INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA INFORMÁTICA




Sistema Inteligente para Análise de Atenção Visual Baseado em Eye Tracking e Inteligência Artificial Focado em UI/UX

DIOGO FERREIRA GONÇALVES

Junho de 2026

POLITÉCNICO
DO PORTO

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

An Intelligent System for Visual Attention Analysis Based on Eye Tracking and Artificial Intelligence Focused on UI/UX

Diogo Ferreira Gonçalves

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science, Specialisation Area of Games, Graphics and Interactive Systems

Advisor: Ana Teixeira
Co-Advisor: Goreti Marreiros

Porto, June 28, 2026

Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I have not plagiarised or applied any form of undue use of information or falsification of results along the process leading to its elaboration.

Therefore the work presented in this document is original and authored by me, having not previously been used for any other end.

I further declare that I have fully acknowledged the Code of Ethical Conduct of P.PORTO.

ISEP, Porto, June 28, 2026

Abstract

The rapid growth of e-commerce has intensified the need for personalised and engaging user experiences. Understanding users' visual attention and emotional responses while interacting with digital interfaces is essential for designing adaptive, effective, and user-centered websites. However, current approaches often address visual attention, emotional analysis, or interface adaptation separately, lacking a unified intelligent system that integrates these components.

A systematic literature review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was conducted to identify existing research on attention metrics, interface design, emotion recognition, and Artificial Intelligence (AI)-driven adaptation. The review revealed a gap that this thesis explores, as few studies link gaze patterns to explicit user preference in a way that could enable adaptive interface personalisation.

The proposed solution is built around two complementary eye-tracking experiments and a multimodal predictive pipeline. The first experiment investigated how spatial position and presentation style influence visual attention within structured product grids. The second, choice-based experiment required participants to select their preferred product from competing alternatives, producing a labelled dataset that combines behavioural, oculomotor, physiological, and contextual features.

Five predictive models were trained and evaluated on the resulting dataset, Random Forest, Extreme Gradient Boost Classifier, Support Vector Machine, Multilayer Perceptron, and Extreme Gradient Boost Ranker. Among these, Random Forest achieved the highest overall performance, correctly identifying the selected product as the top-ranked item in approximately 78% of choice tasks, while the Extreme Gradient Boost Ranker also demonstrated strong ranking performance. These findings demonstrate that eye-tracking data, combined with machine learning, can accurately infer user preferences, providing a foundation for adaptive and personalised e-commerce interfaces.

Keywords: Eye-tracking, AI, Machine Learning, Visual attention, Product Preference, Ranking, UX/UI, E-commerce, Adaptive Interfaces

Acknowledgement

I would like to express my sincere gratitude to everyone who accompanied me throughout the development of this project and my academic journey, without their support this work would not have been possible.

My sincere thanks go to the GECAD research group, the institution where I carried out this work and which welcomed me throughout this entire journey

I am grateful to my supervisors, Prof. Ana Rita Teixeira and Prof. Goreti Marreiros, for their availability, guidance, and constant support at every stage of this project.

I would also like to thank all the participants who took part in the experimental studies, generously contributing their time and attention, without which the data collection underlying this work would not have been possible.

To my parents, I am profoundly thankful for their unconditional support and for making this academic journey possible.

To my girlfriend, for her patience, care, and constant support, especially during the most demanding moments of this project.

To my friends, thank you for your constant presence, encouragement, and support throughout this journey.

Finally, I would like to acknowledge the Instituto Superior de Engenharia do Porto and the Instituto Politécnico do Porto for the academic environment and opportunities provided throughout my studies.

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
AOI	Area of Interest
CV	Cross-Validation
EEG	Electroencephalography
FN	False Negatives
FP	False Positives
GDPR	General Data Protection Regulation
GECAD	Grupo de Estudos em Comunicação e Automação Digital
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
HR	Heart Rate
IPP	Instituto Politécnico do Porto
ISEP	Instituto Superior de Engenharia do Porto
LEDs	Light Emitting Diodes
LOGO	Leave-One-Group-Out
ML	Machine Learning
MLP	Multilayer Perceptron
MRR	Mean Reciprocal Rank
NDCG	Normalised Discounted Cumulative Gain
PREPD	Preparação para Dissertação
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF	Random Forest
ROC-AUC	Area Under the Receiver Operating Characteristic Curve
SVM	Support Vector Machine
TN	True Negatives

TP	True Positives
TTFF	Time to First Fixation
UI	User Interface
UX	User Experience
WBS	Work Breakdown Structure
XGBoost	Extreme Gradient Boosting

Chapter 1

Introduction

This chapter provides an overview of the dissertation by introducing the context and the problem statement that motivated this work. It also presents the objectives, the contributions, and the planning strategy adopted throughout the project. Additionally, the chapter outlines the ethical considerations associated with the study and concludes with a description of the structure of the thesis.

1.1 Context

This report was made in partial fulfilment of the requirements for the degree of Master of Science, Specialisation Area of Games, Graphics and Interactive Systems at Instituto Superior de Engenharia do Porto (ISEP). The project was proposed and developed in collaboration with Grupo de Estudos em Comunicação e Automação Digital (GECAD), in articulation with the CAPE [1] and EARS [2] research projects, which explore complementary approaches to user preference modelling through recommendation systems (content-based filtering and collaborative filtering). While this dissertation adopts a distinct methodological approach centred on eye-tracking and physiological data, it shares with these projects the broader goal of inferring user preference to support adaptive, personalised digital experiences.

GECAD [3] is a research unit at ISEP – Instituto Politécnico do Porto (IPP) in Portugal. It focuses on Intelligent Systems, Power and Energy Systems, and advanced technologies for digital innovation. GECAD is recognized internationally for its contributions to research, development, and industry collaboration.

The rapid expansion of the digital economy has transformed the way users interact with services, particularly within e-commerce and other online platforms. As digital products become increasingly common, the competition among platforms has intensified, making user retention and engagement key challenges for companies. In this environment, the quality of interaction between users and digital systems has become a decisive factor for success, with User Interface (UI) and User Experience (UX) design emerging as a crucial element in achieving that success [4].

In addition to functional efficiency, current UI/UX designs highlight a deeper understanding of user behavior, preferences, and emotional responses. Users' perceptions and emotions during interaction can significantly influence their engagement, trust, and continued use of digital platforms and services. Designing interfaces that align with user expectations and emotional states can increase intuitiveness and foster a sense of comfort and satisfaction [5]. Recent research emphasizes the integration of emotional considerations into interface design, enabling systems to better respond to users' affective states and interaction patterns. These perspectives reinforce the need for advanced methods capable of analysing user attention and emotional engagement to support more responsive and user-centred design solutions [6].

1.2 Problem Statement

Despite the rapid evolution of digital commerce and the increasing focus on engaging and user-centred design, many existing UI/UX approaches remain static. These traditional interfaces do not adapt to individual user characteristics, emotional states, or visual attention patterns. As a result, most current systems are unable to interpret user responses in real time or adjust the interface accordingly to different users [7].

At the same time, recent advances in innovative technologies, such as eye-tracking and Artificial Intelligence (AI), provide new opportunities to better understand how users interact with digital interfaces. Eye-tracking makes it possible to collect data that reveals where users focus their attention and which UI elements attract or fail to attract interest. These techniques, combined with AI-based predictive models, can support decisions on how the UI should be designed to better align with user preferences, enabling the development of intelligent systems capable of inferring product preferences from visual attention and physiological data.

Therefore, the main problem addressed in this dissertation is the development of an intelligent system capable of analysing users' visual attention and inferring product preferences using eye-tracking technology and AI, in order to inform and support adaptive UI/UX design decisions in e-commerce websites and improve the user experience.

1.3 Objectives

As previously described, the main objective of this study is to develop an intelligent system capable of analysing users' visual attention and inferring product preferences from behavioural and physiological responses, using eye-tracking technology and AI. The proposed system aims to support website design decisions, improve user experience, increase interaction effectiveness, and improve businesses' performance.

In order to achieve the main objective, a set of specific objectives was defined:

- **Conduct a literature review:** Establish the current state of research in UI/UX personalisation, eye-tracking, and AI-based preference inference in order to provide a base for the project.
- **Analyse user behavior and emotional responses:** Investigate users' visual attention patterns, emotional reactions, and the interface elements that influence perception and focus, using eye-tracking and physiological data such as Galvanic Skin Response (GSR) and Heart Rate (HR).
- **Construct a labelled dataset:** Design and conduct a choice-based eye-tracking experiment to build a structured dataset pairing gaze and physiological features with explicit product selection outcomes.
- **Develop AI-based predictive models:** Train and evaluate Machine Learning (ML) models capable of predicting product preference from eye-tracking and physiological data.

- **Validate AI models:** Evaluate model performance using classification and ranking metrics and analyse feature importances to interpret model behaviour.
- **Write the dissertation report:** Document the project in a structured and comprehensive dissertation report.

1.4 Project planning

The project plan plays a crucial role in guiding the development and management of this project. To ensure a structured and organised project development, three main tools were created: a project charter to define objectives and scope, a Work Breakdown Structure (WBS) to organise tasks and deliverables, and a Gantt chart to schedule activities and monitor progress throughout the project.

1.4.1 Project Charter

The Project Charter is the document that formally initiates the project, defining its scope, objectives, sponsors and stakeholders, as well as constraints, interdependencies, and risks. Although project management can be a complex and formal discipline, its practical application is essential to the successful execution of technical projects. In this context, the project charter serves as a guiding reference that supports informed decision-making throughout the project lifecycle and helps align expectations with outcomes [8]. The project charter developed for this project is provided in the Appendix (Appendix A).

Within the project charter, the main stakeholders involved in the project were identified. These include end users of e-commerce platforms, who are directly affected by UI/UX design decisions, e-commerce companies, which seek to improve user engagement and business performance, and UI/UX designers and developers, who are responsible for designing and implementing user-centred interfaces.

The primary objective defined in the project charter is the development of an intelligent system capable of analysing users' visual attention and inferring product preferences from behavioural and physiological responses during navigation on e-commerce websites. By combining eye-tracking technologies with AI models, the system aims to support data-driven UI/UX design decisions that promote personalised, efficient, and user-centred experiences. In addition to this objective, the project is expected to contribute to improved user experience, increased customer engagement, enhanced business performance, innovation in personalisation strategies, and contributions to both practical applications and academic knowledge in the UI/UX domain.

Finally, the project charter defines the main deliverables of the project, that includes the project plan, the dissertation report, the presentation and discussion of results, the validation of the obtained results, and the developed software code.

1.4.2 Work Breakdown Structure (WBS)

WBS is a project management tool that organises and groups all the work involved in a project according to its deliverables, defining the total scope of the project. It breaks down the overall work into smaller, well-defined tasks and arranges them in a logical hierarchical structure. By decomposing the project into its component elements, the WBS supports effective planning and control, allowing cost, time, and technical performance to be monitored throughout the entire project lifecycle [9]. Figures 1.1 and 1.2 present the Work Breakdown Structure defined for this project, while Table 1.1 provides a brief description of each main phase of the project.

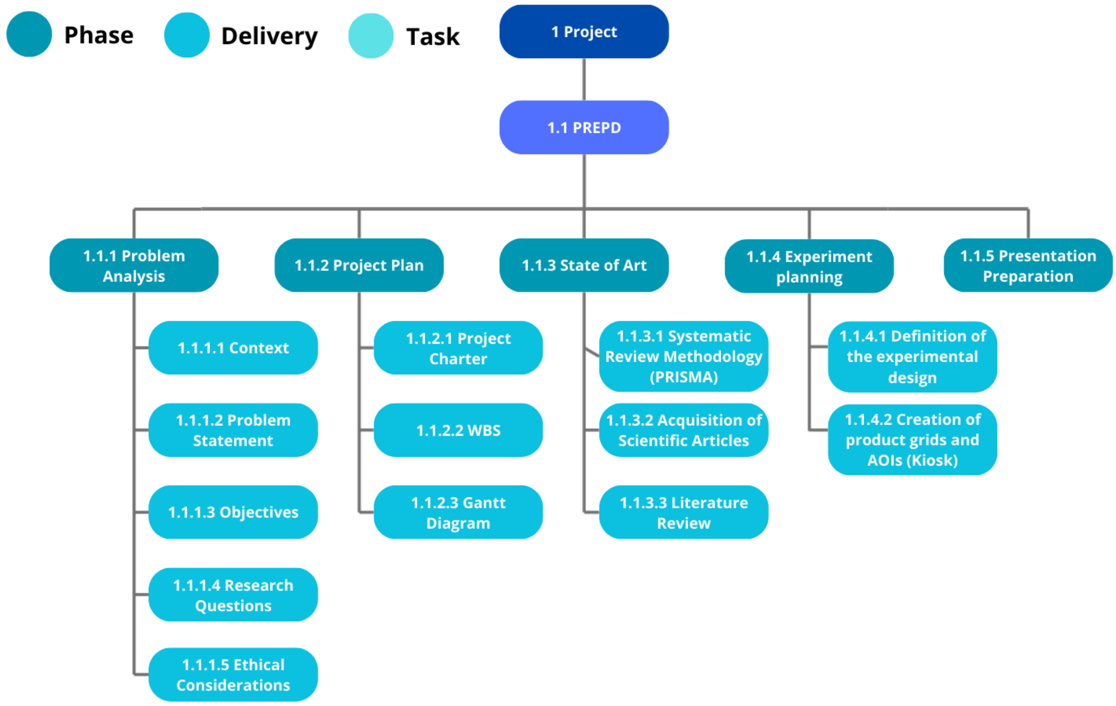

Figure 1.1: Work Breakdown Structure of the Project (Part 1)

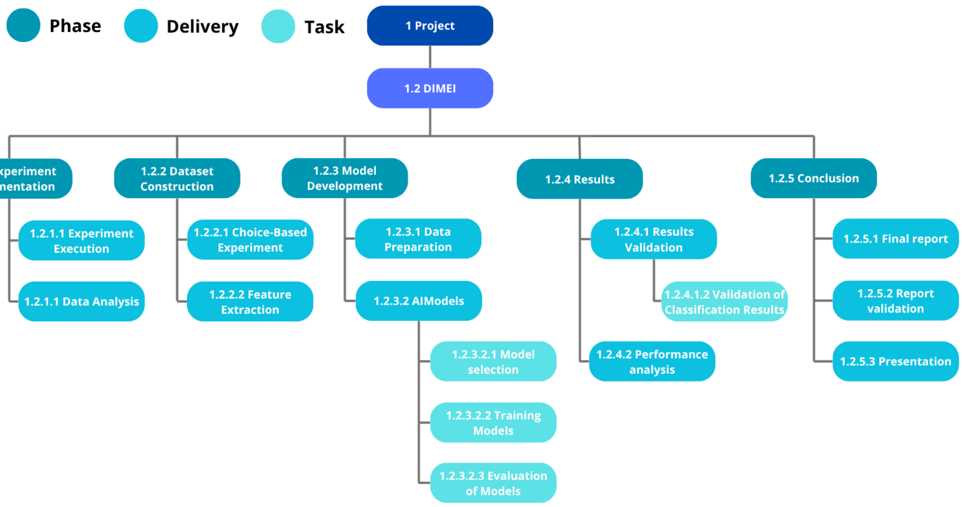

Figure 1.2: Work Breakdown Structure of the Project (Final Part)

Table 1.1: WBS Phases Explained

WBS Phase	Description
Problem Analysis	Understanding the project context, defining the problem statement, establishing objectives, identifying guiding research questions, and considering ethical implications to ensure a structured project development.
Project Plan	Creation of the project charter, development of the WBS, and preparation of the Gantt diagram, providing a clear plan to guide project execution and monitor progress.
State of the Art	Performing a systematic literature review using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), acquiring relevant scientific articles, and creating a clear literature review to provide context for the project.
Experiment Planning	Defining the experimental design and creating product grids and Area of Interest (AOI) for the experimental interface, establishing the structure for data collection and analysis.
Presentation Preparation	Creating a presentation of the project, which will be delivered and presented in the Preparação para Dissertação (PREPD) unit.
Experiment Implementation	Executing the first experimental study, collecting eye-tracking and physiological data from participants interacting with product grids, and analysing the influence of spatial positioning and presentation style on visual attention.
Dataset Construction	Conducting a second choice-based eye-tracking experiment, and processing the resulting recordings into a labelled dataset suitable for supervised machine learning using a target variable.
Model Development	Selecting, implementing, and evaluating ML models to predict product preference from eye-tracking and physiological features.
Results	Evaluating model performance through classification and ranking metrics, analysing feature importances, and interpreting the outputs of the predictive models.
Conclusion	Preparing the final project report, validating its content, and presenting the project.

1.4.3 Gantt chart

The Gantt chart is a widely used project management tool that provides a visual representation of the project schedule by mapping tasks against time. Named after the American engineer Henry L. Gantt (1861–1919), Gantt charts are essentially horizontal bar charts in which each bar represents an activity or task in the project. The position of the bar indicates the start and end dates, while its length reflects the duration of the task. This visual format allows for effective planning, coordination, and monitoring of project progress, making it an essential instrument for managing project execution [10].

Beyond scheduling, Gantt charts provide several practical benefits in project management. They assist in human resource management by optimising task allocation, help reduce costs and material wastage, and contribute to saving time in executing project activities. By enabling clear visualisation of tasks and phases, Gantt charts support both efficient planning and implementation, making sure that project objectives are achieved effectively [11].

The Gantt chart is directly aligned with the WBS, as it schedules and sequences the same phases and tasks defined in the WBS. This ensures consistency between project planning and scheduling and facilitates effective tracking of progress throughout the project lifecycle. The Gantt diagram for this project is presented in the Appendix (Appendix B).

1.5 Research Contributions

Beyond the dissertation itself, this work has produced several contributions intended for dissemination to the wider research community:

- **Contributions directly derived from this dissertation:** A paper summarising the findings of the first experimental study presented in Chapter 3, focusing on the effects of spatial position and presentation style on visual attention, is being prepared for submission to the International Conference on Marketing and Technologies (ICMarkTech'26) [12]. Additionally, the raw and processed datasets obtained from the choice-based eye-tracking experiment described in Chapter 4 are being prepared for submission for publication in *Data in Brief* [13], to support reproducibility and enable reuse by the wider research community.
- **Related contributions in recommendation systems:** Developed in parallel within the scope of the GECAD research unit and the CAPE and EARS projects, these contributions address a related problem, inferring user preference, through a different methodological lens, namely federated learning, content-based and collaborative recommendation systems. As a co-author, the author contributed to two publications in this area: the first, *Federated Learning for Hybrid Recommendation Systems*, presents a federated recommendation architecture that integrates Federated Learning with feature mining derived from a multimodal emotion detection pipeline, and was accepted for publication at the 34th Signal Processing and Communications Applications Conference (SIU 2026) [14] . The second, *Aligning Content-Based Learning with Federated Learning for E-Commerce Recommendation Systems*, investigates the behaviour of content-based filtering under horizontal and vertical federated data-partitioning strategies, and was submitted to the IBERAMIA 2026 conference [15].

1.6 Ethical Considerations

Ethical and professional considerations are fundamental in any research project, particularly when it involves the collection of personal and sensitive data. Careful attention to ethics ensures that participants' rights, privacy, and well-being are fully protected. This project involves gathering sensitive personal information, including users' gaze patterns from eye-tracking devices and emotional signals derived from physiological and behavioural responses, which can reveal aspects of users' cognitive and emotional states, and potentially indicate personal traits or health conditions.

The processing of these data will be conducted in accordance with the General Data Protection Regulation (GDPR), the European legislation that establishes standards for the collection, storage, and processing of personal information. GDPR aims to protect individuals' privacy rights, ensuring that personal data is handled lawfully, transparently, and securely.

In addition, this project will comply with the European Artificial Intelligence Act (Regulation (EU) 2024/1689), which establishes harmonized rules for the development, deployment, and use of AI systems in the EU [16]. The AI Act promotes trustworthy, human-centric AI while protecting fundamental rights, health, and safety. It prohibits certain high-risk and unethical AI practices, such as manipulative behavior, social scoring, and unauthorized biometric inferences, ensuring that AI systems are safe and respectful of users' dignity and autonomy [16].

To comply with these ethical and legal requirements, the following measures will be implemented:

- Obtain informed consent from participants, providing a clear explanation of the study's objectives, procedures, and potential impacts.
- Anonymize and pseudonymize collected data to prevent individual identification.
- Ensure secure data storage, with access restricted exclusively to authorized research team members.
- Use collected data solely for scientific purposes, without commercial exploitation.

These practices ensure that participants' privacy and well-being are respected while enabling the collection of high-quality data to support the development of the intelligent system.

1.7 Document Structure

This dissertation is organised into seven main chapters, each addressing a different aspect of the project.

The first chapter, Introduction, presents the context and problem statement of the project. It also outlines the objectives, describes the project planning approach, discusses ethical considerations, and concludes with a brief overview of the structure of the dissertation.

The second chapter, State of the Art, reviews the relevant literature and foundational concepts related to eye-tracking and AI, describes the systematic literature review methodology (PRISMA) adopted in this study, and concludes with a discussion of the main findings.

The third chapter, Experimental Study, presents the experimental protocol implemented in this study to investigate user attention, visual engagement, and the influence of design factors on gaze behaviour in an e-commerce context. The chapter details the design rationale and prototype development, including the stimuli, product layout, and presentation conditions used, as well as the participants, materials, and procedure followed during data collection. It further describes the data processing and AOI definition applied to the eye-tracking data, and presents the results obtained across product category, spatial position, and presentation style, along with their interaction effects and the correlations observed between eye-tracking metrics.

The fourth chapter, Choice-Based Eye-Tracking Study and Dataset Construction, presents the second experimental study conducted in this research, which extends the eye-tracking analysis from Chapter 3 to examine how gaze behaviour relates to explicit product choice. The chapter details the objective, design, participants, materials, procedure, and AOI definition of this choice-based experiment, in which participants selected the most visually appealing product from sets of grid-displayed products. It then describes the construction of the resulting dataset, including the structure of the raw eye-tracking files, the definition of the behavioural, oculomotor, physiological, and contextual feature groups, and an overview of the final dataset used as the empirical foundation for the predictive models developed in Chapter 5.

The fifth chapter, AI Models, describes the machine learning modelling pipeline developed to predict product preference from eye-tracking and physiological data. It details the validation strategy, class imbalance handling, feature scaling decisions, and hyperparameter tuning procedure, followed by a description of each model implemented: Random Forest (RF), Extreme Gradient Boosting (XGBoost) Classifier, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and XGBoost Ranker. The chapter concludes with a presentation of the classification and ranking evaluation metrics used to assess model performance.

The sixth chapter, Results and Discussion, presents the results obtained from the ML models and discusses their implications. It begins with the feature importance analysis and the incremental feature selection procedure used to identify the final feature subset, followed by the classification and ranking performance of each model. The chapter concludes with a discussion of the best-performing model and the most predictive feature groups.

The seventh and final chapter, Conclusion, reflects on the work developed throughout the project, summarising the main findings in light of the dissertation's objectives. It then identifies the main limitations of the study and outlines directions for future research.

Chapter 2

State of the Art

This chapter presents the state of the art relevant to the development of this thesis. It introduces the foundational concepts, technologies, and methodologies related to eye-tracking and AI. In addition, this chapter describes the systematic literature review methodology adopted to identify and analyse existing studies, discusses the main findings, and identifies the limitations of current approaches.

2.1 Background

This section provides an overview of the key concepts fundamental to this study, focusing on eye-tracking and AI. Eye-tracking is described as a method for capturing human gaze behavior, while AI is presented as a tool for learning patterns from data and supporting decision-making.

2.1.1 Eye-Tracking

An overview of eye-tracking is provided, covering its definition, physiological principles, commonly used metrics, and key applications, forming the basis for understanding how gaze data informs Human Computer Interaction (HCI) and behavioural analysis.

What is Eye-Tracking?

Eye-tracking can be described as the systematic process of monitoring eye movements to identify both the direction and duration of an individual's gaze, typically through the use of specialized hardware and software for recording eye movement and gaze positioning [17]. It involves measuring either the point of gaze, indicating where an individual is looking, or the motion of the eye relative to the head, thereby capturing observable gaze behavior within a visual environment [18].

In practice, eye-tracking measurements may be obtained through various approaches, including tracking pupil movement, analysing the reflection of infrared light on the cornea, or assessing changes in the electrical potentials of the eye muscles [17].

Human Visual System and Eye-Tracking Principles

Vision is one of the most dominant human senses and plays a central role in how individuals perceive and interact with their environment [19]. The effectiveness of visual perception relies primarily on the human eye, which is the primary sensory organ responsible for capturing visual information and transmitting it to the brain for interpretation [20].

The human eye functions as an optical system that captures incoming light and focuses it onto the retina [20, 21]. Light passes through the cornea, pupil, and lens, which together regulate and focus visual information onto the retinal surface [21]. Photoreceptor cells in the retina convert this light into electrical signals that are transmitted through the optic nerve to the brain, where they are interpreted as visual images [20, 21]. Figure 2.1 illustrates the general anatomical configuration of the human eye.

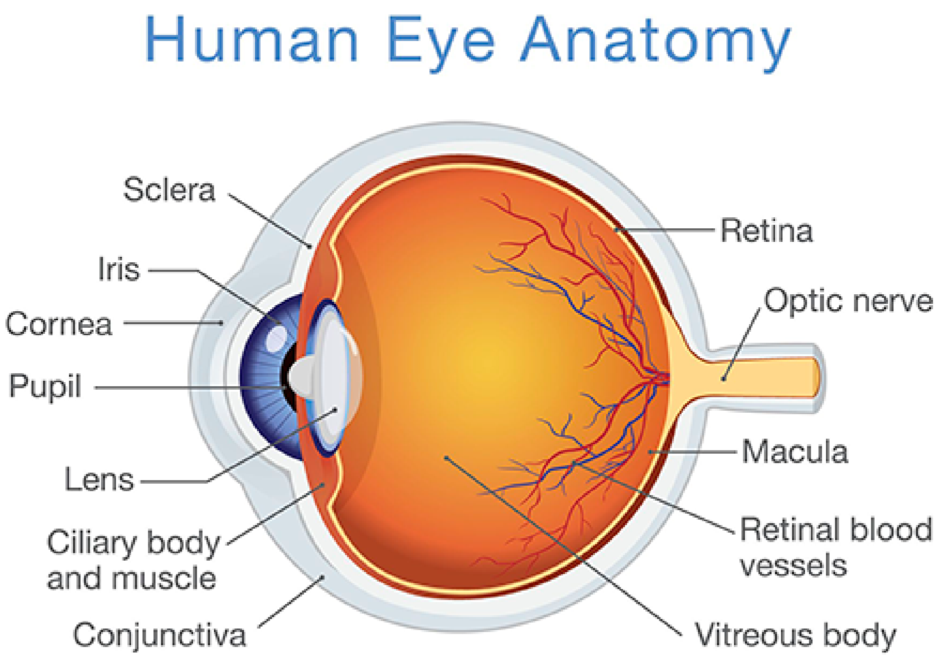

Figure 2.1: Human Eye Composition [22]

Eye-tracking builds upon these physiological characteristics by observing externally measurable features of the eye that change as it moves. A key step in this process is detecting and tracking the pupil, which serves as the primary reference point for gaze estimation [17, 23]. To improve accuracy, eye-tracking systems may also analyse additional features such as the iris, glints (reflections on the cornea or lens), eye corners, and other distinct landmarks [17].

