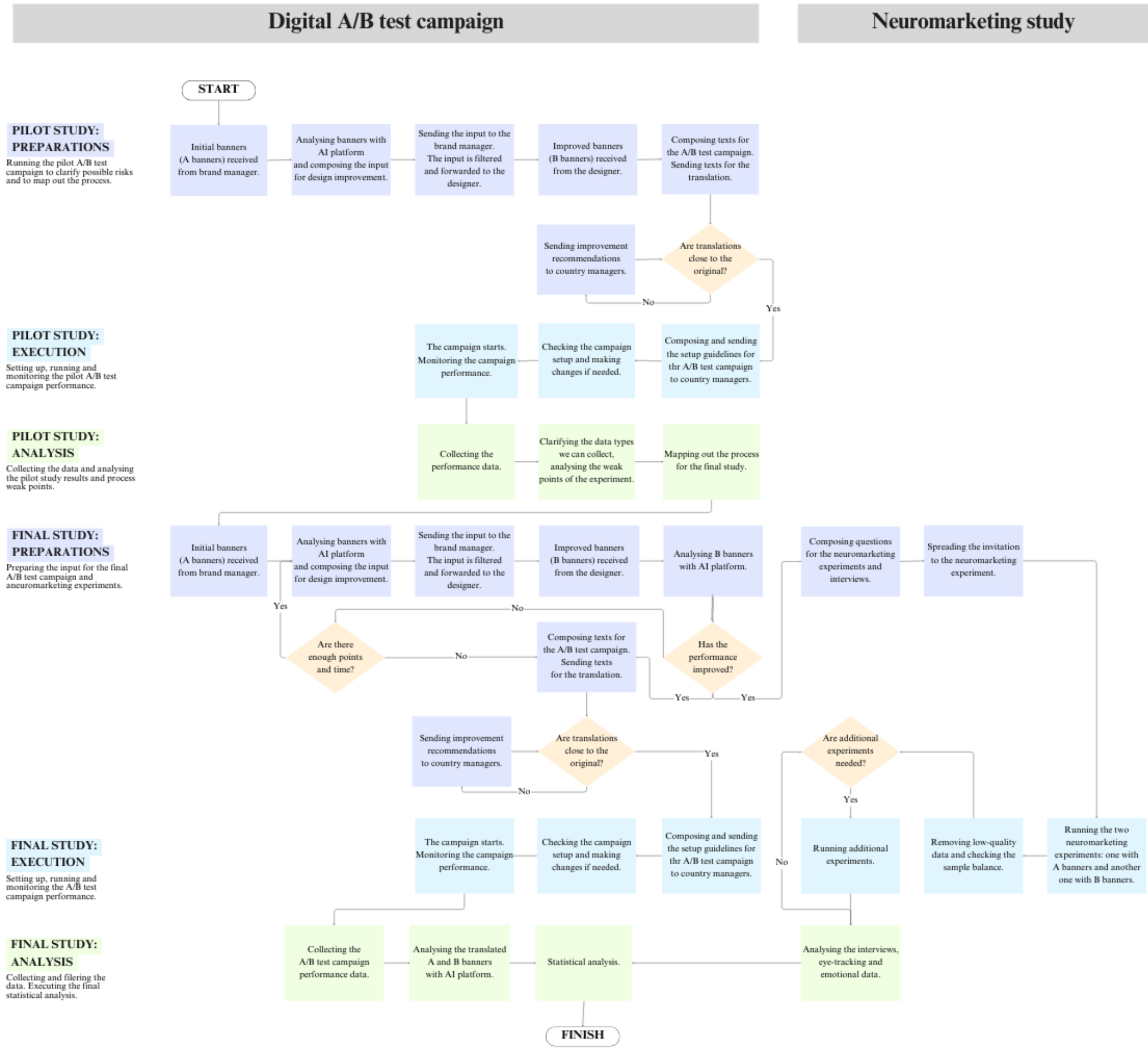
Tobii X2-30 eye-tracking device

SPSS Statistical Package for the Social Sciences

USD United States Dollar

Appendices

Appendix 1. Full Overview of the Research Process.

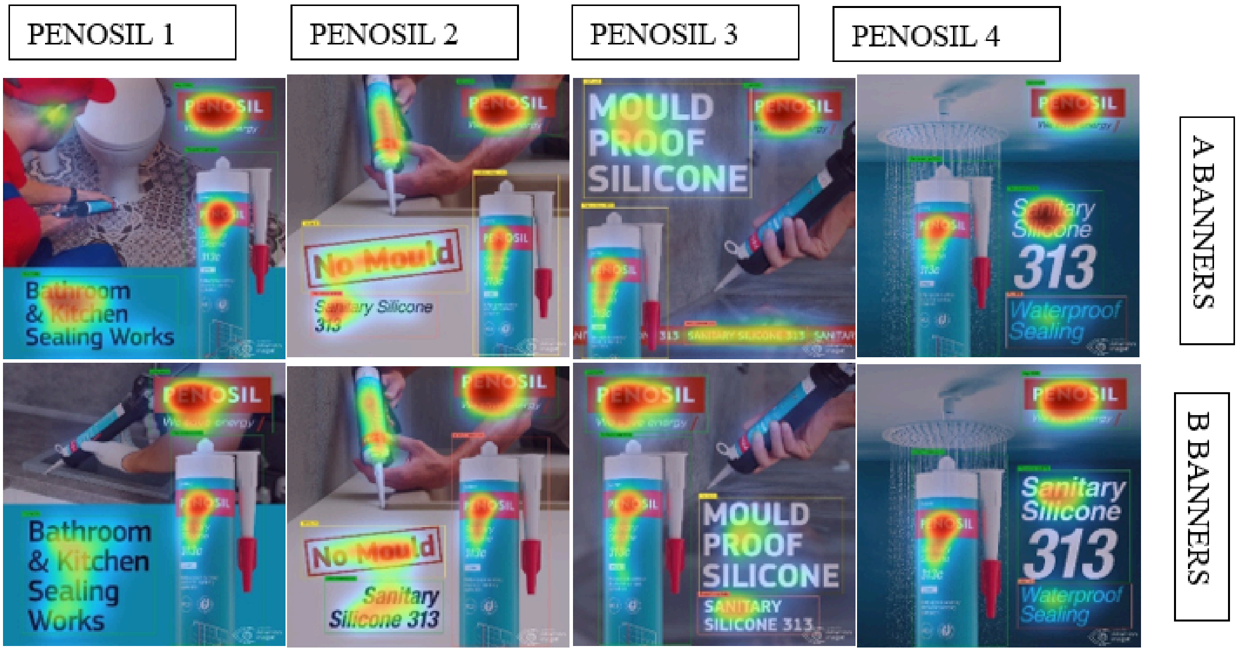


Appendix 2. Non-enhanced (A) and Enhanced (B) Banners.

