THESIS / THÈSE

MASTER IN BUSINESS ENGINEERING PROFESSIONAL FOCUS IN ANALYTICS & DIGITAL BUSINESS

Influence of Explainable Artificial Intelligence (XAI) in the acceptance of online product recommendations

Pé, Margaux

Award date:
2021

Awarding institution:
University of Namur

Influence of Explainable Artificial Intelligence (XAI) in the acceptance of online product recommendations

Margaux PÉ

Directrice Prof. Dr. W. HAMMEDI

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en Analytics & Digital Business

ANNEE ACADEMIQUE 2020-2021

Université de Namur, ASBL
Faculté des Sciences économiques, sociales et de gestion – Département des Sciences de gestion

Rempart de la Vierge 8, B-5000 Namur, Belgique, Tel. +32 [0]81 72 48 41/49 58, Fax +32 [0]81 72 48 40

Summary

Les dernières avancées en matière d'intelligence artificielle ont permis de produire, pour les consommateurs, des recommandations de produits en ligne de plus en plus personnalisées. Cependant, les modèles qui sous-tendent ces recommandations sont souvent perçus comme des "boîtes noires" et posent des problèmes éthiques. Dans la première partie de ce mémoire, après un brève introduction, une revue de la littérature a permis d'identifier que l'utilisation d'explications pouvait aider à résoudre ces problèmes éthiques et de compréhension. Cependant, à ce jour, très peu d'expérimentations ont été menées dans le domaine de l'explication des recommandations de produits en ligne. Pour combler cette lacune, une étude quantitative a été menée auprès de consommateurs belges. L'avant dernière partie de ce mémoire, consacrée à l'analyse des résultats, a permis d'identifier que certains types d'explications peuvent être exploités afin que les consommateurs aient confiance dans le système utilisant une intelligence artificielle et qu'ils acceptent davantage de recevoir des recommandations.

The latest advances in artificial intelligence have produced increasingly personalized online product recommendations for consumers. However, the models underlying these recommendations are often perceived as "black boxes" and raise ethical issues. In the first part of this thesis, after a brief introduction, a review of the literature identified that the use of explanations could help solve these ethical and comprehension problems. However, to date, very few experiments have been conducted in the area of explanation of online product recommendations. To fill this gap, a quantitative study was conducted with Belgian consumers. At the end of this thesis, the analysis of the results has identified that certain types of explanations can be exploited so that consumers trust the system using artificial intelligence and accept to receive recommendations.

Foreword

This thesis was made possible thanks to the help of several people to whom I would like to show my appreciation.

I would first like to express my gratitude to my thesis director, Prof. Dr. Wafa HAMMEDI, whose expertise was invaluable in the formulation of the research question and the hypotheses. I would also like to thank her for her time, her availability and above all her precious advice, which contributed to my reflection.

I would also like to thank the UNamur professors, who gave me the necessary tools to succeed in my academic studies. I would like to thank Prof. Nadia STEILS, whose teaching of market study methodology was of major help in the process of my research. Prof. Isabelle LINDEN and Prof. Bertrand VERLAINE encouraged me to be rigorous in authoring scientific papers, their learning helped me throughout this work, and I would like to thank them.

Finally, I would like to express my gratitude to my brother Guillaume and my friend Gilles for reviewing my thesis, and my parents, for their support and encouragement.

Table of contents

1 Chapter 1: Introduction

The term artificial intelligence (AI) is becoming increasingly popular in the marketing world. However, being able to give a clear definition and understand what is behind this concept can be a challenging task. At the beginning of the research on artificial intelligence, during a conference in 1955, John McCarty, mathematician and computer scientist, proposed a project consisting in simulating human cognitive abilities by a machine (McCarthy et al., 1955a). This constituted one of the first initiatives involving artificial intelligence. Nowadays, this concept has evolved, and has become more complex with concepts such as machine learning (ML) and deep learning (DL), which will be described in the coming pages. Its evolution also has the effect of diversifying its fields of application.

In a report from the European Commission, it is pointed that “in 2020, 7% of enterprises in the EU with at least 10 employees used AI applications” (*Artificial intelligence in EU enterprises*, 2021). In the figure 1-1, four main categories of artificial intelligence are identified: chat services (chatbot or virtual agent) responding to customers, machine learning AIs used to analyse Big Data internally, natural language (processing, generation, and recognition) for Big Data analysis as well, and finally service robots with diverse levels of autonomy.

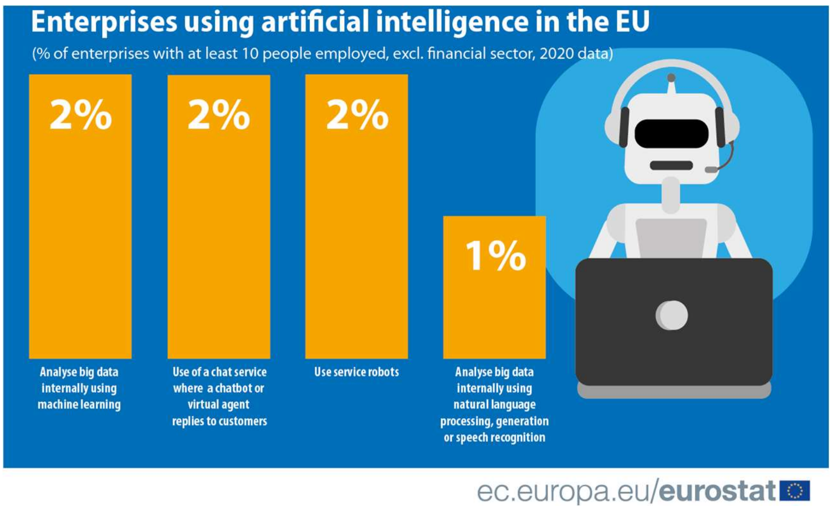


Figure 1-1: Bar chart of AI types used in EU companies in 2020 (taken from Artificial intelligence in EU enterprises*, 2021)*

Among these distinct types of AI, several solutions are applicable to the marketing field. Chatbots to answer customers' questions on websites, in-store robots to advise customers or generate original shopping experiences, machine learning tools to predict customers' buying behaviours and thus, adapt communication tools, and a wide range of other applications are conceivable. Recent progress in the field of artificial intelligence is due notably, according to a report published by the European Commission, to an "increased availability of processing power, improvements in algorithms and the exponential growth in the volume and variety of digital data, and increased funding" (Craglia & Europäische Gemeinschaften, 2018a, p. 16). These changes over the last ten years have turned AI into a more accessible tool for marketing, creating both opportunities and challenges.

Opportunities exist for example in the retail sector, where the use of AI is growing. In an article summarising the "Gondola Day 2018", a conference organized by the Belgian business magazine Gondola, Nikolaos Loutas, director of technology consulting at PwC, presents a number of concrete applications of AI in the retail sector such as: "dynamic pricing, customer feedback and interaction via social media, predictive stock management, customer recommendations..."(*Gondola Day 2018*, 2018). Artificial intelligence is also one of the "9 key trends" expected for retail in 2022, according to an article published in Gondola. In this article, a hypothetical scenario suggests that algorithms will be responsible for handling "non-emotional" purchases so that humans have the time to focus on the purchases that give them feelings (*Retail en 2022*, 2017). This article suggests other trends may also involve the use of artificial intelligence. As an example, some AI technologies could enable direct communication with customers, in order to have an impact on loyalty which will tend to be measured as "experience per square foot" (*Retail en 2022*, 2017). The mentioned trend of smart living also invites retailers to ask themselves the question "what place are we going to take in our customers' homes?" (*Retail en 2022*, 2017). So, there are many interesting opportunities for marketers to undertake, particularly in the retail sector.

However, the existence of these opportunities also implies challenges. On the one hand, there are ethical issues whose concern may vary from one person to another, but on the other hand there are also some regulated by law. In its report "Artificial intelligence, a European perspective", the European Commission highlights the issues of personal data protection (covered by the GDPR), transparency and explainability of algorithms, as well as questions of responsibility of AIs (Craglia & Europäische Gemeinschaften, 2018a, pp. 63–64). A good example of a failure in this field is the intelligent recruitment tool used by Amazon, which had to be stopped because it was biased and the stakeholders did not have sufficient control over its functioning (Pritam Kanti, 2021). This illustrates a profound need for partners to keep an eye on how AI works. The explainability of AI outputs is also "a major lever in ensuring compliance with public expectations and in fostering trust to accelerate adoption" (Burkhardt et al., 2019).

The increasing importance of AI is impacting the way marketing is approached in current business. But marketing itself has also experienced many changes in recent years. First, there is an ever increasing trend towards direct marketing (Sterne, 2017, p. 184,219). From shopping websites, to streaming platforms, or even dating services, personalised recommendations are part of consumers' daily lives. Moreover, the diversity of data available today allows to apply a "customer culture strategy" (Rust et al., 2010, p. 3). Instead of only gathering data on age, gender and previous purchases of consumers, now more complex information is available. In his book on AI applications in marketing, Jim Sterne cites, for example, the various interactions between a customer and the company's website such as clicks or swipes, interactions between a customer and his phone, his real-time location, the weather and other information commonly referred to as "Big Data" (Sterne, 2017, pp. 66–67). Secondly, the emergence of e-commerce has fundamentally impacted marketing. This has brought various challenges, including the fact that in-store visits can be linked to prior website visits (*Gondola Day 2018*, 2018), making website design critical. Finally, key marketing concepts have also evolved. For instance, the concept of loyalty is becoming more and more experience-based and is built through personalization (*Retail en 2022*, 2017).

Problem statement

As just seen, the cross-over between artificial intelligence and marketing produces a wide spectrum of opportunities to address challenges of topical interest. At the beginning of this chapter, it was mentioned that there is a growing interest in AI. However, it is still a subject that may seem difficult to concretise for marketers. This can be explained by the complexity of the topic, which often works as a black box (Craglia & Europäische Gemeinschaften, 2018a, p. 23) and which requires extensive knowledge in computer science. Given this complexity and to meet ethical expectations of the public while also respecting the European legal framework, there is an emerging need to better understand how AI works. The question then arises, from a marketing perspective, whether a better understanding of AI can have an influence on consumers who receive increasingly personalised recommendations produced by AI. This thesis will therefore seek to address the following question:

How the explainability of recommendations made by an Artificial Intelligence (XAI) influence the users' acceptance of AI recommendations?

To answer this general research question, it is first necessary to explore subsidiary questions to fully understand the different concerned topics, such as:

- What exactly constitutes AI?
- How and what to explain about an AI?
- What is a well explainable AI versus a poorly explainable AI?
- Is the explainability of AI linked to concepts of ethics, transparency, and others?

These issues will be addressed in the section reviewing the literature.

Contributions

What managers and marketers may require is a framework to apply an AI solution without having to search for information from scratch. The aim of this thesis is to provide insights into the potential value of explaining artificial intelligence to consumers. The ambition is to provide a better understanding of what these explanations are and how to implement them. Indeed, the use of artificial intelligence in Europe is closely linked to the respect of regulations aimed at ensuring a certain ethical standard (Craglia & Europäische Gemeinschaften, 2018a, Chapter 6). Therefore, the topic of this thesis is expected to be of great interest to managers wishing to implement artificial intelligence solutions.

From an academic point of view, there is a need for explainable AI applications that include evaluations performed by end-users (rather than AI or data experts), and more specifically in the field of retailing, which is not a hot topic in the implementation of XAI. Indeed, the current application domains of XAI are often medical recommendation systems, recruitment tools, self-driving cars, …(Hu et al., 2021, p. 5; Wiegand et al., 2019). In addition, this thesis proposes a novel approach by looking at the acceptance of recommendations.

Approach

In the next part of this thesis, a review of the existing literature on artificial intelligence and explainable artificial intelligence is proposed to better understand the key concepts of this field and to establish the research already conducted as well as their potential limitations. These concepts include artificial intelligence, its origin, and diverse ways of classification, as well as explainable artificial intelligence, its terminology, implementation, and evaluation. At the end of chapter 2, the hypotheses that were revealed by the literature review are described and summarised in a model. Then the chapter 3 describes the methodology used to conduct the research that is designed to evaluate the assumptions. It includes the context of application, the sampling and data collection methods and the measurements employed. The next chapter first describes the results obtained regarding the sample characteristics. Then the selected analysis strategies are presented and applied to give the results of the statistical tests. Finally, the findings are discussed and related to the theoretical concepts previously mentioned. In the last part of this thesis, a conclusion summarises the results and their managerial and academic implications. The limitations of this work are also outlined to make suggestions for future research.

2 Chapter 2: Literature Review

2.1 Artificial Intelligence

2.1.1 Origin and definitions of Artificial Intelligence

There is a fairly straightforward consensus that *artificial intelligence* consists of machines demonstrating intelligence (De Bruyn et al., 2020, p. 92). However, according to Professor Wang, in his contribution to the definition of artificial intelligence, what is complicated in the interpretation of “AI” is not the “A” but the “I” (Wang, 2019, p. 4). Indeed, one of the reasons why there is no widely recognized definition of artificial intelligence is that there are different approaches to the notion of intelligence. A few years before the term artificial intelligence appeared, Alan Turing was already wondering "can machines think?" (Turing, 1950). To answer this question, he offered the now well-known "Turing Test" which suggests that if a respondent cannot distinguish a message written by a human from one written by a machine, the machine can be characterized as intelligent (Haenlein & Kaplan, 2019, p. 3).

The concept of artificial intelligence was introduced in 1955 by John McCarthy and Marvin L. Minsky in a research project proposal for the 1956 Dartmouth Conference (Wang, 2019, p. 7). This innovative topic was motivated by the assumption that "every aspect of learning or any other characteristic of intelligence can in principle be described with such precision that a machine can be made to simulate it"(McCarthy et al., 1955b, p. 12). This illustrates the initial purpose of simulating human cognitive capacities. In the 1960s, this notion of human intelligence was restated by Minsky, for whom AI consisted in “making machines do things that would require intelligence if done by men” (Minsky, 1968 as cited in; Kaplan & Haenlein, 2019, p. 17).

Subsequently, other key concepts come to flesh out the definition of AI, which had been vague until then. An AI should therefore have the ability to adapt to a context (Newel & Simon, 1976) or an environment (Craglia & Europäische Gemeinschaften, 2018b; Wang, 1995), and operate under certain limits of resources, speed, and complexity (Newel & Simon, 1976; Wang, 1995) to achieve some goals (Haenlein & Kaplan, 2019; McCarthy, 1988). These characteristics are indeed shared by some of the definitions from the modern literature listed in Table 2-1.

Artificial Intelligence Definitions	Sources
It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.	McCarthy, 1956 (as cited in Deb et al., 2018)
Artificial intelligence is the science of making machines do things that would require intelligence if done by men.	(Minsky, 1968 as cited in, Kaplan & Haenlein, 2019, p. 17)
By 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action: that in any real situation behaviour appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity.	(Newel & Simon, 1976, p. 116)
AI is concerned with methods of achieving goals in situations in which the information available has a certain complex character. The methods that have to be used are related to the problem presented by the situation and are similar whether the problem solver is human, a Martian, or a computer program.	(McCarthy, 1988 as cited in; Wang, 2019, p. 7)
Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources.	(Wang, 1995 as cited in, 2019, p. 17)
AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action, or proposing decisions.	(Craglia & Europäische Gemeinschaften, 2018b, p. 18)
We define AI as a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.	(Haenlein & Kaplan, 2019, p. 1)
AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity.	(European Parliament, 2020)
Machines that mimic human intelligence in tasks such as learning, planning, and problem-solving through higher-level, autonomous knowledge creation.	(De Bruyn et al., 2020, p. 93)

Table 2-1: List of definitions related to artificial intelligence

As time goes by and technology improves, notions of learning, problem-solving and autonomy increase what is expected from an artificial intelligence (Craglia & Europäische Gemeinschaften, 2018b; De Bruyn et al., 2020; European Parliament, 2020; Haenlein &

Kaplan, 2019). Although new methods such as machine learning or deep learning have deeply impacted the field of application of AI, reducing artificial intelligence to a set of methods would be distancing from the starting point aiming at mimicking human cognitive abilities by machines (Linden, 2021).