Gaze direction is then estimated by considering the anatomical axes of the eye. The visual axis, passing through the fovea and pupil centre, represents the actual gaze direction, while the optical axis, passing through the eyeball centre and pupil centre, forms a three-dimensional vector that facilitates measurement. Most eye-tracking techniques rely on the optical axis, taking into account the individual-specific angle kappa between the visual and optical axes to accurately infer where the user is looking [24]. Once estimated, the gaze direction is translated into a spatial representation within a digital environment. Gaze points are typically mapped to two-dimensional coordinates on a screen or within a visual scene, allowing visualisation as markers that indicate where a user is looking at any given moment. Figure 2.2 illustrates the visual and optical axes of the human eye, highlighting their role in gaze estimation.

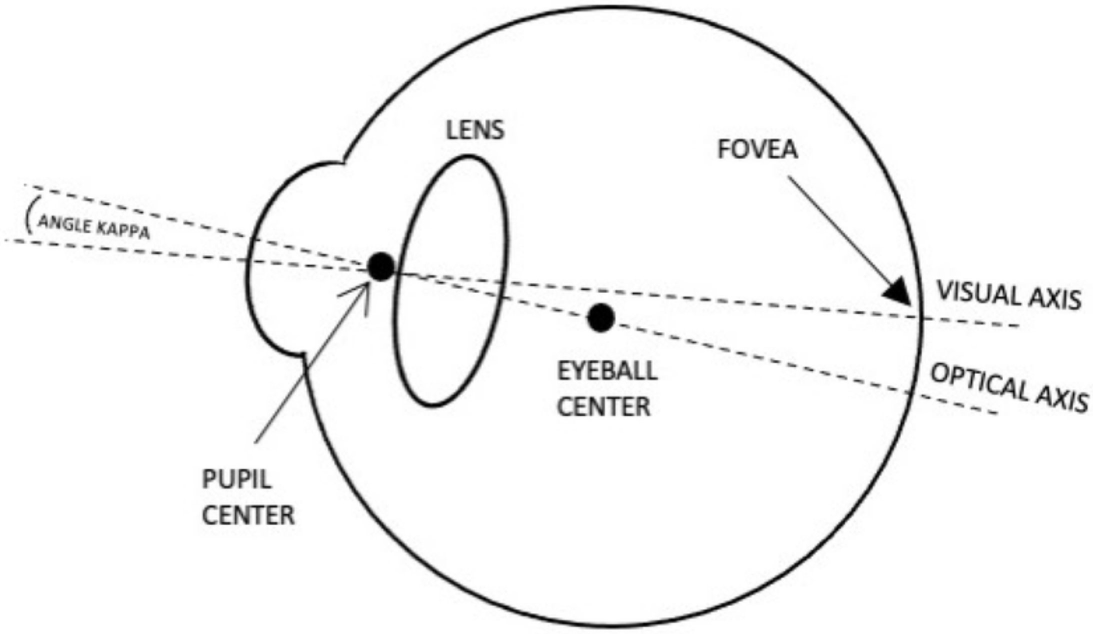

Figure 2.2: Human eye visual axes [25]

Eye-Tracking Metrics

Eye-tracking provides a rich set of metrics that can be analysed to understand user attention, cognitive processes, and interaction behavior across various applications [17, 23, 24]. The most commonly used metrics include fixations, saccades, smooth pursuits, scan paths, heat maps, AOI, gaze duration, fixation sequences, pupil size and blinking, respondent count, and time spent. These metrics capture different aspects of eye movement [17, 23, 24]. Table 2.1 presents the most commonly used metrics in eye-tracking along with a description of their meaning.

Table 2.1: Eye-Tracking Metrics Explained

Metric	Description
Fixation [17, 23, 24]	A period during which the eyes remain relatively stationary on a specific point, allowing visual information acquisition.
Smooth Pursuit[17, 24]	Eye movement that allows the eyes to steadily follow a moving object, used to study visual tracking and attention.
Saccades [17, 23, 24]	Rapid eye movements between fixations that redirect the gaze, often measured in angle velocity and used to study reading or visual scanning.
Scan Paths [17, 24]	Sequences of fixations and saccades that show the path of visual attention, used to analyse user interaction and cognitive behavior.
Heat Maps [24]	Visual representation of aggregated gaze data, highlighting areas of high and low attention using colour-coding.
Gaze Duration [17]	Total time spent fixating within a particular region, reflecting attention and engagement.
Pupil Size and Blinking [17]	Measurements used to assess cognitive load, fatigue, attention, and emotional responses.
AOI [24]	User-defined subregions of a visual stimulus, allowing quantitative analysis of attention distribution within specific areas.
Fixation Sequence [24]	The temporal order of fixations across AOIs, showing the sequence of visual attention.
Respondent Count [24]	The number of participants who fixated on a specific AOI, used to gauge salience and interest.
Time View [24]	The duration respondents spent looking at specific AOIs, often indicating motivation or conscious attention.
Revisits [24]	The number of times a participant returns to a previously viewed AOI, reflecting sustained interest or reconsideration.
Time to First Fixation (TTFF) [24]	The time elapsed from stimulus onset until the first fixation on a specific AOI, indicating how quickly an element captures visual attention.

Applications of Eye-Tracking

Eye-tracking technology is widely applied across various domains due to its ability to objectively capture visual attention, user behavior, and cognitive processes [26, 27]. The main application areas include:

- **Marketing and Design:** Eye-tracking is extensively used to evaluate advertisement effectiveness, website usability, interface layouts, and packaging design [27]. The objective data obtained enable the identification of visual elements that capture users' attention as well as those that are overlooked, supporting workflow optimisation and usability improvements [26].
- **Healthcare and Medical Applications:** Eye-tracking supports diagnostic and rehabilitation processes by identifying abnormal eye movement patterns associated with neurological and developmental conditions, including autism, Alzheimer's disease, Parkinson's disease, and strabismus [26, 27]. It is also employed in assistive technologies, enabling gaze-based communication for individuals with severe motor impairments [27].

- **Automotive Industry:** Eye-tracking is increasingly integrated into driver monitoring systems to detect drowsiness, distraction, and loss of attention through blink rate analysis and gaze tracking. These techniques contribute to improving road safety and reducing driver cognitive load [27].
- **Manufacturing and Industrial Environments:** Eye-tracking is used for skill assessment, process optimisation, and worker training. By analysing visual attention during task execution, organisations can improve ergonomics, reduce errors, and integrate gaze-driven augmented reality assistance for complex operations [27].
- **Research and Education:** Eye-tracking plays a significant role in HCI, cognitive psychology, and behavioural research. It enables the analysis of attention distribution, learning processes, and decision-making, supporting the development of more intuitive and effective digital interfaces and educational tools [26, 27].

2.1.2 AI

An overview of AI is presented, including its foundational principles, core ML mechanisms, and key application areas, providing a basis for understanding how AI processes data, learns patterns, and supports decision-making.

What is AI?

AI has a history much longer than is commonly understood, with its roots tracing back to ancient philosophical inquiries into reasoning and intelligence, but its modern development is largely attributed to Alan Turing and the Dartmouth Conference in 1956, where John McCarthy invented the term AI [28].

Despite decades of research, there is still no universally accepted definition of AI [28, 29]. In general, AI can be described as a technology that enables machines to imitate complex human skills, including perception, decision-making, problem-solving, and learning from experience [29]. Task-based definitions emphasize AI systems' ability to analyse their environment, take actions with some degree of autonomy, and achieve specific goals [29]. However, these definitions can be too broad and might include simple systems, like thermostats, which shows that defining AI precisely is challenging [29].

The recent resurgence of AI research, often referred to as the "current AI", is driven by advances in computing power, the availability of large annotated datasets, and the development of sophisticated learning algorithms. Modern AI systems are therefore better equipped to perform complex tasks in dynamic environments, demonstrating the practical applications of AI across various domains while continuing to evolve conceptually [28].

Machine Learning Principles

ML is a subfield of AI that focuses on the development of algorithms capable of learning patterns from data and improving performance over time without being explicitly programmed [30, 31]. Instead of relying on fixed rules, ML models identify relationships within training data and use this knowledge to make predictions or decisions when exposed to new, unseen data [31].

The core principle of ML lies in model training, where algorithm parameters are optimised using representative datasets so that the learned patterns can generalise to real-world scenarios [31].

ML approaches are commonly categorized into:

- **Supervised Learning**: Supervised learning uses labelled datasets, where each input is associated with a known output, allowing models to learn a mapping function between inputs and outputs [30–32]. The learning process relies on labelled examples as ground truth, enabling the algorithm to recognise patterns and relationships within the data. Common supervised tasks include classification, which assigns data to predefined categories, and regression, which predicts continuous values [30, 32].
- **Unsupervised Learning**: Unsupervised learning analyses unlabelled data to discover hidden structures, patterns, or relationships without prior knowledge of output labels [30–32]. This data-driven approach is commonly used for exploratory analysis and includes techniques such as clustering, dimensionality reduction, anomaly detection, and feature learning [30, 32].
- **Semi-Supervised Learning**: Semi-supervised learning combines elements of supervised and unsupervised learning by leveraging both labelled and unlabelled data during training [31, 32]. This approach is particularly valuable in scenarios where labelled data are scarce or expensive to obtain, but large volumes of unlabelled data are readily available [31]. Common applications include text classification, fraud detection, and behavior modeling tasks [32].
- **Reinforcement Learning**: Reinforcement learning focuses on training autonomous agents to make decisions by interacting with an environment and learning through trial and error [30–32]. The agent receives feedback in the form of rewards or penalties and iteratively optimises its behavior to maximise cumulative reward [30–32]. This environment-driven learning paradigm is well suited for sequential decision-making problems and is widely applied in robotics, game playing, autonomous driving, and autonomous vehicles [30–32].
- **Deep Learning**: Deep learning is a subset of ML based on artificial neural networks with multiple interconnected layers capable of modeling complex, high-dimensional data [30, 31]. Inspired by the structure of the human brain, deep learning models automatically learn hierarchical feature representations from raw data, reducing the need for manual feature engineering. These models underpin many state of the art AI systems, including computer vision, generative AI,self-driving cars and robotics [31].

ML systems follow a common data-driven workflow that is largely independent of the specific algorithm used. At their core, these systems rely on appropriate data representations and a learning model to extract meaningful patterns from input data. The process typically consists of two main stages: a training phase, in which the model learns from historical data, and a prediction phase, where the trained model is applied to new data to generate outputs. During training, raw data are first prepared through preprocessing steps such as cleaning, normalization, and feature selection to improve data quality. The processed data are then used by a learning algorithm to build a model that captures underlying relationships. Once trained, this model can be given new, unseen inputs, and is able to produce outputs such as predictions, classifications, or other decision-support results in real-world applications

[33]. Figure 2.3 illustrates the general workflow of a ML process, from data input and preprocessing to model training, inference, and the generation of outputs.

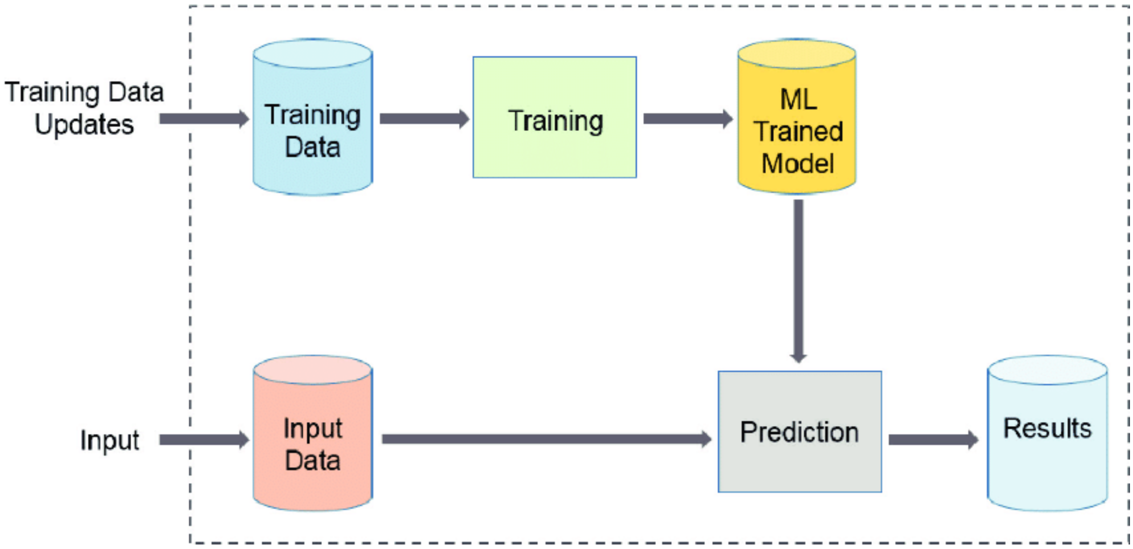

Figure 2.3: ML Workflow

Applications of AI

AI has become a powerful tool across various domains, enabling intelligent decision-making and automation through data-driven models [28]. Some applications include:

- **Healthcare:** AI assists in disease diagnosis, medical image analysis, patient monitoring, and clinical decision support, improving treatment efficiency and patient outcomes [30, 32, 34].
- **Finance:** ML algorithms are widely used for fraud detection, risk assessment, algorithmic trading, and customer behavior analysis, allowing real-time predictions and automation of complex financial processes [30, 32, 34].
- **Transport:** AI supports traffic prediction, route optimisation, and autonomous driving, helping reduce congestion, improve safety, and optimise urban infrastructure [32, 34].
- **E-commerce and Social Media:** AI enables personalised recommendations, trend analysis, and targeted content delivery by analysing user behavior and preferences, increasing user experience and business decision-making [30, 32, 34].
- **Agriculture:** AI-powered solutions assist in crop monitoring, yield prediction, disease detection, and resource optimisation, supporting sustainable farming practices and informed decision-making [32, 34].

2.2 Technologies

This section provides an overview of the key technologies relevant to this study. It introduces the different types of eye-tracking systems for acquiring and analysing eye-tracking data, highlighting their relevance and key characteristics, as well as several commonly used AI models, providing an overview of their main principles and characteristics.

2.2.1 Eye-tracking data retrieval

Eye-tracking data can be acquired using different technological approaches that vary in terms of hardware configuration, user mobility, and accuracy. Among the available solutions, the most widely used eye-tracking systems can be generally classified into screen-based eye trackers, head-mounted eye trackers, and webcam-based eye-tracking systems [17, 23].

Screen-based Eye Trackers

Screen-based eye-tracking measures visual attention by monitoring where and for how long a user's gaze focuses on specific areas of a computer screen. These systems are typically stationary and are either integrated into the display or positioned just below it. Screen-based eye trackers are widely used in controlled environments such as usability testing, market research, and academic and scientific research [35].

Screen-based eye trackers operate by combining infrared illumination with high-speed cameras to capture detailed images of the eyes. Infrared Light Emitting Diodes (LEDs) illuminate the eyes without causing discomfort, creating reflections on the cornea and pupil that are not visible to the user. Cameras positioned around the monitor continuously record eye images, allowing the system to detect pupil position and corneal reflections. Through an initial calibration procedure, during which the user fixates on predefined points on the screen, the system learns the relationship between eye features and screen coordinates. Once calibrated, the eye tracker can accurately map gaze direction onto the display in real time, enabling precise estimation of where the user is looking [35].

Head-mounted Eye Trackers

Head-mounted eye-tracking, or mobile eye-tracking, involves the use of wearable devices, most commonly eye-tracking glasses, to record eye movements while users interact with real-world environments. Unlike stationary screen-based systems, head-mounted eye trackers enable the analysis of gaze behavior in dynamic and unconstrained environments, allowing users to move freely while their visual attention is continuously monitored [36]. This makes mobile eye-tracking particularly suitable for applications such as sports performance analysis, consumer behavior research, psychological and cognitive studies and education and training [36].

Head-mounted eye trackers operate using a combination of infrared illumination, integrated cameras, and real-time processing algorithms. Infrared LEDs embedded in the frame of the device illuminate the eyes, creating reflections on the cornea and pupil that are captured by inward-facing eye cameras. In addition, outward-facing scene cameras record the user's field of view, providing contextual information about the environment. By analysing the relationship between pupil position and corneal reflections, the system computes the user's gaze direction. Through a calibration procedure, gaze data are accurately mapped onto the scene camera view, enabling precise visualisation of where the user is looking within their surroundings. Advanced algorithms process this information in real time, producing a continuous record of gaze behavior during natural movement [36].

Webcam-based Eye-Tracking

Webcam-based eye-tracking is a remote eye-tracking approach that analyses eye movements using a standard computer webcam, eliminating the need for specialized eye-tracking hardware. This technology enables the study of visual attention, engagement, and cognitive processes outside controlled laboratory environments, making it particularly suitable for large-scale, remote, and cost-effective studies [37].

Webcam-based eye trackers operate by applying advanced image-processing and ML algorithms to video data captured by the webcam. The system detects facial landmarks, eye regions, and pupil positions to estimate gaze direction in real time. Through calibration procedures, eye movement data are mapped onto screen coordinates, allowing the identification of fixation points, fixation durations, saccadic movements, and scanpaths [37].

Comparison of Eye-Tracking Technologies

Table 2.2 illustrates the differences between different eye-tracking technologies. Webcam-based eye-tracking offers high mobility and low cost but suffers from limited accuracy and lower sampling rates. Screen-based systems provide the highest precision among the considered technologies, but on the other hand they offer low mobility, making it suitable for laboratory studies. Head-mounted eye trackers offer a balance between accuracy and mobility, enabling gaze analysis in real-world environments at the expense of a higher cost.

Table 2.2: Comparison of eye-tracking technologies [18, 35–39]

Feature	**Webcam-based Eye-tracking**	**Screen-based Eye-tracking**	**Head-mounted Eye-tracking**
Measurement accuracy	Lower accuracy (approx. 0.5–1.5 ° visual angle)	High accuracy (approx. 0.1–0.5 ° visual angle)	High accuracy (approx. 0.25–1.0 ° visual angle)
Sampling rate	Low to moderate (approx. 30–60 Hz)	High (approx. 120–1000 Hz)	Moderate to high (approx. 60–250 Hz)
Saccade analysis	Limited precision for rapid eye movements	Very precise detection of saccades	Good precision for saccades
Mobility	High (remote use supported)	Low (requires stationary setup)	Very high (supports free user movement)
Hardware requirements	Standard webcam only	Specialized infrared cameras and sensors	Wearable device with integrated eye and scene cameras
Sensitivity to lighting	Yes	Yes	Yes
Cost	Low (no specialized hardware required)	High (approx. $2,000–$10,000+)	Very high (approx. $5,000–$20,000+)

Eye-Tracking Systems

Several eye-tracking systems are widely used in research and industry, offering different levels of accuracy, sampling rates, and usability. Among them, the Gazepoint GP3 HD, Tobii Pro Fusion, and Pupil Labs Core represent three distinct eye-tracking examples. These systems differ in their hardware configuration, mobility, and intended use cases.

- **Gazepoint GP3 HD:** The Gazepoint GP3 HD is a screen-based eye tracker developed by Gazepoint, that uses a machine-vision camera at the heart of its imaging and processing system [40].

- **Tobii Pro Fusion:** Tobii Pro Fusion is a versatile, mid-range screen-based eye tracker that captures gaze data at 250 Hz, supporting both fixation and saccade-based studies in and outside the lab [41].
- **Pupil Labs Core:** Pupil Core is an eye-tracking platform consisting of a wearable headset and an open-source software suite, serving as an open platform used by a global community of researchers [42].

Table 2.3 compares three eye-tracking systems with respect to tracking approach, performance, and mobility. Gazepoint GP3 HD and Tobii Pro Fusion are screen-based systems intended for controlled laboratory use, with Tobii Pro Fusion offering higher accuracy, precision, and sampling rates, making it suitable for detailed fixation and saccade analysis. In contrast, Pupil Labs Core employs a wearable headset design, enabling eye-tracking in real-world environments with high precision and mobility, however with slightly lower accuracy.

Table 2.3: Comparison of eye-tracking systems [40–43]

Feature	**Gazepoint GP3 HD**	**Tobii Pro Fusion**	**Pupil Labs Core**
Eye-tracking type	Screen-based	Screen-based	Wearable headset
Accuracy	0.5–1.0°	0.3°	0.6°
Precision	0.1°	0.04°	0.02°
Sampling rate	60 or 150 Hz	30, 60, 120, 250 Hz	200 Hz
Calibration	5- or 9-point	5-point (configurable up to 9)	5-point
Mobility	Low (screen-based)	Low (screen-based)	High (wearable headset)
Hardware requirements	Standard PC, compatible display ≤24"	Standalone monitor or laptop, compatible display ≤24"	Desktop or laptop
Portability	Ultra portable	Standard lab device	Headset

2.2.2 AI models

AI models are built using different algorithms, each with distinct structures, learning mechanisms, and capabilities. Among the most commonly used approaches are RF, XGBoost, SVM, and MLP, each offering unique strategies for handling data, learning patterns, and making predictions.

Random Forest (RF)

Random Forest is an ensemble learning algorithm widely used in ML that builds multiple decision trees during the training process and combines their predictions to produce a final output [32, 44]. The model uses a parallel ensembling strategy in which each decision tree is trained on a different bootstrap sample of the original dataset, while random subsets of features are considered at each split [32]. This combination of bootstrap aggregation and random feature selection introduces diversity among the trees, reducing correlation between individual models and mitigating over-fitting [32]. For classification tasks, the final prediction is typically obtained through majority voting, while regression tasks rely on averaging the outputs of all trees [32]. RF generally achieves higher predictive accuracy and improved generalisation performance compared to single decision-tree models, and it is applicable to both classification and regression problems involving categorical and continuous data [32, 44].

Extreme gradient boosting (XGBoost)

Extreme Gradient Boosting is an ensemble learning algorithm based on the gradient boosting framework, where multiple decision trees are built sequentially to form a strong predictive model [32, 44]. Unlike parallel ensemble methods, gradient boosting trains each new tree to correct the errors made by previous ones by minimizing a specified loss function using gradient-based optimisation [32]. XGBoost extends traditional gradient boosting by incorporating second-order gradient information, allowing more accurate approximations during optimisation and improving convergence efficiency [32]. Additionally, it integrates advanced regularization techniques, including L1 and L2 penalties, which help control model complexity and reduce overfitting [32]. The algorithm is designed for computational efficiency and scalability, supporting fast training, efficient memory usage, and robust handling of large and complex datasets [32, 44].

Support Vector Machine (SVM)

Support Vector Machine is a widely used ML algorithm applicable to classification, regression, and pattern recognition tasks [32]. The fundamental idea of SVM is to construct a decision boundary, known as a hyperplane, in a high-dimensional feature space that best separates the data into distinct classes [32]. This separation is achieved by maximising the margin, defined as the distance between the hyperplane and the nearest data points from each class, referred to as support vectors. Maximising this margin contributes to improved generalisation performance by reducing classification error on unseen data [32].

Multilayer Perceptron (MLP)

Multilayer Perceptron is a fundamental architecture in deep learning and represents a class of feed-forward artificial neural networks [32]. An MLP consists of an input layer, one or more hidden layers, and an output layer, where each neuron in a layer is fully connected to neurons in the subsequent layer through weighted connections [32]. Information flows unidirectionally from the input to the output layer, and each neuron computes a weighted sum of its inputs followed by a nonlinear activation function, enabling the network to model complex relationships within the data [32].

Comparison of AI Models

Table 2.4 summarises the main characteristics of the four AI models discussed in this section. RF and XGBoost are both tree-based ensemble methods capable of handling non-linear relationships, with RF being particularly robust to missing data and unbalanced classes, while XGBoost is optimised for large datasets. SVM is a kernel-based classifier suitable for cleaner, low-noise datasets, whereas MLP is a neural network model designed for large, preprocessed datasets with complex non-linear patterns. All four models have high computational complexity, making them more suitable for tasks where training resources are available.

Table 2.4: Comparison of AI models [32, 45–48]

Feature	RF	XGBoost	SVM	MLP
Model type	Tree-based ensemble	Tree-based ensemble	Kernel-based classifier	Neural network
Handling non-linearity	Yes	Yes	Yes	Yes
Best-suited datasets	Datasets with missing values or unbalanced classes	Large-sized datasets	Datasets with low noise	Large-sized datasets
Computational complexity	High	High	High	High
Problem type	Classification and regression	Classification and regression	Classification and regression	Classification and regression

2.3 Systematic Literature Review Methodology (PRISMA)

This project adopts a Systematic Literature Review methodology based on the PRISMA framework. PRISMA is a widely recognized methodology that provides clear guidelines for identifying, screening, and selecting scientific studies in a transparent and reproducible manner. The adoption of the PRISMA methodology ensures a structured and transparent literature review process, enabling a clear synthesis of the current state of knowledge in the Research area.

2.3.1 Research Questions

The work is guided by the following central research question: **How can eye-tracking technologies integrated with AI be used to analyse users' visual attention and emotional responses on e-commerce websites to enable personalised content and adaptive UI/UX design?**

Seven complementary research questions were defined with the purpose of supporting the main question described above and address specific aspects related to attention metrics, visual design elements, emotional influence, and AI-based emotion recognition. Table 2.5 presents the complete list of research questions that structure the systematic literature review and guide the following development of the proposed system.

Table 2.5: Research Questions

Identifier	Research Question
RQ1	What are the most effective metrics to evaluate users' attention on e-commerce websites?
RQ2	What are the main design elements (colour, contrast, arrangement of images and text) that most influence gaze focus on websites?
RQ3	What visual presentation practices (colours, style, brightness) can promote positive/negative emotions and influence users' decisions?
RQ4	What is the relationship between visual attention and emotions expressed while viewing products online?
RQ5	Which artificial intelligence models are most effective at detecting emotions from eye-tracking data?
RQ6	How can AI algorithms be trained to associate visual patterns (gaze patterns) with emotional responses?
RQ7	How can an intelligent system dynamically adapt the website content based on users' emotions and focus?

2.3.2 Search Strategy

Following the definition of the research questions, a structured search strategy was established to identify relevant scientific literature. A set of keywords aligned with the scope of the study were identified for each domain. These then served as a structure to create the search string applied across the selected scientific article databases, which are described later in this report. Table 2.6 presents the keywords identified for each domain of the study.