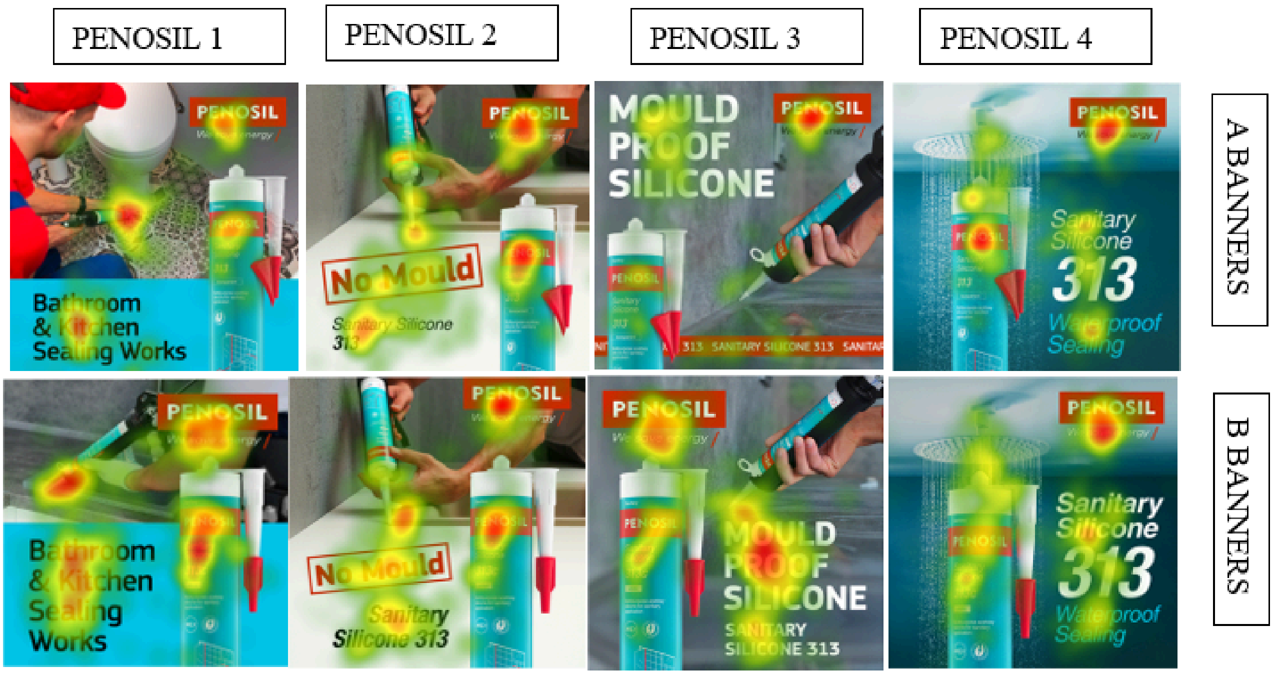
Appendix 2.1. Non-enhanced (A) Banners.

Appendix 2.2. Enhanced (B) Banners.

Appendix 3. AI-generated Visual Attention Predictions.



Appendix 4. Heatmaps Gathered From the Eye-tracking Study.



Appendix 5. Interview Questions of the Eye-tracking Study.

Appendix 5.1. Estonian Questions.

Before the experiment:

1. Sugu
2. Vanus
3. Nägemise kvaliteet
4. Põhiline suhtluskeel
5. Kust oled pärit?
6. Kui hästi oled tuttav ehituskemikaalidega, nagu PU vahud või muud sarnased tooted?

   Vastusevalikud (skaala 1–5):

   1 - Ei ole üldse tuttav ega ole kasutanud

   2 - Olen kuulnud, kuid ei ole kasutanud

   3 - Olen pisut tuttav ja olen näinud neid kasutusel

   4 - Olen tuttav ja oskan ehituskemikaale kasutada

   5 - Olen ekspert
7. Kas Sa oskad nimetada mingeid brände?

After the experiment:

8. Milliseid tooteid näidatud bänneritel reklaamiti?
9. Milliseid bännereid panite tähele?

   9.1. Kas Sulle jäi meelde reklaamitava brändi nimi?
10. Milline bänner meeldis Sulle enim?
11. Milline bänner ei meeldinud?
12. Mis brände bänneritel reklaamiti?
13. Mitu erinevat Penosili bännerit Sa nägid?
14. Mis Penosili bänner on Sulle meelde jäänud?

Appendix 5.2. English Questions.

Before the experiment:

1. Sex
2. Age
3. Vision quality
4. Primary language of communication
5. Where are You from?

6. How familiar are You with construction chemicals such as PU foams or similar products? Answer choices (scale 1-5):

   1 - Not at all familiar or have not used

   2 - I have heard of, but not used

   3 - Somewhat familiar and have seen them in use

   4 - Familiar and know how to use construction chemicals

   5 - I am an expert
7. Can You name any brands?

After the experiment:

8. Which products were advertised on the banners shown?
9. Which banners did You notice?
   9.1. Did You remember the name of the brand advertised?
10. Which banner did You like best?
11. Which banner did You dislike?
12. Which brands were advertised on the banners?
13. How many different Penosil banners did You see?
14. Which Penosil banner do You remember?

Appendix 5.3. Russian Questions.