2.1.2 Types of Artificial Intelligence

As previously discussed, there are many definitions of AI, due to the interpretation of the notion of intelligence and the numerous techniques used to implement such an intelligence. These techniques can be used to establish a classification of the diverse types of AI. However, once again, the classification of this topic varies from one author to another.

2.1.2.1 Artificial Intelligence classification based on techniques and approaches

The Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence proposes in its annexes a classification in three categories of AI techniques and approaches. These approaches include machine learning methods, logic and knowledge-based approaches, and statistical techniques. First, ML approaches include both supervised, unsupervised, and reinforcement learning which can use different methods such as deep learning, (European Commission, 2021). These concepts related to ML are described in table 2-2. The second category includes various approaches of symbolic artificial intelligence designed to mimic human logic. Among these approaches based on symbolic representations, we find for example the *feature model* or *SoaML* (Service Oriented Architecture Modelling Language). Finally, the third category of AI approaches includes statistical techniques, inspired notably by biology (e.g., foraging algorithms) (Linden, 2021).

Deep learning	Deep Learning is a subset of ML that can cope with noisier data by increasing significantly the number of neural layers and neurons and the amount of data used for training. (Craglia & Europäische Gemeinschaften, 2018b, p. 21)
Supervised learning	Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. (Mahesh, 2019)

	In a supervised learning paradigm, a neural network learns from a set of examples (training data) where both inputs (predictors) and outputs (target variables) are known to the analyst, such as the model learns to minimize a loss function (e.g., entropy). (De Bruyn et al., 2020, p. 94)
Unsupervised learning	Unsupervised learning helps find patterns in data without pre-existing labels. (De Bruyn et al., 2020, p. 95)
Reinforcement learning	Reinforcement learning is another set of algorithms that focus on experience-driven sequential decision-making, i.e. they make software agents take action to maximise some notion of cumulative reward. (Mahesh, 2019)

Table 2-2: Glossary of ML concepts

2.1.2.2 Strong vs. weak AI

The theory of strong AI was introduced by the American philosopher Searle in 1980 in an experiment called the *Chinese room* (Cole, 2020). The experiment starts by placing an English-speaking person who does not speak Chinese in a room with a book and a Chinese person outside the room. The two people exchange messages through the door, the English one follows the instructions of the book which tells him what to answer according to the symbols he receives from the Chinese speaker. Together, they hold a conversation, which proceeds in perfect Chinese. However, at the end of the experiment the English-speaking person still does not understand Chinese.

This aims to demonstrate that an intelligent system (represented by the English speaker) can "give the impression that it understands the language, but does not really understand it" (Cole, 2020). The Chinese room contradicts Alan Turing's definition of an intelligent computer since a machine can pass the Turing Test without being considered as intelligent because the machine is not able to understand the meaning of what it writes (Cole, 2020).

The goal of strong AI is “to develop Artificial Intelligence to the top point where machine’s intellectual capability is functionally equal to a human’s”(*Strong AI*, n.d.). Existing AIs to date fall into the category of weak (or narrow) AIs since they are intended to respond to relatively specific tasks for which they only simulate human capabilities without understanding them (Cole, 2020; Sterne, 2017, p. 71). However, referring back to the basic principle of AI that “any cognitive activity can be learned”, De Bruyn et al. (2020) suggest that machines might, one day, learn to understand.

2.1.2.3 Artificial Narrow, General, and Super Intelligence

One approach that comes close to the strong vs. weak AI view is a categorization of three generations of AI. Indeed, it is common to speak of *Artificial Narrow Intelligence* (ANI) if the scope of application is restricted to specific tasks; this corresponds to the first generation of AI (Kaplan & Haenlein, 2019, p. 16). This type of AI is identical to the weak AI, discussed above, however the term narrow may seem more appropriate as it can consist in " some very robust applications, such as Apple's Siri, Amazon's Alexa, IBM Watson, and autonomous vehicles" (IBM Cloud Education, 2020). As explained previously, even the most powerful AIs currently available fall into this ANI category “in the sense that they operate strictly within the confine of the scenarios for which they are programmed” (Miailhe & Hodes, 2017, p. 9). The second and third generation of AI relate more to strong AI; there is the *Artificial General Intelligence* (AGI) and the *Artificial Super Intelligence* (ASI) (IBM Cloud Education, 2020). The former kind of AI can "solve problems autonomously for tasks they were never even designed for" (Kaplan & Haenlein, 2019, p. 16). The latter type of AI exceeds human cognitive abilities (IBM Cloud Education, 2020). The figure 2-1 below summarizes the three types of artificial intelligence according to Kaplan and Haenlein (2019).

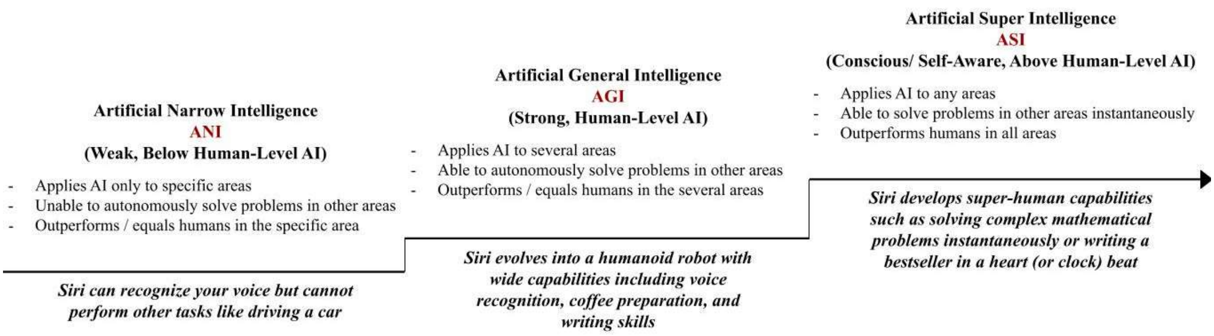


Figure 2-1: Stages of AI from (Kaplan & Haenlein, 2019)

2.1.2.4 AI classification based on skills

In the management literature, artificial intelligence may also be classified according to three skills: *cognitive* intelligence, *emotional* intelligence, and *social* intelligence. The skill of cognitive intelligence is related to "pattern recognition and systematic thinking" (Kaplan & Haenlein, 2019, p. 17). Emotional intelligence, a concept introduced by Peter Salovey and John D. Mayer in the 1990s, refers to "the ability to perceive, understand and manage emotions in the self and others" (Mayer et al., 1997). Social intelligence is, as its name suggests, the ability to interact with the social environment with qualities such as "empathy, team work, and leadership" (Kaplan & Haenlein, 2019, p. 18). According to Boyatzis (2008) findings, the acquisition of such competencies enables a higher efficiency in the professional environment.

Kaplan and Haenlein (2019) suggest to transpose this theory from humans to AI. For this purpose, they distinguish three categories of AI systems that demonstrate characteristics of one or a combination of the above-mentioned skills. First, *analytical* AI only borrows features from cognitive intelligence, which is the case for most artificial intelligences implemented nowadays. Then comes the *human-inspired* AI, which is characterized by cognitive intelligence but also emotional intelligence allowing the recognition of emotions to take decisions. Finally, *humanized* AI, which remains a future project, possesses all three types of skills and would be able to "be self-conscious and self-aware in their interactions with others" (Kaplan & Haenlein, 2019, p. 19). Miao et al. also refer to humanised AI as cognitive and social intelligences that can “establish ties with humans over the long term” (Miao et al., 2018, p. 1).

2.1.2.5 AI as an ecosystem of capabilities

In an article addressing the interrelationships between AI and consumers, Puntoni et al. (2021, p. 2) provide a view of AI as an ecosystem of three fundamental elements associated with four capabilities. The three core elements of this ecosystem are: the data collection and storage, the statistical and computational techniques, and the output system. The authors associate these elements with capabilities such as listening, predicting, producing, and communicating. In the remainder of this article, the authors focus on the experiences that consumers have with these three key elements and their associated capabilities. These experiences are in turn organized into four chronological categories. The *data capture experience* corresponds to the AI's "endowment with personal data." The *classification experience* addresses personalized predictions provided by an AI, which is interesting in the context of this thesis. Then there is the *delegation experience* in which consumers let an AI perform tasks for them. And finally, consumers can have a *social experience* by communicating with an AI.

Each of these experiences can be perceived differently by the user. For example, sometimes the access to data is accepted; in a study proposed by Gogus and Saygin (2019), it was found that students do not actually perceive photos, friends list, and other information shared on social media like Facebook as personal data. But users' data can also be collected and used without consent with considerable impact, as in the case of the Cambridge Analytica scandal, a company accused of stealing personal data in the context of the Trump campaign in 2016 (Audureau, 2018).

2.2 Explainable Artificial Intelligence

2.2.1 Why are explanations needed?

As already discussed, artificial intelligence has greatly improved over the last few years, in particular thanks to new techniques and approaches such as machine learning (Gunning & Aha, 2019, p. 45). This has also raised questions about the ethical aspect of these progresses. Indeed, ML models are becoming more and more a part of the daily life of people, but these models, despite being very efficient, have their own disadvantages (Mohseni et al., 2020, p. 3). These include the risk of discrimination and unfair decisions, as AIs have the unfortunate propensity to replicate human biases and exploit deceptive correlations (De Bruyn et al., 2020, p. 99; Mohseni et al., 2020, p. 3). There are also concerns about the accountability of decisions provided by AIs, as well as the transparency of processes (Haenlein & Kaplan, 2019, p. 7). What brings these various risks together is the problem of opacity (Gunning & Aha, 2019, p. 45). Machine learning models work as black boxes, where we have information about the inputs and outputs but not about what happens in between. These complex models perform well and provide accurate results, however there is a trade-off between interpretability performance and these models become less and less interpretable as their complexity[1] increases (Mohseni et al., 2020, p. 9). According to Arrieta et al. (2019), the inability to explain how ML algorithms work also relates to the gap between researchers and the business sector. Since ML models are increasingly being used for decision making, “there is an emerging need for understanding how such decisions are furnished by AI methods” (Arrieta et al., 2019, p. 2). Furthermore, in Europe, the General Data Protection Regulation has established a legal right to an explanation for a decision taken involving the evaluation of certain personal data (Hoffman et al., 2018, p. 2; Mohseni et al., 2020, p. 2; Radley-Gardner et al., 2016).

---

[1] By making its structure more complex, a model can become more performant.

2.2.2 XAI terminology

To address the previously mentioned issues, a solution could be the auditing of algorithms. It consists of "a mechanism for investigating algorithms' functionality to detect bias and other unwanted algorithm behaviours without the need to know about its specific design details" (Mohseni et al., 2020, p. 3). This is an interesting process for error detection, but it is far from solving the problem of understanding models considered as black box.

A new field, known as *explainable artificial intelligence* (XAI), trying to address these concerns of transparency, accountability, fair decision making, and other challenges towards a more ethical AI, is now emerging (De Bruyn et al., 2020, p. 99). In May 2017, the Defense Advanced Research Projects Agency (DARPA) launched its XAI program in the United States in collaboration with universities across the country. The definition of an explainable AI, according to the DARPA, is the following: "AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future" (Gunning & Aha, 2019, p. 44). The figure 2-2 from the DARPA, represents their view of XAI concept.

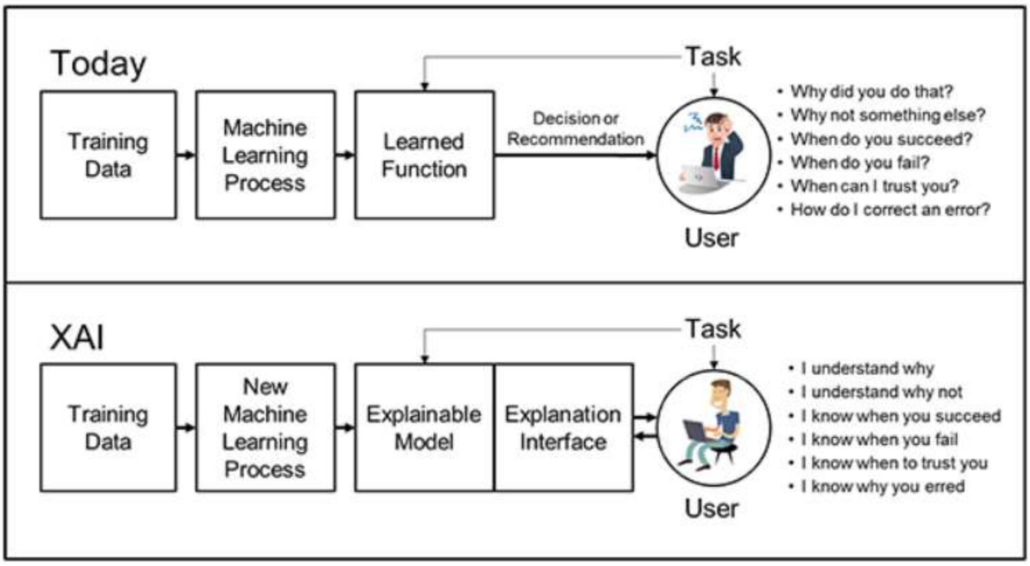


Figure 2-2: XAI Concept (Turek, 2016)

However, as in the case of artificial intelligence, the definitions vary from one author to another, particularly regarding the beneficiaries of these explanations. In the opinion of the DARPA (2019, p. 45), the target of XAI are end-users impacted by the recommendations produced by an AI. De Bruyn et al. (2020, p. 99) rather consider that the target of explanations are human experts. Arrieta et al. suggest that the definition of XAI should go further and include the notion of audience: “given an audience, an explainable Artificial Intelligence is one that produces details or reasons to make its functioning clear or easy to understand” (Arrieta et al., 2019, p. 6). Hoffman et al. (2018, p. 3) also support this view, arguing that explanations depend mainly on what users need to know, their current knowledge and their goals.

Under this logic, the goals of explainable AI design are linked to the audience. In this thesis, the focus is on XAI from a user-oriented point of view. Therefore, the objectives related to this audience are identified here.

Arrieta et al. (2019, p. 8) identify five main objectives for users of a model affected by its decisions: trustworthiness, accessibility, fairness, interactivity and privacy awareness. The authors define trustworthiness as the certainty that a model will function as expected when confronted with a given problem and they specify that it is a property which is not easy to quantify. The accessibility and interactivity objectives are quite evident, as explanations of how an AI works allow users to become more involved in the process of ML model design. With regard to fairness, the authors argue that explainability should be seen as "a bridge to avoid unfair or unethical use of the algorithm's outputs" by controlling the relations that affect the results of the algorithm (Arrieta et al., 2019, p. 9). Finally, understanding how a model works, even a complex one, allows a certain control to be maintained so that confidentiality and privacy can be respected (Arrieta et al., 2019, p. 10).