Table 2.6: Key Words Identified by Domain

Domain	Key-Words
E-commerce	e-commerce, website, UX/UI
Eye-Tracking	eye-tracking, gaze tracking
AI	ML, AI
Emotions	emotion, emotional response, cognitive data, physiological measure, behavioral metric

The identified keywords were combined using Boolean operators to define a relevant search string aimed at retrieving scientific articles that cover all the relevant domains of the project. Table 2.7 presents the search string used to retrieve the research articles considered in this review.

Table 2.7: Search String Used in the Literature Review

("e-commerce" OR "website" OR "UX" OR "UI" OR "User Interface" OR "User Experience")
AND ("eye-tracking" OR "gaze tracking")
AND ("ML" OR "AI" OR "Machine Learning" OR "Artificial Intelligence")
AND ("emotion" OR "emotional response" OR "cognitive data" OR "physiological measure" OR "behavioral metric*")

A systematic literature review was conducted across the article databases listed in Table 2.8. The search for relevant research articles was performed using the search string defined previously.

Table 2.8: Articles Database

Identifier	**Source**	**URL**
AD1	ACM Digital Library	`https://dl.acm.org/`
AD2	SpringerLink	`https://link.springer.com/`
AD3	PubMed	`https://pubmed.ncbi.nlm.nih.gov/`

To ensure the quality and relevance of the articles included in this literature review, a set of inclusion and exclusion criteria was applied to the research articles retrieved from the selected databases. These criteria helped in selecting studies that are most relevant to the objectives of this project and in defining the body of research that forms the foundation of this thesis. Table 2.9 presents the inclusion criteria applied to the retrieved articles, and Table 2.10 presents the exclusion criteria.

Table 2.9: Inclusion Criteria

Identifier	**Inclusion Criteria**
IC1	Articles whose primary focus is on analysing user emotions in digital content.
IC2	Articles that use eye-tracking technologies or AI methods to study emotions.

Table 2.10: Exclusion Criteria

Identifier	**Exclusion Criteria**
EC1	Articles that do not provide an adequate description of methodologies or experiments.
EC2	Articles published before 2016.

2.3.3 Literature Selection

The literature selection process involved screening the retrieved articles to ensure they met the predefined inclusion and exclusion criteria. This screening was performed in multiple stages, including title and abstract review, full-text assessment, and final selection of studies for inclusion in the systematic review.

The retrieval of research articles from the selected databases resulted in a total of 198 studies: 96 from SpringerLink, 101 from ACM Digital Library, and 1 from PubMed.

The first step involves applying a publication date filter, retaining only articles published from 2016 onwards, in line with the exclusion criteria (EC2). This reduced the pool to 155 studies. These studies formed the initial pool for the systematic screening process according to the predefined inclusion and exclusion criteria.

Next, titles and abstracts of these 155 articles were screened to verify compliance with the inclusion criteria (IC1 and IC2), resulting in 49 studies considered relevant.

Of these, 13 articles were inaccessible due to paywalls. Following a thorough full-text reading of the remaining studies, 21 articles were excluded because they did not provide a clear description of methodologies or experiments, in accordance with exclusion criterion EC1.

After completing all screening steps, 14 articles remained eligible and were included in the literature review. Figure 2.4 illustrates the flow of the article selection process.

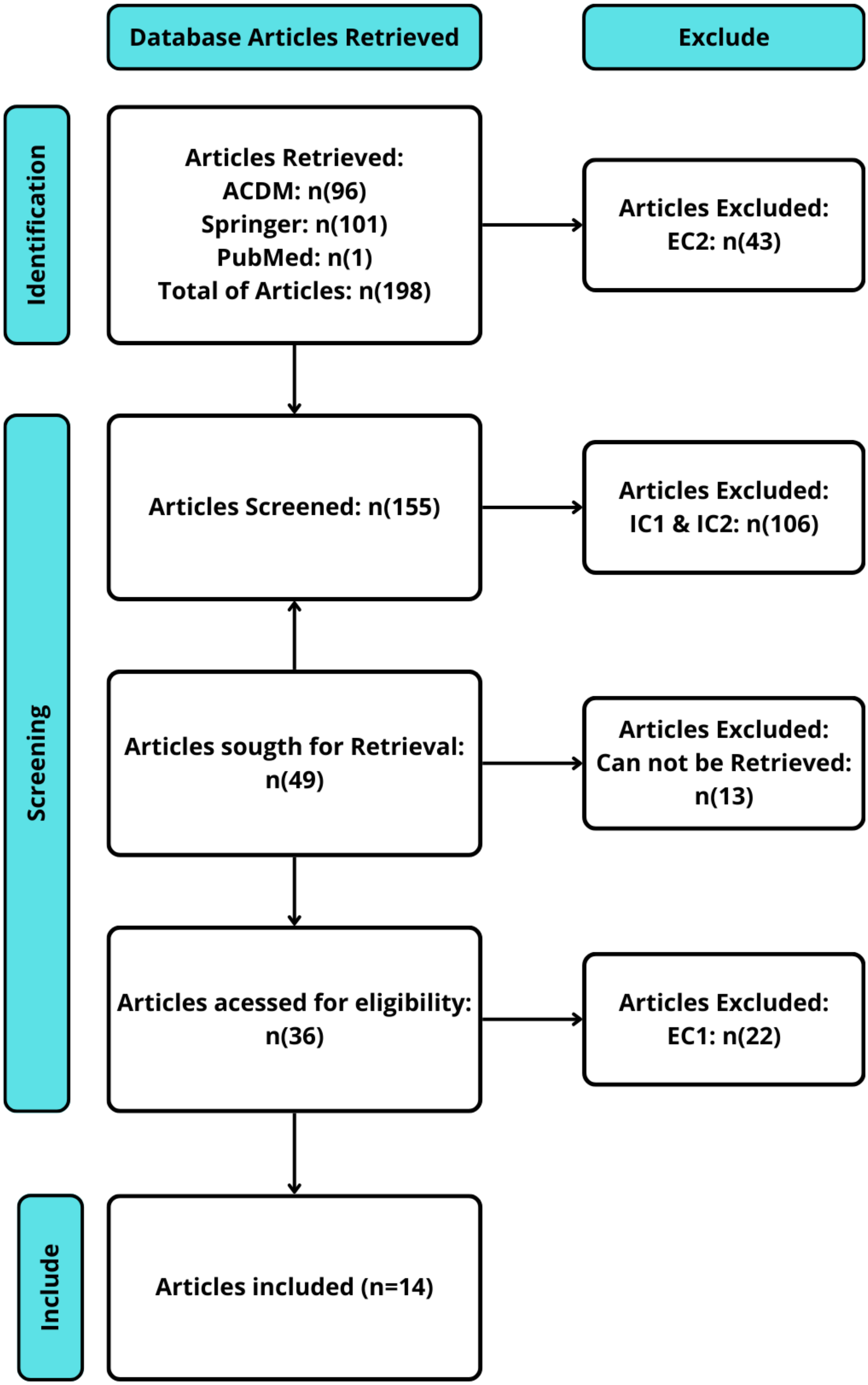

Figure 2.4: PRISMA Flow

2.4 Discussion of Findings

This section discusses the findings derived from the literature review. It first provides an analysis of the retrieved literature, organising existing evidence according to a respective research question. Based on this analysis, a research gap is identified, highlighting limitations in existing approaches and motivating the focus and contribution of the present thesis.

2.4.1 Analysis of Retrieved Literature

The following analysis addresses the research questions by synthesizing findings from the scientific literature identified in the review process. Relevant studies are examined in relation to each research question.

RQ1: What are the most effective metrics to evaluate users' attention on e-commerce websites?

Eye-tracking studies consistently identify fixation-based metrics as the most effective indicators of users' visual attention, with measures such as fixation duration, total fixation duration, fixation count, and dwelling time widely used to determine where and for how long users focus on interface elements [49–51]. Physiological indicators such as pupil size are also employed to capture early cognitive and perceptual responses, revealing differences in user engagement and processing intensity [52]. These metrics are particularly valuable because they reflect cognitive processing intensity and selective attention, revealing how users allocate visual resources across different interface regions or content types [50, 53]. Complementing fixation measures, temporal metrics such as TTFF is effective for identifying which elements capture attention earliest [53, 54]. Saccade-based metrics, including saccade length, scan distance, and gaze transitions between predefined AOI, further enrich attention analysis by capturing visual exploration behavior and navigation patterns across interface elements [51, 55]. Visualisation-based representations such as heatmaps and scan paths are also extensively used to summarise collective attention patterns and reveal differences in visual focus across interface designs, even when simple fixation counts are insufficiently discriminative [49, 52].

RQ2: What are the main design elements (colour, contrast, arrangement of images and text) that most influence gaze focus on websites?

Studies indicate that various visual factors, including the number of objects, overall clutter, openness, organisation, and colour variety, affect user cognitive load and guide gaze behavior, shaping how users scan and process information on a webpage [49]. Well-structured and symmetrical interface elements tend to support more consistent visual scanning, whereas cluttered or poorly organised layouts can interfere with attention and disrupt gaze patterns [52]. Specific design choices such as colour palettes, contrast, images, font style, and font size can evoke emotional responses that interact with cognitive load, further influencing where users focus and how they engage with the interface [56]. Furthermore, bottom-up visual factors, including colour contrast, orientation, and intensity, play a key role in guiding gaze allocation and attention distribution, often acting automatically to capture user focus [57].

RQ3: What visual presentation practices (colours, style, brightness) can promote positive/negative emotions and influence users' decisions?

Studies show that interface elements such as colours, brightness, font style, font size, and typographic markup can be linked to specific emotional states, often visualised through emotion heatmaps that highlight where positive or negative reactions occur [49, 58]. Simple, clear, and aesthetically pleasing designs are associated with higher valence and arousal, promoting positive emotions, whereas cluttered, complex, or inconsistent layouts tend to promote negative emotional responses [52]. Colour is particularly influential, with schemes mapped to emotional states, for example, through models like Bianchi's colour circle, and adaptations such as background colour changes can directly adjust user mood and engagement [56]. Additionally, primitive visual shapes influence affective reactions, with rounded or oval shapes generally evoking more positive responses compared to sharp or angular forms [58]. Cultural symbols and colour features such as lightness, saturation, and overall colourfulness further contribute to emotional valence. [54, 59].

RQ4: What is the relationship between visual attention and emotions expressed while viewing products online?

Visual attention and emotional responses are connected, with gaze data often synchronized with physiological and behavioural signals, such as pupil size, electrodermal activity, and HR, to infer users' emotional and cognitive states during interface interaction [49, 50, 60]. Heatmaps and affective saliency maps can highlight specific areas of a visual interface or product associated with emotional reactions [49, 58].

RQ5: Which artificial intelligence models are most effective at detecting emotions from eye-tracking data?

Several studies have leveraged ML techniques to analyse eye-tracking data, often in combination with physiological signals, to infer cognitive or affective states [49, 52]. Commonly employed algorithms in related eye-tracking applications include RF, SVM, k-Nearest Neighbor, XGBoost, and deep learning methods, like convolutional neural networks used for tasks such as classifying, which have been used to classify user expertise, preferences, or affective responses based on eye movement patterns [53, 58, 61].

RQ6: How can AI algorithms be trained to associate visual patterns (gaze patterns) with emotional responses?

AI algorithms are trained to associate visual patterns, such as gaze features, with cognitive or emotional responses by using eye-tracking metrics, including fixations, saccades, scan paths, and TTFF, as input features for ML models [49, 51]. In many approaches, physiological signals such as pupil dilation, Electroencephalography (EEG), and facial expressions are segmented based on gaze behavior and incorporated into classifiers to infer cognitive load or affective states [49, 52]. Deep learning methods, particularly convolutional neural networks, can be trained on large-scale datasets where aggregated user reactions to visual stimuli serve as soft target labels, enabling the prediction of emotional responses to new stimuli [58]. Additionally, clustering techniques like K-means can be applied to fixation-based metrics, such as total fixation duration and TTFF, to uncover patterns of attention that correlate with emotional states, forming a foundation for supervised training of AI models [54].

RQ7: How can an intelligent system dynamically adapt the website content based on users' emotions and focus?

Intelligent systems can dynamically adapt website content by using user emotional states and behavioural patterns to personalise interface elements in real time. Adaptive UIs, such as UIFlex 2.0, modify features like background colour, font size, audio, and screen brightness in response to detected emotions, using an Adaptation Engine to apply changes automatically based on classified emotional data [56]. Similarly, intelligent systems can utilize eye-tracking metrics and user profiles to adjust visual characteristics in order to increase user engagement and guide attention [54]. These adaptations can be further optimised using real-time algorithms, to maximise user focus and interaction outcomes [54]. In general, frameworks that incorporate saliency prediction or demographic-aware attention models can support the design of adaptive, user-centred web interfaces, laying the groundwork for dynamic personalisation [60, 62].

2.4.2 Research Gap

To sum up, although existing studies address visual attention, emotional responses, and AI-based adaptation independently, there is a clear lack of research integrating these components into a unified framework for e-commerce. In particular, few studies establish a direct link between gaze patterns and implicit user preferences, treating product selection as a behavioural signal for positive emotional engagement. Even fewer investigate how multimodal data combining eye-tracking, physiological signals, and contextual design variables can be used to train predictive models capable of inferring product preference. This thesis aims to fill this gap by developing and evaluating ML models that predict product selection from visual attention and physiological data, contributing towards the longer-term goal of supporting adaptive and user-centred UI/UX design in e-commerce.

Chapter 3

Experimental Study

This chapter presents the experimental design implemented in this study, detailing the methodological framework used to investigate user attention, cognitive effort, and emotional responses in an e-commerce context.

3.1 Experimental Design

The following section presents a comprehensive overview of the experimental protocol implemented in this study, including the design rationale, procedural structure, and key methodological components that guided the data collection.

3.1.1 Objective

The present experiment was designed to investigate how visual layout and product presentation style influence visual attention in an e-commerce environment. More specifically, the study aims to understand how users allocate gaze when interacting with structured product grids and whether certain design factors, such as spatial position within a grid and contextual versus isolated product presentation, affect early attention, fixation behaviour, and overall visual engagement.

This research is based on the fact that visual attention plays a central role in online consumer behaviour, particularly in e-commerce platforms where users are continuously exposed to multiple competing visual stimuli. In such environments, subtle design decisions, such as product positioning or the inclusion of contextual product presentation, may significantly influence what users notice first, what they explore longer, and potentially what they consider for purchase.

Therefore, the experiment was developed to simulate a realistic online shopping experience while maintaining strict experimental control. The interface was based on the e-commerce platform Sanimaia, a website that offers a wide range of household products across multiple categories, including bathroom, kitchen, living, and tableware. This approach allowed for the systematic measurement of attention related metrics using eye-tracking and physiological data within a realistic yet controlled environment.

3.1.2 Prototype Development

Experimental Interface and Presentation Format

The experimental interface was conceived and developed using Figma with the objective of replicating a realistic e-commerce browsing experience based on the Sanimaia website. The initial design goal was to create an interactive prototype that would allow participants to navigate the interface freely in a manner similar to real-world online shopping behaviour.

This approach was motivated by the need to maximise ecological validity while maintaining a familiar shopping environment, thereby encouraging participants to engage with the stimuli in a natural and intuitive way. To support this objective, key structural components of the original website, such as the header and sidebar, were replicated. These elements were included to preserve the original structure and visual design of the website, ensuring that participants experienced an environment closely resembling the actual Sanimaia online shopping interface. A visual comparison between one of the experimental frames and the original website interface is presented in Figures 3.1 and 3.2.

Although the initial objective was to develop a fully interactive prototype that would replicate natural online shopping behaviour, this approach was ultimately deemed unsuitable for the requirements of the experimental design. While interactivity would enhance ecological validity, it would also introduce significant methodological limitations, particularly in terms of uncontrolled exposure time, inconsistent experimental conditions across participants, and difficulties in defining and standardising AOIs for eye-tracking analysis. These factors would reduce the comparability and reliability of the collected data. Therefore, an alternative controlled presentation format was required to ensure methodological rigour and experimental consistency. Several options were considered, including PowerPoint-based slides and a video-based approach. However, both interactive and slide-based approaches still allowed participant-driven navigation or timing variability.

As a result, a video-based presentation format was selected as the final experimental delivery method. This approach ensured full control over stimulus timing, sequence, and exposure duration while maintaining the visual structure originally designed in Figma.

Stimuli Design and Product Layout

A total of 16 images containing different products were created and distributed evenly across four product categories: Bathroom, Tableware, Kitchen, and Living. Each category included four frames, resulting in a balanced experimental structure that ensured equal representation of product domains throughout the study. This balance was important for avoiding category-driven biases in visual attention and for ensuring that the observed effects could be attributed to the experimental manipulations rather than to thematic differences between product types.

Each frame consisted of eight products arranged in a 2×4 grid layout. All products were real items displayed on the Sanimaia e-commerce website. Within each frame, products were selected to ensure consistency in terms of product category and functional purpose.

A key manipulation in the stimulus design concerned product presentation style. Each frame included two distinct types of visual representation:

- **Contextualised products**: products shown in real-world usage contexts
- **Isolated products**: products presented on a neutral white background

This manipulation was introduced to investigate whether contextual information influences the allocation of visual attention. More specifically, it enables the analysis of whether products presented in realistic usage scenarios attract more initial attention, longer fixation durations, or more frequent revisits than visually isolated product representations.

In addition to presentation style, the study followed a controlled within-subjects experimental design in which participants were exposed to different variations of product positioning within the 2×4 grid layout. This manipulation enabled the systematic investigation of spatial attention effects, particularly whether certain positions within a structured visual grid naturally attract earlier or stronger visual engagement.

An example of an experimental frame and the original interface of the *Sanimaia* website are presented in Figures 3.1 and 3.2, respectively.

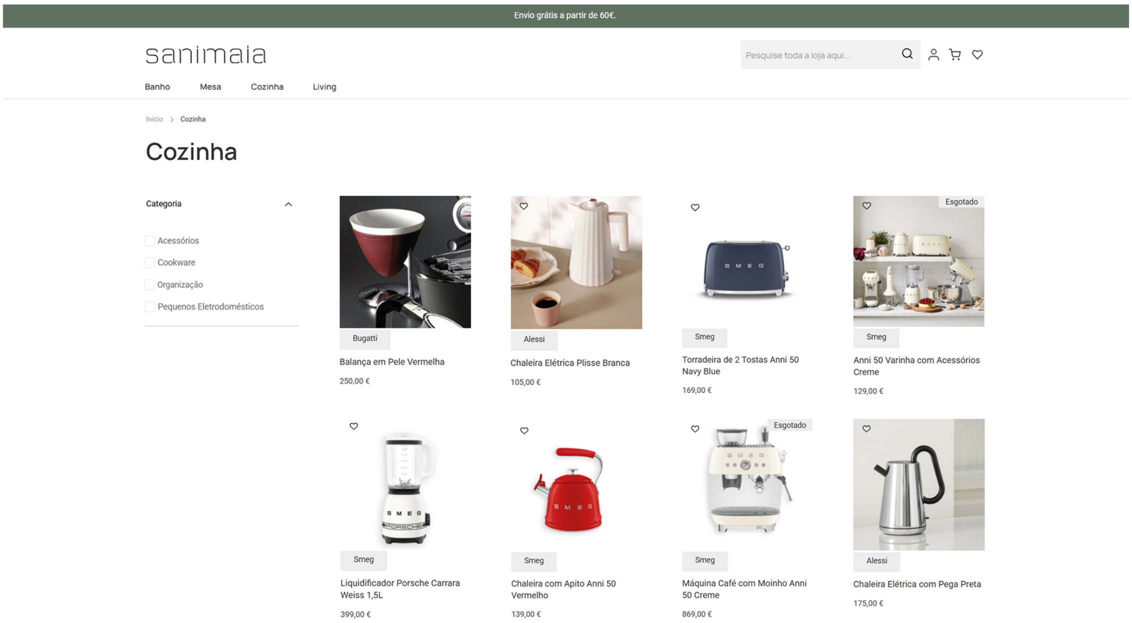

Figure 3.1: A Frame retrieved from the experiment designed in Figma

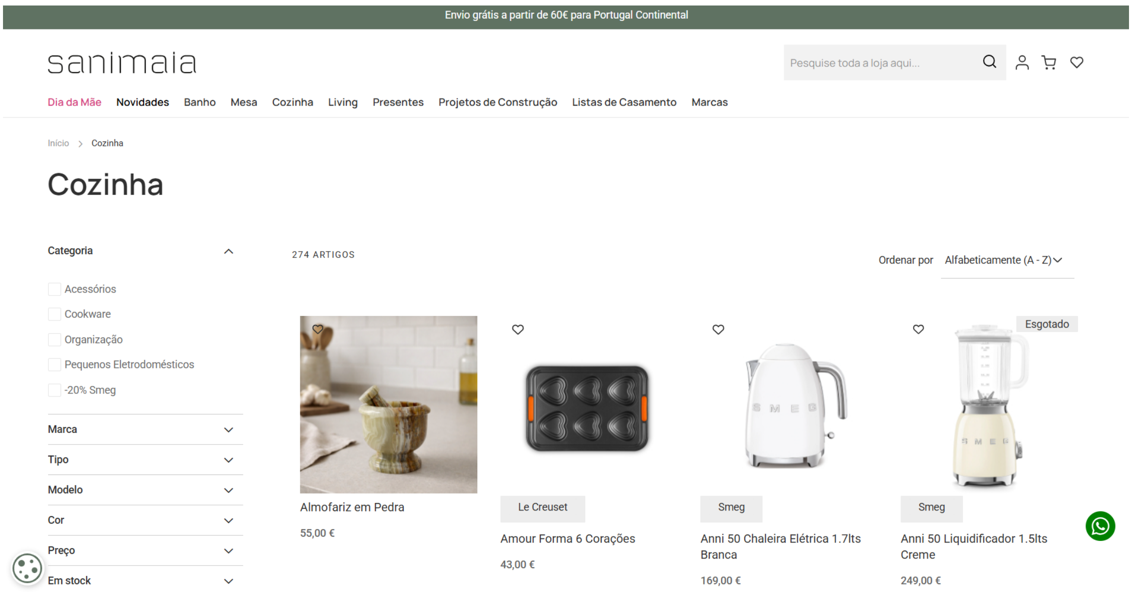

Figure 3.2: The Sanimaia Website Interface

Video Structure and Pilot Study Refinement

The final video followed a strictly standardised temporal structure designed to optimise data quality and minimise carry-over effects between frames. The sequence began with a 20-second neutral screen consisting of a plain white background, serving as a baseline period to stabilise participants' gaze and reduce residual visual activation from environmental distractions. Following this, each of the 16 frames was displayed for 15 seconds. This duration allowed sufficient time for natural visual exploration while ensuring consistent exposure across all conditions.

After the initial design and creation of the video, a pilot study was conducted with four participants to identify potential methodological issues prior to the main data collection phase. During this pilot study, a key limitation was identified in the transition between consecutive frames, as participants did not experience a sufficient visual reset. As a result, gaze position tended to carry over from one frame to the next. In practice, if a participant was fixating on a product at the moment of a frame transition, the first fixation in the subsequent frame often occurred at the same spatial location. This introduced a bias in early attention metrics, particularly TTFF, as initial gaze behaviour was no longer independent of the preceding frame.

To address this issue, a gaze reset mechanism was introduced in the final version of the video by inserting a 3 second fixation cross between each of the 16 experimental frames. The cross was positioned centrally within the header area, outside the product grid, and served to redirect participants' attention away from the product region before the onset of each new frame. This controlled reset of visual attention reduced carry-over effects between frames and improved the validity of early fixation-based metrics, such as TTFF.

3.1.3 Participants

The final study included 30 participants, divided into five experimental groups of six participants each. Participants were recruited on a voluntary basis and were all aged 18 or older. All participants reported normal or corrected-to-normal vision and no known visual impairments that could interfere with eye-tracking measurements. Each group was exposed to a different variation of product positioning within the grid layout, enabling the analysis of position effects on visual attention.

3.1.4 Materials

The experiment was conducted using a Gazepoint GP3 HD eye-tracking system, together with physiological sensors measuring HR and GSR. The experimental video, developed from the Figma prototype, was displayed on an external monitor.

3.1.5 Experiment Procedure

A standard procedure was used across all experimental sessions to ensure consistency between every participant.

At the beginning of each session, participants were informed about the purpose and procedures of the study in a clear and standard manner. They then provided informed consent through a Google Forms, which also collected basic demographic information. Data were collected in a pseudonymised manner and no directly identifying personal information was stored alongside the recorded data.

The setup phase followed, during which the eye-tracking system and physiological sensors were calibrated individually to ensure accurate data collection. Eye-tracker calibration was performed using the standard 5-point procedure in the Gazepoint Analysis software. This process was repeated until both the participant and the researcher confirmed satisfactory tracking accuracy. Participants were positioned approximately 55 cm from the eye-tracking device to maintain optimal recording conditions, and readiness to begin was confirmed once calibration was completed.

The experimental video was then presented under controlled laboratory conditions. Participants were instructed to observe the frames freely, without completing any task, to preserve natural viewing behaviour. During the session, the researcher remained behind the participant to minimise potential influence or distraction.

All sessions were conducted under controlled environmental conditions, including standardised lighting and the absence of external distractions.

Each session lasted approximately seven minutes and concluded with a debriefing, during which participants had the opportunity to ask questions about the study.

3.1.6 Data Processing and AOI Definition

A structured AOI framework was defined for each experimental condition (gaze project) to support the statistical analysis of eye-tracking data. For each individual product within every frame, an AOI was created, allowing systematic quantification of fixation-based metrics such as fixation count, dwell time, and TTFF.

To ensure consistency across the five experimental groups, a unified naming convention was applied to all AOIs, following the simple format FrameX_ProductX_Product. This convention enabled clear identification of each AOI based on its corresponding frame and product position, facilitating comparisons across experimental conditions.

Once AOI definitions were verified, all eye-tracking datasets were exported for further analysis. The resulting data were processed using Python-based analytical tools to examine the effects of spatial positioning and presentation style on visual attention.

3.1.7 Results

This section presents the results obtained from the eye-tracking analysis conducted across the 16 experimental frames. The analysis focused on four primary gaze-related metrics: Time Viewed, Fixation Count, Revisits, and TTFF. The results were analysed by structuring the dataset according to three main grouping factors: product category, spatial position within the grid, and presentation style (contextualised versus isolated products). In addition, interaction effects between positioning and presentation style, as well as correlations between eye-tracking metrics, were examined to provide a more comprehensive understanding of user visual behaviour.

AOI-based Analysis

Given the large number of individual AOIs, a direct product-by-product interpretation of the results would not provide meaningful insights at this stage. Therefore, heatmaps were generated to facilitate the visualisation of eye-tracking metrics across all products, enabling a clearer identification of general attention patterns. Figure 3.3 presents a heatmap of the Time Viewed metric across all AOIs, while the remaining heatmaps for the other metrics are included in C.