Before the experiment:

1. Пол
2. Возраст
3. Качество зрения
4. Основной язык общения
5. Откуда Вы родом?
6. Насколько Вы знакомы со строительной химией, такой как пенополиуретан или аналогичные продукты? Варианты ответов (шкала 1-5):

   1 - Совсем не знаком или не пользовался

   2 - Слышал, но не использовал

   3 - В некоторой степени знаком и видел их в использовании

   4 - Знаком и знаю, как использовать строительную химию

   5 - Я эксперт
7. Можете ли Вы назвать какие-либо бренды?

After the experiment:

8. Какие продукты рекламировались на баннерах?
9. Какие баннеры Вы заметили?
    9.1. Запомнили ли Вы название рекламируемого бренда?
10. Какой баннер Вам понравился больше всего?
11. Какой баннер Вам не понравился?
12. Какие бренды рекламировались на баннерах?
13. Сколько разных баннеров Penosil Вы видели?
14. Какой баннер Penosil Вам запомнился?

Appendix 6. Socio-demographic Data of the Interview Participants.

https://docs.google.com/spreadsheets/d/1GFk8c7pEoaR2lP1nPg9q0GVzo_OkrztW0kvFFNrBNco/edit?gid=0#gid=0

Appendix 7. Digital Campaign Guidelines.

Appendix 7.1. Pilot Digital Campaign Guidelines.

https://docs.google.com/document/d/1IyuLmdr7Wm2c89K1jAP5opb4txaeVRoI/edit?usp=sharing&ouid=109553874519111849946&rtpof=true&sd=true

Appendix 7.2. Final Digital Campaign Guidelines.

https://docs.google.com/document/d/1zM4_nCSpLiQ3KCLv2s_XH9--Ve-0Wr25/edit?usp=sharing&ouid=109553874519111849946&rtpof=true&sd=true

Appendix 8. The Results of the Final International Digital Campaign.

Country	Design Type	Clarity Score	Focus Score	CTR
Estonia	1A	52	67	0,73%
Estonia	2A	56	56	3,26%
Estonia	3A	48	76	0,48%
Estonia	4A	60	68	0,35%
Estonia	1B	57	77	1,03%
Estonia	2B	59	61	1,41%
Estonia	3B	54	67	0,47%
Estonia	4B	58	68	1,17%
France	1A	45	76	1,83%
France	2A	57	60	1,52%
France	3A	44	76	1,60%
France	4A	63	68	1,39%
France	1B	54	67	1,47%
France	2B	57	60	1,47%
France	3B	56	67	1,72%
France	4B	59	78	1,42%
Latvia	1A	55	67	0,38%
Latvia	2A	62	56	1,25%
Latvia	3A	29	66	0,66%
Latvia	4A	64	68	1,06%
Latvia	1B	64	68	0,97%
Latvia	2B	60	53	0,83%
Latvia	3B	57	77	0,94%

Latvia	4B	62	68	1,45%
Spain	1A	61	66	1,89%
Spain	2A	51	56	2,45%
Spain	3A	43	60	1,86%
Spain	4A	55	57	2,97%
Spain	1B	58	67	1,47%
Spain	2B	51	52	1,70%
Spain	3B	46	60	1,76%
Spain	4B	50	67	2,45%
Lithuania	1A	55	67	1,86%
Lithuania	2A	60	56	2,11%
Lithuania	3A	30	66	1,32%
Lithuania	4A	64	68	2,82%
Lithuania	1B	61	77	3,51%
Lithuania	2B	61	56	0,54%
Lithuania	3B	57	67	3,84%
Lithuania	4B	61	68	2,76%

Appendix 9. The Descriptive Statistics of the Digital campaign

Appendix 9.1. Clarity Score by Country

Clarity Score by Country	Mean	Median	Variance	Standard deviation
Estonia	55,50	56,50	16,00	4,00
France	54,38	56,50	43,98	6,63
Latvia	56,63	61,00	134,84	11,61
Spain	51,88	51,00	35,55	5,96
Lithuania	56,13	60,50	118,98	10,91
All countries combined	54,90	57,00	65,63	8,10

Appendix 9.2. Focus Score by Country

Focus Score by Country	Mean	Median	Variance	Standard deviation
Estonia	67,50	67,50	48,29	6,95
France	69,00	67,50	50,00	7,07
Latvia	65,38	67,50	57,13	7,56
Spain	60,63	60,00	31,41	5,60
Lithuania	65,63	67,00	47,13	6,86
All countries combined	65,63	67,00	50,19	7,08

Appendix 9.3. Click-through-rate by Country

CTR by Country	Mean	Median	Variance	Standard deviation
Estonia	1.1125%	0.8800%	0.0089%	0.9450%
France	1.5525%	1.4950%	0.0002%	0.1540%
Latvia	0.9425%	0.9550%	0.0011%	0.3329%
Spain	2.0688%	1.8750%	0.0025%	0.5026%
Lithuania	2.3450%	2.4350%	0.0123%	1.1077%
All countries combined	1.6043%	1.4700%	0.0075%	0.8642%

Appendix 10. ANOVA Single-Factor Test

Appendix 10.1. ANOVA Test for Clarity Score Means By Country

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Estonia	8	444	55,5	16
France	8	435	54,375	43,9821429
Latvia	8	453	56,625	134,839286
Spain	8	415	51,875	35,5535714
Lithuania	8	449	56,125	118,982143

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	114,1	4	28,525	0,40824985	0,80143328	2,641465186
Within Groups	2445,5	35	69,8714286			
Total	2559,6	39				

Appendix 10.2. ANOVA Test for Focus Score Means By Country

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Estonia	8	540	67,5	48,2857143
France	8	552	69	50
Latvia	8	523	65,375	57,125
Spain	8	485	60,625	31,4107143
Lithuania	8	525	65,625	47,125

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	319,75	4	79,9375	1,70845737	0,17017778	2,64146519
Within Groups	1637,625	35	46,7892857			
Total	1957,375	39				

Appendix 10.3. ANOVA Test for Click-through-rate Means By Country

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Estonia	8	0,089	0,011125	8,9299E-05
France	8	0,1242	0,015525	2,37E-06
Latvia	8	0,0754	0,009425	1,1079E-05
Spain	8	0,1655	0,0206875	2,5261E-05
Lithuania	8	0,1876	0,02345	0,0001227