Mohseni et al. (2020), who draw a distinction between AI novices, data experts and AI experts, identifies as goals for AI novices: algorithm transparency, trust and reliance, error mitigation and privacy awareness. Similarities between the two views are observable, with trust and privacy forming recurrent elements of XAI goals. The authors further note that transparency constitute an "immediate goal to help the end-user to understand how the intelligent system works" and the importance of insight into the details of the process to detect and remove bias (Mohseni et al., 2020, p. 14). In section 2.2.3.2, we will see how the goals of XAI need to be measured in order to assess the quality of the explanations, and that the design goals and evaluation methods are strongly interdependent (Mohseni et al., 2020, p. 23).

Regarding the other parts of the definition of XAI, the several authors agree that it involves various techniques and approaches to improve the explainability of AI models, in terms that are understandable to humans, while maintaining a high level of performance (Arrieta et al., 2019; De Bruyn et al., 2020; Gunning & Aha, 2019). A distinction remains to be made between an XAI and an XAI system, which consists of an intelligent, self-explanatory system that will meet the requirements of an XAI (Mohseni et al., 2020, p. 2).

Before looking in the next section at the various approaches to XAI, it is worth noting the elements that need to be explained and the challenges involved. De Bruyn et al. (2020, p. 99) describe three main areas which need to be explained: the intention behind the use of AI, the data collected for this purpose and how the inputs and outputs of the AI are connected. Mohseni et al. (2020) identify six types of explanations answering the question "what to explain?". These include:

- The "*how*" explanations that attempt to explain how the AI model works.
- The “*why*" explanations that communicate the logic behind the use of inputs to make recommendations.
- The "*why not*" explanations address the potential divergence between an output and the one expected by the user.

- The "*what if*" explanations are about how manipulations in the inputs can impact the outputs.
- The "*how to*" explanations are intended to enable the user to adjust the inputs to achieve a targeted output.
- And finally, the "*what else*" explanations are a kind of "explaining by example" to show how similar inputs can generate the same kind of outputs.

Reach satisfying type of explanation requires: the production of more explainable models, the design and testing of user-friendly explanation interfaces, and an understanding of what is required to provide a quality explanation that meets the expectations of users (Gunning & Aha, 2019, p. 45). These different challenges, identified by Gunning and Aha (2019) in the context of the DARPA XAI program, have the characteristic of being user-oriented.

Finally, to conclude this section on XAI terminology, it is of paramount to review the concepts that are closely related to XAI. Arrieta et al. (2019) define very clearly these concepts listed in the table 2-3.

Understandability = Intelligibility	Characteristic of a model to make a human understand its function (how it works, not internal design explanation).
Comprehensibility	Ability of a learning algorithm to represent its learned knowledge in a human understandable fashion.
Interpretability	Ability to explain or to provide the meaning in understandable terms to a human.
Explainability	Notion of explanation as an interface between humans and a decision maker.
Transparency	A model is transparent if by itself is understandable.

Table 2-3: Concepts related to XAI (Arrieta et al., 2019, p. 5)

According to Arrieta et al. (2019), the most closely related concept is that of understandability, which is linked to transparency. Indeed, the transparency of a model consists in being understandable by itself, but the understandability will measure the degree to which a human understands the model and so in a way it helps to assess the quality of the explanation.

The authors insist on the difference between, on the one hand, interpretability and transparency, which are rather passive characteristics of a model, and, on the other hand, explainability, which is a much more active characteristic, since it involves the techniques deployed to improve the understanding of the model's functions (Arrieta et al., 2019, p. 4-5). In terms of ethics, explainability is actually part of the properties of an ethical AI as well as fairness, robustness, transparency and privacy.(IBM, 2021b).

2.2.3 How to make AI explainable?

2.2.3.1 How to Explain?

Several classifications of XAI implementation techniques exist (IBM, 2021; Mohseni et al., 2020; Towards Data Science, 2021).

2.2.3.1.1 Explanations for transparent and black-box models

The classification of XAI approaches proposed by Arrieta et al. (2019) in their article on XAI concepts, opportunities and challenges is based on the distinction between *transparent models* and *post-hoc explainability models*. The former are interpretable by design, while the latter require techniques to clarify the models (Arrieta et al., 2019, p. 10). This can be compared to solving problems of understanding either transparent-box or black-box models (Guidotti et al., 2019). In the table 2-4 is a listing of the different ML models divided into transparent models and those that require post-hoc explainability.

Transparent ML models	Non transparent ML models
Linear/ Logistic regression	Trees Ensembles
Decision Trees	Support Vector Machines
K-Nearest Neighbours	Multi-layer Neural Network
Rule Based Learners	Convolutional Neural Network
General Additive Models	Recurrent Neural Network
Bayesian Models	

Table 2-4 : Classification of ML models (Arrieta et al., 2019)

Transparent models

Models considered to be transparent may have a varying degree of interpretability and are classified in terms of the “domain in which they are interpretable”(Arrieta et al., 2019, p. 10). This includes the *simulatability*, *decomposability,* and *transparency of algorithm* and each of these areas contains its predecessors. Indeed, a simulatable model, i.e., one that can be designed by a human, is both decomposable and algorithmically transparent. First, it is important to note that by simulatable the authors mean a model that may be complex but whose data processing is humanly manageable. Then a model is considered decomposable under the authors' criterion if each of its components, namely the inputs, the parameters, and the calculation, can be understood and explained by a human. Finally, algorithmic transparency refers to models in which the mathematical process of mapping inputs and outputs is understandable to the user.

Post-hoc explainability

Concerning models which require an "interpretability layer" (Towards Data Science, 2021) to be understood by end-users, Arrieta et al. (2019) identify six explainability techniques detailed below.

- *Text explanation*

The results and functioning of a ML model can be detailed by means of textual explanations. These can take the form of properly constructed sentences but also of semantic mappings including symbols (see figure 2-3).

- *Visual explanation*

Visual explanations are intended to explain visually the behaviour of the model by means of one or more techniques such as dimensionality reduction. However, by offering simple and human-friendly visualisations, the explanations may need to be reinforced with additional information, for example through written material.

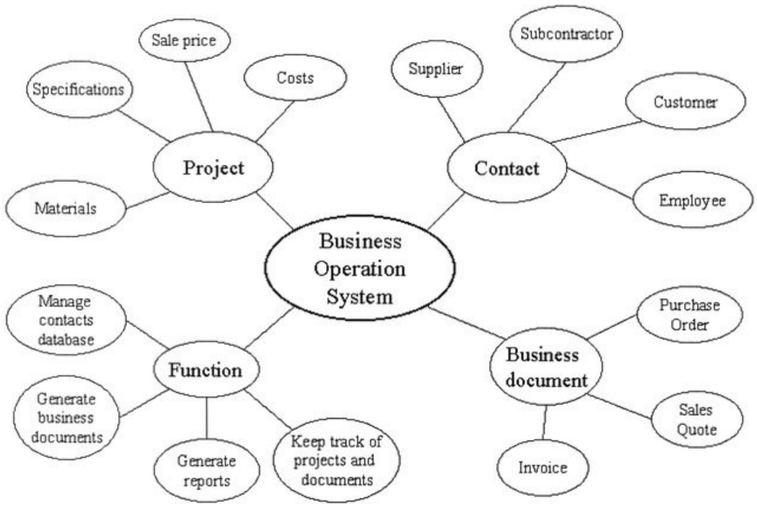


Figure 2-3: Semantic map of the business operation system from (Siau & Tan, 2005)

- *Local explanation*

These explanations are close to the domain of transparent decomposable models. Indeed, the focus is on the explanation of the model by segmentation, to deal with less complex processes.

- *Explanation by example*

This type of explanation consists of extracting examples of outcomes to explain how the model works. However, these type of explanations "only make sense if we can represent an instance of the data in a humanly understandable way […]. This works for images […]. It is more challenging to represent tabular data in a meaningful way" (Molnar, 2021).

- *Explanation by simplification*

This category of explanation covers the different methods by which a new, simpler model with similar performance can be created. The aim is to explain a less complex model, but as seen above, there is a trade-off between interpretability and performance, therefore attention must be paid to keeping an acceptable level of performance.

- *Feature Relevance Explanation*

Feature relevance methods, considered as indirect explainability techniques by the authors, establish the contribution of variables to the generation of an outcome.

2.2.3.1.2 Explanation format and reference

Another classification of explanations that comes up regularly is the distinction between the *format* of the explanation ("how it is expressed") and the *reference* of the explanation ("what it is about") (Mueller et al., 2019, p. 84).

The format of explanation

A basic differentiation can be made between two major types of explanation format, textual explanations and visual explanations (Hu et al., 2021, p. 4).

According to Hu et al. (2021, p. 4), “people usually give each other explanations verbally”. According to Kouki et al. (2019), in their survey involving an online music recommender, textual explanations are perceived as more persuasive than visual explanations. In addition, Hohman et al. (2019) note that textual explanations may be easier to understand than visualisations, especially if the textual explanations are short. However, according to Hu et al. (2021), visual explanations provide also a meaningful insight into the AI's reasoning. By testing the explanations of their credit scoring model Demajo et al. (2021) found that the majority of experts preferred a visual explanation (in the form of a decision tree) rather than a textual (rule-based) explanation. This raises the question of whether this result is specific to an expert audience or whether it would be transferrable to an end-user audience. In their study applied to the explanation of music recommendations, Millecamp et al. (2019) raise that one point to consider with visual explanation is the "visualisation literacy" which is characterised by the ability to interpret information presented in a visual format (image, graph,...). The visualisation literacy could indeed be linked to the audience type of the explanations.

The reference of explanation

The reference, i.e., the focus of the explanation, can relate to examples, features ranking, strategies and goals, cause-effect relationships, ... Actually, as the reference of an explanation corresponds to "what to explain", it can be linked to the six types of explanations (how/ why/

why not/ what if/ how to/ what else) outlined by Mohseni et al. (2020, pp. 9–10). For instance, explanations with an example-based reference correspond to "what else" explanations since "*What-else* explanations pick samples from the model's training dataset that are similar to the original input in the model representation space" (Mohseni et al., 2020, p. 10).

2.2.3.1.3 Explanation style for recommender system

A final classification that is very interesting when it comes to explaining recommendations made by an AI to a user is the one based on the "style" of explanation (Papadimitriou et al., 2012). In their article on the taxonomy of explanation styles, Papadimitriou et al. (2012) propose three individual styles of explanation and many 'hybrid' styles. However, by reviewing the literature three main styles of recommendation explanations can be distinguished There are content-based explanations, those based on collaborative filtering and finally there are also "hybrid" explanations (Naveed et al., 2020).

Content-based explanations, also called "item-based" explanations, consist in explaining how a recommended item is relevant according to its characteristics directly related to the user's profile (Naveed et al., 2020, p. 2). *Collaborative filtering* is used to explain a recommendation based on collaborative information gathered by multiple similar users, through explicit or implicit feedback (Naveed et al., 2020; Pfaff, 2021, p. 3). This style of explanation is more user-oriented than content-oriented. Finally, a *hybrid* explanation involves both styles of explanation.

In a study evaluating user perceptions of explanations of a music recommendation system, Kouki et al. (2019) identified that users preferred recommendations based on content rather than on other users' preferences. But Herlocker et al. (2000), in an experimental test of movie recommendations, identified that explanations based on feedback from other users were perceived as useful. Another study also identified that users who receive explanations based on similar consumers (versus content-based explanations) perceive additional information which makes the recommendation more accurate (Gai & Klesse, 2019). It remains particularly important to point out that the style of explanation must be distinguished from the recommendation process itself.

To briefly summarise, Mohseni et al. (2020) and Mueller et al. (2019) share the same view by making the distinction between "what to explain" and "how to explain". Arrieta et al. (2019) do not make this distinction but focus on the explainability techniques applied to transparent models and black box models, which are different. Finally, a classification of explanation style can help to differentiate between user-oriented and content-oriented explanations.

2.2.3.2 Measures for evaluation

One of the challenges of XAI research and development is the strong need for common measures to assess the validity of explanations (Arrieta et al., 2019; Hoffman et al., 2018; Mohseni et al., 2020). Explainable AI is a truly interdisciplinary field. It involves data science, engineering, human-computer relations, and psychology, and its applications extend into many domains (Mohseni et al., 2020). Therefore, although there is no clear consensus, some proposals for evaluation measures from different domains can be found in the literature.

As previously mentioned, the goals of XAI are linked to the audience. Similarly, the evaluation measures verify the validity of the explanations in relation to their purpose (Mohseni et al., 2020, p. 16). Mohseni et al. (2020) suggest the evaluation of the model understanding, the user perceived -usefulness, -satisfaction and -trust, and the performance of the human-AI task. In terms of comprehension of the model, the authors use what is called in psychology the *mental model*. This is basically a "representation of how users understand a system" (Mohseni et al., 2020, p. 16). In their paper, the authors do not provide a way to assess these measures properly but refer the reader to questionnaires and interviews used in practical cases of the literature.

An other paper, entirely focused on XAI metrics, is proposed by Hoffman et al. (2018) In this paper, the authors propose to evaluate the goodness of the explanation, the user's satisfaction, the user's mental model, the influence of curiosity on the search for an explanation, the user's trust in the AI and the performance of the human-XAI system.

A checklist for assessing the *goodness* of explanations is provided in the appendix of their paper with close-ended questions (i.e. the answer can only be yes or no) such as: "the explanation helps me **understand** how the [software, algorithm, tool] works?", “the explanation of the [software, algorithm, tool] sufficiently **detailed?**”, or “the explanation is **actionable**, that is, it helps me know how to use the [software, algorithm, tool]?” (Hoffman et al., 2018, p. 35). The authors argue that this evaluation should be conducted a priori to the deployment of the explanations, while satisfaction should be measured a posteriori.

Once again, to assess satisfaction, the authors submit a questionnaire in the appendix, but this time the answers are on a Likert scale. As previously suggested, the authors evaluate the understanding of the system with a mental model.

Twelve methods are identified to obtain such a mental model, for instance by asking a user to solve a problem aloud, or to sort cards according to the semantic similarities between the concepts of the system. The different methods are taken from the literature (from the field of psychology, according to experts in knowledge acquisition) and are associated with reference articles.

Concerning curiosity, it is the subject of a short checklist, reproduced in the table 2-5, to understand what the users' real expectations for explanations are. The aim is to respond to the need in the best conceivable way and to avoid overwhelming someone who is not curious with information.

Finally, the performance of the XAI system is evaluated as a function of the benefits perceived by the user, whether it concerns satisfaction, perceived trust, or other qualities such as the completeness of the explanations. In practice, as performance can vary from one goal to another and from one user to another, the authors do not propose an evaluation questionnaire. Instead, they suggest that "the most powerful and direct way of evaluating the performance of a work system that includes an XAI is to evaluate how easy or difficult it is to get prospective users (stakeholders) to adopt the XAI system" (Hoffman et al., 2018, p. 23).

Why have you asked for an explanation? Check all that apply.	
	I want to know what the AI just did.
	I want to know that I understand this AI system correctly.
	I want to understand what the AI will do next.
	I want to know why the AI did not make some other decision.
	I want to know what the AI would have done if something had been different.
	I was surprised by the AI’s actions and want to know what I missed.