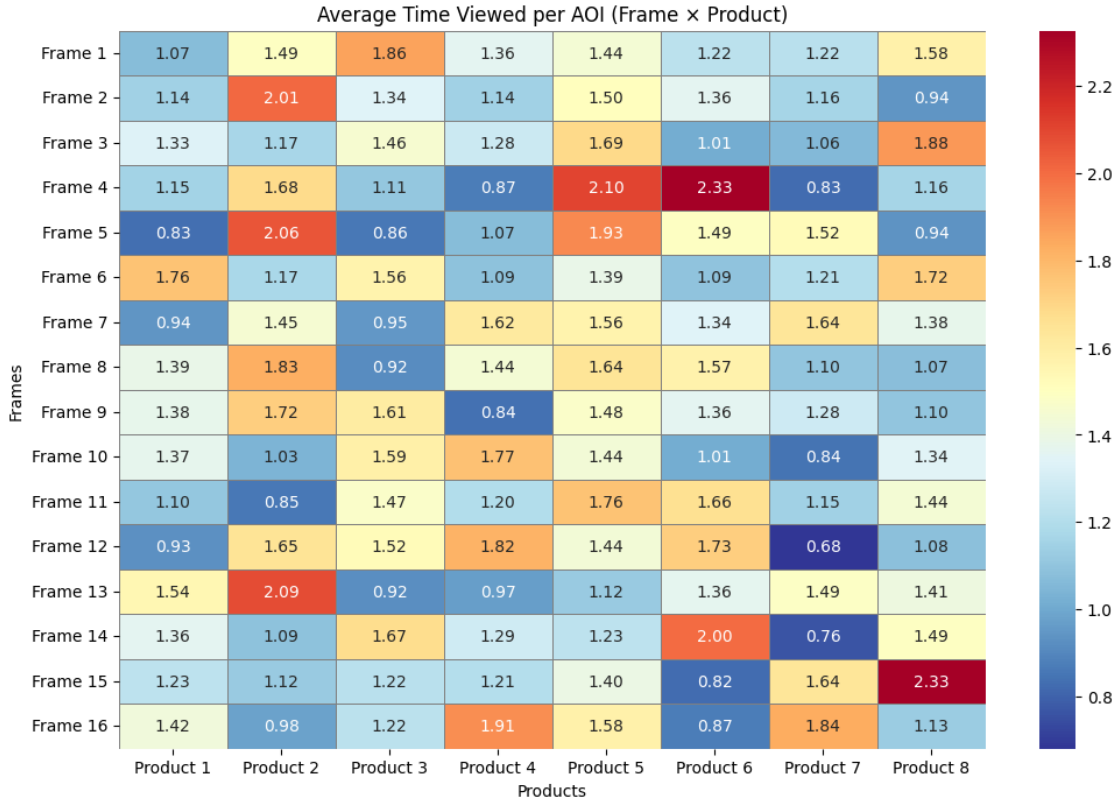

Figure 3.3: Heatmap of Time Viewed across all AOIs

These visualisations allowed for a more intuitive understanding of how attention was distributed across the interface, making it easier to compare engagement levels between products and identify areas of higher or lower visual interest.

Category-Based Analysis

The comparison between product categories revealed only minor variations across the analysed eye-tracking metrics. A vertical bar chart was used to support the visual comparison of average values across categories. Figure 3.4 presents the distribution of Time Viewed, while the remaining metrics are included in D.

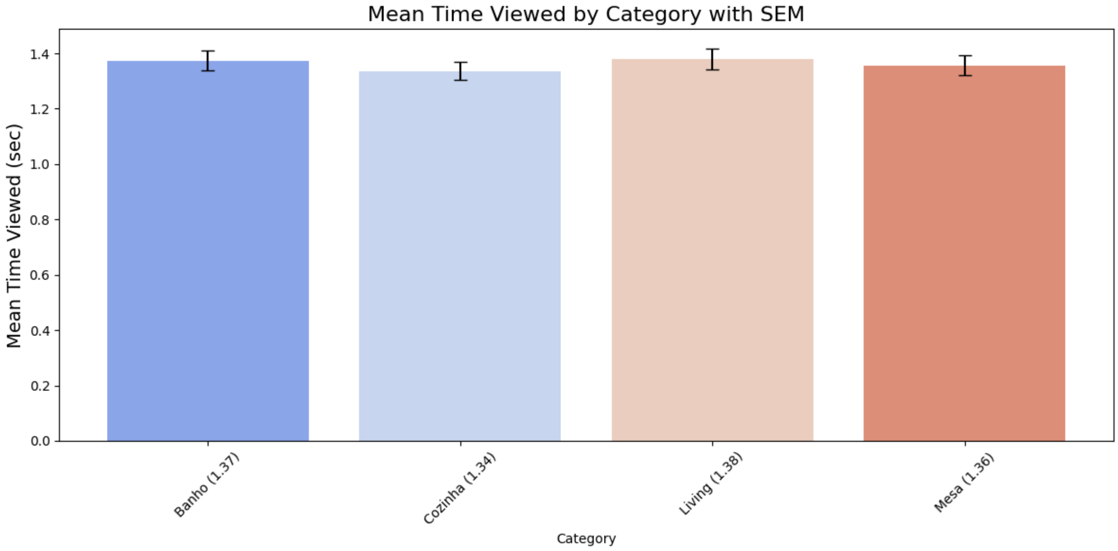

Figure 3.4: Bar chart of Time Viewed across product categories

To complement the descriptive analysis, a one-way ANOVA was conducted to assess whether statistically significant differences existed between product categories across the main eye-tracking metrics. The results are summarised in Table 3.1.

Metric	**F-value**	**p-value**
Time Viewed	0.302	0.824
Fixations	0.464	0.708
Revisits	2.210	0.085
TTFF	5.120	0.0016

Statistical significance when $p < 0.05$.

Table 3.1: One-way ANOVA results for product categories

The ANOVA results indicate that product category does not have a statistically significant effect on most visual attention metrics. No significant differences were found for Time Viewed, Fixations, or Revisits. A statistically significant effect was only observed for TTFF, suggesting differences in initial attentional capture between categories. However, this result should be interpreted with caution, as it may be influenced by an order effect in the experimental design. Since the experiment always began with a 20-second neutral white screen followed by the first stimulus frame, the initial exposure may have benefited from a transition effect, leading to systematically lower TTFF values for the first presented condition. As a result, this effect may reflect stimulus order rather than inherent differences between product categories.

The results suggest that product category plays a limited role in shaping visual attention and visual engagement.

Position-Based Analysis

To analyse the influence of spatial positioning on visual attention, vertical bar charts were created for each position within the 2×4 grid layout. These visualisations compared the average values obtained for the five analysed metrics: Time Viewed, Fixations, Revisits, TTFF, and Percentage Seen. Figure 3.5 presents the Time Viewed distribution across positions, while the remaining metric visualisations are included in E.

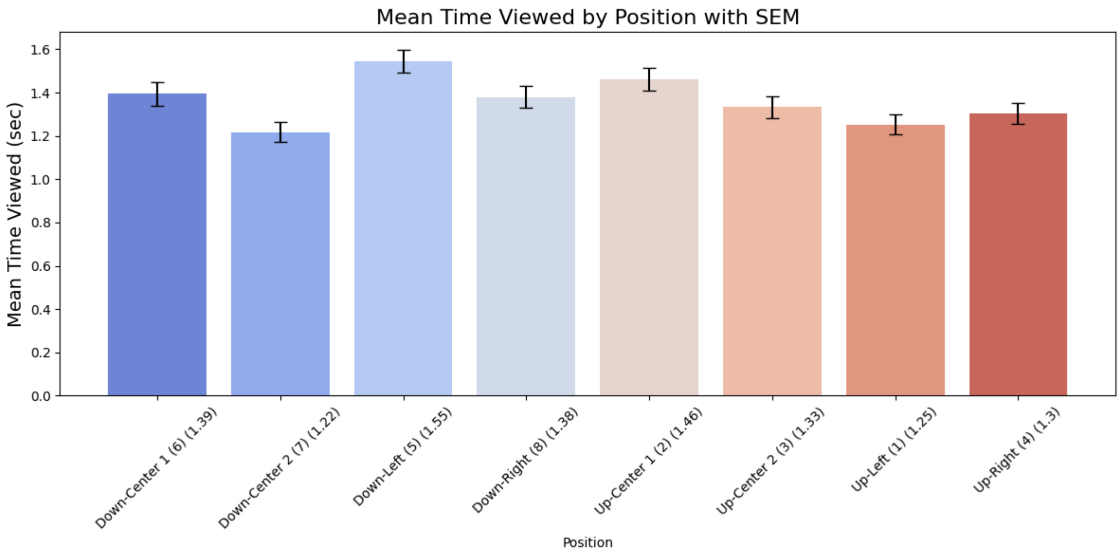

Figure 3.5: Bar chart of Time Viewed across grid positions

The visual analysis revealed noticeable differences between positions across multiple eye-tracking metrics. Certain positions consistently attracted higher levels of visual engagement, while others received comparatively lower attention. In particular, products positioned in the upper-central left region (Position 2) and lower-left (Position 5) regions of the interface generally demonstrated stronger visual engagement patterns across several metrics. However products positioned in the down-center right region (Position 7) consistently demonstrate lower levels of engagement, standing out as one of the least effective positions in attracting visual attention.

To evaluate whether these differences were statistically significant, one-way ANOVA tests were conducted for Time Viewed, Fixations, Revisits, and TTFF. The results are summarised in Table 3.2.

Metric	**F-value**	**p-value**
Time Viewed	4.756	< 0.001
Fixations	4.004	< 0.001
Revisits	1.784	0.086
TTFF	11.136	< 0.001

Statistical significance when $p < 0.05$.

Table 3.2: One-way ANOVA results for position-based analysis

The ANOVA results indicate significant differences in positional effects for Time Viewed, Fixations, and TTFF. In contrast, Revisitation behaviour did not present significant differences across positions, suggesting that spatial positioning primarily influences initial and sustained visual attention rather than repeated visual returns. Following the ANOVA, post-hoc Tukey HSD tests were conducted in order to identify which specific position pairs differed significantly for each metric. The results can be summarised by metric as follows:

- **Time Viewed:** Position 5 showed significantly higher values compared to Positions 1, 3, and 4. In contrast, Position 7 consistently exhibited lower values, particularly when compared with Positions 5 and 2.
- **Fixation Count:** A similar pattern was observed, with Positions 5 and 2 attracting significantly more fixations than Positions 1 and 7.
- **TTFF:** Position 2 was consistently associated with earlier visual attention compared to most other positions, whereas Positions 4 and 8 required longer time intervals before being initially fixated.

Overall, these findings indicate statistically significant positional effects on visual attention within the e-commerce layout, with Positions 2 and 5 emerging as the most effective regions in terms of user engagement.

Presentation-Style Analysis

The analysis of presentation style focused on comparing contextualised product images (with background) versus isolated product images (without background) across the main eye-tracking metrics. To support visual interpretation, a horizontal bar chart was used to compare the average values of Time Viewed, Fixations, Revisits, and TTFF between the two conditions. Figure 3.6 illustrates these differences.

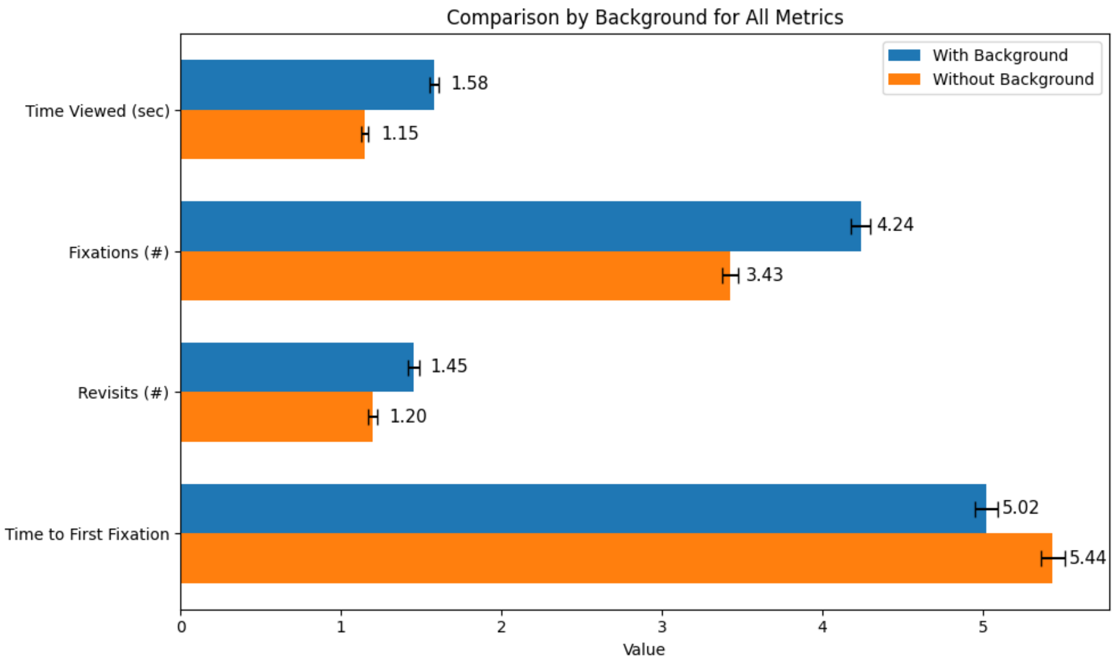

Figure 3.6: Comparison of eye-tracking metrics between background and no background conditions

In addition, a vertical bar chart was used to compare the percentage of products viewed under each condition. Figure 3.7 presents the percentage of viewed products for both presentation styles.

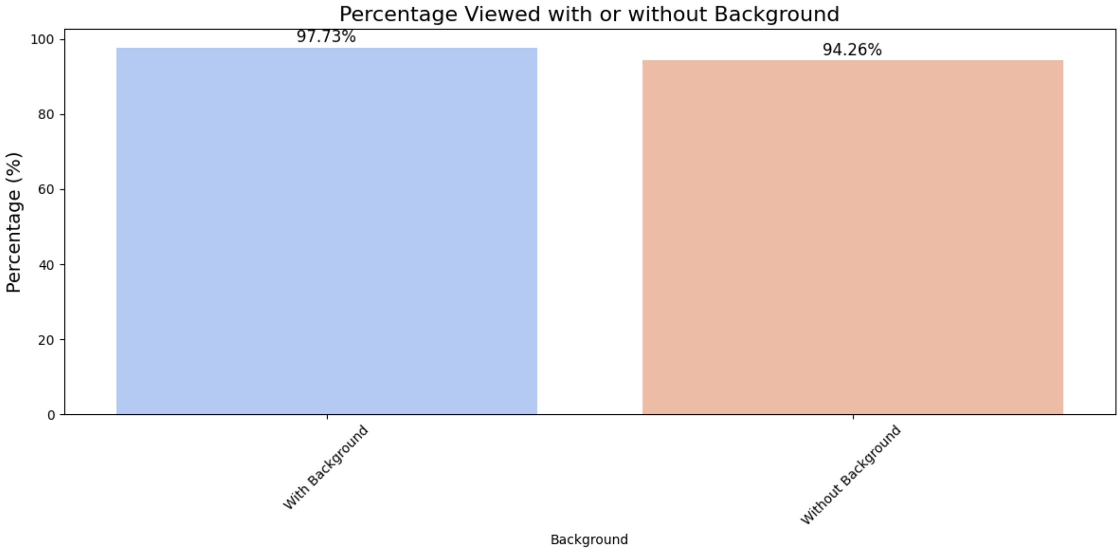

Figure 3.7: Percentage of products viewed in background and no background conditions

To assess whether the observed differences between conditions were statistically significant, independent samples t-tests were conducted for each eye-tracking metric. The results are summarised in Table 3.3.

Metric	t-value	p-value
Time Viewed	12.48	$p < 0.0001$
Fixations	10.55	$p < 0.0001$
Revisits	5.87	$p < 0.0001$
TTFF	-3.96	$p < 0.0001$

Statistical significance when $p < 0.05$.

Table 3.3: Independent samples t-test results for background vs no background conditions

The results indicate statistically significant differences between presentation styles across all analysed metrics. Products presented with a contextual background consistently received higher Time Viewed, more fixations, and more revisits compared to isolated product presentations. In contrast, TTFF was lower for the background condition. Additionally, the percentage of viewed products was higher in the background condition, indicating a broader distribution of visual attention across stimuli.

In summary, the findings suggest that contextualised product presentation is more effective at capturing and maintaining visual attention, leading to increased engagement compared to isolated product presentations.

Interaction Between Position and Presentation Style

A combined analysis was performed to evaluate the interaction between spatial positioning and presentation style on the main eye-tracking metrics. Line graphs were used to visualise differences between contextualised and isolated product presentations across all positions, with Figure 3.8 illustrating the distribution of Time Viewed. The remaining metrics are included in F.

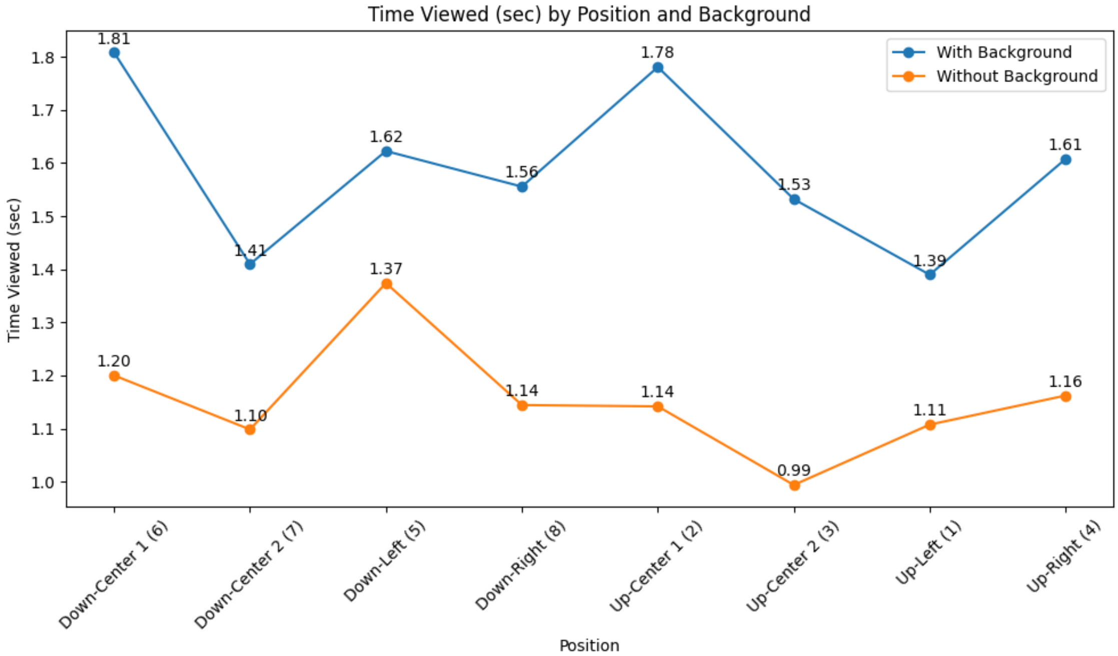

Figure 3.8: Comparison of eye-tracking metric Time Viewed across grid positions for contextualised and isolated product presentations

The visual analysis revealed a highly consistent pattern across positions, where products presented with contextual backgrounds generally achieved higher values for Time Viewed, Fixations, and Revisits when compared to isolated product presentations. Furthermore, contextualised products generally exhibited lower TTFF values, indicating earlier visual attention across the interface.

To complement the visual analysis, independent samples t-tests were conducted for each position and metric to assess whether the observed differences between presentation styles were statistically significant, with the results summarised in Table 3.4.

Position	**Time Viewed**	**Fixations**	**Revisits**	**TTFF**
Up-Left (1)	$p = 0.0017$	$p = 0.0885$	$p = 0.3489$	$p = 0.2273$
Up-Center 1 (2)	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p = 0.3499$
Up-Center 2 (3)	$p < 0.0001$	$p < 0.0001$	$p = 0.0009$	$p = 0.0200$
Up-Right (4)	$p < 0.0001$	$p = 0.0011$	$p = 0.3439$	$p = 0.0941$
Down-Left (5)	$p = 0.0183$	$p = 0.0715$	$p = 0.4462$	$p = 0.0288$
Down-Center 1 (6)	$p < 0.0001$	$p < 0.0001$	$p = 0.0029$	$p = 0.0106$
Down-Center 2 (7)	$p = 0.0022$	$p = 0.0085$	$p = 0.4618$	$p = 0.9428$
Down-Right (8)	$p < 0.0001$	$p = 0.0067$	$p = 0.0262$	$p = 0.4517$

Statistical significance when $p < 0.05$.

Table 3.4: Independent t-test results comparing contextualised and isolated product presentation across grid positions

The statistical analysis largely confirmed the patterns observed visually, with significant differences between contextualised and isolated product presentations identified across most positions for Time Viewed and Fixation Count. On the other hand, Revisits showed significant differences in a smaller number of positions, suggesting a weaker influence of presentation style on repeated visual returns compared to initial and sustained attention. Regarding TTFF, contextualised products tended to attract attention earlier in several positions, although this effect was less consistent across the entire layout.

Taken together, these findings reinforce the statistical results presented in the previous section, suggesting that the effect of contextualised product presentation remains relatively strong across different spatial positions within the interface and consistently contributes to stronger visual engagement and visual attention.

Correlation Between Eye-Tracking Metrics

The relationships between eye-tracking metrics were analysed using Spearman correlation coefficient. The results are presented in Table 3.5.

	Time Viewed	**Fixations**	**Revisits**	**TTFF**
Time Viewed	1.000	0.828	0.464	-0.197
Fixations	0.828	1.000	0.694	-0.322
Revisits	0.464	0.694	1.000	-0.406
TTFF	-0.197	-0.322	-0.406	1.000

Table 3.5: Spearman correlation matrix between eye-tracking metrics

The results indicate a strong positive association between Time Viewed and Fixation Count. Moderate positive correlations were also observed between Revisits and both Time Viewed and Fixation Count. In contrast, TTFF shows weak negative associations with the remaining metrics, suggesting that earlier visual attention is generally associated with higher levels of engagement.

Chapter 4

Choice-Based Eye-Tracking Study and Dataset Construction

This chapter presents the second experimental study conducted in this research and describes the construction of the dataset used for machine learning modelling. The first section introduces the choice-based eye-tracking experiment, detailing its objective, design, participants, materials, procedure, and AOI definition. The second section documents the dataset construction process, covering the structure of the raw data files, the definition and extraction of the feature set, and an overview of the final dataset. Together, these two components form the empirical and analytical foundation for the predictive models developed in Chapter 5.

4.1 Choice-Based Eye-Tracking Experiment

This section presents the second experimental study conducted in this research, which extends the eye-tracking analysis introduced in Chapter 3. The focus shifts from understanding how visual attention is influenced by interface design to examining how gaze behaviour relates to explicit product choice decisions. It is organised into the following components: objective, experimental design, participants, materials, procedure, and data processing.

4.1.1 Objective

Building on the eye-tracking study presented in Chapter 3, a second experiment was conducted with the primary objective of constructing a labelled dataset pairing eye-tracking features with user-selected outcomes. This dataset subsequently served as the basis for developing and evaluating ML models capable of predicting product selection from visual attention behaviour alone. While the previous experiment focused on how layout characteristics influence gaze behaviour, this study examines whether eye-tracking patterns can be associated with user product preferences.

To achieve this, participants were presented with sets of products and instructed to select the item they found most visually appealing or most likely to purchase. For each product, eye-tracking metrics were recorded during visual exploration and later linked to the participant's final selection, enabling a direct analysis of the relationship between visual attention and purchasing preference.

4.1.2 Experimental Design

The experimental design followed a repeated-measures structure in which each participant completed four product selection tasks. In each task, participants were presented with a 2×4 grid containing eight products and asked to select the item they considered most visually appealing or most likely to purchase.

Each choice task used the same set of products, while the spatial arrangement of items was varied across groups through position randomisation. This ensured that participants within the same group were exposed to identical layouts, while allowing controlled variation in product positioning across experimental conditions.

After each exposure, participants performed a single-choice task by selecting one product from the eight displayed items. As a result, each participant provided four independent selections, with all non-selected items within each choice task treated as alternative options.

This design produces a structured dataset in which each row corresponds to a specific product within a given choice task and participant. Eye-tracking metrics are associated with each product, alongside a binary label indicating whether it was selected or not, enabling a supervised learning formulation of the choice prediction task. A visual overview of the experimental flow is presented in Figure 4.1.

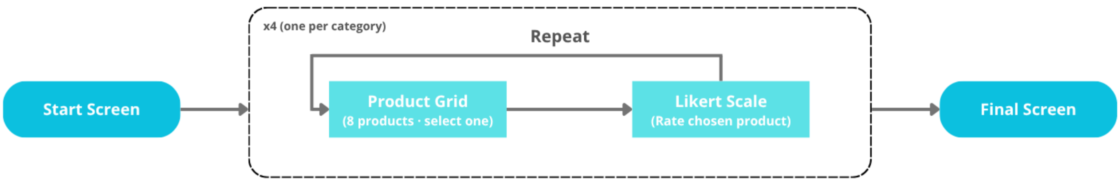

Figure 4.1: Flowchart of the choice-based experiment procedure

4.1.3 Participants

The participant sample followed a similar structure to the experiment presented in Chapter 3, with a total of 30 participants divided into five groups of six. All participants were aged 18 or older and reported normal or corrected-to-normal vision. As in the previous experiment, each group was shown the same set of products, while the positions varied across groups to ensure controlled variation in layout while keeping the stimuli consistent within each condition.

4.1.4 Materials

The materials used in this experiment were identical to those described in Chapter 3. Eye-tracking data were collected using the Gazepoint GP3 HD eye tracker, while the experimental interface was presented on an external monitor.

4.1.5 Experiment Procedure

The experimental procedure followed a structure similar to that described in Chapter 3. At the start of each session, participants were informed about the purpose of the study and provided informed consent via a Google Forms questionnaire, through which basic demographic information was also collected.

Following consent, the eye-tracking system and physiological sensors were calibrated individually for each participant using the standard 5-point calibration procedure available in the Gazepoint Analysis software. The experiment proceeded only once satisfactory calibration accuracy had been achieved.

Participants then completed the interactive product selection task, which began with a start screen allowing them to proceed when ready. Four choice tasks were subsequently presented, one per product category, in which participants viewed a grid of eight products and selected the item they considered most visually appealing or most likely to purchase. Following each selection, participants rated the attractiveness of their chosen product on a Likert scale. Once all four choice tasks were completed, the session concluded with a final end screen.