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0,0011575	4	0,00028938	5,77105189	0,00112412	2,64146519
Within Groups	0,00175499	35	5,0143E-05			
Total	0,0029125	39				

Appendix 11. Regression Analysis between AI metrics and CTR by Country

Appendix 11.1. Regression Analysis between AI Metrics and CTR for All Countries

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,14962664
R Square	0,02238813
Adjusted R Square	-0,0304558
Standard Error	0,00877233
Observations	40

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	6,5205E-05	3,2603E-05	0,4236655	0,65777841
Residual	37	0,00284729	7,6954E-05		
Total	39	0,0029125			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,01799282	0,01650921	1,08986582	0,28281966	-0,015458	0,05144366	-0,015458	0,05144366
Clarity Score	0,00011271	0,00017354	0,6494891	0,52003421	-0,0002389	0,00046434	-0,0002389	0,00046434
Focus Score	-0,000124	0,00019845	-0,624902	0,53587045	-0,0005261	0,00027808	-0,0005261	0,00027808

RESIDUAL OUTPUT

PROBABILITY OUTPUT

Observation	*Predicted CTR*	*Residuals*
1	0,01554512	-0,0082451
2	0,01736009	0,01523991
3	0,01397817	-0,0091782
4	0,01632281	-0,0128228
5	0,01486857	-0,0045686
6	0,01707817	-0,0029782
7	0,01577054	-0,0110705
8	0,01609738	-0,0043974
9	0,01364003	0,00465997
10	0,01697676	-0,0017768
11	0,01352732	0,00247268
12	0,01666094	-0,0027609
13	0,01577054	-0,0010705
14	0,01697676	-0,0022768
15	0,01599597	0,00120403
16	0,01496998	-0,00077
17	0,01588326	-0,0120833
18	0,01803637	-0,0055364
19	0,01307674	-0,0064767
20	0,01677366	-0,0061737
21	0,01677366	-0,0070737
22	0,01818298	-0,009883
23	0,01486857	-0,0054686
24	0,01654823	-0,0020482
25	0,01668354	0,00221646

Percentile	*CTR*
1,25	0,0035
3,75	0,0038
6,25	0,0047
8,75	0,0048
11,25	0,0054
13,75	0,0066
16,25	0,0073
18,75	0,0083
21,25	0,0094
23,75	0,0097
26,25	0,0103
28,75	0,0106
31,25	0,0117
33,75	0,0125
36,25	0,0132
38,75	0,0139
41,25	0,0141
43,75	0,0142
46,25	0,0145
48,75	0,0147
51,25	0,0147
53,75	0,0147
56,25	0,0152
58,75	0,016
61,25	0,017

26	0,01679653	0,00770347	63,75	0,0172
27	0,01539878	0,00320122	66,25	0,0176
28	0,01712337	0,01257663	68,75	0,0183
29	0,01622139	-0,0015214	71,25	0,0186
30	0,01729257	-0,0002926	73,75	0,0186
31	0,01573692	0,00186308	76,25	0,0189
32	0,01531969	0,00918031	78,75	0,0211
33	0,01588326	0,00271674	81,25	0,0245
34	0,01781094	0,00328906	83,75	0,0245
35	0,01318945	1,0546E-05	86,25	0,0276
36	0,01677366	0,01142634	88,75	0,0282
37	0,01531942	0,01978058	91,25	0,0297
38	0,01792365	-0,0125237	93,75	0,0326
39	0,01610868	0,02229132	96,25	0,0351
40	0,01643552	0,01116448	98,75	0,0384

Appendix 11.2. Regression Analysis between AI Metrics and CTR for Estonia

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,71156001
R Square	0,50631764
Adjusted R Square	0,3088447
Standard Error	0,00785619
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0,0003165	0,00015825	2,56398489	0,17124544
Residual	5	0,0003086	6,172E-05		
Total	7	0,0006251			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,08496897	0,06290899	1,35066507	0,23471012	-0,0767437	0,24668168	-0,0767437	0,24668168
Clarity Score	-0,0001242	0,00079907	-0,1554598	0,88254001	-0,0021783	0,00192984	-0,0021783	0,00192984
Focus Score	-0,0009918	0,00045997	-2,1563109	0,08357565	-0,0021742	0,00019055	-0,0021742	0,00019055

RESIDUAL OUTPUT

Observation	*Predicted CTR*	*Residuals*
1	0,0120557	-0,0047557
2	0,02246912	0,01013088
3	0,00362598	0,00117402
4	0,01007008	-0,0065701
5	0,00151613	0,00878387
6	0,01713722	-0,0030372
7	0,01180726	-0,0071073
8	0,01031852	0,00138148

PROBABILITY OUTPUT

Percentile	*CTR*
6,25	0,0035
18,75	0,0047
31,25	0,0048
43,75	0,0073
56,25	0,0103
68,75	0,0117
81,25	0,0141
93,75	0,0326

Appendix 11.3. Regression Analysis between AI Metrics and CTR for France

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,7178426
R Square	0,515298
Adjusted R Square	0,3214172
Standard Error	0,00126836
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	8,5514E-06	4,2757E-06	2,65780828	0,16356374
Residual	5	8,0436E-06	1,6087E-06		
Total	7	1,6595E-05			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,02618051	0,00840188	3,11602994	0,02637031	0,00458279	0,04777824	0,00458279	0,04777824
Clarity Score	-0,0001744	8,2003E-05	-2,1271838	0,08671283	-0,0003852	3,636E-05	-0,0003852	3,636E-05
Focus Score	-1,696E-05	7,691E-05	-0,2205757	0,83414669	-0,0002147	0,00018074	-0,0002147	0,00018074

RESIDUAL OUTPUT

Observation	*Predicted CTR*	*Residuals*
1	0,01704159	0,00125841
2	0,01521979	-1,979E-05
3	0,01721602	-0,001216

PROBABILITY OUTPUT

Percentile	*CTR*
6,25	0,0139
18,75	0,0142
31,25	0,0147

4	0,01403745	-0,0001375	43,75	0,0147
5	0,01562434	-0,0009243	56,25	0,0152
6	0,01521979	-0,0005198	68,75	0,016
7	0,01527547	0,00192453	81,25	0,0172
8	0,01456555	-0,0003656	93,75	0,0183