Table 2-5: Curiosity checklist (Hoffman et al., 2018)

2.2.4 Benefits and practical applications

2.2.4.1 Benefits

The benefits of making AI explainable are numerous. First of all, from an AI design point of view, it is very valuable to have a complete understanding of how the system works in order to detect possible errors, to ensure that the system only uses meaningful variables, and to improve robustness by identifying potential perturbations (Arrieta et al., 2019; Mohseni et al., 2020). Performance may not be so incompatible with the interpretability of the system (Arrieta et al., 2019, p. 2). Secondly, from the end-user's perspective, it meets an urgent need for comprehension and trust (Shin, 2021). It also ensures that the decisions made by an AI are fair, impartial and secure (Arrieta et al., 2019; Mohseni et al., 2020). Finally, it allows to comply with the right to an explanation as regulated by the GDPR in Europe.

2.2.4.2 Practical applications

Although the subject is quite recent, there are already many applications of AI explanations. For example, Coppers et al. (2018) carried out a study to evaluate the effects of a translation aid that provides details of how suggestions are made, particularly in relation in the context of the text. The study conducted with professional translators showed that this additional information did not necessarily improve the user experience but that translators preferred to have explanations when it improved the quality of the translation.

On another topic, Rader et al. (2018) conducted a study on the influence that explanations on how Facebook news feed algorithms work can have on user awareness, accuracy of recommendations, and accountability. They found that user awareness was significantly impacted but that the explanations were not sufficient to help users assess the correctness of the algorithm's output. In addition, the explanations create a negative feeling among users towards Facebook with the feeling that the feed algorithm is unfair.

The measurement of justice perception is also proposed by Binns et al. (2018), in an experiment with different types of explanations for different types of scenarios (applying for financial loan, promotion at work, car insurance,...), with the aim of assessing the perceived justice of the algorithms. The results of their research show that in general people are overly concerned about the fairness of decisions made by both humans and machines, but that the automatic nature of intelligent system, even with additional explanation, reinforces the feeling of unfairness.

Holliday et al. (2016) conducted a quantitative and qualitative study to assess user trust in a qualitative data coding support system. They found that when explanations were provided, trust in the system first increased and then returned to its basic state, whereas without an explanation, trust in the system only decreased.

A large number of other papers on the evaluation of XAI measures can be found , Mohseni et al. draw up a non-exhaustive list in their paper (Mohseni et al., 2020, p. 13). However, articles interested in applications in the retail sector are quite rare, the recurrent themes are rather: loan rates (Binns et al., 2018), self-driving cars (Wiegand et al., 2019) and the health sector (Tjoa & Guan, 2020).

2.3 Hypotheses and model development

This literature review has highlighted several needs for explanations of AI models. As previously noted, these needs are related to the audience receiving the explanations, which has its own objectives in terms of understanding AI models. This section has also detailed several ways of assessing the quality and impact of an explanation. This thesis focuses on end users, novices in AI, receiving explanations. It was discussed in the introduction that recent developments in marketing and artificial intelligence have led to the emergence of many highly personalised recommendations towards consumers (Sterne, 2017).

The need for consumers to trust AI models was outlined by several authors (Arrieta et al., 2019; Burkhardt et al., 2019; Holliday et al., 2016; Mohseni et al., 2020). Therefore, a first assumption is made that trust in an AI recommendation increases the user's acceptance of that AI recommendation.

H1: Trust in AI recommendation improves the user’s acceptance of AI recommendation.

Secondly, reviewing the literature has shown that trust in a AI recommendation can be improved by explaining how the AI works (Burkhardt et al., 2019; Holliday et al., 2016; Shin, 2021). This information leads to a second hypothesis that an explanation that is perceived as of superior quality by the user increases the user's trust in the AI recommendation.

H2: An explanation perceived as of superior quality improves user’s trust in AI recommendation.

Finally, the users' perceived quality of the explanation can be assessed using a scale proposed by Hoffman et al. (2018) based on several criteria such as understanding, satisfaction, usefulness, level of detail of the explanation. Two formats (visual and textual) and styles (content-oriented and user-oriented) of explanations were proposed in this section. Previous studies allow to suggest that these differences it the way of explaining a model may have an impact on the users' perception of explanation quality. Indeed, explanations presented in a textual format are easier to understand (Hohman et al., 2019; Kouki et al., 2019). While visual explanations, in order to be easily understood, are no longer complete enough (Arrieta et al., 2019). Regarding the "style" of the explanation, Kouki et al. (2019) identified that users were more satisfied with content-oriented explanations. Regarding explanations based on similar users, they provide the perception that the recommendation is more accurate, which is one of the criteria for evaluating a good explanation (Gai & Klesse, 2019; Hoffman et al., 2018). Therefore, the following assumptions are made.

H3A: Textual explanations improve the user’s perceived understanding of the recommendation.

H3B: Visual explanations decrease the user’s perception of sufficient detail in the explanation.

H3C: Content-oriented explanations increase the user's perceived satisfaction of the explanation.

H3D: User-oriented explanations improve the user's perception that the recommendation is accurate.

Figure 2-4 summarises the model, assessed in this thesis, of the influence of explanations of IA recommendation on users' acceptance of that recommendation.

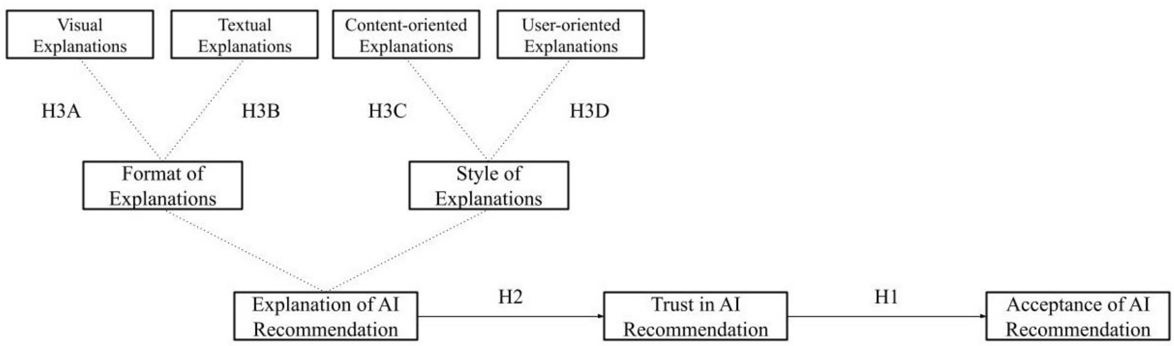


Figure 2-4: Explainable Artificial Intelligence - conceptual model

3 Chapter 3: Research Design

To test the hypotheses described in section 2.3, causal research is conducted. Indeed, this type of research focuses on assessing the cause and effect relationships between variables (Malhorta et al., 2014, p. 66). The research is punctual, and not longitudinal, because the interest of this work is in the influence of explanations on the acceptance of recommendations at a given time and not in the factors that could change this acceptance over time. For this purpose, an online questionnaire is used to collect quantitative data.

3.1 Scenario and context

This research takes place in the context of recommendation systems using artificial intelligence to generate personalised recommendations on online shopping websites. More specifically, the focus is on supermarket websites. These companies collect some data on consumer behaviour, both online and in stored through loyalty cards, which allow them to offer highly personalised recommendations. This research is conducted in this specific context because, as previously noted, there is a lack of experimentation in this area.

The causal research is based on individual evaluations of a given scenario. Participants in this experiment are asked to imagine themselves in a situation and are given a product recommendation with a plausible explanation for this recommendation. As explained in the model of the previous section, four types of explanations are developed:

- a visual content-oriented explanation,
- a textual content-oriented explanation,
- a visual user-oriented explanation, and
- a textual user-oriented explanation.

The visual content-oriented explanation consists of a horizontal bar chart showing the main personal characteristics of the consumer that explain why the product is recommended. The visual user-oriented explanation represents a five-star rating given by consumers with a similar profile to the one receiving the recommendation. Textual explanations describe, in the form of a short text with key elements highlighted, why the product is recommended. In the case of the content-oriented explanation, the information is based on the respondent's own characteristics. Whereas for the user-oriented explanation, the text illustrates the behaviour of a similar consumer.

These explanations were developed based on data collected by the Belgian shop Delhaize from their consumers (Delhaize Le Lion SCA, 2020), as well as existing interfaces from the literature (Binns et al., 2018, p. 5; Herlocker et al., 2000, p. 6; Ramon, Vermeire, Toubia, et al., 2021, p. 12). The interfaces to these four explanations are available in appendix A.

It is important to note that a true recommendation system based on respondents' characteristics is not used because the data are not available as for supermarket companies. Therefore, all this research is based on the idea that participants should imagine themselves in a scenario where the information would be personalised.

Each scenario corresponds to a type of explanation and each respondent assesses only one scenario, but the questionnaire is the same for all four scenarios. The choice was made not to expose each respondent to the four explanations, as this would make the questionnaire long to fill in (about 15-20 minutes), so it would risk increasing the number of drop-outs from the questionnaire and decreasing the quality of the answers (less concentration after a certain time) (Malhorta, 2006, p. 31; Žmuk, 2017, p. 51).

3.2 Scales and variables

The questionnaire is built in three parts (see appendix B for full questionnaire). The 25 questions are listed following a funnel sequence (Steils, 2018). First, a series of socio-demographic questions are asked to obtain more information about the respondents (i.e., their age, gender, education level, and employment status). Then, after receiving a description of the scenario, participants answer a series of questions in relation to the product recommendation and its associated explanation. Finally, a multiple-choice question aims to establish the consumer's expectations in terms of a recommendation's explanation in general. For this purpose, the curiosity checklist proposed by Hoffman et al (2018), already mentioned, is used.

The second part or the questionnaire was designed by adapting existing multi-item scales from the literature. To evaluate the model developed in this thesis, questions are asked about three constructs: the perceived quality of explanations, the trust in recommendations, and the acceptance of these recommendations. The details of these constructs and their items are shown in table 3-1. Each of the questions (one per item) are evaluated by the participants on a 5-points Likert scale indicating their degree of agreement or disagreement. Besides the fact that the authors providing these questions use a Likert scale format for the responses, Likert scales have the benefit to be easily understood by the respondents (Malhorta et al., 2014, p. 214).

CONSTRUCT	QUESTION	CODE	SOURCE
Explanation Perceived Quality	From the explanation, I understand how the recommender system works.	EPQ1	(Hoffman et al., 2018, p. 39)
	This explanation of how the recommender system (AI) works is satisfying.	EPQ2	
	This explanation of how the recommender system (AI) works has sufficient detail.	EPQ3	
	This explanation of how the recommender system (AI) works seems complete.	EPQ4	
	This explanation of how the recommender system (AI) works is useful to my goals.	EPQ5	
	This explanation of the recommender system shows me how accurate the recommender is.	EPQ6	
	This explanation lets me judge when I should trust and not trust the recommender system (AI).	EPQ7	
Recommendation Perceived Trust	I am confident in the recommender system (AI). I feel that it works well.	RPT1	(Hoffman et al., 2018, p. 49)
	The outputs of the recommender system (AI) are very predictable.	RPT2	
	The recommender system (AI) is very reliable. I can count on it to be correct all the time.	RPT3	
	I feel safe that when I rely on the recommender system (AI) I will get the right recommendations.	RPT4	
	I am wary of the recommender system (AI).	RPT5	
	The recommender system (AI) can perform the task better than a novice human user.	RPT6	
	I like using the recommender system (AI) for decision making.	RPT7	
Recommendation Acceptance	I intend to use the recommendation system (AI) in the future.	RA1	(Sohn & Kwon, 2020, p. 7)
	I intend to use the recommendation system (AI) frequently.	RA2	
	I intend to recommend that other people use the recommendation system (AI).	RA3	
	I am willing to use this recommendation system (AI) as an aid with my decision about which product to buy.	RA4	(Komiak & Benbasat, 2006, p. 11)
	I am willing to let this recommendation system (AI) assist me in deciding which product to buy.	RA5	
	I am willing to use this recommendation system (AI) as a tool that suggests to me a number of products from which I can choose.	RA6	

Table 3-1:Items and their source

3.3 Participants and distribution

The target population is broad as it is any consumer likely to receive a product recommendation online. The study aims to assess the model in Belgium and the scenario interfaces are inspired by what could be found on the website of a Belgian shop, Delhaize. Therefore, the main selection criterion for this survey is to live in Belgium. Another restriction is that respondents must be at least 18 years old, which is in any case a criterion for joining the panel website used in this survey. Regarding the respondent's level of education and employment status, there is no restriction, to have responses from various profiles.

The survey method is an online questionnaire. This choice is due to its ease of design, the speed with which responses are obtained, but also because it is perfectly adapted to the context of recommendations given online. The questionnaire is designed on Sphinx Declic. A pre-test is conducted with a small sample to assess understanding of the scenarios and questions. Following some minor modifications, the questionnaire is distributed on Prolific, a data collection service.

The sampling technique is stratified. Among all respondents in the Prolific database, adults who reside in Belgium are selected. This represents about 900 participants out of the 227,000 available on Prolific. Then, a male-female parity is selected to improve representativeness. Finally, the participants are selected randomly from each of the two stratifications (men living in Belgium and women living in Belgium).

There are four studies, each evaluating a different scenario, 90 respondents per study, 360 in total. The determination of sample size is complex task. It is a question of having enough respondents to test hypotheses while limiting the cost of using a paid panel. Since this study is being conducted as part of a master thesis, the impact of the study's findings is limited. In the context of a company wishing to evaluate the same hypotheses for future managerial decisions, it would be preferable to increase the sample size.

3.4 Analysis method

Data collection and cleaning

As soon as 90 people have completed the questionnaire, the study on Prolific stops. Responses can be accepted or rejected. For example, when the questionnaire is completed in a few seconds instead of 3 minutes (the average time), the answer is deleted. A questionnaire response that includes missing values would also be deleted, however this is not possible due to the mandatory nature of each question. Once the 90 responses are accepted, the survey is closed, and the respondents are paid. This process is repeated for each of the four surveys.

After this data cleaning, the responses are extracted from Sphinx and analysed on the SPSS software. The data are first coded. For example, for the Likert scale, each answer corresponds to a code: from 1 for "I disagree strongly" to 5 for "I agree strongly". The purpose of this coding is mainly to facilitate data processing.

Once the data has been coded, a descriptive analysis of the sample and the 'curiosity' (expectations of explanations) is performed, and then the hypotheses are evaluated following the selected analysis strategy described below.

Analysis strategy

The selection of analysis strategies depends on "the characteristics of the data as well as the properties of the statistical techniques" (Malhorta et al., 2014, p. 351).

Firstly, since the variables involved in the hypotheses H1 and H2 are some constructs assessed by a series of items, a factor analysis must be performed to reduce these different items into a single variable. The steps of this analysis are described in section 4.2.

In hypotheses H1 and H2, the relationships between an independent variable and a dependent variable are investigated. Indeed, the intention is to assess whether trust in a recommendation increases acceptance of that recommendation (H1), and whether the perceived quality of a recommendation explanation increases trust in that recommendation (H2).

Therefore, the use of bivariate statistical techniques is appropriate. Furthermore, given that the independent variable is unique, a simple regression can be performed.

Hypotheses H3A H3B H3C and H3D, each assess a dependant variable[2] respectively: understanding of the recommendation system, level of detail of the explanation, satisfaction of the explanation and accuracy of the recommendation system. For each of these variables, the intention is to find out whether their mean varies when the explanation of the recommendation changes. To do so, the sample is divided into two independent samples for each hypothesis, comparing either visual and textual explanations, or content and user-oriented explanations[3]. To evaluate these hypotheses, it is therefore appropriate to employ univariate statistical techniques. Since the dependent variable is metric (interval) and the independent variable non-metric (nominal), the choice is to use a test of variance comparison (Levene's test) and then a test of means comparison (Anova if the variances are equal and Welch if the variances are unequal).