The complete session, including setup and calibration, lasted approximately four minutes per participant. Upon completion, participants were invited to raise any questions and were debriefed on the study's objectives and procedures.

4.1.6 AOI Definition

Following data collection, the eye-tracking recordings were exported from the Gazepoint Analysis software for further processing. As in the previous experiment, AOIs were defined for each individual product displayed within the experimental frames, enabling the extraction of product-level eye-tracking metrics for all items presented during the selection task. The same standardised AOI naming convention adopted in Chapter 3 was applied, ensuring each product retained a consistent identifier regardless of its position within the grid.

4.2 Dataset Construction

This section describes the construction of the dataset used to train and evaluate the ML models developed in Chapter 5. Following the completion of data collection, the raw eye-tracking recordings were processed and transformed into a structured dataset suitable for supervised learning. This involved organising the exported output files, defining and extracting the feature set, and assembling the final dataset at the participant–choice–product level. Each of these steps is described in the following subsections.

4.2.1 Data Structure

Following AOI definition, the eye-tracking recordings were exported from Gazepoint Analysis using the two standard output files generated for each participant: the *All_Gaze* file and the *Data_Summary* file. Each file was processed independently, as they provide different levels of granularity required for distinct stages of feature extraction.

The *Data_Summary* file provided pre-aggregated, product-level eye-tracking metrics per AOI, including fixation count, total time viewed, revisits count, and TTFF. These metrics constitute the primary behavioural features used in the modelling stage, as they directly quantify patterns of attention directed towards each product.

The *All_Gaze* file contained continuous, sample-level gaze data, including raw gaze coordinates and event timestamps, as well as physiological signals such as HR, GSR, and pupil-related measures recorded throughout the experiment. Due to its full temporal resolution, this file was used to compute additional features that could not be obtained from the *Data_Summary* file.

As the objective of the study was to analyse attention directed specifically towards products, a filtering step was applied exclusively to the *All_Gaze* file. Since this file records data at the sample level, the majority of rows correspond to gaze points falling outside any product region. Only rows associated with predefined product AOIs were retained, discarding all rows where gaze was directed elsewhere in the interface. The *Data_Summary* file required no such filtering, as it is natively structured around AOIs and already contains only product-level aggregates.

4.2.2 Feature Definition

The feature set integrates four groups of variables, behavioural eye-tracking metrics, oculomotor features, physiological signals, and contextual variables, each capturing a distinct dimension of the participant–product interaction. This multimodal design allows the models to capture not only where participants direct their attention, but also how they visually explore stimuli, how their physiological state responds during product evaluation, and how presentation conditions shape perception. Together, these features form the input to the supervised classification models developed in Chapter 5.

Behavioural Eye-Tracking Features

Behavioural features consist of AOI-level metrics that directly quantify the degree and pattern of visual attention directed towards each product, constituting the primary behavioural descriptors in the modelling stage. Table 4.1 lists the five features in this group along with their source and description.

Feature	Source	Description
`FixationsCount`	`Data_Summary`	Number of fixations recorded on the product AOI
`TimeViewedAOI`	`Data_Summary`	Total time spent viewing the product AOI, in seconds
`RevisitsCount`	`Data_Summary`	Number of times the participant returned to the AOI
`TTFF`	`Data_Summary`	Time elapsed until the first fixation on the AOI after the stimulus is presented
`TTFF_Missing`	`Data_Summary`	Binary indicator flagging observations where the AOI was never fixated
`FixationDurationMax`	`All_Gaze`	Duration of the longest individual fixation recorded on the AOI

Table 4.1: Behavioural eye-tracking features

Oculomotor Features

Oculomotor features capture saccadic eye movement patterns during product observation. These features were derived from the *All_Gaze* file based on saccade-related measures and are detailed in Table 4.2.

Feature	Source	Description
MeanSaccadeMag	All_Gaze	Average amplitude of saccadic movements during product observation
MaxSaccadeMag	All_Gaze	Largest saccade amplitude recorded during product observation
SaccadeCount	All_Gaze	Total number of saccades detected during product observation
SaccadeVariance	All_Gaze	Variability in saccade amplitude

Table 4.2: Oculomotor features extracted from All_Gaze

Physiological Features

Physiological features capture autonomic responses during product observation and were aggregated within each AOI observation window, corresponding to the time interval in which a participant viewed a given product. HR and GSR were recorded directly by the Gazepoint system, while pupil diameter was computed at each sample as the average of the left and right eye measurements. The full set of physiological features is presented in Table 4.3.

Feature	Source	Description
MeanPupil	All_Gaze	Average pupil diameter
MaxPupil	All_Gaze	Largest pupil diameter
MinPupil	All_Gaze	Smallest pupil diameter
PupilVariance	All_Gaze	Variability in pupil diameter
MeanGSR	All_Gaze	Average GSR
MaxGSR	All_Gaze	Highest GSR
MinGSR	All_Gaze	Lowest GSR
GSRVariance	All_Gaze	Variability in GSR
MeanHR	All_Gaze	Average HR
MaxHR	All_Gaze	Highest HR recorded
MinHR	All_Gaze	Lowest HR
HRVariance	All_Gaze	Variability in HR

Table 4.3: Physiological features extracted from All_Gaze

Contextual Features

Two contextual variables derived from the experimental design were included to capture structural influences on visual attention identified in Chapter 3. The *Position* feature captures the spatial location of each product within the display grid, while *Contextualised* indicates whether the product was presented in a contextualised or isolated condition. Both are described in Table 4.4.

Feature	Source	Description
Contextualised	Experimental design	Presentation style of the product (1 = contextualised, 0 = isolated)
Position	Experimental design	Spatial position of the product within the grid layout

Table 4.4: Contextual features

Target Variable

The target variable encodes the participant's product selection within each choice as a binary outcome: the chosen product is labelled 1 and all remaining products within the same choice are labelled 0. This formulation defines the task as a supervised binary classification problem at the product level, as summarised in Table 4.5.

Feature	Source	Description
`Choice`	Choice-Based Experiment	Binary indicator of whether the participant selected the product during the choice task

Table 4.5: Target variable for supervised classification

4.2.3 Final Dataset Overview

The final dataset is structured at the participant–choice–product level, where each row represents one participant's interaction with a specific product within a given choice task. With 30 participants, four choices per participant, and eight products to choose from, the dataset comprises 960 rows in total.

Each row of the final dataset combines 23 predictor variables alongside the binary target variable *Choice*, comprising 5 behavioural, 4 oculomotor, 12 physiological, and 2 contextual features, together with a binary missingness indicator for *TTFF*. Missing values arose in two distinct situations within the dataset. For most features, missing values resulted from AOIs with no recorded fixations, reflecting the absence of measurable engagement rather than data loss, and were therefore set to zero. The *TTFF* feature, however, required a different treatment, since a value of zero would be indistinguishable from an extremely fast first fixation and would therefore misrepresent the absence of engagement as immediate attention. To address this, a binary missingness indicator, *TTFF_ Missing*, was first created to explicitly flag observations in which the corresponding AOI was never fixated, as described in Table 4.1. The underlying *TTFF* values for these flagged observations were then imputed using the global median of the feature, rather than zero, to avoid introducing an artificial and misleading value that could compromise model training and downstream decisions. This approach preserves all participant–product rows, including products that received no visual attention during a given choice task.

The target variable is imbalanced by design: since participants selected exactly one product per choice task, only 120 rows carry a positive label (12.5%), with the remaining 840 labelled as not selected (87.5%). This class imbalance reflects the structure of the choice task and is accounted for in the modelling decisions described in Chapter 5. A summary of the dataset structure is provided in Table 4.6.

Component	**Value**
Participants	30
Choices per participant	4
Products per choice	8
Total rows	960
Positive labels (selected)	120
Negative labels (not selected)	840
Predictor variables	23
Missingness indicator	`TTFF_Missing` (binary)
Target variable	`Choice` (binary)

Table 4.6: Summary of the final dataset structure

The resulting dataset combines behavioural, physiological, and contextual information at the product level, providing a rich multimodal representation of the decision-making process. By linking these features to explicit product selections, it enables supervised learning models to learn patterns associated with user preference and product choice, serving as the foundation for the models developed and evaluated in the following chapter.

Chapter 5

AI Models

This chapter presents the ML approach developed to predict product selection. It begins by describing the modelling pipeline, covering the validation strategy, the treatment of class imbalance, feature scaling, and the hyperparameter tuning procedure adopted across all models. The chapter then introduces the five models used in this study, detailing the rationale for their selection and their specific implementation within the experimental pipeline. Finally, it presents the evaluation metrics used to assess model performance, covering both classification and ranking perspectives.

5.1 Modelling Pipeline

This section describes the methodological decisions underlying the development and evaluation of the ML models presented in this chapter. It covers four key aspects of the experimental pipeline: the Cross-Validation (CV) strategy used to evaluate model generalisation across unseen participants, the approach adopted to handle the class imbalance inherent to the choice task structure, the feature scaling decisions applied to ensure comparability across the multimodal feature set, and the hyperparameter tuning procedure used to optimise each model's configuration. Together, these decisions form the foundation of a rigorous and reproducible modelling pipeline.

5.1.1 Validation Strategy

CV strategies differ in how data are partitioned into training and testing sets, with the choice depending on the dataset structure and the study's generalisation objective [63]. Common approaches include standard k-fold and holdout validation, each involving different trade-offs between bias, variance, and computational cost [64]. However, these methods generally assume that observations are independently and identically distributed [65]. This assumption is violated in the present study, where multiple rows correspond to the same participant.

To account for this dependency structure, Leave-One-Group-Out (LOGO) CV was employed, with each participant treated as a separate group [65]. In each of the 30 folds, data from 29 participants were used for training, while all rows from the remaining participant were held out for testing. This procedure was repeated until each participant had served as the held-out test participant exactly once, and performance metrics were averaged across all folds.

LOGO was chosen for two primary reasons. First, it aligns directly with the study's objective of evaluating model performance on unseen participants rather than on new observations from participants already represented in the training data. Second, it prevents participant-level data leakage, a critical concern given that observations from the same participant share underlying gaze and physiological characteristics that could otherwise lead to overly optimistic performance estimates [63].

A limitation of this approach is that each test fold contains only 32 rows (four choice tasks $\times$ eight products), which may result in substantial variability in fold-level performance estimates. Averaging the results across all 30 folds mitigates this variability, providing a more stable estimate of overall model performance.

5.1.2 Class Imbalance

Class imbalance occurs when one outcome category is substantially less frequent than the other, which can cause models to favour the majority class and report misleadingly high accuracy while failing to identify the minority class. Common approaches to address this include resampling techniques such as SMOTE, which generate synthetic minority-class samples, and cost-sensitive learning, which adjusts the relative penalty for misclassifying each class during training [66].

In the present study, the target variable is imbalanced by design, as participants selected a single product from a set of eight, meaning that only 12.5% of observations correspond to the chosen products. Resampling was not adopted, as each row represents a real combination of behavioural, oculomotor, and physiological signals recorded from a participant. Synthetic resampling or interpolation between these observations was therefore avoided, since it could generate physiologically implausible combinations and potentially distort the relationships that the models are intended to learn.

Instead, class imbalance was addressed at the model level using cost-sensitive learning, which adjusts the relative penalty assigned to each class during training in proportion to its frequency. The specific parameters applied varied by model and are detailed further in the document.

5.1.3 Feature Scaling

Feature scaling transforms numeric features onto a comparable range or distribution, preventing variables with larger numeric ranges from disproportionately influencing model training and often improving model performance [67]. Common approaches include *StandardScaler*, which transforms features to zero mean and unit variance, and *MinMaxScaler*, which rescales features to a fixed range such as $[0, 1]$ [68].

In the present study, features span very different scales, for example, *TimeViewedAOI* is measured in seconds while *MeanHR* is measured in beats per minute. To address this, *StandardScaler* was applied to the models that require feature scaling. Importantly, scaling was implemented within a pipeline alongside each model, ensuring that the scaler was fitted exclusively on the training fold and applied to the test fold within each CV iteration. This prevents information from the test set, such as its mean and variance, from influencing the scaling parameters, avoiding data leakage [67].

5.1.4 Hyperparameter Tuning

Hyperparameter tuning involves searching over different configurations of a model's settings to identify the combination that provides the best performance on a given objective [69]. Common approaches include grid search, which exhaustively evaluates every combination in a predefined space, and randomised search, which samples a fixed number of combinations, offering a more computationally efficient alternative when the search space is large [70].

In the present study, randomised search was applied to all models, sampling 10 hyperparameter combinations per model. For each fold in the LOGO CV, an inner 5-fold GroupKFold CV was performed on the training participants to evaluate each combination, using Normalised Discounted Cumulative Gain (NDCG)@1 as the tuning objective, selected to align with the study's emphasis on correctly identifying the chosen product at the top of the ranking. The hyperparameter configuration achieving the highest mean inner NDCG@1 was then used to train the final model for that fold.

This nested structure ensures that hyperparameter selection introduces no information leakage from the held-out test participant, preserving the validity of the LOGO evaluation.

5.2 Models

This section presents the ML models selected for this study. The theoretical foundations of the classification models used in this work are presented in Chapter 2. Specifically, these comprise RF, XGBoost, SVM, and MLP. This section focuses on why these models were selected and how they were implemented in the experimental pipeline. In addition to these four classification models, an XGBoost Ranker was included to provide a dedicated ranking formulation of the prediction task.

5.2.1 Random Forest

RF is a tree-based ensemble classification model. It was selected for this study due to its robustness to noisy and correlated features, its ability to handle mixed feature types without requiring scaling, and its native support for feature importance estimation [32], properties particularly suited to the multimodal nature of the dataset, which combines behavioural, oculomotor, physiological, and contextual variables. As a tree-based model, it is invariant to feature scale, so no scaling was applied [67]. Class imbalance was addressed through the *class_weight="balanced"* parameter, which adjusts the weight assigned to each class in inverse proportion to its frequency. An overview of its implementation in the present study is provided in Table 5.1.

Configuration	Value
Task	Binary classification
Class imbalance	`class_weight="balanced"`
Feature scaling	Not applied
Hyperparameter tuning	Randomised search (10 combinations)
Tuning objective	NDCG@1 (inner 5-fold GroupKFold)
Parameters tuned	`n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, `max_features`

Table 5.1: Random Forest implementation overview

5.2.2 XGBoost Classifier

XGBoost is a gradient boosting ensemble classification model. It was selected for its strong empirical performance across a wide range of classification tasks, its regularisation mechanisms that reduce overfitting, and its sequential boosting strategy [32], which makes it well suited to capturing complex non-linear relationships between eye-tracking and physiological features and product selection outcomes. As a tree-based model, it is invariant to feature scale, so no scaling was applied [67]. Class imbalance was handled through the *scale_pos_weight* parameter, computed within each training fold as the ratio of negative to positive samples. An overview of its implementation in the present study is provided in Table 5.2.

Configuration	Value
Task	Binary classification
Class imbalance	`scale_pos_weight`
Feature scaling	Not applied
Hyperparameter tuning	Randomised search (10 combinations)
Tuning objective	NDCG@1 (inner 5-fold GroupKFold)
Parameters tuned	`n_estimators`, `max_depth`, `learning_rate`, `subsample`, `colsample_bytree`

Table 5.2: XGBoost Classifier implementation overview

5.2.3 Support Vector Machine

SVM is a kernel-based classification model. It was selected for its effectiveness in high-dimensional feature spaces and its ability to model non-linear decision boundaries using the RBF kernel, allowing it to capture complex patterns in the feature space [32]. Unlike tree-based models, SVM relies on distance calculations to define the decision boundary and is therefore sensitive to feature scale [67]. To address this, *StandardScaler* was applied within the pipeline, ensuring that scaling parameters were fitted exclusively on the training data. Class imbalance was addressed through the *class_weight="balanced"* parameter. An overview of its implementation in the present study is provided in Table 5.3.

Configuration	Value
Task	Binary classification
Class imbalance	`class_weight="balanced"`
Feature scaling	`StandardScaler` (within pipeline)
Hyperparameter tuning	Randomised search (10 combinations)
Tuning objective	NDCG@1 (inner 5-fold GroupKFold)
Parameters tuned	`C`, `gamma`, `kernel`

Table 5.3: Support Vector Machine implementation overview

5.2.4 Multilayer Perceptron

MLP is a feed-forward neural network classification model. It was selected for its capacity to learn complex non-linear mappings between the multimodal input features and the binary choice outcome. As a gradient-based model, MLP is sensitive to feature scale, since unscaled features can cause unstable or slow convergence during training. *StandardScaler* was therefore applied within the pipeline [67]. Unlike the other models, MLP does not provide an equivalent class weighting parameter and was therefore trained without explicit class imbalance correction. An overview of its implementation in the present study is provided in Table 5.4.

Configuration	Value
Task	Binary classification
Class imbalance	Not explicitly addressed (no equivalent parameter)
Feature scaling	`StandardScaler` (within pipeline)
Hyperparameter tuning	Randomised search (10 combinations)
Tuning objective	NDCG@1 (inner 5-fold GroupKFold)
Parameters tuned	`hidden_layer_sizes`, `activation`, `alpha`, `learning_rate`

Table 5.4: Multilayer Perceptron implementation overview

5.2.5 XGBoost Ranker

Unlike the four classification models described above, the XGBoost Ranker adopts a learning-to-rank formulation of the prediction task. Rather than predicting a binary label for each product independently, learning-to-rank models are trained to optimise the relative ordering of items within a defined group, making them particularly suited to tasks where the goal is to identify the most preferred item within a set of alternatives [71].

The XGBoost Ranker extends the gradient boosting framework of XGBoost to this ranking setting. Rather than minimising a classification loss, it optimises a ranking-specific objective that penalises incorrect orderings within each group. In the present study, the *rank:pairwise* objective was used, which trains the model by comparing pairs of products within the same choice task and learning to assign higher scores to the chosen product [71].

This formulation is more naturally aligned with the structure of the task, in which the goal is not to classify each product in isolation but to identify the most preferred item within a set of eight alternatives. Groups were defined at the participant–frame level, ensuring that ranking was performed within each choice task independently. An overview of its implementation is provided in Table 5.5.

Configuration	**Value**
Task	Learning to rank
Ranking objective	`rank:pairwise`
Group definition	Participant × Frame
Class imbalance	Not applicable (ranking formulation)
Feature scaling	Not applied
Hyperparameter tuning	Randomised search (10 combinations)
Tuning objective	NDCG@1 (inner leave-one-participant-out)
Parameters tuned	`n_estimators`, `max_depth`, `learning_rate`, `subsample`, `colsample_bytree`

Table 5.5: XGBoost Ranker implementation overview

5.3 Evaluation Metrics

The evaluation of ML models requires performance metrics that are appropriate to the underlying learning task. Different problem formulations, including classification, regression, and ranking, require distinct evaluation criteria to provide meaningful and reliable assessments of model performance [72]. In the present study, the prediction task can be framed in two complementary ways. On one hand, each product within a choice task can be treated as an independent binary classification instance, where the model predicts whether that specific product was selected or not. On the other hand, the underlying experimental design is inherently a forced-choice ranking problem: participants were presented with eight products and instructed to select exactly one, meaning that the products within a given choice task are not independent alternatives but a competing set from which a single preferred item must be identified.

For this reason, the performance of the classification models, RF, XGBoost, SVM, and MLP, was assessed using both classification and ranking metrics, while the XGBoost Ranker, given its learning-to-rank formulation, was evaluated exclusively using ranking metrics. Classification metrics provide a standard assessment of how well the model distinguishes between selected and non-selected products at the level of an individual prediction, whereas ranking metrics evaluate how well the model orders the eight products within each choice task relative to one another, which more closely reflects the decision-making process being modelled. The two types of metrics are described in turn in the following subsections.

5.3.1 Classification Metrics

Classification metrics quantify the agreement between the labels predicted by a model and the true labels of the target variable, typically derived from the four possible outcomes of a binary classification problem: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) [73]. In the present study, a positive instance corresponds to a product that was selected by the participant, while a negative instance corresponds to a product that was not selected. Five classification metrics were used to evaluate model performance: Accuracy, Precision, Recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC):

- **Accuracy** measures the overall proportion of correct predictions made by the model, irrespective of class [73]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

  Although widely used, Accuracy can be misleading in imbalanced classification settings, as a model that predicts the majority class for every instance can still achieve a high Accuracy score while failing entirely to identify the minority class [66]. Given that only 12.5% of the observations in the present dataset correspond to selected products, Accuracy was retained as a reference metric but was not the primary criterion for assessing model performance.

- **Precision** measures the proportion of instances predicted as positive that were actually positive, reflecting how reliable a positive prediction made by the model is [74]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2}$$

- **Recall**, also referred to as sensitivity, measures the proportion of actual positive instances that were correctly identified by the model, reflecting the model's ability to detect selected products without missing them [74]:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.3}$$

- **F1-score** addresses this trade-off by computing the harmonic mean of Precision and Recall, providing a single balanced measure of classification performance that is particularly informative in imbalanced settings [73]:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.4}$$

- **ROC-AUC** evaluates the model's ability to distinguish between the positive and negative classes across all possible classification thresholds [73]. The ROC curve plots the TP rate against the FP rate as the decision threshold varies, and the ROC-AUC summarises this curve as a single value. A higher ROC-AUC indicates better predictive performance, meaning the model is more effective at correctly distinguishing positive from negative instances and assigning higher scores to positive cases [74].

5.3.2 Ranking Metrics

While classification metrics evaluate predictions independently for each product, they do not capture whether the model is able to correctly identify the selected product relative to the other alternatives presented within the same choice task. Since the experimental design always involves a single selection among eight competing products, ranking metrics were introduced to directly assess this relative ordering, by evaluating, for each choice task, how the model ranks all eight products against one another according to their predicted relevance scores.

For the classification-based models, the ranking of products within each choice task was obtained by using the predicted probability of selection as a relevance score. Specifically, the output probability from each classifier was used to rank the eight products in descending order, with the highest probability corresponding to the model's predicted top choice [75]. In contrast, the XGBoost Ranker was trained directly within a learning-to-rank framework, which is specifically designed to optimise the ordering of items within a group. As a result, it natively produces ranking scores for each product within a choice task, without requiring any post-processing step to convert predictions into a ranked list.

Ranking performance was evaluated using rank-aware metrics, specifically NDCG@k, Precision@k, Recall@k, and Mean Reciprocal Rank (MRR), which capture both the position of the relevant item and the quality of the overall ranking:

- **NDCG@k** measures the quality of the ranked list by assigning higher weight to relevant items appearing at higher positions, while applying a logarithmic discount to lower ranks. Scores are normalised against an ideal ranking, producing values in the range [0, 1], where higher values indicate better ranking performance. The NDCG@k is defined as [76, 77]:

$$\text{DCG@}k = \sum_{i=1}^{k} \frac{rel_i}{\log_2(i+1)} \tag{5.5}$$

$$\text{NDCG@}k = \frac{\text{DCG@}k}{\text{IDCG@}k} \tag{5.6}$$

- **Precision@k** measures the proportion of relevant items within the top-k results. Since each choice task contains exactly one relevant item (the selected product), Precision@1 corresponds to the proportion of tasks where the correct product is ranked first. The Precision@k is defined as [76, 77]:

$$\text{Precision@}k = \frac{|\{\text{Relevant items}\} \cap \{\text{Top-}k\text{ items}\}|}{k} \tag{5.7}$$

- **Recall@k** measures whether the relevant item appears within the top-k positions. Given the single-relevant-item structure of each task, Recall@k simplifies to the proportion of tasks in which the selected product is ranked within the top-k positions. The Recall@k is defined as [76, 77]:

$$\text{Recall@}k = \frac{1}{N} \sum_{j=1}^{N} \frac{\left|\{\text{Relevant items}\}_j \cap \{\text{Top-}k\text{ items}\}_j\right|}{\left|\{\text{Relevant items}\}_j\right|} \tag{5.8}$$

- **MRR** evaluates ranking quality by considering the exact position of the selected product within the ordered list. It assigns higher scores when the correct item appears at higher ranks and progressively lower scores as its position decreases. The MRR is defined as [76]:

$$\text{MRR} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{\text{rank}_j} \tag{5.9}$$

At k = 1, Precision@1, Recall@1, and NDCG@1 are equivalent, as all three reduce to measuring whether the selected product is ranked first.

Overall, these metrics provide a comprehensive evaluation of ranking performance, capturing both whether the model identifies the correct product and how highly it ranks it relative to the remaining alternatives.

Chapter 6

Results and Discussion

This chapter presents the results obtained from the ML models described in Chapter 5, together with a discussion of their implications. It begins with the feature importance analysis, which supports the incremental feature selection procedure used to identify the final feature subset. It then presents the classification and ranking performance of each model, evaluated on this selected feature subset. The chapter concludes with a discussion of the best-performing model and the most predictive feature groups.

6.1 Results

This section presents the results obtained from the ML models described in Chapter 5. It begins with the feature importance analysis, which informed the incremental feature selection procedure used to identify the final feature subset. It then presents the classification and ranking performance of each model, evaluated on this selected feature subset.

6.1.1 Feature Importance and Selection

This subsection presents the feature importance analysis used to identify the most predictive variables in the dataset, the ranking strategy derived from this analysis, and the incremental feature selection procedure used to determine the final feature subset used for model evaluation.

Feature Importance

The complete set of 23 candidate features, with *TTFF* and *TTFF_Missing* jointly counted as a single logical feature, was used to initially train all five models (RF, XGBoost, SVM, MLP, and XGBoost Ranker). Permutation importance was then computed to assess the contribution of each individual feature to model performance. Permutation importance measures the change in a model's performance when the values of a single feature are randomly shuffled, thereby breaking its relationship with the target variable while leaving all other features intact [78]. In the present study, this procedure was applied on the held-out test fold of each Leave-One-Group-Out iteration, using NDCG@1 as the scoring metric, and the resulting importance values were averaged across the 30 folds for each model.

A positive permutation importance indicates that shuffling the feature decreases the model's performance, suggesting that the feature contributes useful predictive information. On the other hand, a negative importance indicates that shuffling the feature does not degrade, and may even slightly improve, model performance, suggesting that the feature provides little useful information or may introduce noise for that particular model [78]. Table 6.1 presents the permutation importance values obtained for each feature across the five models.