Appendix 11.4. Regression Analysis between AI Metrics and CTR for Latvia

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,50962226
R Square	0,25971485
Adjusted R Square	-0,0363992
Standard Error	0,00338859
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2,0142E-05	1,0071E-05	0,87707706	0,47151726
Residual	5	5,7413E-05	1,1483E-05		
Total	7	7,7555E-05			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,00131964	0,01330146	0,09921011	0,92482628	-0,0328729	0,03551213	-0,0328729	0,03551213
Clarity Score	0,00014593	0,00011074	1,31780539	0,24471546	-0,0001387	0,0004306	-0,0001387	0,0004306

Focus Score	-2,418E-06	0,00017014	-0,0142127	0,98920993	-0,0004398	0,00043493	-0,0004398	0,00043493

RESIDUAL OUTPUT

Observation	*Predicted CTR*	*Residuals*
1	0,00918393	-0,0053839
2	0,01023206	0,00226794
3	0,0053921	0,0012079
4	0,01049491	0,00010509
5	0,01049491	-0,0007949
6	0,00994745	-0,0016474
7	0,00945161	-5,161E-05
8	0,01020304	0,00429696

PROBABILITY OUTPUT

Percentile	*CTR*
6,25	0,0038
18,75	0,0066
31,25	0,0083
43,75	0,0094
56,25	0,0097
68,75	0,0106
81,25	0,0125
93,75	0,0145

Appendix 11.5. Regression Analysis between AI Metrics and CTR for Spain

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,22079442
R Square	0,04875018
Adjusted R Square	-0,3317498
Standard Error	0,00580014
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	8,6204E-06	4,3102E-06	0,12812138	0,88254422
Residual	5	0,00016821	3,3642E-05		

Total	7	0,00017683

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,02878836	0,02629187	1,09495282	0,32345627	-0,0387971	0,09637377	-0,0387971	0,09637377
Clarity Score	9,1277E-05	0,00039508	0,23103351	0,82644507	-0,0009243	0,00110686	-0,0009243	0,00110686
Focus Score	-0,0002117	0,00042033	-0,5037154	0,63586478	-0,0012922	0,00086876	-0,0012922	0,00086876

RESIDUAL OUTPUT

Observation	*Predicted CTR*	*Residuals*
1	0,02038238	-0,0014824
2	0,02158686	0,00291314
3	0,02000975	-0,0014097
4	0,02174024	0,00795976
5	0,01989682	-0,0051968
6	0,02243376	-0,0054338
7	0,02028358	-0,0026836
8	0,01916661	0,00533339

PROBABILITY OUTPUT

Percentile	*CTR*
6,25	0,0147
18,75	0,017
31,25	0,0176
43,75	0,0186
56,25	0,0189
68,75	0,0245
81,25	0,0245
93,75	0,0297

Appendix 11.6. Regression Analysis between AI Metrics and CTR for Lithuania

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,77108713
R Square	0,59457537

Adjusted R Square	0,43240552
Standard Error	0,00834539
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0,00051069	0,00025535	3,66637424	0,1046587
Residual	5	0,00034823	6,9645E-05		
Total	7	0,00085892			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	-0,0699253	0,03460906	-2,0204334	0,09931537	-0,1588907	0,01904012	-0,158890703	0,01904012
	0,00037599	0,00028921	1,30003651	0,25028906	-0,0003675	0,00111943	-0,000367456	0,00111943
	0,0011013	0,00045955	2,39649735	0,06188825	-8E-05	0,00228261	-8,00001E-05	0,00228261

RESIDUAL OUTPUT

Observation	*Predicted*	*Residuals*
1	0,02454131	-0,0059413
2	0,01430689	0,00679311
3	0,01404037	-0,0008404
4	0,02902648	-0,0008265
5	0,03781027	-0,0027103
6	0,01468287	-0,0092829

PROBABILITY OUTPUT

Percentile	
6,25	0,0054
18,75	0,0132
31,25	0,0186
43,75	0,0211
56,25	0,0276
68,75	0,0282

7	0,02529328	0,01310672	81,25	0,0351
8	0,02789853	-0,0002985	93,75	0,0384

Appendix 12. Mean TFD Values for PenosilB Banners Compared to the Overall Noise Environment

AOI_clean	Mean	N	Std. Deviation	Minimum	Maximum
GESealants1	0,32	10	0,18	0,09	0,67
GESealants2	0,37	14	0,21	0,09	0,82
Mortafil1	0,27	10	0,13	0,09	0,49
Onebod1	0,50	6	0,39	0,09	1,12
OneBond1	0,32	12	0,32	0,08	1,22
Onebond2	0,51	5	0,17	0,23	0,66
OneBond2	0,29	2	0,06	0,24	0,33
Onebond3	0,27	6	0,13	0,08	0,43
OneBond3	0,32	9	0,16	0,15	0,60
Penosil1B	0,33	18	0,26	0,06	1,09
Penosil2B	0,33	19	0,18	0,09	0,81
Penosil3B	0,61	12	0,34	0,06	1,25
Penosil4B	0,57	20	0,47	0,16	2,07
Pioner1	0,35	16	0,20	0,09	0,70
Pioner2	0,31	11	0,11	0,17	0,54
Sealants1	0,39	7	0,28	0,12	0,80
Sealants2	0,30	6	0,18	0,08	0,55
Soudal2	0,65	17	0,37	0,10	1,29
Soudal3	0,37	12	0,16	0,18	0,71
Soudal4	0,24	16	0,16	0,06	0,58
Total	0,39	228	0,28	0,06	2,07