In summary, hypotheses 1 and 2 are used to assess the relationships between "perceived quality of explanation" and "recommendation perceived trust", and between "recommendation perceived trust" and "recommendation acceptance". These relationships are tested by means of simple linear regressions. Hypotheses 3 are used to assess whether the mean of certain items measuring the quality of an explanation vary according to the style or format of the explanation. This is achieved by means of a comparison of means test, which requires a comparison of variances test beforehand.

---

[2] Each of these variables are items that constitute the construct "quality of an explanation". In hypotheses H3A,B,C and D these items are assessed separately.

[3] The format/style of the explanation is the independent variable.

4 Chapter 4: Results and Discussion

4.1 Descriptive analysis

4.1.1 Description of the sample

In this section, socio-demographic data are analysed for the aggregate sample. However, the four separate samples (one per scenario) are also analysed to identify variations.

The aggregate sample size is 360 elements, respecting the requirement of being resident in Belgium and at least 18 years old. Of this sample, 48.49% are women, 50.83% are men and 0.28% are of other gender (see appendix C, figure 1).

Regarding the age of the participants in the overall sample, 43.6% are between 18 and 24 years old, 32.8% between 25 and 34 years old, 17.2% between 35 and 44 years old, 4.7% between 45 and 54 years old and 1.7% are over 54 years old. This age distribution varies slightly in the samples from one scenario to another (see appendix C, figures 2 and 3).

Most participants have at least a secondary school education. Indeed, in the overall sample, the highest level of education of the participants corresponds for 1.1% to primary school, for 25.6% to secondary school, for 38.1% to a bachelor's degree, for 32.2% to a master's degree and for 3.1% to a PhD. Again, this distribution varies a little from the sample of one scenario to another (see appendix C, figures 4 and 5).

Finally, concerning the employment status of the participants, most of them are employees (46.94%) or students (36.94%), but there are also 4.44% self-employed, 5.56% workers, 5.56% unemployed, and less than 1% retired or disabled (see appendix C figure 6).

4.1.2 Curiosity analysis

Another remarkably interesting point, although it was not part of the hypotheses tested, is the question of consumer curiosity. The literature review showed that the evaluation of an explanation is closely linked to the expectations of the person receiving it (Hoffman et al., 2018). Respondents were therefore asked about their expectations in the form of a multiple-choice question. The top three or expectations are: knowing what the AI just did (in 71.4% of responses), making sure to understand the system well (in 53.9% of responses) and knowing why another decision was not made (in 40% of responses). Then slightly more than a third of respondents want to know what the AI would have recommended if things had been different and what it will recommend next. In only 15% of cases, participants say they are surprised by the AI's action and want to know what they missed (see details in figure 4-1).

However, it should be noted that this question was mandatory. Therefore, we can ask the question if the participants answered by being constrained or if they are really interested in receiving information about artificial intelligence.

Curiosity Frequencies

		Responses		Percent of Cases
		N	Percent	
Curiosity[a]	I want to know what the AI just did	257	28,2%	71,4%
	I want to know that I understand this AI system correctly	194	21,3%	53,9%
	I want to understand what the AI will do next	128	14,1%	35,6%
	I want to know why the AI did not make some other decision	144	15,8%	40,0%
	I want to know what the AI would have done if something had been different	134	14,7%	37,2%
	I was surprised by the AI's actions and want to know what I missed	53	5,8%	14,7%
Total		910	100,0%	252,8%

a. Dichotomy group tabulated at value 1.

Figure 4-1: Curiosity frequencies

4.2 Factor analysis and scale reliability

In this section, a factor analysis is conducted for the following three constructs: the perceived quality of explanations (EPQ), the perceived trust in recommendations (RPT), and the acceptance of these recommendations (RA). Indeed, "factor analysis is a method of data summarisation, which reduces a large number of items into a smaller number of variables that measure the same dimension", thus allowing the computation of a single score per construct (Steils, 2018, p. 92). This confirmatory[4] factor analysis is carried out in several steps.

First, the sample size must be at least four to five times larger than the number of items. This condition is met because there are a maximum of 7 items per construct and the sample size is 360 (Steils, 2018, p. 93).

Secondly, it is necessary to verify the relevance of the factor analysis, i.e., to check that the items are highly correlated with each other. To do this, two tests can be used: Bartlett's sphericity test and the Kaiser-Meyer-Olkin (KMO) test. A Bartlett test explores the matrix of correlations between items, a significant test ($p\text{-value} < 0.05$) rejects the null hypothesis that the items are not correlated. The KMO index varies between 0 and 1 and increases the higher the inter-item relations, a value above 0.8 is considered very good (Steils, 2018, p. 94).

Thirdly, following a principal component analysis, factors with an eigenvalue greater than 1 and which explain at least 50% of the initial information are selected. In the case of items with a communality of less than 0.5, they are removed, and the analysis must be rerun.

Then, the reliability (internal consistency) of the scale is measured with Cronbach's alpha, for which a value greater than 0.6 is recommended. The closer the alpha value is to 1, the more homogeneous the set of items is.

---

[4] The factors linking the items are known a priori.

Finally, based on the mean of the items retained in the factor analysis and in the reliability evaluation, a new single variable is created.

EPQ – Explanation perceived quality

To assess this variable, 7 items are considered. The KMO and Bartlett tests are performed with SPSS. The KMO index has a value greater than 0.8 and the Bartlett test is significant (details in figure 4-2). The items are therefore significantly correlated. Then, in figure 4-3, we can see that all the communalities are greater than 0.5, thus the 7 items are retained. Then, only one factor is retained, and this factor explains more than 50% (64.828%) of the variance (see appendix D). Finally for the analysis of the scale's internal consistency, Cronbach's alpha value is 0.909 (see appendix D), which means that the score of the new unique variable "EPQ" can be calculated from the mean of the 7 items.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,885
Bartlett's Test of Sphericity	Approx. Chi-Square	1564,468
	df	21
	Sig.	,000

Figure 4-2: KMO and Bartlett's test for EPQ

Communalities

	Initial	Extraction
EPQ1 Explanation understanding	1,000	,607
EPQ2 Explanation satisfaction	1,000	,724
EPQ3 Explanation detailed	1,000	,757
EPQ4 Explanation completeness	1,000	,725
EPQ5 Explanation usefulness	1,000	,614
EPQ6 Explanatuon accuracy	1,000	,609
EPQ7 Explanation trustworthiness	1,000	,502

Extraction Method: Principal Component Analysis.

Figure 4-3: Communalities for EPQ

RPT – Recommendation perceived trust

To assess this variable, 7 items are also considered. The KMO index has a value of 0.855 and the Bartlett test is significant because the p-value is <0.001. The items are therefore significantly correlated. However, not all communalities are above 0.50 (see figure 4-4). These items are therefore removed, and the analysis is rerun. The new KMO is above 0.8 and the Bartlett test is significant (see figure 4-5). This time, the communalities are all above 0.5 (see figure 4-6). Only one factor is retained, and this factor explains more than 50% (70.655%) of the variance (see appendix E). The Cronbach's alpha value is 0.859 (see appendix E). This means that the score of the new unique variable "RPT" can be computed from the mean of the 7 items.

Communalities

	Initial	Extraction
RPT1 Recommendation confidence	1,000	,713
RPT2 Recommendation predictability	1,000	,398
RPT3 Recommendation reliability	1,000	,650
RPT4 Right recommendation	1,000	,740
RPT6 Recommendation perfomance	1,000	,243
RPT7 Recommendation satisfaction	1,000	,571
RPT5_Corr Recommendation wariness	1,000	,208

Extraction Method: Principal Component Analysis.

Figure 4-4: Communalities for RPT

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,802
Bartlett's Test of Sphericity	Approx. Chi-Square	687,116
	df	6
	Sig.	<,001

Figure 4-5: New KMO and Bartlett's test for RPT

Communalities

	Initial	Extraction
RPT1 Recommendation confidence	1,000	,742
RPT3 Recommendation reliability	1,000	,708
RPT4 Right recommendation	1,000	,796
RPT7 Recommendation satisfaction	1,000	,581

Extraction Method: Principal Component Analysis.

Figure 4-6: New Communalities for RPT

RA – Recommendation acceptance

To assess this variable, 6 items are considered. The KMO index has a value greater than 0.8 and the Bartlett test is significant (details in figure 4-7). All communalities are above 0.5

(see figure 4-8). Only one factor is retained, and this factor explains more than 50% (71.505%) of the variance (see appendix F). The Cronbach's alpha value is 0.920 (see appendix F), which means that the score of the new unique variable "RA" can be calculated from the mean of the 6 items.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,874
Bartlett's Test of Sphericity	Approx. Chi-Square	1606,528
	df	15
	Sig.	,000

Figure 4-7: KMO and Bartlett's test for RA

Communalities

	Initial	Extraction
RA1 I intend to use the recommendation system (AI) in the future	1,000	,784
RA2 I intend to use the recommendation system (AI) frequently	1,000	,740
RA3 I intend to recommend that other people use the recommendation system (AI)	1,000	,603
RA4 I am willing to use this recommendation system (AI) as an aid with my decision about which product to buy	1,000	,790
RA5 I am willing to let this recommendation system (AI) assist me in deciding which product to buy	1,000	,718
RA6 I am willing to use this recommendation system (AI) as a tool that suggests to me a number of products from which I can choose	1,000	,655

Extraction Method: Principal Component Analysis.

Figure 4-8: Communalities for RA

4.3 Hypotheses testing

4.3.1 Hypothesis 1: Trust in AI recommendation improves the user's acceptance of AI recommendation

To evaluate this first hypothesis, a simple regression is used. Four conditions must be met: linearity, homoscedasticity, independence of the error terms and normal distribution of the error terms (Steils, 2018).

To assess *linearity*, a scatterplot is produced with the newly computed variables RPT and RA. The scatter plot shows that as the trust increases the acceptance of recommendation appears to increase, so we can assume that there is linearity (see appendix G, figure 1). *Homoscedasticity* is observed on the scatterplot between the standardised predicted values and the standardised residuals. No pattern (e.g., triangle) seems to emerge, for all values of RA, the variance of the residuals is therefore assumed to be homogeneous (see appendix G, figure 2). The *independence of the error terms* can be verified by the Durbin-Watson test. The independence of the error term is assumed when the test statistic, which can vary between 0 and 4, is close to 2, which is the case here as it is 1.933 (see appendix G, figure 3). The *normality of the distribution of the error term* is also validated as looking at the Gaussian plot of the residuals, we can see that most of the points are close to the diagonal (see appendix G, figure 4).

Since the conditions are met, we can look at the results of the simple regression. The model summary table gives the value of the Pearson correlation coefficient. The coefficient is significant because the p-value is smaller than 0.5, so the null hypothesis that there is no linear relationship between the two variables is rejected. A consumer's acceptance of recommendations is significantly correlated with his or her trust in the recommendation. Person's coefficient is positive, meaning that the two variables tend to increase together and decrease together. Moreover, the absolute value of the coefficient is close to 1 (more than 0), so the linear relationship between RA and RPT is strong (details in figure 4-9).

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,743[a]	,552	,551	,63580	,552	441,476	1	358	<,001	1,933

a. Predictors: (Constant), RPT

b. Dependent Variable: RA

Figure 4-9: Simple Regression - Model Summary

In the analysis of variance table (Anova), the significance level is less than 0.5 (details in appendix G, figure 5). It can be concluded that the independent variable (RPT) explains the dependent variable (RA) and it is possible to predict the value of recommendation acceptance based on the value of trust in the recommendation. However, in this case predicting an exact value is of no great interest since the variables are initially evaluated based on several items and on a Likert scale. What is interesting is to understand the relationship between these two variables and the related managerial implications.

4.3.2 Hypothesis 2: An explanation perceived as of good quality improves user's trust in AI recommendation

For this second hypothesis, the same steps of simple linear regression are followed. This time, the purpose is to analyse the relationship between the independent variable "perceived quality of the explanation" (EPQ) and the dependent variable " trust in the recommendation " (RPT).

To assess *linearity*, a new scatterplot is produced with EPQ and RPT. The scatter plot shows that EPQ and RPT increase together, so we can assume that there is linearity (see appendix H, figure 1). No pattern seems to emerge from the scatterplot between the standardised predicted values and the standardised residuals, *homoscedasticity* is therefore verified (see appendix H, figure 2). The *independence of the error terms* is supported by a Durbin-Watson value of 2.041 (see appendix H, figure 3). The *normality of the distribution of* *the error term* is also validated as looking at the Gaussian plot of the residuals, we can see that most of the points are close to the diagonal (see appendix H, figure 4).

The Person coefficient is analysed. It is significant (p value < 0.05), and its value is high (see figure 4-10). The relationship between RPT and EPQ is therefore positively correlated and quite strong.

Model Summary[b]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	,708[a]	,501	,500	,64450	,501	359,808	1	358	<,001	2,041

a. Predictors: (Constant), EPQ
b. Dependent Variable: RPT

Figure 4-10: Simple Regression - Model Summary

The analysis of variance, which is significant because the p-value is less than 0.001 (and therefore <0.05), leads to the conclusion that the independent variable "perceived quality of the explanation" predicts the dependent variable " trust in the recommendation" (details in appendix H, figure 5).

4.3.3 Hypotheses 3

To assess the four hypotheses 3, the overall sample is split into two independent samples. For hypotheses 3A and 3B, one sample consists of respondents who evaluated a visual explanation, and the other consists of respondents who evaluated a textual explanation. For hypotheses 3C and 3D, one sample consists of respondents who evaluated a content-oriented explanation, and the other consists of respondents who evaluated a user-oriented explanation.
To establish whether there is a significant difference in the mean of the dependant variable from one sample to the other, a comparison of means is performed.

This is conducted in two steps. **Step 1**: the equality of variances must be verified using a Levene's test. The hypotheses evaluated are the following:

H0: There is no difference in the variance of the dependent variable between the two samples.
H1: There is a significant difference in the variance of the dependent variable between the two samples.

Step 2: A means comparison test is performed. The hypotheses evaluated are the following:

H0: There is no difference between the means of the two samples. If a difference is observed, it is random.
H1: There is a significant difference between the means of the two samples (one is higher/lower than the other).

4.3.3.1 Hypothesis 3A: Textual explanations improve the user's perceived understanding of the recommendation

To assess whether the understanding of the recommendation varies from one explanation format to another (textual vs. visual), the sample is divided in two (180 respondents per sample). Levene's test is conducted and of significance as it has a p-value<0.05, thus the variances are significantly different (see details in appendix I). Therefore an Anova test for comparison of means cannot be used, instead the Welch's F test must be applied (Datanovia, 2018; Steils, 2018). The null hypothesis and alternative hypothesis of this means comparison test remain the same as for an Anova test. This test is significant (details in figure 4-11), leading to the conclusion that the average understanding varies between the two samples. In Figure 4-12, we can see that the mean understanding of the recommendation is higher for textual explanations.

Robust Tests of Equality of Means

Recommendation understanding

	Statistic[a]	df1	df2	Sig.
Welch	11,889	1	316,086	<,001
Brown-Forsythe	11,889	1	316,086	<,001

a. Asymptotically F distributed.