Feature	RF	XGB	SVM	MLP	XGBRanker
TimeViewedAOI	0.148	0.354	0.261	0.182	0.337
FixationDurationMax	0.073	0.102	0.092	0.013	0.064
MaxSaccadeMag	0.055	0.089	0.047	0.026	0.073
MeanSaccadeMag	0.026	0.058	0.097	0.089	0.040
TTFF	0.022	0.024	-0.004	-0.023	0.063
TTFF_Missing	0.000	0.000	0.033	0.002	0.000
FixationsCount	0.026	0.037	0.037	0.044	0.055
SaccadeCount	0.079	0.039	0.083	-0.003	0.041
MaxPupil	0.007	0.007	0.104	0.116	0.020
RevisitsCount	0.029	0.031	0.017	0.004	0.060
PupilVariance	0.007	0.015	0.024	0.023	0.020
Contextualised	-0.004	0.011	0.076	0.021	0.016
MeanPupil	0.009	-0.003	0.040	0.004	0.026
MeanHR	0.008	0.010	0.044	0.003	0.015
MaxGSR	0.009	-0.004	0.071	0.006	0.002
Position	0.002	0.010	0.040	0.000	0.027
MeanGSR	0.012	0.004	0.068	-0.001	0.001
MaxHR	0.006	0.005	0.026	0.013	0.003
MinHR	0.004	0.014	0.005	0.009	0.010
MinPupil	0.003	0.001	0.031	0.009	0.012
MinGSR	0.008	0.000	0.027	0.002	0.001
HRVariance	0.006	0.012	-0.001	-0.022	0.015
SaccadeVariance	0.005	0.000	0.017	0.001	0.009
GSRVariance	0.004	-0.006	0.010	-0.006	0.023

Table 6.1: Permutation importance values per feature across all models

As shown in Table 6.1, *TimeViewedAOI* emerged as by far the most predictive feature across all five models, with an average importance more than three times higher than the second-ranked feature, *FixationDurationMax*. This pattern held consistently across algorithm types, tree-based (RF, XGBoost, XGBoost Ranker), kernel-based (SVM), and neural network (MLP), suggesting that the total time spent viewing a product is the single strongest behavioural signal of selection in this dataset.

Beyond *TimeViewedAOI*, *FixationDurationMax*, *MaxSaccadeMag*, and *MeanSaccadeMag* formed a second tier of consistently informative predictors, indicating that both sustained visual attention and eye movement characteristics contribute to modelling user preference. In contrast, most pupil and GSR related variables exhibited importance values close to zero, suggesting that these physiological measures provided comparatively limited predictive information.

Some features exhibited negative permutation importance for specific models, indicating that shuffling them slightly improved rather than degraded predictive performance. Such behaviour suggests that these features contribute little useful information or introduce noise for the corresponding model. This pattern was most evident for *TTFF*, which held a negative importance under both the SVM and MLP. As described in Chapter 4, missing *TTFF* values were imputed using the global median rather than zero, accompanied by the *TTFF_Missing* indicator. This strategy appears to benefit tree-based models more than SVM and MLP. *HRVariance* followed a similar pattern, also negative under the same two models, while *GSRVariance* was negative under XGBoost and MLP.

Feature Ranking Strategy

Since some features negatively affected model performance, a ranking of features was constructed to guide a subsequent feature selection procedure, with the goal of identifying the combination of features that maximises model performance, rather than relying on the full feature set by default. To obtain a single aggregated ranking, the per-model rankings were combined by averaging the rank position of each feature across the different models. As *TTFF_Missing* represents the missingness pattern of *TTFF*, both variables were treated jointly when constructing the aggregated ranking and in all subsequent analyses. Their combined ranking was computed using only the tree-based models, excluding SVM and MLP given the detrimental importance values observed for these models. The resulting aggregated ranking is presented in Table 6.2.

Rank	Feature
1	TimeViewedAOI
2	FixationDurationMax
3	MaxSaccadeMag
4	MeanSaccadeMag
5	TTFF / TTFF_Missing
6	FixationsCount
7	SaccadeCount
8	MaxPupil
9	RevisitsCount
10	PupilVariance
11	Contextualised
12	MeanPupil
13	MeanHR
14	MaxGSR
15	Position
16	MeanGSR
17	MaxHR
18	MinHR
19	MinPupil
20	MinGSR
21	HRVariance
22	SaccadeVariance
23	GSRVariance

Table 6.2: Aggregated feature importance ranking across models

Incremental Feature Selection

To determine the optimal feature subset, models were retrained iteratively, starting with only the top-ranked feature and progressively adding one feature at a time according to the ranking presented in Table 6.2, until all 23 features were included. At each step, all five models were retrained and evaluated using the same Leave-One-Group-Out cross-validation procedure described in Chapter 5, with hyperparameters re-tuned at each step.

Figure 6.1 presents the resulting NDCG@1 scores for each of the five models as a function of the number of features included.

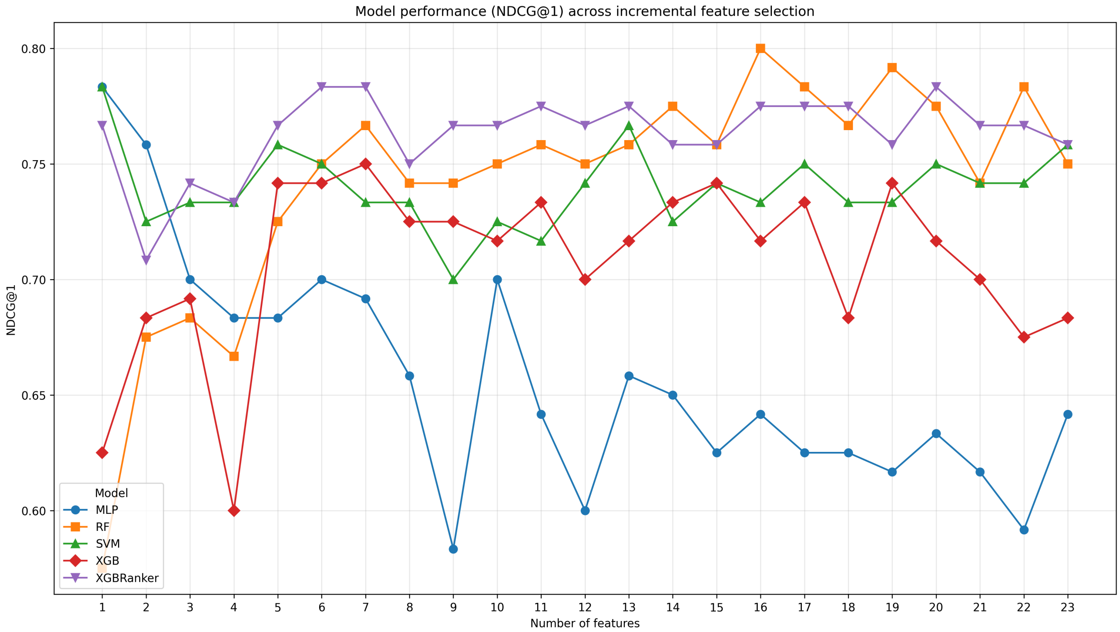

Figure 6.1: NDCG@1 performance across RF, XGBoost, SVM, MLP, and XGBoost Ranker during incremental feature selection

As shown in Figure 6.1, the MLP exhibited a markedly different behaviour from the remaining models. While RF, XGBoost, SVM, and XGBoost Ranker maintained relatively stable or improving performance as features were added, the MLP showed a consistent decline in NDCG@1 with increasing feature set size, suggesting that it was unable to benefit from the additional variables and was instead negatively affected by the growing input dimensionality. Given this divergent behaviour, the MLP was excluded from the feature subset selection procedure to avoid distorting the aggregated signal used to identify the optimal number of features.

To identify the optimal feature subset size, the mean NDCG@1 across the remaining four models was computed at each step. Figure 6.2 presents this mean performance curve, from which the number of features corresponding to the highest average score was selected as the final feature subset.

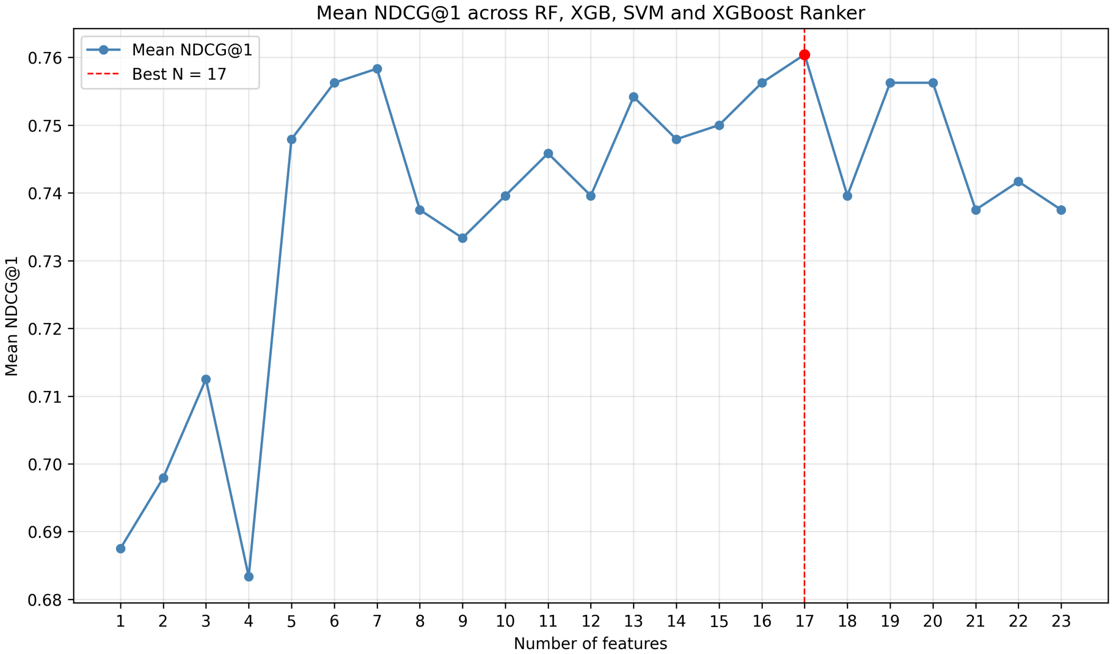

Figure 6.2: Mean NDCG@1 across RF, XGBoost, SVM, and XGBoost Ranker during incremental feature selection

As shown in Figure 6.2, the mean NDCG@1 across the four models increased substantially up to approximately six features, after which performance stabilised within a narrow range. The global peak was observed at 17 features, corresponding to the top-17 features listed in Table 6.2. This configuration was therefore adopted as the final feature subset for model evaluation.

6.1.2 Classification Performance

Table 6.3 presents the classification performance of the four classification models, averaged across the 30 Leave-One-Group-Out folds, evaluated on the selected subset of 17 features.

Model	Accuracy	Precision	Recall	F1	ROC-AUC
RF	0.886	0.577	0.483	0.474	0.920
XGBoost	0.860	0.535	0.650	0.534	0.901
SVM	0.815	0.445	0.800	0.524	0.909
MLP	0.881	0.560	0.475	0.485	0.875

Table 6.3: Classification performance across models (selected feature subset, mean across 30 LOGO folds)

Observing Table 6.3, all models achieved high accuracy values, ranging from 0.815 to 0.886. However, as discussed in Chapter 5, high accuracy alone is not a reliable indicator of performance under the class imbalance inherent to this task, where only 12.5% of observations correspond to selected products and a model predicting the majority class in every instance would already achieve approximately 87.5% accuracy. The ROC-AUC values, which are less sensitive to class imbalance, provide a more informative picture, with RF achieving the highest value (0.920), followed by SVM (0.909) and XGBoost (0.901), and MLP obtaining the lowest (0.875). In terms of the trade-off between Precision and Recall, notable differences were observed across models. The SVM achieved the highest Recall (0.800), correctly identifying a larger proportion of selected products at the cost of lower Precision (0.445), reflecting a tendency to favour sensitivity over specificity. RF and MLP showed the opposite pattern, with higher Precision but lower Recall, while XGBoost achieved a more balanced trade-off. F1-scores, which summarise this trade-off, were moderate across all models, ranging from 0.474 to 0.534, consistent with the difficulty of the task under strong class imbalance.

6.1.3 Ranking Performance

Table 6.4 presents the ranking performance of all five models, including the XGBoost Ranker, evaluated on the selected subset of 17 features.

Model	NDCG@1	NDCG@3	NDCG@5	MRR	P@1	R@3	R@5
RF	0.783	0.872	0.890	0.861	0.783	0.942	0.983
XGBoost	0.733	0.843	0.860	0.830	0.733	0.925	0.967
SVM	0.750	0.861	0.875	0.845	0.750	0.942	0.975
MLP	0.625	0.778	0.806	0.760	0.625	0.892	0.958
XGBoost Ranker	0.775	0.866	0.888	0.858	0.775	0.933	0.983

P@1 = Precision@1, R@3 = Recall@3, R@5 = Recall@5

Table 6.4: Ranking performance across models (selected feature subset, mean across 30 LOGO folds)

The ranking results presented in Table 6.4 reveal that all models substantially outperformed the random baseline, where a random ranking of eight products would yield an expected NDCG@1 and Precision@1 of 0.125. RF achieved the highest NDCG@1 (0.783) and Precision@1 (0.783), correctly identifying the selected product as the top-ranked item in approximately 78% of choice tasks. The XGBoost Ranker, despite being specifically designed for ranking tasks, obtained a slightly lower NDCG@1 of 0.775, followed by SVM (0.750) and XGBoost (0.733). MLP consistently underperformed across all ranking metrics, with a NDCG@1 of 0.625. Regarding broader ranking quality, all models achieved high Recall@3 and Recall@5 values, indicating that the selected product was ranked within the top three in over 89% of tasks and within the top five in over 95% of tasks across all models. The MRR values followed a similar pattern, with RF achieving the highest score (0.861), suggesting that on average the correct product appeared very close to the top of the ranked list.

6.2 Discussion

This section discusses the main findings derived from the results presented above. It first examines the performance of the models, identifying the best-performing approach and discussing the factors that may explain the observed differences. It then analyses the contribution of the different feature groups to model performance, linking the importance of individual features to the experimental findings reported in Chapter 3.

6.2.1 Best-Performing Model

Among the five models evaluated, RF achieved the strongest overall performance, attaining the highest NDCG@1 (0.783) and ROC-AUC (0.920) across all models. A notable observation is that the ranking of models by ROC-AUC is consistent with their ranking by NDCG@1, with RF leading in both metrics, followed by SVM and XGBoost, and MLP consistently underperforming. This alignment suggests that the models' ability to discriminate between selected and non-selected products at the binary classification level is reflected in their capacity to correctly order products within each choice task, reinforcing the validity of the evaluation across both perspectives.

A noteworthy finding is that the XGBoost Ranker, despite being specifically designed to optimise ranking objectives, did not outperform the RF classifier. While the difference in NDCG@1 was modest (0.775 versus 0.783), the result suggests that, for this dataset, a well-calibrated classification model can produce ranking scores that are at least as effective as those obtained from a dedicated learning-to-rank approach. One possible explanation is the relatively small group size of eight products per choice task, which limits the number of pairwise comparisons available during training and may reduce the advantage of pairwise ranking objectives.

The MLP consistently underperformed across all metrics, both in classification and ranking, and exhibited a systematic decline in performance as features were added during the incremental feature selection procedure. This behaviour suggests that the MLP was unable to effectively leverage the multimodal feature set, possibly due to the relatively small dataset size, which may have been insufficient for the neural network to learn stable representations from the combined behavioural, oculomotor, physiological, and contextual inputs.

6.2.2 Predictive Feature Groups

The feature importance analysis revealed a clear hierarchy among the feature groups, with behavioural eye-tracking features emerging as the most predictive of product selection. *TimeViewedAOI* was by far the dominant feature across all models, with permutation importance values substantially higher than any other variable. This finding is consistent with the results reported in Chapter 3, where Time Viewed was one of the most discriminative metrics across spatial positions and presentation styles, suggesting that the total time a participant spends looking at a product is a strong and reliable behavioural signal of visual preference, reflecting a positive emotional engagement with the product.

Beyond *TimeViewedAOI*, oculomotor features, particularly *FixationDurationMax*, *MaxSaccadeMag*, and *MeanSaccadeMag*, formed a secondary tier of consistently informative features across models. These results indicate that not only the duration of visual attention but also the nature of eye movements during product observation, specifically the amplitude and variability of saccadic movements, carry meaningful information about product preference.

In contrast, physiological features, including HR, GSR, and pupil-related measures, contributed comparatively little to model performance, with importance values close to zero across most models. Although physiological signals are theoretically linked to emotional arousal and cognitive engagement, their marginal predictive contribution in this study may reflect the difficulty of capturing reliable autonomic responses within the short observation windows associated with individual product AOIs, as well as the inherent noisiness of such signals in a controlled laboratory setting.

Contextual features, *Position* and *Contextualised*, showed moderate importance, particularly for the SVM. This is consistent with the findings of Chapter 3, which demonstrated statistically significant effects of both spatial position and presentation style on visual attention metrics. The inclusion of these variables therefore added meaningful structural information to the models, capturing design-driven influences on visual engagement that complement the behavioural and oculomotor signals.

Chapter 7

Conclusion

This chapter concludes the dissertation by reflecting on the work developed throughout the project, identifying the main limitations encountered, and outlining directions for future research.

7.1 General Considerations

The rapid growth of e-commerce has placed user experience at the centre of digital competitiveness, with UI/UX design playing a decisive role in user engagement, trust, and continued use of digital platforms. However, most existing approaches remain static, unable to adapt to individual users' emotional states or visual attention patterns in real time.

This dissertation addressed this gap by developing an intelligent system capable of inferring product preferences from users' visual attention and physiological responses, using eye-tracking and AI, with the goal of supporting adaptive, data-driven UI/UX design decisions in e-commerce.

To achieve this objective, the dissertation combined controlled experimentation with multimodal machine learning. Two complementary user studies were designed. The first investigated how interface layout influences visual attention, while the second linked gaze behaviour and physiological responses to explicit product selection. Together, these studies produced a multimodal dataset that enabled the training and evaluation of five predictive models using participant-level cross-validation. RF emerged as the strongest model, correctly identifying the selected product as the top-ranked item in approximately 78% of choice tasks (NDCG@1 = 0.783, ROC-AUC = 0.920), with behavioural features, particularly *TimeViewedAOI*, proving far more predictive than physiological signals.

These results support the dissertation's central hypothesis, that visual attention captured through eye-tracking, when combined with AI-based modelling, can be used to infer product preference with high predictive performance. By demonstrating that gaze behaviour alone carries a reliable signal of user preference, this work provides a validated foundation for adaptive UI/UX systems that respond to user attention and preference in real time, contributing to the development of intelligent e-commerce interfaces capable of adapting dynamically to users' attention and preferences, ultimately supporting more personalised, responsive, and engaging online shopping experiences.

7.2 Limitations

Despite the results obtained, several limitations should be acknowledged when interpreting them.

- **Sample size:** The final dataset comprised 30 participants and 960 participant–choice–product observations, of which only 120 corresponded to selected products. This relatively small scale likely limited the performance of data-intensive models such as the MLP, which consistently underperformed and declined as additional features were introduced. It also meant that the LOGO CV strategy produced test folds containing only 32 observations each (four choice tasks $\times$ eight products), introducing greater variability into fold-level performance estimates.
- **Ecological validity of the stimulus presentation:** A video-based presentation format was adopted instead of a fully interactive prototype to standardise stimulus timing and support reliable AOI-based analysis. This reduced ecological validity, as participants could not freely browse, click, or interact with the interface as they would during real online shopping behaviour.
- **Single platform and product domain:** The experimental interface was based exclusively on the Sanimaia e-commerce website, limited to household product categories (bathroom, kitchen, living, and tableware). Findings and trained models may not generalise to other platforms, visual design languages, or product domains such as fashion, electronics, or grocery retail.

7.3 Future Work

Building on the findings and limitations discussed above, several directions for future research are proposed.

- **Expanding the dataset:** Increase the number of participants and choice tasks to improve statistical power and provide data-intensive models, particularly the MLP, with sufficient training data to learn more robust representations. A larger, more diverse sample would also support subgroup analyses, such as differences in attention-preference relationships across demographic groups.
- **Increasing ecological validity:** Use fully interactive prototypes or live e-commerce websites, combined with mobile or webcam-based eye-tracking, to capture gaze behaviour under more naturalistic browsing conditions, alongside strategies to manage variable exposure duration and dynamic AOI definition.
- **Generalising across platforms and product domains:** Replicate the experimental design across different e-commerce platforms, visual styles, and product categories, such as fashion, electronics, or groceries, to test whether the predictive relationships identified here, particularly the dominance of *TimeViewedAOI*, generalise across consumer contexts or are specific to the household product domain examined.
- **Towards real-time adaptive systems:** Integrate the predictive models developed in this dissertation into a real-time adaptive UI/UX framework that dynamically adjusts interface elements, such as product layout, positioning, or presentation style, based on inferred user preference, and evaluate the impact on user engagement and satisfaction in live e-commerce environments.

Bibliography

[1] *https://cape.ftpporto.com/*. pt-PT. url: https://cape.ftpporto.com/ (visited on 06/27/2026).

[2] *EARS*. pt-PT. url: https://ears.ftpporto.com/ (visited on 06/27/2026).

[3] *gecad – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development*. en-US. url: https://www.gecad.isep.ipp.pt/ (visited on 12/27/2025).

[4] Tsabbita Isnina Hasan et al. "UI/UX Design Impact on E-Commerce Attracting Users". In: *Procedia Computer Science*. 9th International Conference on Computer Science and Computational Intelligence 2024 (ICCSCI 2024) 245 (Jan. 2024), pp. 1075–1082. issn: 1877-0509. doi: 10.1016/j.procs.2024.10.336. url: https://www.sciencedirect.com/science/article/pii/S1877050924031442 (visited on 12/23/2025).

[5] Daglous Masveta and Masimba Elvis Manyangara. "The UX-UI continuum: exploring the interplay between user experience and user interface in e-learning platforms". In: *Cogent Education* 12.1 (Dec. 2025), p. 2536531. issn: null. doi: 10.1080/2331186X.2025.2536531. url: https://doi.org/10.1080/2331186X.2025.2536531 (visited on 12/23/2025).

[6] Munim Ahsan Chowdhury Primon. *Emotionally Intelligent User Interfaces of Social Media: The Role of Emotional Intelligence in User Interface Design and UX Pattern of Social Media on Human Behaviour*. Tech. rep. May 2024. doi: 10.13140/RG.2.2.17632.08964.

[7] *(PDF) Designing Personalized User Interfaces using Artificial Intelligence-Driven Behavioral Analysis for Enhanced User Experience*. en. url: https://www.researchgate.net/publication/387971161_Designing_Personalized_User_Interfaces_using_Artificial_Intelligence-Driven_Behavioral_Analysis_for_Enhanced_User_Experience (visited on 12/24/2025).

[8] *Project Charter - an overview | ScienceDirect Topics*. url: https://www.sciencedirect.com/topics/computer-science/project-charter (visited on 12/25/2025).

[9] Mukul Burghate. "Work Breakdown Structure: Simplifying Project Management". In: (June 2018).

[10] Francesco Bianconi. "Gantt Charts". en. In: *Data and Process Visualisation for Graphic Communication: A Hands-on Approach with Python*. Ed. by Francesco Bianconi. Cham: Springer Nature Switzerland, 2024, pp. 175–181. isbn: 978-3-031-57051-3. doi: 10.1007/978-3-031-57051-3_13. url: https://doi.org/10.1007/978-3-031-57051-3_13 (visited on 12/26/2025).

[11] "(PDF) Gantt Chart: An Important Tool of Management". en. In: *ResearchGate* (Aug. 2025). url: https://www.researchgate.net/publication/358234055_Gantt_Chart_An_Important_Tool_of_Management (visited on 12/26/2025).

[12] *Call for Papers — International Conference on Marketing and Technologies (ICMarkTech)*. en-US. ICMarkTech. url: https://icmarktech.com/submission/ (visited on 06/27/2026).

[13] *Data in Brief*. ScienceDirect, Elsevier. url: https://www.sciencedirect.com/journal/data-in-brief (visited on 06/27/2026).

[14] Piri Reis Üniversitesi. *SIU 2026 - Piri Reis University*. tr. url: https://siu2026.pirireis.edu.tr/en/ (visited on 06/28/2026).

[15] *IBERAMIA'2026 – IBERAMIA'2026, Asunción, Paraguay*. es. url: https://www.iberamia.org/iberamia/iberamia2026/ (visited on 06/28/2026).

[16] *AI Act | Shaping Europe's digital future*. en. June 2026. url: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (visited on 06/22/2026).

[17] Govind Ram Chhimpa et al. "A Comprehensive Framework for Eye Tracking: Methods, Tools, Applications, and Cross-Platform Evaluation". In: *Journal of Eye Movement Research* 18.5 (Sept. 2025), p. 47. issn: 1995-8692. doi: 10.3390/jemr18050047. url: https://pmc.ncbi.nlm.nih.gov/articles/PMC12564957/ (visited on 12/29/2025).

[18] *What is eye tracking? | How does eye tracking work*. en. url: https://www.tobii.com/learn-and-support/get-started/what-is-eye-tracking (visited on 12/28/2025).

[19] Smart Vision Labs. *Why Vision Is the Most Important Sense Organ*. en. Jan. 2017. url: https://medium.com/@SmartVisionLabs/why-vision-is-the-most-important-sense-organ-60a2cec1c164 (visited on 12/29/2025).

[20] *Eyes: A Window to the World for Your Brain*. en. url: https://my.clevelandclinic.org/health/body/21823-eyes (visited on 12/29/2025).

[21] *How the Eyes Work | National Eye Institute*. url: https://www.nei.nih.gov/eye-health-information/healthy-vision/how-eyes-work (visited on 12/29/2025).

[22] *The Anatomy of the Eye*. en-US. url: https://www.usqeyecare.com/conditions-treatments/the-anatomy-of-the-eye/ (visited on 06/28/2026).

[23] "Exploring the potential of eye tracking on personalized learning and real-time feedback in modern education". en-US. In: *Progress in Brain Research*. Vol. 282. ISSN: 0079-6123. Elsevier, Jan. 2023, pp. 49–70. doi: 10.1016/bs.pbr.2023.09.001. url: https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S0079612323000936 (visited on 12/29/2025).

[24] Moayad Mokatren, Tsvi Kuflik, and Ilan Shimshoni. "3D Gaze Estimation Using RGB-IR Cameras". In: *Sensors* 23.1 (Dec. 2022), p. 381. issn: 1424-8220. doi: 10.3390/s23010381. url: https://doi.org/10.3390/s23010381.

[25] Moayad Mokatren, Tsvi Kuflik, and Ilan Shimshoni. "3D Gaze Estimation Using RGB-IR Cameras". en. In: *Sensors* 23.1 (Jan. 2023), p. 381. issn: 1424-8220. doi: 10.3390/s23010381. url: https://www.mdpi.com/1424-8220/23/1/381 (visited on 06/28/2026).

[26] *Applications & History Of Eye Tracking Technology*. en-US. url: https://www.gazept.com/eye-tracking/ (visited on 12/29/2025).

[27] Justyna. *What is Eye Tracking? Definition and Applications*. en-US. Mar. 2025. url: https://dac.digital/what-is-eye-tracking-technology-and-how-does-it-transform-different-industries/ (visited on 12/29/2025).