Appendix 13. Descriptive Statistics of Eye-Tracking Data

Appendix 13.1. Descriptive Statistics of A (original) Banners

AOI	*N*	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
TFD_Penosil1.png_Brand_Mean	26	0,06	1,41	0,563	0,326
TFD_Penosil2.png_Brand_Mean	25	0,08	2,17	0,622	0,459
TFD_Penosil3.png_Brand_Mean	25	0,03	3,5	0,754	0,704
TFD_Penosil4.png_Brand_Mean	25	0,11	1,52	0,726	0,395
TFF_Penosil1.png_Brand_Mean	26	0,00	4,54	1,533	1,260
TFF_Penosil1.png_Product image_Mean	30	0,00	4,97	1,831	1,157
TFF_Penosil1.png_Text_Mean	29	0,00	3,12	1,336	0,940
TFF_Penosil2.png_Brand_Mean	25	0,00	4,84	2,177	1,485
TFF_Penosil2.png_Product image_Mean	29	0,00	4,65	2,170	1,296
TFF_Penosil2.png_Product name_Mean	27	0,67	4,76	2,511	1,273
TFF_Penosil2.png_Text_Mean	33	0,01	3,11	0,888	0,894
TFF_Penosil3.png_Brand_Mean	25	0,38	5	1,942	1,313
TFF_Penosil3.png_Product image_Mean	28	0,00	3,83	1,972	1,043
TFF_Penosil3.png_Product name_Mean	19	1,57	4,79	3,151	0,872
TFF_Penosil3.png_Text_Mean	33	0,00	2,63	0,219	0,527
TFF_Penosil4.png_Brand_Mean	25	0,00	4,37	1,707	1,352

TFF_Penosil4.png_Text_Mean	22	1,23	3,99	2,664	0,842
TFD_Penosil1.png_Product image_Mean	30	0,05	2,14	0,718	0,467
TFD_Penosil1.png_Text_Mean	29	0,09	2,29	0,933	0,562
TFD_Penosil2.png_Product image_Mean	29	0,06	1,71	0,689	0,378
TFD_Penosil2.png_Product name_Mean	27	0,08	1,71	0,589	0,403
TFD_Penosil2.png_Text_Mean	33	0,06	2,06	0,798	0,530
TFD_Penosil3.png_Product image_Mean	28	0,09	2,14	0,696	0,532
TFD_Penosil3.png_Product name_Mean	19	0,08	1,33	0,453	0,404
TFD_Penosil4.png_Product name_Mean	28	0,15	1,9	0,793	0,476
TFD_Penosil4.png_Text_Mean	22	0,10	1,43	0,575	0,327
TFD_Penosil4.png_Product image_Mean	33	0,15	4,73	1,158	0,858
TFD_Penosil3.png_Text_Mean	33	0,08	2,71	1,203	0,692
TFF_Penosil4.png_Product name_Mean	28	0,00	4,34	1,331	1,021
TFF_Penosil4.png_Product image_Mean	33	0,00	3,94	0,733	0,932

Appendix 13.2. Descriptive Statistics of the B Banners

AOI	*N*	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
TFF_Penosil1B.png_Brand_Mean	31	0,28	4,81	1,882	1,405
TFF_Penosil1B.png_Product iimage_Mean	32	0,00	4,81	1,313	1,342
TFF_Penosil1B.png_Text_Mean	32	0,00	4,84	1,144	1,201
TFF_Penosil2B.png_Brand_Mean	24	0,40	4,64	2,815	1,139
TFF_Penosil2B.png_Product image_Mean	32	0,00	4,44	1,955	1,213
TFF_Penosil2B.png_Product name_Mean	30	0,70	4,6	2,060	1,023
TFF_Penosil2B.png_Text_Mean	32	0,00	3,5	0,776	0,747
TFF_Penosil3B.png_Brand_Mean	28	0,00	4,76	1,606	1,396
TFF_Penosil3B.png_Product image_Mean	30	0,00	2,73	0,407	0,757
TFF_Penosil3B.png_Product name_Mean	20	1,32	4,24	3,113	0,815
TFF_Penosil3B.png_Text_Mean	32	0,00	3,77	1,192	1,052

TFF_Penosil4B.png_Brand_Mean	25	0,46	4,47	2,184	1,515
TFF_Penosil4B.png_Product image_Mean	33	0,00	3,39	0,524	0,890
TFF_Penosil4B.png_Product name_Mean	32	0,00	4,46	1,190	1,081
TFF_Penosil4B.png_Text_Mean	20	0,65	4,96	2,607	1,340
TFD_Penosil1B.png_Brand_Mean	31	0,08	2,1	0,599	0,483
TFD_Penosil1B.png_Product iimage_Mean	32	0,06	2,44	1,024	0,656
TFD_Penosil1B.png_Text_Mean	32	0,09	3,47	1,311	0,719
TFD_Penosil2B.png_Brand_Mean	24	0,08	1,94	0,668	0,473
TFD_Penosil2B.png_Product image_Mean	32	0,14	1,91	0,838	0,456
TFD_Penosil2B.png_Product name_Mean	30	0,08	2,11	0,698	0,462
TFD_Penosil2B.png_Text_Mean	32	0,12	1,8	0,746	0,407
TFD_Penosil3B.png_Brand_Mean	28	0,04	2,05	0,776	0,454
TFD_Penosil3B.png_Product image_Mean	30	0,09	2,35	0,826	0,617

TFD_Penosil3B.png_Product name_Mean	20	0,15	1,1	0,486	0,255
TFD_Penosil3B.png_Text_Mean	32	0,33	4,09	1,493	0,806
TFD_Penosil4B.png_Brand_Mean	25	0,12	3,43	0,913	0,769
TFD_Penosil4B.png_Product image_Mean	33	0,19	3,06	1,310	0,774
TFD_Penosil4B.png_Product name_Mean	32	0,16	3,44	0,895	0,776
TFD_Penosil4B.png_Text_Mean	20	0,07	0,98	0,456	0,303

Appendix 14. Tests of normality

Katse		Kolmogorov-Smirnov[b]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
TFD_Penosil4B.png _Product image_Mean	B	0,132	33	0,156	0,929	33	0,033

a. There are no valid cases for TFD_Penosil4B.png_Product image_Mean when Katse = ,000. Statistics cannot be computed for this level.