Figure 4-11: Welch Test - Explanation understanding

Descriptive Statistics

Format of explanation		N	Minimum	Maximum	Mean	Std. Deviation
Visual	Recommendation understanding	180	1	5	3,74	1,375
	Valid N (listwise)	180				
Textual	Recommendation understanding	180	1	5	4,17	,939
	Valid N (listwise)	180				

Figure 4-12: Mean score of explanation understanding

4.3.3.2 Hypothesis 3B: Visual explanations decrease the user's perception of sufficient detail in the explanation

To assess whether the explanation's level of detail perceived varies from one explanation format to another (textual vs. visual), the sample is divided in two. The Levene's test is conducted and of significance as it has a p-value<0.05, the variances are therefore not homogeneous (see details in appendix I). The Welch's F test is conducted and of significance (details in figure 4-13). It can therefore be concluded that the average level of detail perceived by users varies according to the format of the explanation. In Figure 4-14, we can see that the mean is higher for textual explanations.

Robust Tests of Equality of Means

Explanation detailed

	Statistic[a]	df1	df2	Sig.
Welch	27,427	1	340,344	<,001
Brown-Forsythe	27,427	1	340,344	<,001

a. Asymptotically F distributed.

Figure 4-13: Welch Test - Details of the explanation

Descriptive Statistics

Format of explanation		N	Minimum	Maximum	Mean	Std. Deviation
Visual	Explanation detailed	180	1	5	2,99	1,416
	Valid N (listwise)	180				
Textual	Explanation detailed	180	1	5	3,70	1,123
	Valid N (listwise)	180				

Figure 4-14: Mean score of explanation details

4.3.3.3 Hypothesis 3C: Content-oriented explanations increase the user's perceived satisfaction of the explanation

To assess whether the perceived satisfaction of the explanation varies from one explanation style to another (content-oriented vs user-oriented), the sample is divided in two. Levene's test is conducted and of significance as it has a p-value<0.05, the variances are therefore not homogeneous (see details in appendix I). The Welch's F test is conducted and of significance (details in figure 4-15). We can conclude that the perceived satisfaction of the explanation varies according to the style of explanation. In Figure 4-16, we can see that the average satisfaction of explanation is higher for content-oriented explanations.

Robust Tests of Equality of Means

Explanation satisfaction

	Statistic[a]	df1	df2	Sig.
Welch	54,761	1	345,067	<,001
Brown-Forsythe	54,761	1	345,067	<,001

a. Asymptotically F distributed.

Figure 4-15: Welch Test - Explanation satisfaction

Descriptive Statistics

Style of explanation		N	Minimum	Maximum	Mean	Std. Deviation
Content-oriented	Explanation satisfaction	180	1	5	3,94	1,023
	Valid N (listwise)	180				
User-oriented	Explanation satisfaction	180	1	5	3,06	1,245
	Valid N (listwise)	180				

Figure 4-16: Mean score of explanation satisfaction

4.3.3.4 Hypothesis 4D: User-oriented explanations improve the user's perception that the recommendation is accurate

To assess whether the perceived accuracy of the recommendation changes from one explanation style to another, the sample is divided in two. Levene's test is performed and of significance as it has a p-value<0.05, the variances are therefore significantly different (see details in appendix I). The Welch's F test is conducted and of significance (details in figure 4-17). We can conclude that the recommendation accuracy perceived by users varies according to the style of the explanation. In Figure 4-18, we can see that the mean is higher for content-oriented explanations.

Robust Tests of Equality of Means

Recommendation accuracy

	Statistic[a]	df1	df2	Sig.
Welch	61,575	1	348,383	<,001
Brown-Forsythe	61,575	1	348,383	<,001

a. Asymptotically F distributed.

Figure 4-17: Welch Test - Explanation accuracy

Descriptive Statistics

Style of explanation		N	Minimum	Maximum	Mean	Std. Deviation
Content-oriented	Recommendation accuracy	180	1	5	3,62	1,037
	Valid N (listwise)	180				
User-oriented	Recommendation accuracy	180	1	5	2,68	1,226
	Valid N (listwise)	180				

Figure 4-18: Mean score of explanation accuracy

4.4 Discussion

In this section, the results of this study are discussed in relation to the research question and assumptions made. As a reminder, the research question investigates the influence of AI-generated recommendation explanations on consumers' acceptance of these recommendations. This question, based on a literature review, was decomposed into a model suggesting:

1. that consumers who trust a recommendation will be more likely to accept it,
2. that this trust can be increased by a good quality explanation,
3. and that the quality of the explanation will vary according to its format and style.

First, hypotheses H1 and H2 demonstrated a relationship between trust and acceptance of a recommendation, as well as a relationship between a good-quality explanation and trust in the explained recommendation. In concrete terms, this implies that explanations can be used as a tool to leverage users' trust in recommendations and thus enable greater acceptance of such recommendations in consumers' daily lives. The findings of these hypotheses therefore confirm the concepts found in the existing literature. For example, it was noted that trust and the existence of an explanation were "keys to promoting technology acceptance" (Shin, 2021). The research conducted in this thesis confirms that acceptance of a recommendation increases when the recommendation is perceived as trustworthy. Holliday et al (2016) identified that, in their study, participants who received explanations trusted the intelligent system more. The hypothesis tests of H2 also concluded that trust in a recommendation was strongly correlated with receiving a good explanation of how the recommendation works. However, it is also necessary to be aware that an inadequate explanation, which does not satisfy consumers and is not perceived as trustworthy, may make the recommendations no longer acceptable.

Secondly, the research focused on the impact of a change in the style or format of the explanation on the items that can be used to establish the quality of the explanation (Hoffman et al., 2018).

The first three hypotheses, H3A, H3B, and H3C were confirmed. Indeed, the textual explanations showed better results in the users' understanding of the recommendation. This first hypothesis seems to be linked to the second, which revealed that visual explanations were perceived as less detailed than the textual ones. These two hypotheses therefore support the idea identified in the literature that visual explanations are perceived as incomplete and that textual explanations allow for a better level of understanding (Arrieta et al., 2019; Kouki et al., 2019). The 3C hypothesis was also confirmed. Content-oriented explanations satisfy more than user-oriented explanations. Although the means are quite close (3.94 and 3.06), the variances are quite different. Content-oriented explanations received positive evaluations while user-oriented explanations received more distributed evaluations (see appendix J). It can be concluded from this assumption that, in general, it is preferable to explain a recommendation based on content (product characteristics) rather than based on a match with consumers with similar profiles.

The only assumption that did not give the expected result was the H3D. Indeed, the hypothesis tests showed that there was a difference in the mean of the perceived accuracy of the recommendations, but contrary to what was expected, it is the content-based recommendations that are perceived as more accurate. If we look in more detail at the frequencies of recommendation accuracy, but this time at the scenario level, we observe that the user-based visual explanation was rated as particularly inaccurate (see appendix K). This may be related to the lack of detail. The amount of detail should not be too high so as not to make the explanation too complex, but it should be high enough to prove that the recommendation is not due to " randomness ". Looking at the literature, it has been stated that consumers "prefer a medium rather than a low level of complexity" (Ramon, Vermeire, Martens, et al., 2021, p. 10). This could explain why the user-based visual explanation, which is quite simple, has been very negatively evaluated in a global way.

5 Chapter 5: Conclusion

Artificial intelligences allow to generate highly personalised product recommendations, which is especially useful for the current direct marketing trends. However, these recommendations are generated by models acting as black boxes and requiring the collection of a vast amount of personal data. To gain insight into the functioning of intelligent systems and to meet ethical needs, explainable artificial intelligences (XAI) are emerging.

A review of the existing literature identified the methods of implementing such a solution by detailing what can be explained by an AI and how it can be explained according to the audience and its needs. This information led to the development of five hypotheses for addressing the research question: “how the explainability of recommendations made by an Artificial Intelligence (XAI) influence the users’ acceptance of AI recommendations?”. It was assumed that a recommendation that is perceived as trustworthy is more likely to be accepted by consumers, and that this trust can be increased by explaining how the recommendation system works. Several types of explanation were presented and a classification was suggested with two formats (visual or textual) and two styles (content-based or user-based) explanations. It was noted that these formats and styles of explanation could have an impact on the items used to measure the perceived quality of the explanation.

To assess these hypotheses, quantitative research was conducted. Four scenarios were created, one for each pair of different format and style of explanation. For each of the four scenarios evaluated, participants of this survey received a product recommendation with an explanation of how this recommendation is generated. They were then asked to answer a series of questions, the results were collected and analysed.

First, the analyses showed that the consumer’s trust in a recommendation and the acceptance of that recommendation were strongly correlated. The validation of this hypothesis reinforces the experimental framework of the literature, as the study was conducted in a different area (retail sector in Belgium) than those tested in general for XAIs.

Secondly, the relationship between the trustworthiness of a recommendation and the perceived quality of that recommendation was also supported. This implies that XAI, with an appropriate explanation, can really be a lever to generate recommendation acceptance. XAI is therefore a tool that managers should integrate when they are planning recommendations involving artificial intelligence.

Then, the analysis of the results showed that textual explanations were better understood and seen as more detailed than visual explanations. The motivations of this thesis, from a managerial point of view, was to provide insights for the development of explainable artificial intelligence solutions. The results of this hypothesis allow to advise managers to prefer textual explanations when it comes to explaining the functioning of an intelligent recommendation system.

Finally, the last hypotheses also revealed that consumers are more satisfied with content-oriented explanations and see them as more accurate than explanations involving the behaviour of similar consumers. Managers should then provide explanations based on the content of the recommendation rather than explanations based on implicit or explicit feedback from similar consumers. This type of explanation, found on many shopping websites in the form of "consumers rated this product X stars" or "consumers who bought this item also bought that one", should be replaced by explanations that describe how the product's features match the user's needs and thus justify the recommendation.

5.1 Contributions

From a theoretical research point of view, this thesis had the ambition to answer the need for explainable artificial intelligence evaluations by end-users, as the literature was mainly built on assessments made by experts and in recurrent domains (medicine, self-driving cars, recruitment tools, ...). This thesis therefore provides an experiment in a new domain, explanations of product recommendations to consumers. In addition, this thesis simultaneously investigates the implication of the trust generated by explanations on recommendation acceptance and the impact of explanation style and format on the perceived quality of the explanation, this new approach also contributes to XAI research.

From a managerial perspective, this thesis contributes to a better understanding of what explainable artificial intelligence is and how it can be implemented in a concrete case. The research part allows managers to understand the challenge of developing explanations in line with consumer expectations. Furthermore, the analysis of the results shows that XAI can improve trust and acceptance of recommendations. This thesis is therefore evidence that such explanations should be part of the managers' strategy when developing product recommendations that use artificial intelligence.

5.2 Limitations and future research

Although great care has been put into the methodology, this research is not without its limitations.

Firstly, the study conducted for this thesis was published on a panel data collection website, four different studies were conducted rather than a long one which would have resulted in dropouts. However, it is possible that the same person may have responded to more than one questionnaire. Therefore, the use of independent sample tests may be questioned. In addition, the order in which participants completed the surveys may have had an impact on their perception of the explanations.

Secondly, as seen in the literature, explanations must be adapted to the needs of the person receiving them. The results showed that the explanations were not at an equivalent level of detail and a complexity variable should have been included in the model. It would therefore be interesting to reconduct this type of research but with a preliminary phase to determine the requirements of the recommendations and the expected level of complexity.

Thirdly, Holliday et al. (2016) identified that explanations increase users' trust in intelligent systems for a while and then the trust level returns to its initial level (without explanations, trust only decreases). This study was conducted on a one-time basis, but it would be interesting to investigate longitudinally how trust and acceptance of recommendations with explanations change over time.

Finally, this research cannot be seen as sufficient to develop explainable artificial intelligence from a managerial point of view. These are the first steps, but other issues need to be addressed. For example, the feasibility of such explanations and the way to integrate them into a website must be considered: the explanation can be visible for each recommendation, but it could also be accessible only on "demand". Furthermore, this study focuses on the acceptance of recommendations, but if we look at the acceptance models of technologies, we can see that other concepts are involved, such as consumer behaviour (Sohn & Kwon, 2020). Whether these explanations lead consumers to buy the recommended products is an interesting approach for future research.

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *ArXiv:1910.10045 [Cs]*. http://arxiv.org/abs/1910.10045

Artificial intelligence in EU enterprises. (2021, April 13). https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210413-1

Audureau, W. (2018, March 22). Ce qu'il faut savoir sur Cambridge Analytica, la société au cœur du scandale Facebook. *Le Monde.fr*. https://www.lemonde.fr/pixels/article/2018/03/22/ce-qu-il-faut-savoir-sur-cambridge-analytica-la-societe-au-c-ur-du-scandale-facebook_5274804_4408996.html

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173951

Boyatzis, R. E. (2008). Competencies in the 21st century. *Journal of Management Development*, *27*(1), 5–12. https://doi.org/10.1108/02621710810840730

Burkhardt, R., Hohn, N., & Wigley, C. (2019, May 2). Leading your organization to responsible AI. *McKinsey*. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/leading-your-organization-to-responsible-ai

Cole, D. (2020). The Chinese Room Argument. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/chinese-room/

Coppers, S., Van den Bergh, J., Luyten, K., Coninx, K., van der Lek-Ciudin, I., Vanallemeersch, T., & Vandeghinste, V. (2018). Intellingo: An Intelligible Translation Environment. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3174098

Craglia, M., & Europäische Gemeinschaften (Eds.). (2018a). *Artificial intelligence: A european perspective*. Publications Office of the European Union. https://doi.org/10.2760/936974

Craglia, M. & Europäische Gemeinschaften (Eds.). (2018b). *Artificial intelligence: A european perspective*. Publications Office of the European Union. https://doi.org/10.2760/936974

Datanovia. (2018). Test T de Welch. *Datanovia*. https://www.datanovia.com/en/fr/lessons/types-de-test-t/test-t-non-apparie/test-t-de-welch/

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K.-U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, *51*, 91–105. https://doi.org/10.1016/j.intmar.2020.04.007

Deb, S., Jain, R., & Deb, V. (2018). *Artificial Intelligence —Creating Automated Insights for Customer Relationship Management* (p. 764). https://doi.org/10.1109/CONFLUENCE.2018.8442900

Delhaize Le Lion SCA. (2020, September 21). *Déclaration de vie privée*. https://www.delhaize.be/fr-be/CP-gdprprivacy

Demajo, L. M., Vella, V., & Dingli, A. (2021). An Explanation Framework for Interpretable Credit Scoring. *International Journal of Artificial Intelligence & Applications*, *12*(1), 19–38. https://doi.org/10.5121/ijaia.2021.12102

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Parliament. (2020, September 4). *What is artificial intelligence and how is it used? | News | European Parliament*. https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used

Gai, P. J., & Klesse, A.-K. (2019). Making Recommendations More Effective Through Framings: Impacts of User- Versus Item-Based Framings on Recommendation Click-Throughs. *Journal of Marketing*, *83*(6), 61–75. https://doi.org/10.1177/0022242919873901

Gogus, A., & Saygın, Y. (2019). Privacy perception and information technology utilization of high school students. *Heliyon*, *5*(5), e01614. https://doi.org/10.1016/j.heliyon.2019.e01614

Gondola Day 2018: Intelligence artificielle, e-commerce et opportunités liées. (2018, June 7). https://www.gondola.be/fr/society/news/gondola-day-2018-intelligence-artificielle-e-commerce-et-opportunites-liees

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, 15.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*, 000812561986492. https://doi.org/10.1177/0008125619864925

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work - CSCW '00*, 241–250. https://doi.org/10.1145/358916.358995

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for Explainable AI: Challenges and Prospects*. 50.