[28] Christopher Collins et al. "Artificial intelligence in information systems research: A systematic literature review and research agenda". In: *International Journal of Information Management* 60 (Oct. 2021), p. 102383. issn: 0268-4012. doi: 10.1016/j.ijinfomgt.2021.102383. url: https://www.sciencedirect.com/science/article/pii/S0268401221000761 (visited on 12/30/2025).

[29] Haroon Sheikh, Corien Prins, and Erik Schrijvers. “Artificial Intelligence: Definition and Background”. en. In: *Mission AI: The New System Technology*. Ed. by Haroon Sheikh, Corien Prins, and Erik Schrijvers. Cham: Springer International Publishing, 2023, pp. 15–41. isbn: 978-3-031-21448-6. doi: 10.1007/978-3-031-21448-6_2. url: https://doi.org/10.1007/978-3-031-21448-6_2 (visited on 12/30/2025).

[30] *(PDF) The Role of Machine Learning in Data-Driven Decision- Making*. en. url: https://www.researchgate.net/publication/387365444_The_Role_of_Machine_Learning_in_Data-Driven_Decision-_Making (visited on 12/30/2025).

[31] *What Is Model Training? | IBM*. en. Feb. 2025. url: https://www.ibm.com/think/topics/model-training (visited on 12/30/2025).

[32] Iqbal H. Sarker. “Machine Learning: Algorithms, Real-World Applications and Research Directions”. In: *Sn Computer Science* 2.3 (2021), p. 160. issn: 2662-995X. doi: 10.1007/s42979-021-00592-x. url: https://pmc.ncbi.nlm.nih.gov/articles/PMC7983091/ (visited on 12/30/2025).

[33] Sumit K. Bishnu, Sabla Y. Alnouri, and Dhabia M. Al-Mohannadi. “Computational applications using data driven modeling in process Systems: A review”. In: *Digital Chemical Engineering* 8 (Sept. 2023), p. 100111. issn: 2772-5081. doi: 10.1016/j.dche.2023.100111. url: https://www.sciencedirect.com/science/article/pii/S2772508123000297 (visited on 01/03/2026).

[34] “(PDF) Research Paper On Artificial Intelligence And It’s Applications”. en. In: *ResearchGate* (Dec. 2025). doi: 10.53555/jaz.v44iS8.3544. url: https://www.researchgate.net/publication/377897946_Research_Paper_On_Artificial_Intelligence_And_It's_Applications (visited on 12/30/2025).

[35] *Understanding Screen-Based Eye Trackers: How They Work and Their Applications - iMotions*. en-US. Section: Uncategorized. May 2024. url: https://imotions.com/blog/learning/product-guides/understanding-screen-based-eye-trackers-how-they-work-and-their-applications/ (visited on 12/30/2025).

[36] *Exploring Mobile Eye Trackers: How Eye Tracking Glasses Work and Their Applications - iMotions*. en-US. Section: Academia. May 2024. url: https://imotions.com/blog/learning/exploring-mobile-eye-trackers-how-eye-tracking-glasses-work-and-their-applications/ (visited on 12/30/2025).

[37] *Webcam-based eyetracking, the analysis of eye movements using a webcam – Blog Laboratorium Badań Medioznawczych UW*. url: https://blog.lbm.uw.edu.pl/en/webcam-based-eyetracking/ (visited on 12/31/2025).

[38] Matteo Cognolato, Manfredo Atzori, and Henning Müller. “Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances”. In: *Journal of Rehabilitation and Assistive Technologies Engineering* 5 (June 2018). issn: 2055-6683. doi: 10.1177/2055668318773991. url: https://pmc.ncbi.nlm.nih.gov/articles/PMC6453044/ (visited on 12/31/2025).

[39] Peter Lowe. *4 Eye Tracking Systems: How They Work & What They Cost*. en-US. May 2025. url: https://www.crazyegg.com/blog/eye-tracking/ (visited on 12/31/2025).

[40] *GP3 HD Eye-Tracking Device - Take Advantage Of Research-Grade Equipment*. en-US. url: https://www.gazept.com/product/gp3hd/ (visited on 01/06/2026).

[41] *Pupil Core - Open source eye tracking platform*. en-US. url: https://pupil-labs.com/products/core (visited on 01/06/2026).

[42] *Reach further with your research | Tobii Pro Fusion*. en. url: https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-fusion (visited on 01/06/2026).

[43] *Gazepoint GP3 Eye Tracker*. en. Gazepoint. url: https://www.gazept.com/product/gazepoint-gp3-eye-tracker/ (visited on 01/06/2026).

[44] Ayesha Alharthi, Meera Alaryani, and Sanaa Kaddoura. "A comparative study of machine learning and deep learning models in binary and multiclass classification for intrusion detection systems". In: *Array* 26 (July 2025), p. 100406. issn: 2590-0056. doi: 10.1016/j.array.2025.100406. url: https://www.sciencedirect.com/science/article/pii/S2590005625000335 (visited on 12/31/2025).

[45] Sarvandani. *Top 11 algorithms of non-linear regression in machine learning + Proposed Python library+Python...* en. Jan. 2024. url: https://medium.com/@mohamadhasan.sarvandani/top-algorithms-of-non-linear-regression-in-machine-learning-proposed-python-library-and-python-c871752ddd81 (visited on 01/01/2026).

[46] "(PDF) Random Forest Algorithm Overview". en. In: *ResearchGate* (Oct. 2025). doi: 10.58496/BJML/2024/007. url: https://www.researchgate.net/publication/382419308_Random_Forest_Algorithm_Overview (visited on 01/01/2026).

[47] Rafay Qayyum. *Computational Complexity of Machine Learning Algorithms*. en. Aug. 2022. url: https://rafayqayyum.medium.com/computational-complexity-of-machine-learning-algorithms-254c275de84 (visited on 01/01/2026).

[48] Muhammad Zeshan Arshad and Ali Algarni. "Exploring time complexity and machine learning scalability for COVID-19 Predictions: A case study from Saudi Arabia". In: *Journal of Computational and Applied Mathematics* 474 (Mar. 2026), p. 116933. issn: 0377-0427. doi: 10.1016/j.cam.2025.116933. url: https://www.sciencedirect.com/science/article/pii/S0377042725004479 (visited on 01/01/2026).

[49] Vanessa Georges et al. "UX Heatmaps: Mapping User Experience on Visual Interfaces". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. event-place: San Jose, California, USA. New York, NY, USA: Association for Computing Machinery, 2016, pp. 4850–4860. isbn: 978-1-4503-3362-7. doi: 10.1145/2858036.2858271. url: https://doi.org/10.1145/2858036.2858271.

[50] Nattapat Boonprakong et al. "Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases when Facing Different Opinions". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. event-place: Hamburg, Germany. New York, NY, USA: Association for Computing Machinery, 2023. isbn: 978-1-4503-9421-5. doi: 10.1145/3544548.3580917. url: https://doi.org/10.1145/3544548.3580917.

[51] Mansi Sharma et al. "Implicit Search Intent Recognition using EEG and Eye Tracking: Novel Dataset and Cross-User Prediction". In: *Proceedings of the 25th International Conference on Multimodal Interaction*. ICMI '23. event-place: Paris, France. New York, NY, USA: Association for Computing Machinery, 2023, pp. 345–354. isbn: 979-8-4007-0055-2. doi: 10.1145/3577190.3614166. url: https://doi.org/10.1145/3577190.3614166.

[52] Syrine Haddad et al. "Good GUIs, Bad GUIs: Affective Evaluation of Graphical User Interfaces". In: *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '24. event-place: Cagliari, Italy. New York, NY, USA: Association for Computing Machinery, 2024, pp. 232–243. isbn: 979-8-4007-0433-8. doi: 10.1145/3627043.3659549. url: https://doi.org/10.1145/3627043.3659549.

[53] Zubair Ahsan and Unaizah Obaidellah. "Eye-Tracking Indicators of Novice Programmers' Proficiency: A Machine Learning Approach". In: *ACM Trans. Comput. Educ.* (Oct. 2025). Place: New York, NY, USA Publisher: Association for Computing Machinery. doi: 10.1145/3773898. url: https://doi.org/10.1145/3773898.

[54] Xiaojun Ye et al. "Research on the Design of Multimodal AI-Driven Visual Narrative Recommendation System for Cross-border E-commerce". In: *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*. DEAI '25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 1374–1379. isbn: 979-8-4007-1279-1. doi: 10.1145/3745238.3745452. url: https://doi.org/10.1145/3745238.3745452.

[55] Janet H. Hsiao et al. "Eye movement analysis with hidden Markov models (EMHMM) with co-clustering". en. In: *Behavior Research Methods* 53.6 (Dec. 2021), pp. 2473–2486. issn: 1554-3528. doi: 10.3758/s13428-021-01541-5. url: https://doi.org/10.3758/s13428-021-01541-5 (visited on 11/30/2025).

[56] Patrícia Deud Guimarães et al. "A mixed factorial experiment with colors and adaptive web user interfaces to change emotions". In: *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*. IHC '23. event-place: Maceió, Brazil. New York, NY, USA: Association for Computing Machinery, 2024. isbn: 979-8-4007-1715-4. doi: 10.1145/3638067.3638092. url: https://doi.org/10.1145/3638067.3638092.

[57] Thomas Le Bras, Benoit Allibe, and Karine Doré-Mazars. "The way we look at an image or a webpage can reveal personality traits". en. In: *Scientific Reports* 14.1 (July 2024), p. 15488. issn: 2045-2322. doi: 10.1038/s41598-024-62210-z. url: https://doi.org/10.1038/s41598-024-62210-z (visited on 11/30/2025).

[58] Ilia Zaitsev. "Affect-Predictive Models: Predicting Emotional Responses Directly to Stimuli". In: *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. WI-IAT '21. event-place: Melbourne, VIC, Australia. New York, NY, USA: Association for Computing Machinery, 2022, pp. 263–267. isbn: 978-1-4503-9187-0. doi: 10.1145/3498851.3498962. url: https://doi.org/10.1145/3498851.3498962.

[59] Mengyue Wang, Xin Li, and Patrick Y. K. Chau. "Leveraging Image-Processing Techniques for Empirical Research: Feasibility and Reliability in Online Shopping Context". en. In: *Information Systems Frontiers* 23.3 (June 2021), pp. 607–626. issn: 1572-9419. doi: 10.1007/s10796-020-09981-8. url: https://doi.org/10.1007/s10796-020-09981-8 (visited on 11/30/2025).

[60] Pablo Villanueva González et al. "A gender-aware saliency prediction system for web interfaces using deep learning and eye-tracking data". en. In: *Brain Informatics* 12.1 (Oct. 2025), p. 25. issn: 2198-4026. doi: 10.1186/s40708-025-00274-x. url: https://doi.org/10.1186/s40708-025-00274-x (visited on 11/30/2025).

[61] Mashael Aldayel, Mourad Ykhlef, and Abeer Al-Nafjan. "Consumers' Preference Recognition Based on Brain–Computer Interfaces: Advances, Trends, and Applications". en. In: *Arabian Journal for Science and Engineering* 46.9 (Sept. 2021), pp. 8983–8997. issn: 2191-4281. doi: 10.1007/s13369-021-05695-4. url: https://doi.org/10.1007/s13369-021-05695-4 (visited on 11/30/2025).

[62] Shuning Han et al. "Enhanced Scene-Specific Saliency Prediction Incorporating Scene Information". en. In: *Cognitive Computation* 17.6 (Oct. 2025), p. 155. issn: 1866-9964. doi: 10.1007/s12559-025-10506-1. url: https://doi.org/10.1007/s12559-025-10506-1 (visited on 11/30/2025).

[63] Eneko Lopez et al. "The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples". In: *Analytica Chimica Acta* 1275 (Sept. 2023), p. 341532. issn: 0003-2670. doi: 10.1016/j.aca.2023.341532. url: https://www.sciencedirect.com/science/article/pii/S0003267023007535 (visited on 06/15/2026).

[64] Mehmet Ali TOR. *Top 5 Validation Methods In Machine Learning*. en. July 2024. url: https://medium.com/@mehmetalitor/top-5-model-validation-methods-in-machine-learning-77eed8d08937 (visited on 06/15/2026).

[65] *User Guide*. en. url: https://scikit-learn/stable/user_guide.html (visited on 06/15/2026).

[66] Azal Ahmad Khan, Omkar Chaudhari, and Rohitash Chandra. "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation". In: *Expert Systems with Applications* 244 (June 2024), p. 122778. issn: 0957-4174. doi: 10.1016/j.eswa.2023.122778. url: https://www.sciencedirect.com/science/article/pii/S0957417423032803 (visited on 06/15/2026).

[67] João Manoel Herrera Pinheiro et al. *The Impact of Feature Scaling In Machine Learning: Effects on Regression and Classification Tasks*. arXiv:2506.08274 [cs.LG] version: 1. June 2025. doi: 10.48550/arXiv.2506.08274. url: http://arxiv.org/abs/2506.08274 (visited on 06/15/2026).

[68] Messaoud Ahmed Ouameur, Manouane Caza-Szoka, and Daniel Massicotte. "Machine learning enabled tools and methods for indoor localization using low power wireless network". In: *Internet of Things* 12 (Dec. 2020), p. 100300. issn: 2542-6605. doi: 10.1016/j.iot.2020.100300. url: https://www.sciencedirect.com/science/article/pii/S2542660520301323 (visited on 06/15/2026).

[69] Luca Franceschi et al. "Hyperparameter Optimization in Machine Learning". In: *Foundations and Trends® in Machine Learning* 18.6 (2025). arXiv:2410.22854 [stat.ML], pp. 1054–1201. issn: 1935-8237, 1935-8245. doi: 10.1561/2200000088. url: http://arxiv.org/abs/2410.22854 (visited on 06/15/2026).

[70] *3.2. Tuning the hyper-parameters of an estimator*. en. url: https://scikit-learn/stable/modules/grid_search.html (visited on 06/15/2026).

[71] Nunung Nurul Qomariyah, Dimitar Kazakov, and Ahmad Fajar. "Predicting User Preferences with XGBoost Learning to Rank Method". In: Dec. 2020, pp. 123–128. doi: 10.1109/ISRITI51436.2020.9315494.

[72] Tobias Schlosser et al. *A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision*. Jan. 2024. doi: 10.13140/RG.2.2.14331.69928.

[73] Oona Rainio, Jarmo Teuho, and Riku Klén. "Evaluation metrics and statistical tests for machine learning". In: *Scientific Reports* 14 (Mar. 2024), p. 6086. issn: 2045-2322. doi: 10.1038/s41598-024-56706-x. url: https://pmc.ncbi.nlm.nih.gov/articles/PMC10937649/ (visited on 06/17/2026).

[74] *Classification Model Evaluation Metrics*. en. url: https://www.researchgate.net/publication/352902406_Classification_Model_Evaluation_Metrics (visited on 06/17/2026).

[75] Ping Li, Christopher J C Burges, and Qiang Wu. "Learning to Rank Using Classification and Gradient Boosting". en. In: ().

[76] Aryan Jadon and Avinash Patil. "A Comprehensive Survey of Evaluation Techniques for Recommendation Systems". en. In: *Computation of Artificial Intelligence and Machine Learning*. Ed. by Amit Kumar Bairwa et al. Cham: Springer Nature Switzerland, 2025, pp. 281–304. isbn: 978-3-031-71484-9. doi: 10.1007/978-3-031-71484-9_25.

[77] Daniel Valcarce et al. "Assessing ranking metrics in top-N recommendation". en. In: *Information Retrieval Journal* 23.4 (Aug. 2020), pp. 411–448. issn: 1573-7659. doi:

10.1007/s10791-020-09377-x. url: https://doi.org/10.1007/s10791-020-09377-x (visited on 06/18/2026).

[78] *5.2. Permutation feature importance*. en. url: https://scikit-learn/stable/modules/permutation_importance.html (visited on 06/20/2026).

Appendix A

Project Charter

PREPD 2025/2026
Gestão de Projetos

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

PROJECT CHARTER

1. Informação geral	
Nome do Projeto:	An Intelligent System for Visual Attention Analysis Based on Eye Tracking and Artificial Intelligence Focused on UI/UX
Sponsor:	Ana Teixeira Goreti Marreiros
Instituição / Departamento	GECAD

2. Equipa do Projeto				
Cargo	**Nome**	**Departamento**	**Contact Tel**	**E-mail**
Student	Diogo Gonçalves	DEI/GECAD	969614404	1211509@isep.ipp.pt
Advisor	Ana Teixeira	GECAD	------------------	art@isep.ipp.pt
Co-advisor	Goreti Marreiros	GECAD	------------------	mgt@isep.ipp.pt

3. Stakeholders		
Nome	Poder	Interesse
End users	Medium	High
E-commerce companies	High	High
UI/UX developers	Medium	High

Figure A.1: Project Charter Part1

4. Âmbito

Problema / justificação

The growth of the digital market has intensified competition among companies, making the capture and retention of user attention a decisive factor for the success of e commerce platforms, where UX/UI design plays a crucial role in differentiation and in influencing user purchasing behavior. In this context, companies are constantly seeking strategies that allow them to better understand how users interact with their websites, in order to optimize the experience, increase engagement, and foster loyalty.

However, many websites still adopt static and inflexible structures, failing to effectively consider visual focus, navigation patterns, and users' emotional responses during interaction. This limitation hinders the creation of truly personalized and adaptive experiences, undermining the effectiveness of visual communication and reducing the impact of design decisions on conversion and user satisfaction.

In this context, it becomes necessary to develop a technological solution capable of objectively analyzing users' visual attention and emotions, contributing to more effective, intuitive, and individually aligned digital experiences.

This research is situated within several fields of study:

- Human–Computer Interaction
- Visual Attention Analysis with Eye Tracking
- User Experience (UX) and User Interface (UI)
- Artificial Intelligence applied to human behavior

Figure A.2: Project Charter Part2

Objetivos do projeto

General Objective:

- The project aims to develop an intelligent system capable of analyzing users' visual attention and predicting product preferences during navigation on e-commerce websites, by combining eye-tracking technologies with Artificial Intelligence models. The system seeks to inform and support adaptive UX/UI design decisions, contributing to more personalized and effective user experiences.

Specific Objectives:

- Collect and analyze visual attention data using eye-tracking devices.
- Identify design elements (e.g. layout, and presentation style) that influence user focus and visual engagement.
- Investigate the relationship between gaze and physiological responses and explicit product selection in e-commerce interactions.
- Develop and evaluate AI models capable of predicting product preference from eye-tracking, oculomotor, and physiological data.
- Analyze feature importance to identify which behavioural and physiological signals most strongly contribute to user engagement and preference.
- Design mechanisms for dynamic content adaptation and personalized recommendations.
- Assess the potential of the resulting models to inform future adaptive UX/UI personalization strategies.

Global Research Question:

- How can the integration of eye-tracking technologies and Artificial Intelligence be used to analyse users' visual attention and emotions on e-commerce websites, achieving dynamic content personalisation and better adaptation of UX/UI design?

Sub-Research Questions:

- What are the most effective metrics to evaluate users' attention on e-commerce websites?
- What are the main design elements (color, contrast, arrangement of images and text) that most influence gaze focus on websites?
- What visual presentation practices (colors, style, brightness) can promote positive/negative emotions and influence users' decisions?
- What is the relationship between visual attention and emotions expressed while viewing products online?
- Which Artificial Intelligence models are most effective at detecting emotions from eye-tracking data?
- How can AI algorithms be trained to associate visual patterns (gaze patterns) with emotional responses?
- How can an intelligent system dynamically adapt the website content based on users' emotions and focus?

Figure A.3: Project Charter Part3

Benefícios
• Improved user experience • Increase in customers • Increase commercial performance • Innovation in personalization strategies • Contribution to research and practice
Entregáveis
Project plan Dissertation report Presentation and discussion Results validation Software code

Figure A.4: Project Charter Part4

5. Tempo	
Milestones / Datas	
Committed Dates 12/09/2025 – First draft of the Extended Abstract 12/13/2025 – Completion of the Extended Abstract 01/02/2026 – First draft of the Final Document 01/05/2026 – Completion of the Final Document Mandatory Dates 12/15/2025 – Submission of the Extended Abstract to PREPD 01/07/2026 – Submission of the Final Document to PREPD	
6. Custo	
Fontes de custo	
Does not apply	
7. Pressupostos	**8. Restrições**
Does not apply	Does not apply
Notas	

Figure A.5: Project Charter Final

Appendix B

Gantt Diagram

	Nome da Tarefa	Duração	Início	Conclusão
1	**Project**	**188 dias**	**Qui 9/25/25**	**Seg 6/15/26**
2	**PREPD**	**75 dias**	**Qui 9/25/25**	**Qua 1/7/26**
3	**Problem Analysis**	**68 dias**	**Qui 9/25/25**	**Sáb 12/27/25**
4	Context	15 dias	Qui 9/25/25	Qua 10/15/25
5	Problem Statement	15 dias	Qui 9/25/25	Qua 10/15/25
6	Objectives	9 dias	Qui 10/16/25	Ter 10/28/25
7	Research Questions	19 dias	Qua 10/29/25	Sáb 11/22/25
8	Ethical Considerations	6 dias	Dom 11/23/25	Sex 11/28/25
9	**Project Plan**	**21 dias**	**Sex 10/31/25**	**Sex 11/28/25**
10	Project Charter	6 dias	Sex 10/31/25	Sex 11/7/25
11	WBS	12 dias	Sáb 11/8/25	Dom 11/23/25
12	Gantt Diagram	4 dias	Seg 11/24/25	Qui 11/27/25
13	**State of Art**	**46 dias**	**Qua 11/5/25**	**Qua 1/7/26**
14	Systematic Review Methodology (PRISMA)	7 dias	Sex 11/14/25	Dom 11/23/25
15	Acquisition of Scientific Articles	12 dias	Seg 11/24/25	Ter 12/9/25
16	Literature Review	17 dias	Qua 12/10/25	Qui 1/1/26
17	**Experiment planning**	**13 dias**	**Seg 12/15/25**	**Qua 12/31/25**
18	Definition of the experimental design	4 dias	Ter 12/16/25	Sex 12/19/25
19	Creation of product grids and AOIs (Kiosk)	8 dias	Seg 12/22/25	Qua 12/31/25
20	Presentation Preparation	3 dias	Qui 1/1/26	Dom 1/4/26

Figure B.1: Diagrama Gantt Part1

21	**DIMEI**	**118 dias**	**Qui 1/15/26**	**Seg 6/29/26**
22	**Experiment implementation**	**48 dias**	**Qui 1/22/26**	**Seg 3/30/26**
23	Experiment Execution	14 dias	Qui 1/22/26	Ter 2/10/26
24	Data Analysis	27 dias	Ter 2/17/26	Qua 3/25/26
25	**DataConstruction**	**15 dias**	**Ter 3/31/26**	**Dom 4/19/26**
26	Choice-Based Experiment	6 dias	Ter 3/31/26	Ter 4/7/26
27	Feature Extraction	9 dias	Qua 4/8/26	Dom 4/19/26
28	**Model Development**	**31 dias**	**Seg 4/20/26**	**Dom 5/31/26**
29	DataPreparation	7 dias	Seg 4/20/26	Ter 4/28/26
30	**AIModels**	**25 dias**	**Ter 4/28/26**	**Dom 5/31/26**
31	Model selection	5 dias	Ter 4/28/26	Sáb 5/2/26
32	Training Models	16 dias	Seg 5/4/26	Dom 5/24/26
33	Evaluation of Models	6 dias	Seg 5/25/26	Dom 5/31/26
34	**Results**	**14 dias**	**Ter 6/2/26**	**Sex 6/19/26**
35	**Results Validation**	**14 dias**	**Ter 6/2/26**	**Sex 6/19/26**
36	Validation of Ranking Results	7 dias	Ter 6/2/26	Qua 6/10/26
37	Performance analysis	7 dias	Qui 6/11/26	Sex 6/19/26
38	**Conclusion**	**32 dias**	**Dom 5/17/26**	**Seg 6/29/26**
39	Final report	32 dias	Dom 5/17/26	Seg 6/29/26
40	Report validation	7 dias	Sáb 6/20/26	Seg 6/29/26
41	Presentation	3 dias	Qui 6/25/26	Seg 6/29/26
42	**Milestones**	**151 dias**	**Seg 12/1/25**	**Seg 6/29/26**