Appendix 15. Mann–Whitney U test

Ranks

	Katse_kood	N	Mean Rank	Sum of Ranks
TFD	A	33	31,58	1042,00
	B	33	35,42	1169,00
	Total	66		

Test Statistics[a]

	TFD
Mann-Whitney U	481,000
Wilcoxon W	1042,000
Z	-,814
Asymp. Sig. (2-tailed)	,415

a. Grouping Variable: Katse_kood

Appendix 16. Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,194[a]	,038	,031	,60847

a. Predictors: (Constant), Age, AOI_kood, Banner, Katse_kood, Language_kood, Familiarity

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12,158	6	2,026	5,473	<,001[b]
	Residual	310,259	838	,370		
	Total	322,417	844			

a. Dependent Variable: TFD

b. Predictors: (Constant), Age, AOI_kood, Banner, Katse_kood, Language_kood, Familiarity

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,402	,135		2,969	,003
	Katse_kood	,083	,047	,067	1,755	,080
	AOI_kood	,066	,018	,121	3,569	<,001
	Familiarity	,068	,028	,105	2,422	,016
	Banner	,029	,019	,051	1,511	,131
	Language_kood	,036	,025	,050	1,434	,152
	Age	-,006	,003	-,081	-2,050	,041

a. Dependent Variable: TFD

Appendix 17. Correlations

Correlations

[DataSet28]

Correlations

		TFD %	AI Predicted Attention (%)
TFD %	Pearson Correlation	1	,536**
	Sig. (2-tailed)		,002
	N	30	30
AI Predicted Attention (%)	Pearson Correlation	,536**	1
	Sig. (2-tailed)	,002	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

```
NONPAR CORR
  /VARIABLES=TFD AIPredictedAttention
  /PRINT=SPEARMAN TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
```

Nonparametric Correlations

Correlations

			TFD %	AI Predicted Attention (%)
Spearman's rho	TFD %	Correlation Coefficient	1,000	,458*
		Sig. (2-tailed)	.	,011
		N	30	30
	AI Predicted Attention (%)	Correlation Coefficient	,458*	1,000
		Sig. (2-tailed)	,011	.
		N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		FC (%)	AI Predicted Attention (%)
FC (%)	Pearson Correlation	1	-,329
	Sig. (2-tailed)		,076
	N	32	30
AI Predicted Attention (%)	Pearson Correlation	-,329	1
	Sig. (2-tailed)	,076	
	N	30	30

Banner	Vers	AOI	TFD (s)	TF (s) full Banner	TFD %	Count (N33)	FC (%)	AI Predicted Attention	Absolute Differenc	Relative Difference
PENOSIL 1	A	Brand	0,56	2,21	25,34	26,00	78,79	26,50	1,16	4,58
PENOSIL 1	B	Brand	0,60	2,93	20,48	31,00	93,94	34,40	13,92	67,99
PENOSIL 1	A	Product Name								
PENOSIL 1	B	Product Name								
PENOSIL 1	A	Product Image	0,72	2,21	32,58	29,00	87,88	19,20	13,38	41,07
PENOSIL 1	B	Product Image	1,02	2,93	34,81	32,00	96,97	21,70	13,11	37,67
PENOSIL 1	A	Text	0,93	2,21	42,08	25,00	75,76	40,00	2,08	4,95
PENOSIL 1	B	Text	1,31	2,93	44,71	24,00	72,73	42,00	2,71	6,06
PENOSIL 2	A	Brand	0,62	2,70	22,96	27,00	81,82	21,20	1,76	7,68
PENOSIL 2	B	Brand	0,67	2,96	22,64	30,00	90,91	24,60	1,96	8,68
PENOSIL 2	A	Product Name	0,59	2,70	21,85	29,00	87,88	12,70	9,15	41,88
PENOSIL 2	B	Product Name	0,70	2,96	23,65	32,00	96,97	19,30	4,35	18,39
PENOSIL 2	A	Product Image	0,69	2,70	25,56	33,00	100,00	13,60	11,96	46,78
PENOSIL 2	B	Product Image	0,84	2,96	28,38	32,00	96,97	12,60	15,78	55,60
PENOSIL 2	A	Text	0,80	2,70	29,63	25,00	75,76	32,10	2,47	8,34
PENOSIL 2	B	Text	0,75	2,96	25,34	28,00	84,85	23,60	1,74	6,86
PENOSIL 3	A	Brand	0,75	3,10	24,19	19,00	57,58	22,50	1,69	7,00
PENOSIL 3	B	Brand	0,78	3,59	21,73	20,00	60,61	35,80	14,07	64,77
PENOSIL 3	A	Product Name	0,45	3,10	14,52	28,00	84,85	13,20	1,32	9,07
PENOSIL 3	B	Product Name	0,49	3,59	13,65	30,00	90,91	12,20	1,45	10,62
PENOSIL 3	A	Product Image	0,70	3,10	22,58	33,00	100,00	16,70	5,88	26,04
PENOSIL 3	B	Product Image	0,83	3,59	23,12	32,00	96,97	22,80	0,32	1,38
PENOSIL 3	A	Text	1,20	3,10	38,71	25,00	75,76	36,20	2,51	6,48
PENOSIL 3	B	Text	1,49	3,59	41,50	25,00	75,76	24,20	17,30	41,69
PENOSIL 4	A	Brand	0,73	3,25	22,46	28,00	84,85	32,80	10,34	46,03
PENOSIL 4	B	Brand	0,91	3,57	25,49	32,00	96,97	35,50	10,01	39,27
PENOSIL 4	A	Product Name	0,79	3,25	24,31	33,00	100,00	31,90	7,59	31,23
PENOSIL 4	B	Product Name	0,89	3,57	24,93	33,00	100,00	23,70	1,23	4,93
PENOSIL 4	A	Product Image	1,16	3,25	35,69	22,00	66,67	20,60	15,09	42,28
PENOSIL 4	B	Product Image	1,31	3,57	36,69	20,00	60,61	26,20	10,49	28,60
PENOSIL 4	A	Text	0,57	3,25	17,54	26,00	78,79	11,50	6,04	34,43
PENOSIL 4	B	Text	0,46	3,57	12,89	31,00	93,94	11,50	1,39	10,75