Hohman, F., Srinivasan, A., & Drucker, S. M. (2019). *TELEGAM: Combining Visualization and Verbalization for Interpretable Machine Learning*. 5.

Holliday, D., Wilson, S., & Stumpf, S. (2016). User Trust in Intelligent Systems: A Journey Over Time. *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 164–168. https://doi.org/10.1145/2856767.2856811

Hu, Z. F., Kuflik, T., Mocanu, I. G., Najafian, S., & Shulner Tal, A. (2021). Recent Studies of XAI - Review. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 421–431. https://doi.org/10.1145/3450614.3463354

IBM. (2021a, March 26). *Explainable AI*. https://www.ibm.com/watson/explainable-ai

IBM. (2021b, October 1). *AI Ethics*. https://www.ibm.com/artificial-intelligence/ethics

IBM Cloud Education. (2020, June 3). *What is Artificial Intelligence (AI)?* IBM. https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Komiak & Benbasat. (2006). The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly*, *30*(4), 941. https://doi.org/10.2307/25148760

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., & Getoor, L. (2019). Personalized explanations for hybrid recommender systems. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 379–390. https://doi.org/10.1145/3301275.3302306

Linden, I. (2021, May 5). *Course of Artificial Intelligence and Decision Support System*.

Mahesh, B. (2019). *Machine Learning Algorithms -A Review*. https://doi.org/10.21275/ART20203995

Malhorta, N. (2006). *Marketing Research—An Applied Orientation* (6th ed.). Pearson Education Australia.

Malhorta, N., Décaudin, J.-M., Bouguerra, A., & Bories, D. (2014). *Etudes marketing* (6ème édition). Pearson.

Mayer, J. D., Salovey, P., Caruso, D. R., & Cherkasskiy, L. (1997). *CHAPTER 26 Emotional Intelligence*.

McCarthy, J. (1988). Mathematical Logic in Artificial Intelligence. *Daedalus*, *117*(1), 297–311.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955a). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–12. https://doi.org/10.1609/aimag.v27i4.1904

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955b). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–12. https://doi.org/10.1609/aimag.v27i4.1904

Miailhe, N., & Hodes, C. (2017). *The Third Age of Artificial Intelligence*. 7.

Miao, C., Zeng, Z., Wu, Q., Yu, H., & Leung, C. (2018). *Humanized Artificial Intelligence: What, Why and How*. *24*(2), 21.

Millecamp, M., Htun, N. N., Conati, C., & Verbert, K. (2019). To explain or not to explain: The effects of personal characteristics when explaining music recommendations. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 397–407. https://doi.org/10.1145/3301275.3302313

Minsky, M. (editor). (1968). Semantic Information Processing. *Viii, 438 p. : Ill. Cambridge, Mass.: The MIT Press, 1982, [C1968]. Includes Bibliography.--- 3rd. Printing*. http://papers.cumincad.org/cgi-bin/works/BrowseAZname=authors/Show?0c01

Mohseni, S., Zarei, N., & Ragan, E. D. (2020). A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ArXiv:1811.11839 [Cs]*. http://arxiv.org/abs/1811.11839

Molnar, C. (2021). *Chapter 7 Example-Based Explanations | Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/example-based.html

Mueller, S. T., Hoffman, R. R., Clancey, W., & Emrey, A. (2019). *Explanation in Human-AI Systems: A Literature Meta-Review Synopsis of Key Ideas and Publications and Bibliography for Explainable AI*. 204.

Naveed, S., Loepp, B., & Ziegler, J. (2020). On the Use of Feature-based Collaborative Explanations: An Empirical Comparison of Explanation Styles. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 226–232. https://doi.org/10.1145/3386392.3399303

Newel, A., & Simon, H. A. (1976). *Completer Science as Empirical Inquiry: Symbols and Search*. *19*(3), 14.

Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2012). A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, *24*(3), 555–583. https://doi.org/10.1007/s10618-011-0215-0

Pfaff, H. (2021). *A Survey on Current Recommender Systems*. 14.

Pritam Kanti, P. (2021, October 25). Driving AI Adoption With Explainability At The Core. *Forbes*. https://www.forbes.com/sites/forbesbusinesscouncil/2021/10/25/driving-ai-adoption-with-explainability-at-the-core/

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. *Journal of Marketing*, *85*(1), 131–151. https://doi.org/10.1177/0022242920953847

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3173677

Radley-Gardner, O., Beale, H., & Zimmermann, R. (Eds.). (2016). *Fundamental Texts On European Private Law*. Hart Publishing. https://doi.org/10.5040/9781782258674

Ramon, Y., Vermeire, T., Martens, D., Evgeniou, T., & Toubia, O. (2021). How Should Artificial Intelligence Explain Itself? Understanding Preferences for Explanations Generated by XAI Algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3877426

Ramon, Y., Vermeire, T., Toubia, O., Martens, D., & Evgeniou, T. (2021). Understanding Consumer Preferences for Explanations Generated by XAI Algorithms. *ArXiv:2107.02624 [Cs]*. http://arxiv.org/abs/2107.02624

Retail en 2022: Les neuf tendances clés. (2017, August 14). https://www.gondola.be/fr/news/retail-en-2022-les-neuf-tendances-cles

Rust, R. T., Moorman, C., & Bhalla, G. (2010). Rethinking marketing. *Harvard Business Review*, *88*(1/2), 94–101.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Siau, K., & Tan, X. (2005). Technical Communication in Information Systems Development: The Use of Cognitive Mapping. *IEEE Transactions on Professional Communication*, *48*(3), 269–284. https://doi.org/10.1109/TPC.2005.853933

Sohn, K., & Kwon, O. (2020). Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. *Telematics and Informatics*, *47*, 101324. https://doi.org/10.1016/j.tele.2019.101324

Steils, N. (2018). *Etudes de marché Partie 2—Approche Quantitative*.

Sterne, J. (2017). *Artificial intelligence for marketing: Practical applications*. Wiley.

Strong AI. (n.d.). Retrieved 22 September 2021, from https://www.ocf.berkeley.edu/~arihuang/academic/research/strongai3.html

Tjoa, E., & Guan, C. (2020). A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. https://doi.org/10.1109/TNNLS.2020.3027314

Towards Data Science. (2021, September 26). *Explainable Artificial Intelligence*. Medium. https://towardsdatascience.com/explainable-artificial-intelligence-14944563cc79

Turek, M. (2016). *Explainable Artificial Intelligence*. Defense Advanced Research Projects Agency. https://www.darpa.mil/program/explainable-artificial-intelligence

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Wang, P. (1995). *Non-Axiomatic Reasoning System – Exploring the Essence of Intelligence*.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Wiegand, G., Schmidmaier, M., Weber, T., Liu, Y., & Hussmann, H. (2019). I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312817

Žmuk, B. (2017). Impact of questionnaire length and complexity on survey time: Comparison of two business web questionnaire versions. *Metodoloski Zvezki*, *14*, 49–74.

Appendices

Appendix A – The four explanatory scenarios

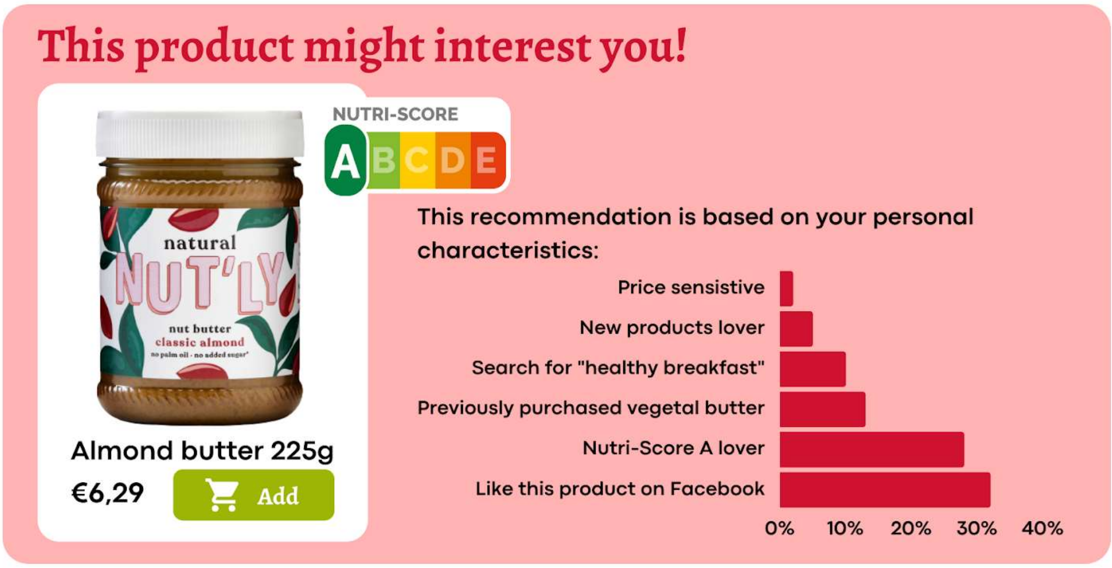


Figure 1: Visual content-oriented explanation scenario

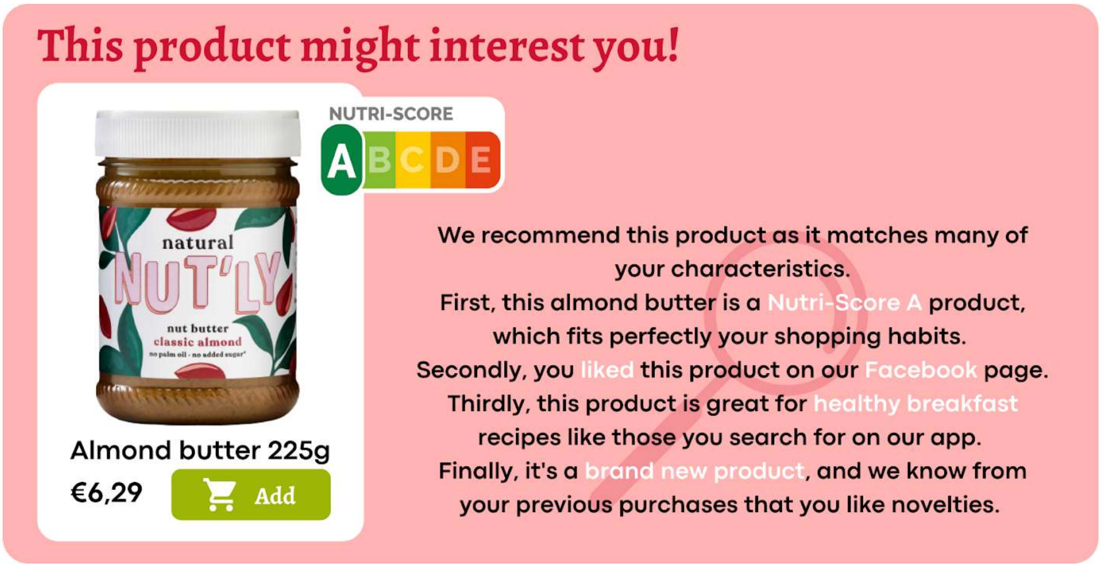


Figure 2: Textual content-oriented explanation scenario

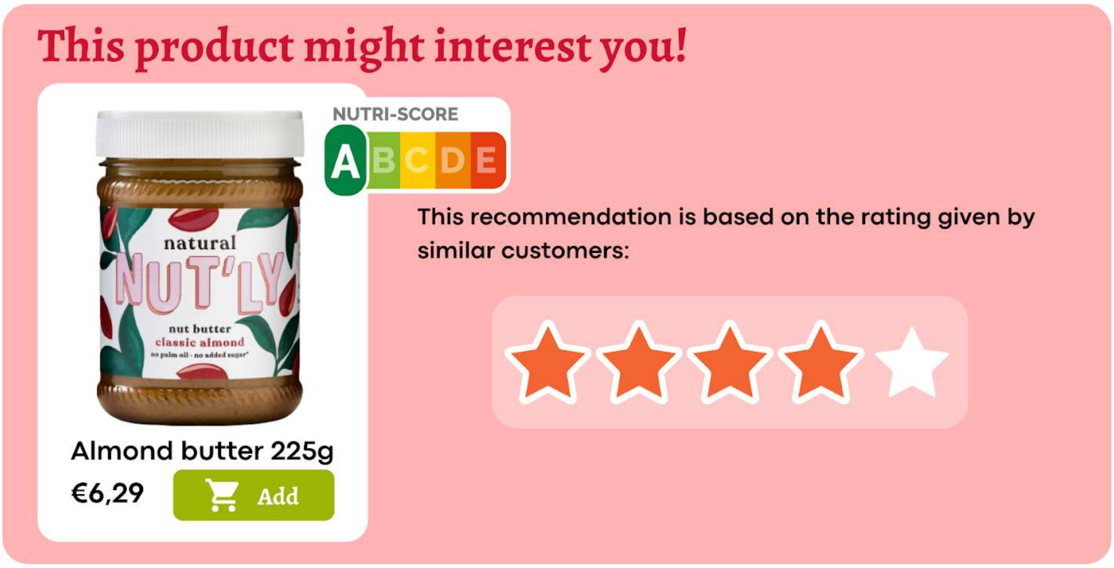


Figure 3: Visual user-oriented explanation scenario

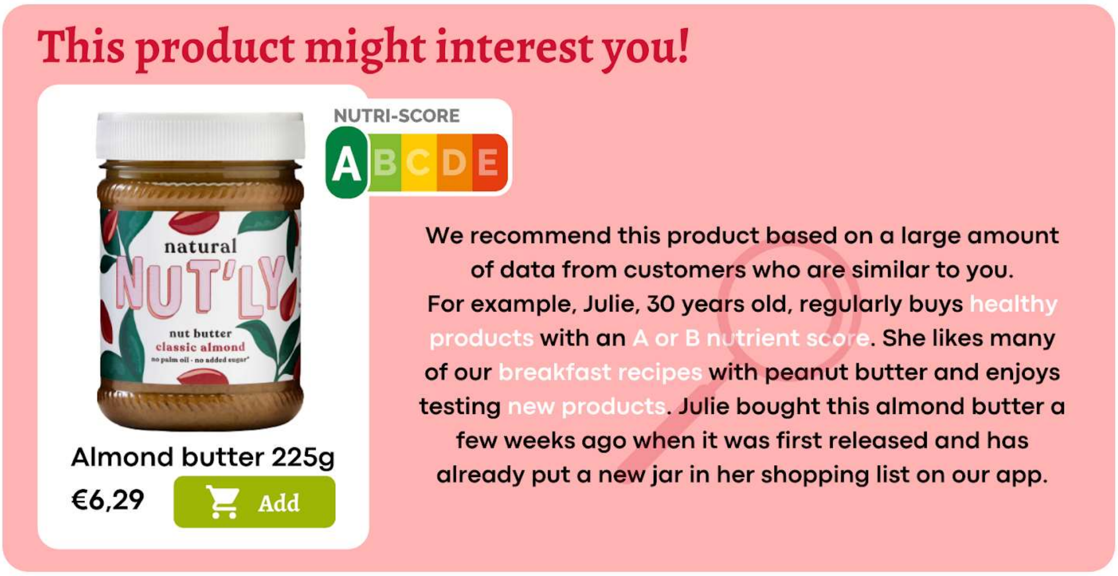


Figure 4: Textual user-oriented explanation scenario

Appendix B – Quantitative Research Questionnaire

Investigation into the explanation of the recommendation system

Dear participant,
As part of my master's thesis in Business Engineering at the University of Namur, Belgium, I am conducting a survey concerning product recommendations on shopping websites.
More specifically, I focus on recommender systems that provide an explanation of how Artificial Intelligence (AI) produces recommendations.
Please answer the questions honestly, there are no right or wrong answers, I am interested in your opinion.
Please also note that your answers will remain confidential and will only be used for the purposes of this thesis.