Figure B.2: Diagrama Gantt Part2

Appendix C

Supplementary AOI Heatmaps for Remaining Metrics

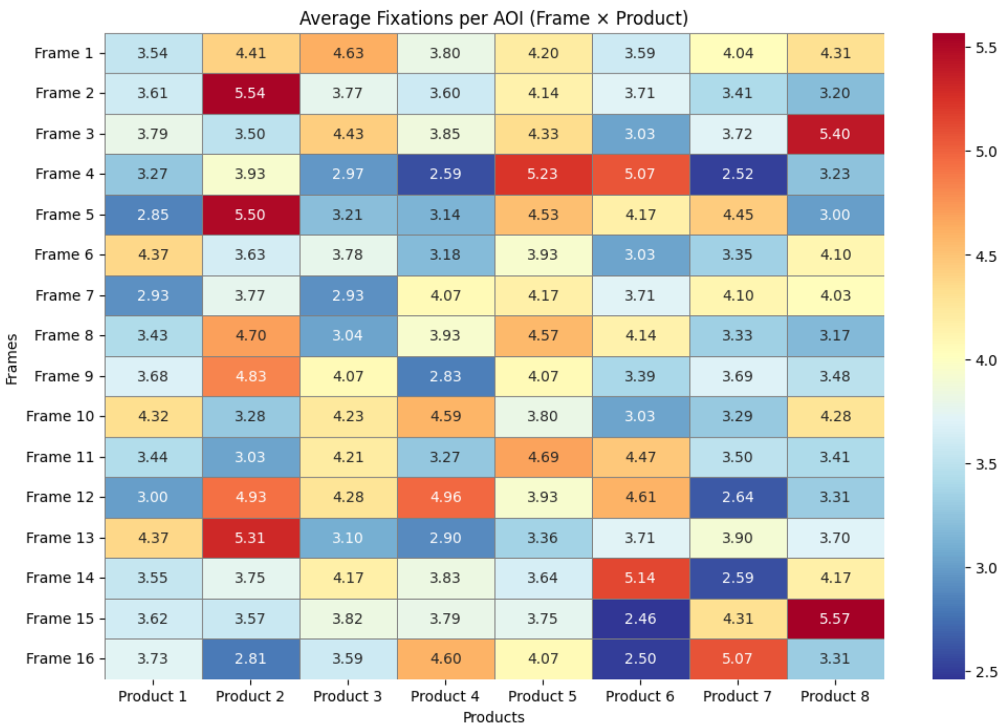

Figure C.1: Heatmap presenting the fixation for every AOI

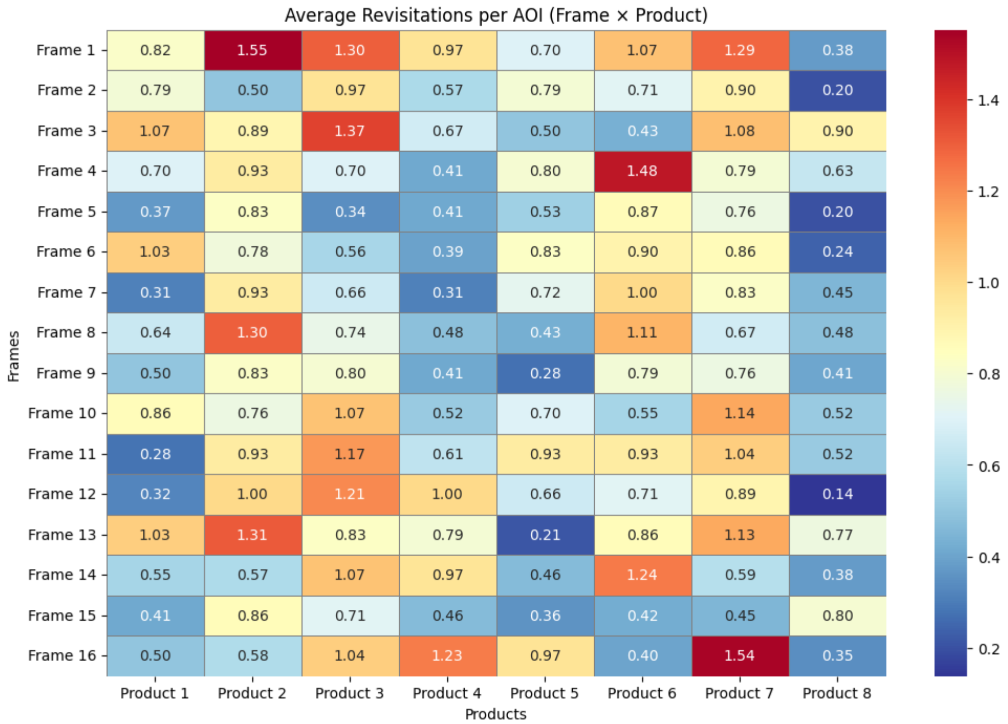

Figure C.2: Heatmap presenting the revisits for every AOI

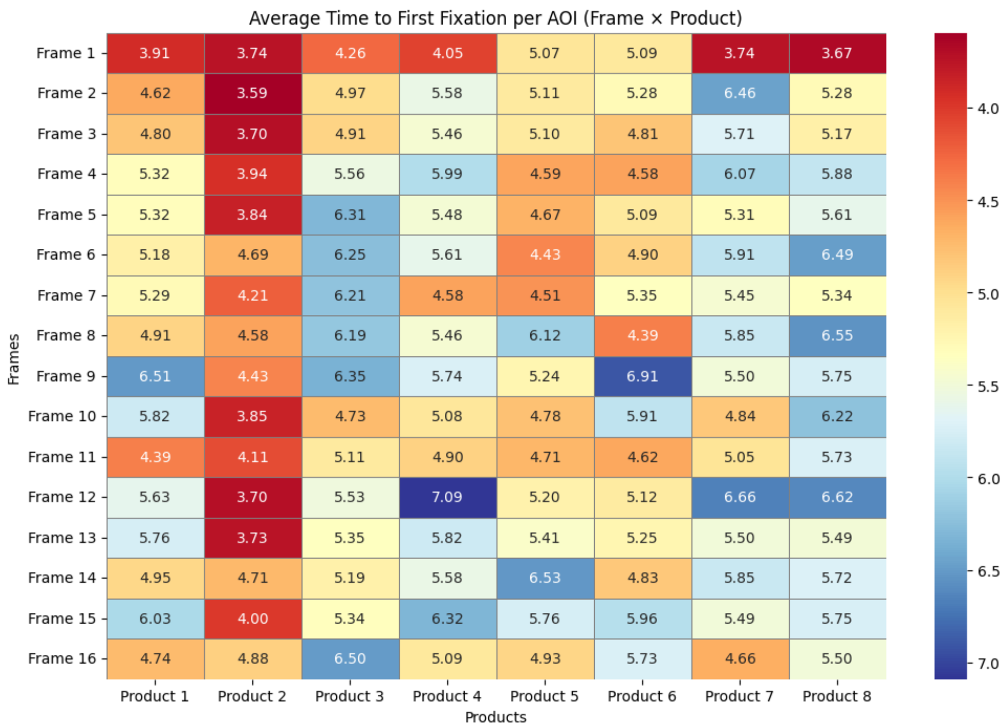

Figure C.3: Heatmap presenting the first time to view for every AOI

Appendix D

Supplementary Category-based Vertical Graphs for Remaining Metrics

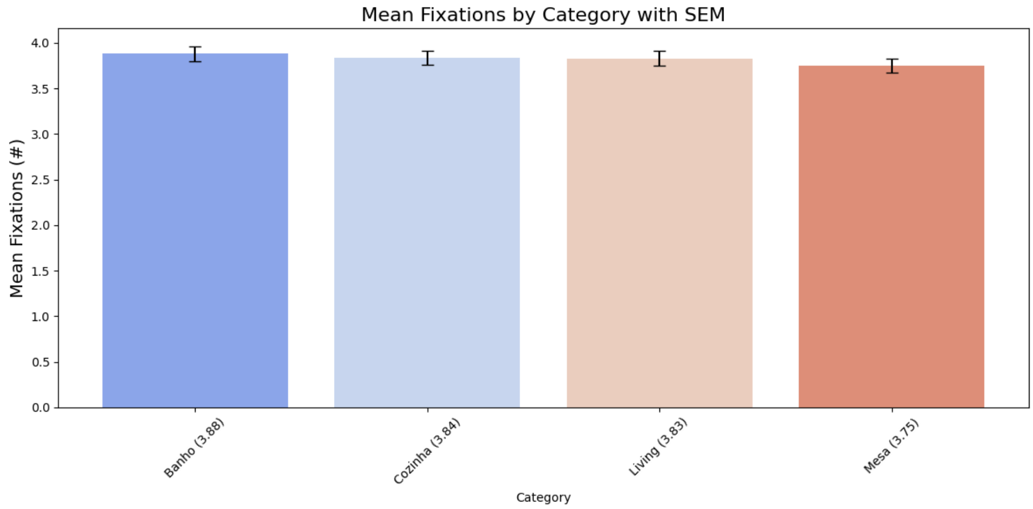

Figure D.1: Vertical Graph presenting the fixations for every Category

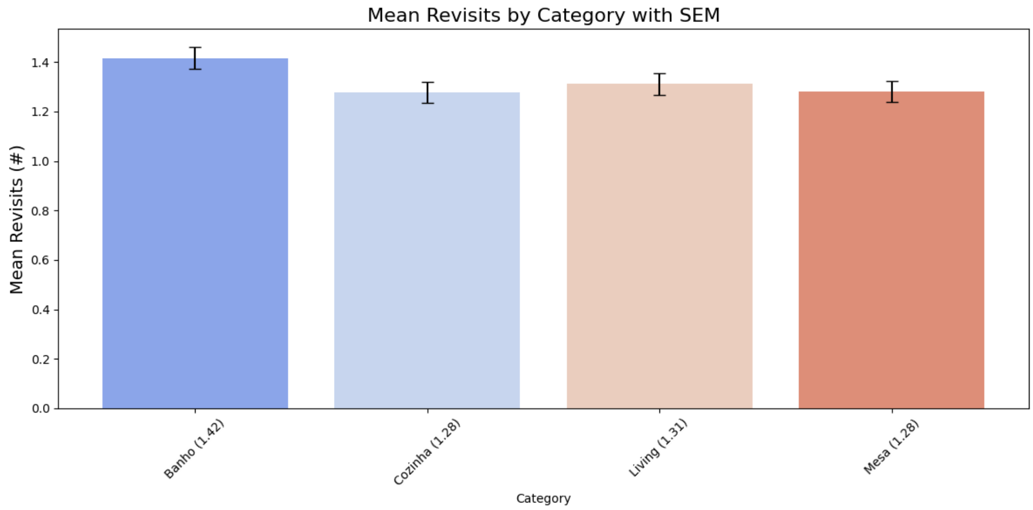

Figure D.2: Vertical Graph presenting the revisits for every Category

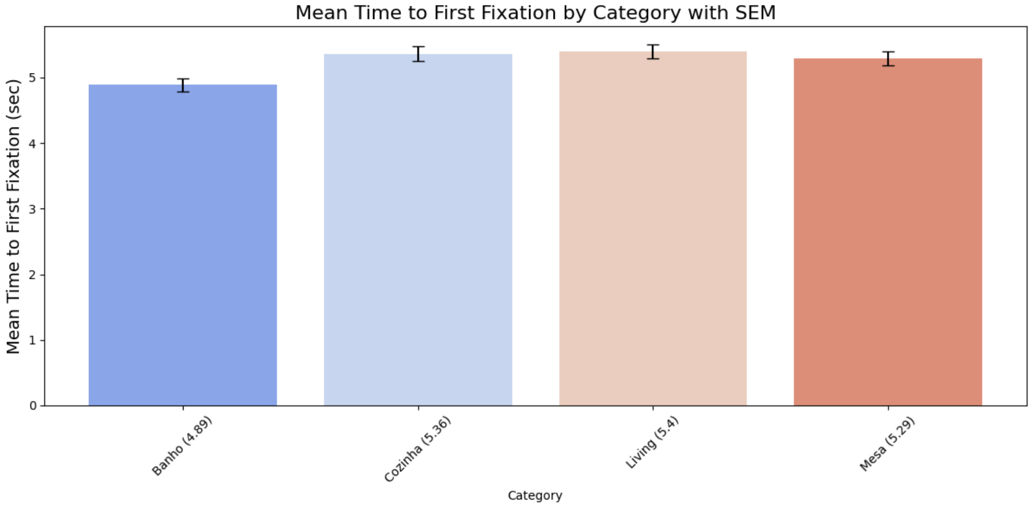

Figure D.3: Vertical Graph presenting the first time to view for every Category

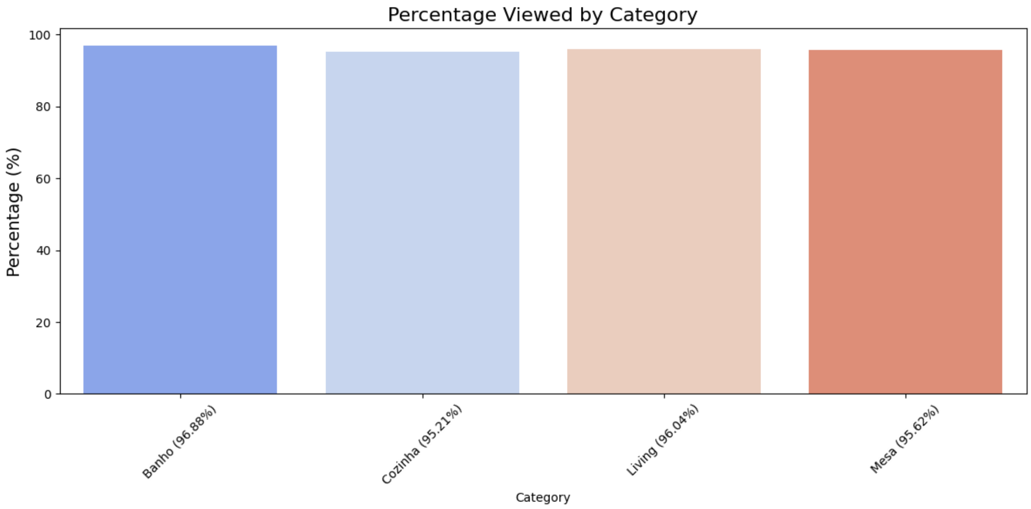

Figure D.4: Vertical Graph presenting the percentage of products seen for every Category

Appendix E

Supplementary Position-based Vertical Graphs for Remaining Metrics

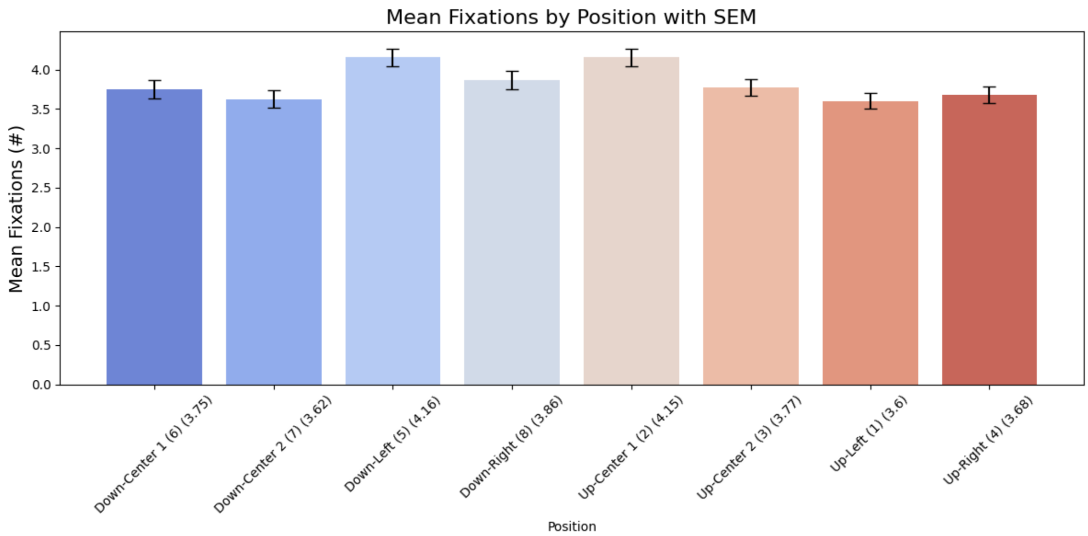

Figure E.1: Vertical Graph presenting the fixations for every Position

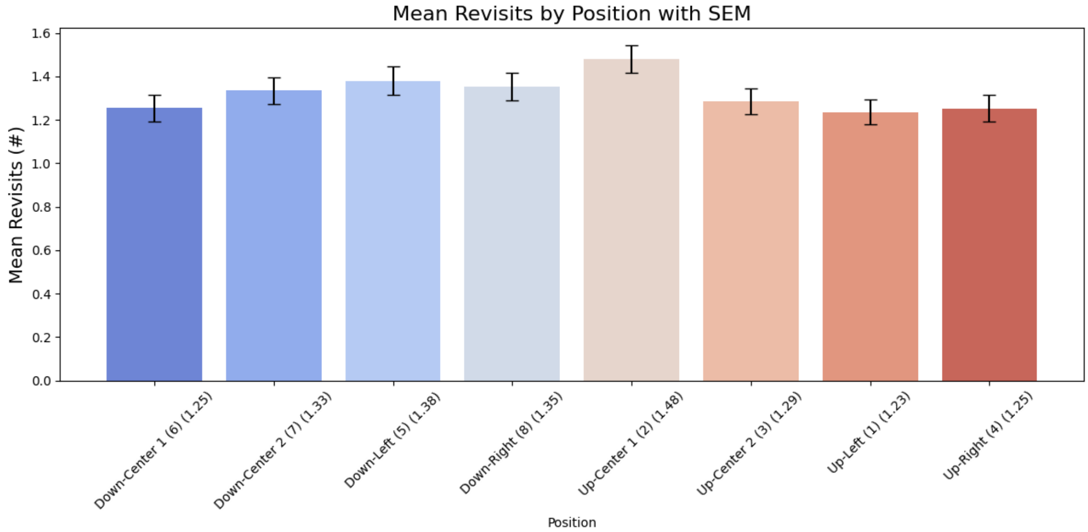

Figure E.2: Vertical Graph presenting the revisits for every Position

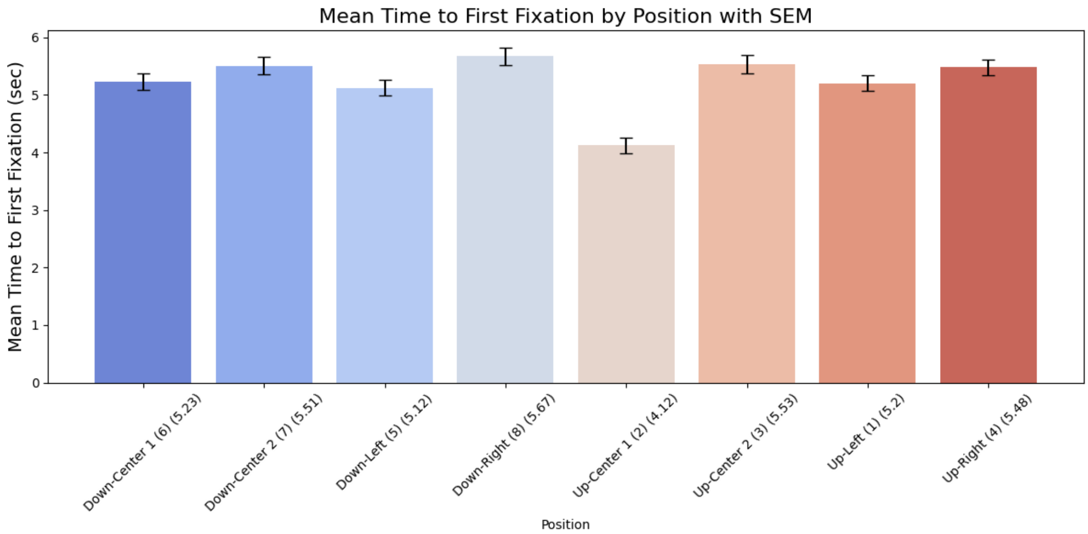

Figure E.3: Vertical Graph presenting the first time to view for every Position

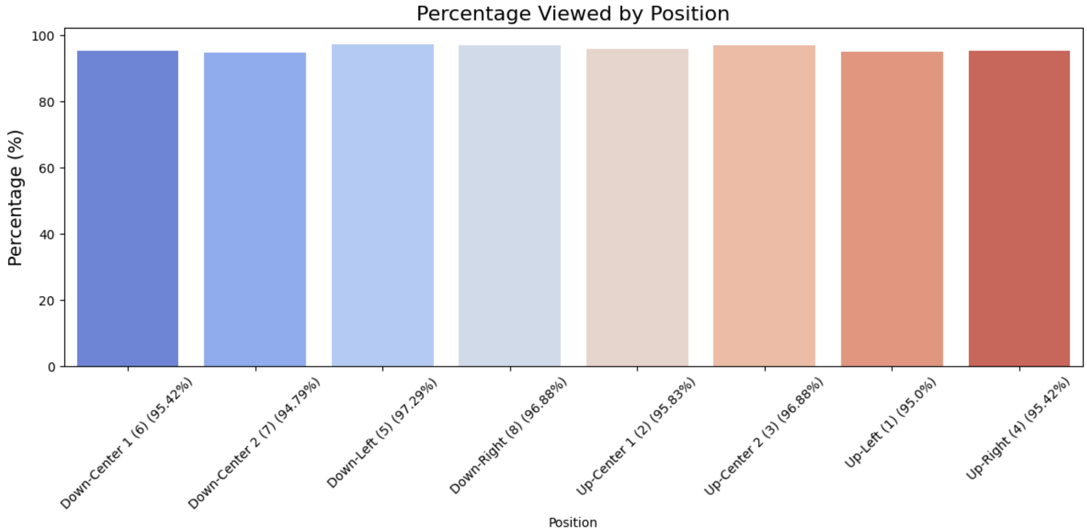

Figure E.4: Vertical Graph presenting the percentage of products seen for every Position

Appendix F

Supplementary Position X Background Line Graphs for Remaining Metrics

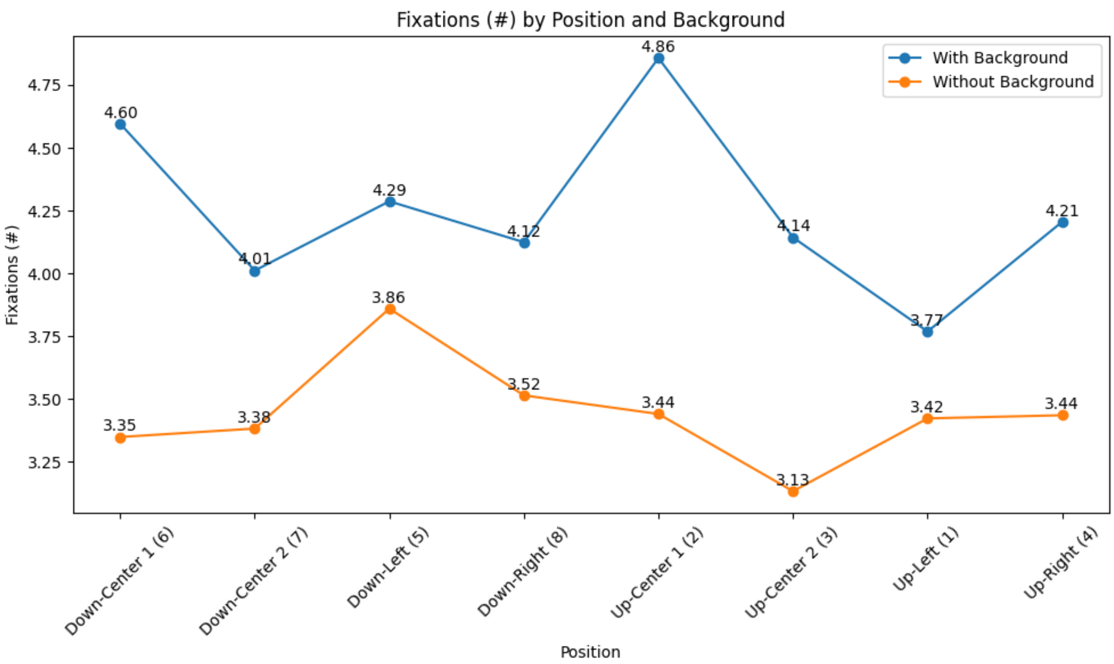

Figure F.1: Line graph of Fixations across grid positions with and without contextual background

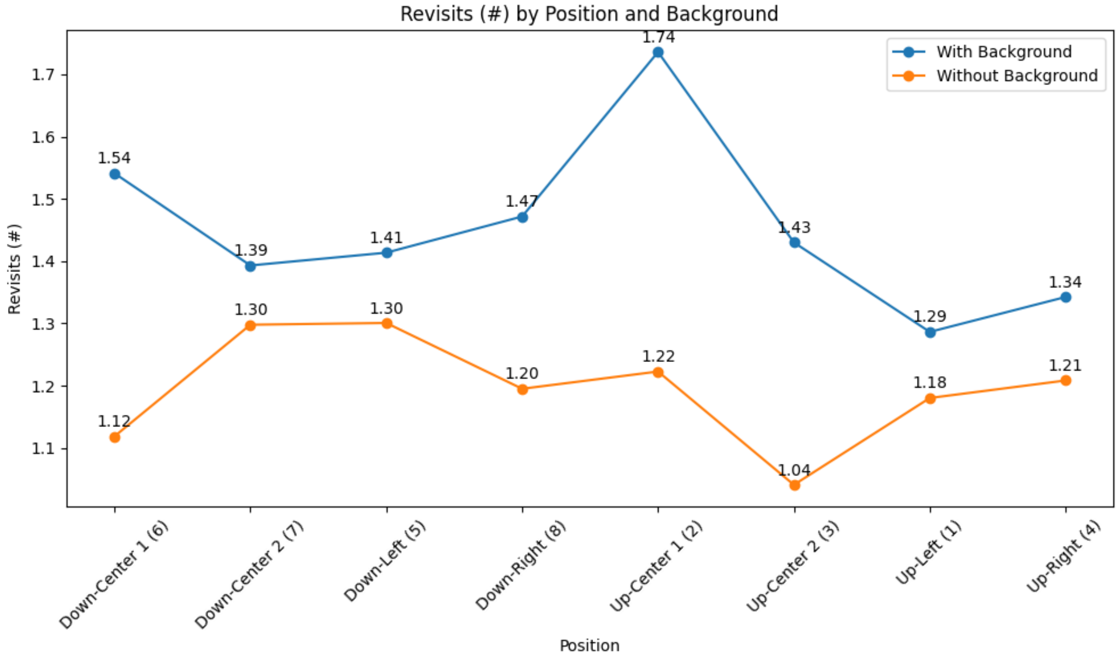

Figure F.2: Line graph of Revisits across grid positions with and without contextual background

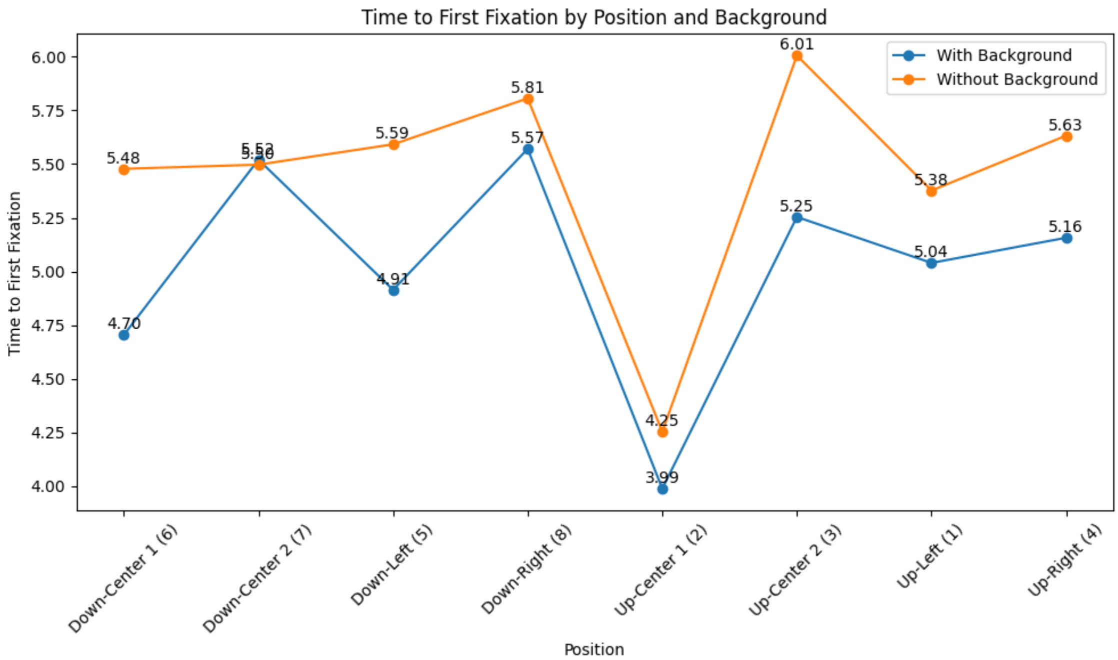

Figure F.3: Line graph of First Time to View across grid positions with and without contextual background