Appendix 18. Traditional Eye-Tracking and AI Eye-Tracking Differences

			Viewers		Difference	
Banner	**Versio**	**AOI**	**Real Eye tracking**	**AI Eye tracking**	**Absolute Difference**	**Relative Difference**
PENOSIL 1	A	Brand	78,79%	26,5%	52,29%	66,37%
PENOSIL 1	B	Brand	93,94%	34,4%	59,54%	63,38%
PENOSIL 1	A	Product Name				
PENOSIL 1	B	Product Name				
PENOSIL 1	A	Product Image	87,88%	19,2%	68,68%	78,15%
PENOSIL 1	B	Product Image	96,97%	21,7%	75,27%	77,62%
PENOSIL 1	A	Text	75,76%	40,0%	35,76%	47,20%
PENOSIL 1	B	Text	72,73%	42,0%	30,73%	42,25%
PENOSIL 2	A	Brand	81,82%	21,2%	60,62%	74,09%
PENOSIL 2	B	Brand	90,91%	24,6%	66,31%	72,94%
PENOSIL 2	A	Product Name	87,88%	12,7%	75,18%	85,55%
PENOSIL 2	B	Product Name	96,97%	19,3%	77,67%	80,10%
PENOSIL 2	A	Product Image	100,00%	13,6%	86,40%	86,40%
PENOSIL 2	B	Product Image	96,97%	12,6%	84,37%	87,01%
PENOSIL 2	A	Text	75,76%	32,1%	43,66%	57,63%
PENOSIL 2	B	Text	84,85%	23,6%	61,25%	72,19%
PENOSIL 3	A	Brand	57,58%	22,5%	35,08%	60,92%
PENOSIL 3	B	Brand	60,61%	35,8%	24,81%	40,93%
PENOSIL 3	A	Product Name	84,85%	13,2%	71,65%	84,44%
PENOSIL 3	B	Product Name	90,91%	12,2%	78,71%	86,58%
PENOSIL 3	A	Product Image	100,00%	16,7%	83,30%	83,30%
PENOSIL 3	B	Product Image	96,97%	22,8%	74,17%	76,49%
PENOSIL 3	A	Text	75,76%	36,2%	39,56%	52,22%
PENOSIL 3	B	Text	75,76%	24,2%	51,56%	68,06%
PENOSIL 4	A	Brand	84,85%	32,8%	52,05%	61,34%
PENOSIL 4	B	Brand	96,97%	35,5%	61,47%	63,39%
PENOSIL 4	A	Product Name	100,00%	31,9%	68,10%	68,10%
PENOSIL 4	B	Product Name	100,00%	23,7%	76,30%	76,30%
PENOSIL 4	A	Product Image	66,67%	20,6%	46,07%	69,10%
PENOSIL 4	B	Product Image	60,61%	26,2%	34,41%	56,77%
PENOSIL 4	A	Text	78,79%	11,5%	67,29%	85,40%
PENOSIL 4	B	Text	93,94%	11,5%	82,44%	87,76%

Summary

THE COMPARATIVE STUDY OF TRADITIONAL AND AI EYE-TRACKING: PRACTICAL UTILITY OF AI RECOMMENDATIONS AND VISUAL ATTENTION PREDICTIONS IN DIGITAL BANNER ADVERTISING

Arina Zaviyalova, Irina Kalinina

The present master's thesis aimed to investigate the practical utility and accuracy of artificial intelligence (AI) eye-tracking tools in digital banner advertising, by comparing them with traditional eye-tracking data and digital campaign performance. The central objective of this study was to examine whether design improvement recommendations for AI eye-tracking softwares can enhance the performance of digital banner advertisements and visual attention, as well as to what extent AI predictions correspond to actual human visual behaviour.

The study adopted a two-part methodological design. The initial phase of the study entailed the execution of a Google Demand Gen A/B testing campaign across multiple countries, wherein original banner designs (A banners) were contrasted with banners improved according to AI recommendations (B banners). The creation of these modified banners were based Attention Insight AI eye-tracking software predictions, which generated design quality metrics (Clarity and Focus scores) and predicted heatmaps. The click-through rate (CTR) was utilised as the performance indicator of digital banner campaigns. The second component involved a traditional eye-tracking experiment using a Tobii X2-30 device. Participants were shown the same set of A and B banners while their gaze data were recorded. The study's primary focus was on two key visual attention metrics: The Time to First Fixation (TFF) and Total Fixation Duration (TFD) for predefined Areas of Interest (AOIs).

The literature review covered theories of visual attention and advantages and limitations of traditional and AI eye-tracking. While traditional eye-tracking offers high precision and insight into gaze patterns, it is costly and time-consuming. AI tools present a scalable and accessible alternative, but their predictive accuracy and practical impact have not yet been sufficiently validated across different cultural and real-world settings.

The results of the digital campaign showed that the banners improved according to AI suggestions did not significantly outperform original versions in overall CTR. However, moderate correlations were found between AI design quality metrics and CTR in specific

countries. These findings suggest that the impact of AI recommendations may be influenced by cultural context and campaign execution variables.

The traditional eye-tracking experiment revealed partial alignment between AI-predicted and real gaze patterns. In some cases, AI-predicted attention hotspots corresponded to areas of high TFD, but differences emerged in TFF and scanpath order. This indicates that AI models, which are primarily bottom-up and visually driven, may not fully capture top-down attention processes shaped by goals, language, or cultural cues.

In conclusion, AI-based attention prediction can be a helpful complementary tool in the early design phase of banner advertising, particularly for streamlining pre-testing and detecting major attention bottlenecks. However, these tools should not replace traditional eye-tracking methods and human interpretation, especially in complex or cross-cultural contexts. The thesis emphasizes the need for critical and contextual use of AI predictions and encourages further research that incorporates emotional responses, longer exposure durations, and a broader range of stimuli types to build more robust understanding of attention modelling in advertising.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Meie, Arina Zaviyalova ja Irina Kalinina,

anname Tartu Ülikoolile tasuta loa (lihtlitsentsi) meie loodud teose
“The Comparative Study Of Traditional And AI Eye-Tracking: Practical Utility Of AI Recommendations And Visual Attention Predictions In Digital Banner Advertising”
(“Traditsioonilise ja AI pilgujälgimise võrdlev uuring: AI soovituste ja visuaalse tähelepanu prognooside praktiline kasulikkus digitaalses bännerreklaamis”),

mille juhendajad on Kristian Pentus ja Kerli Ploom,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

Anname Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

Kinnitame, et lihtlitsentsi andmisega ei riku me teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Arina Zaviyalova
Irina Kalinina
22.05.2025