Thank you for your participation.

Margaux Pé

If you have a Prolific account, please enter your unique Prolific ID.

Enter your text here

Next

What age group do you belong to?

- 18-24
- 25-34
- 35-44
- 45-54
- >54

What is your gender?

- Women
- Man
- Other

What is the highest level of education you completed?

- Primary school
- Secondary school
- Bachelor's degree or equivalent
- Master's degree
- Ph.D. or higher

Which category best describes your employment status?

- Student
- Employee
- Worker
- Self-employed
- Unemployed
- Retired
- Other

Previous Next

Product recommendation - Scenario

The website of a supermarket chain provides its customers, who identify themselves using the e-mail address linked to their loyalty card, some recommended products by means of artificial intelligence (AI).

For this study, you are invited to picture yourself in the following scenario:

You shop regularly in this supermarket (online and in-store) and scan your loyalty card every time you check out. You sometimes search for recipes on the supermarket's app and also subscribe to their Facebook page, where news and promotions are posted. Therefore, the company has some information about your shopping behaviour.

When you receive a recommendation on the shop's website for a product that might interest you, you will also receive an explanation of how the recommendation is generated.

You are asked to evaluate this explanation by answering the questions below as truthfully as possible.

**Here are some statements about the above explanation.
For each, please indicate your level of agreement.**

	I disagree strongly	I disagree somewhat	I am neutral about it	I agree somewhat	I agree strongly
From the explanation, I understand how the recommendation system works.	○	○	○	○	○
This explanation of how the recommender system (AI) works is satisfying.	○	○	○	○	○
This explanation of how the recommender system (AI) works has sufficient detail.	○	○	○	○	○
This explanation of how the recommender system (AI) works seems complete.	○	○	○	○	○
This explanation of how the recommender system (AI) works is useful to my goals.	○	○	○	○	○
This explanation of the recommender system shows me how accurate the recommender is.	○	○	○	○	○
This explanation lets me judge when I should trust and not trust the recommender system (AI).	○	○	○	○	○

**Here are some statements about the above explanation.
For each, please indicate your level of agreement.**

	I disagree strongly	I disagree somewhat	I am neutral about it	I agree somewhat	I agree strongly
I am confident in the recommender system (AI). I feel that it works well.	○	○	○	○	○
The outputs of the recommender system (AI) are very predictable.	○	○	○	○	○
The recommender system (AI) is very reliable. I can count on it to be correct all the time.	○	○	○	○	○
I feel safe that when I rely on the recommender system (AI) I will get the right recommendations.	○	○	○	○	○
I am wary of the recommender system (AI).	○	○	○	○	○
The recommender system (AI) can perform the task better than a novice human user.	○	○	○	○	○
I like using the recommender system (AI) for decision making.	○	○	○	○	○

**Here are some statements about the above explanation.
For each, please indicate your level of agreement.**

	I disagree strongly	I disagree somewhat	I am neutral about it	I agree somewhat	I agree strongly
I intend to use the recommendation system (AI) in the future.	○	○	○	○	○
I intend to use the recommendation system (AI) frequently.	○	○	○	○	○
I intend to recommend that other people use the recommendation system (AI).	○	○	○	○	○
I am willing to use this recommendation system (AI) as an aid with my decision about which product to buy.	○	○	○	○	○
I am willing to let this recommendation system (AI) assist me in deciding which product to buy.	○	○	○	○	○
I am willing to use this recommendation system (AI) as a tool that suggests to me a number of products from which I can choose.	○	○	○	○	○

**When you receive such an explanation about the functioning of a recommendation system, what would you expect to know?
Please select one or more options.**

- ☐ I want to know what the AI just did.
- ☐ I want to know that I understand this AI system correctly.
- ☐ I want to understand what the AI will do next.
- ☐ I want to know why the AI did not make some other decision.
- ☐ I want to know what the AI would have done if something had been different.
- ☐ I was surprised by the AI's actions and want to know what I missed.

Previous | Save

Appendix C - Descriptive statistics of socio-demographic data

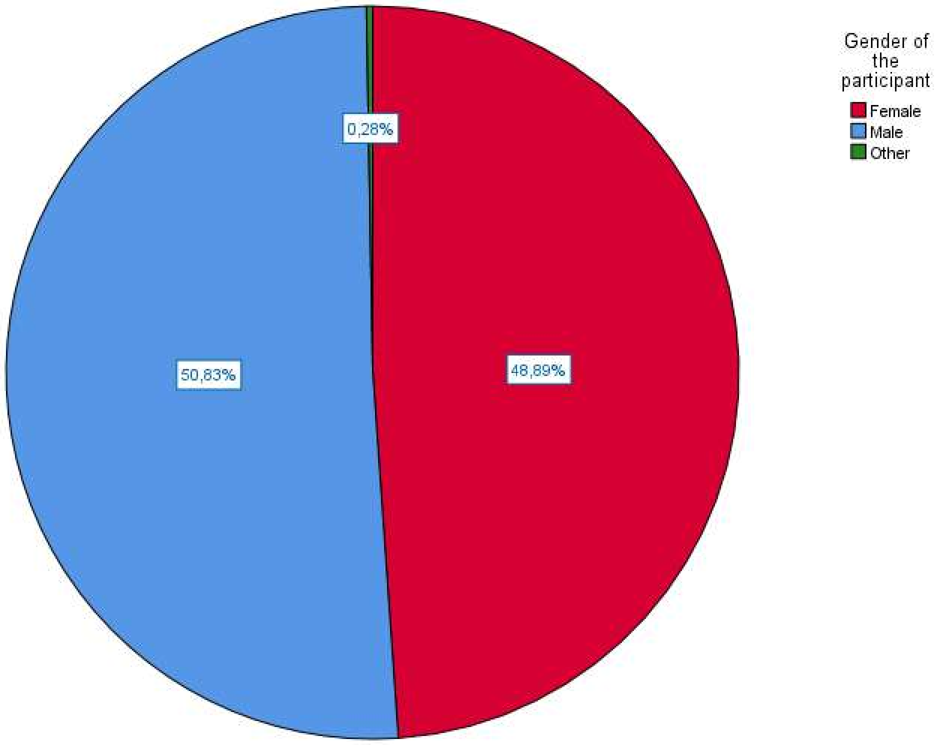


Figure 1: Pie chart of the percentage of men and women in the overall sample

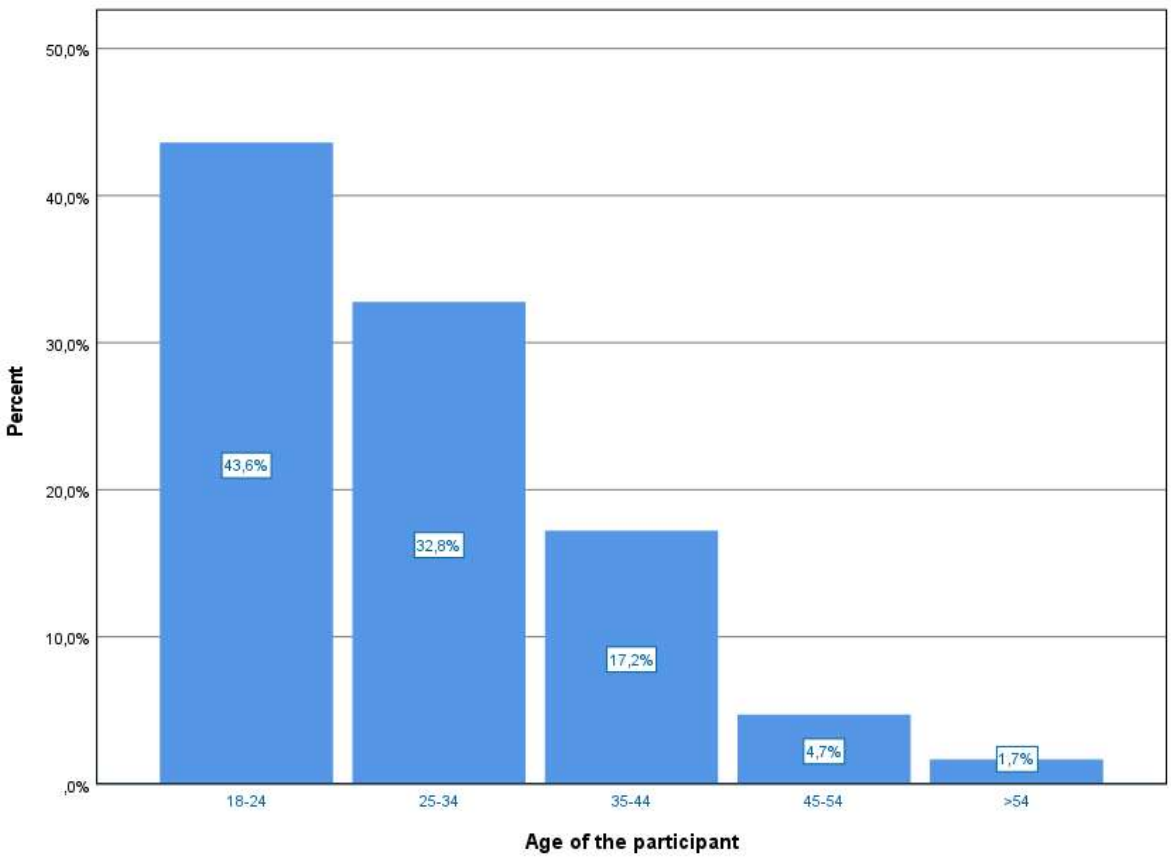


Figure 2: Bar chart of the distribution (in %) of age groups of the overall sample

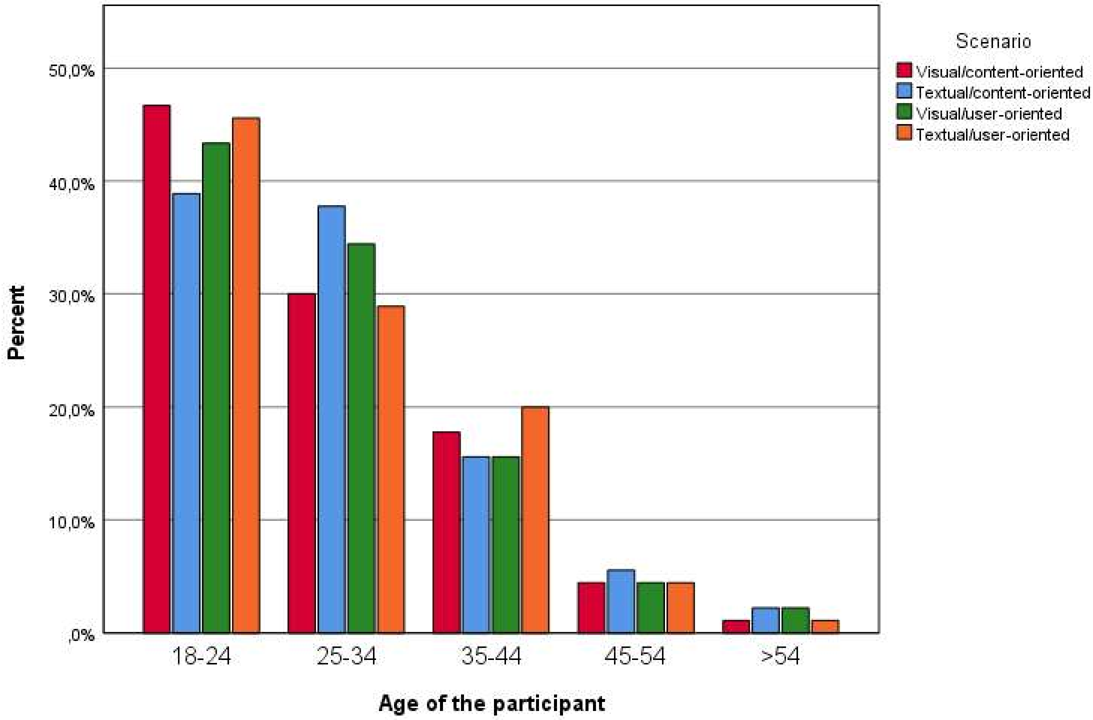


Figure 3: Bar chart of the distribution (in %) of age groups of the samples by scenario

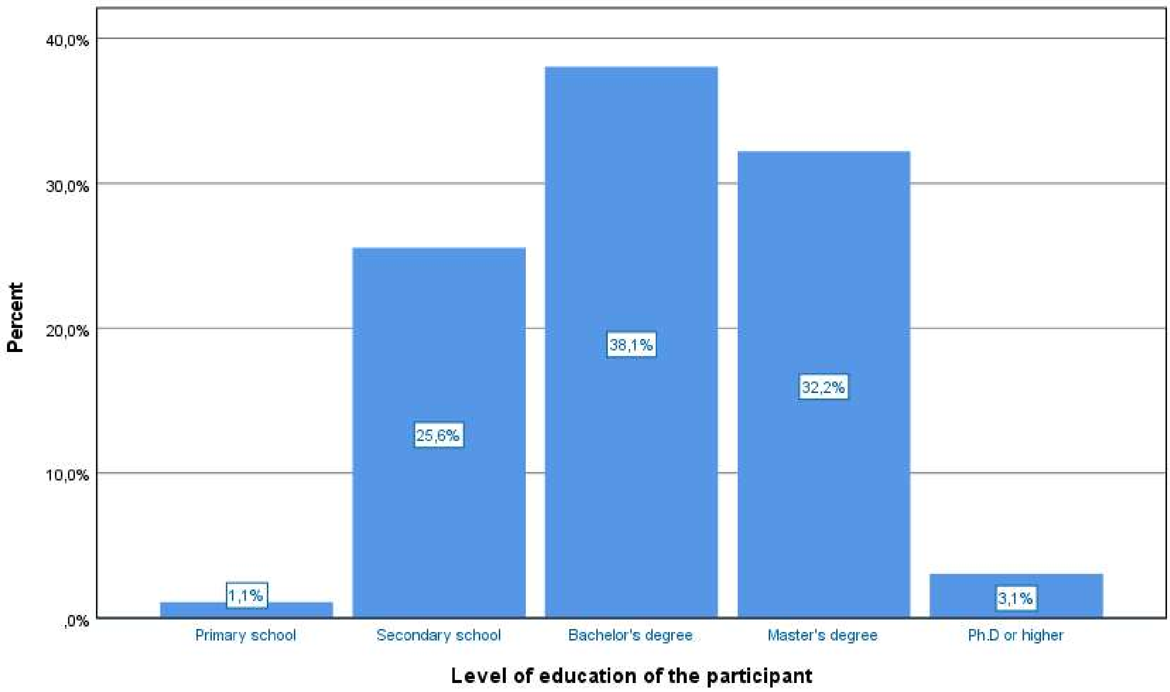


Figure 4: Bar chart of the distribution (in %) of education level of the overall sample

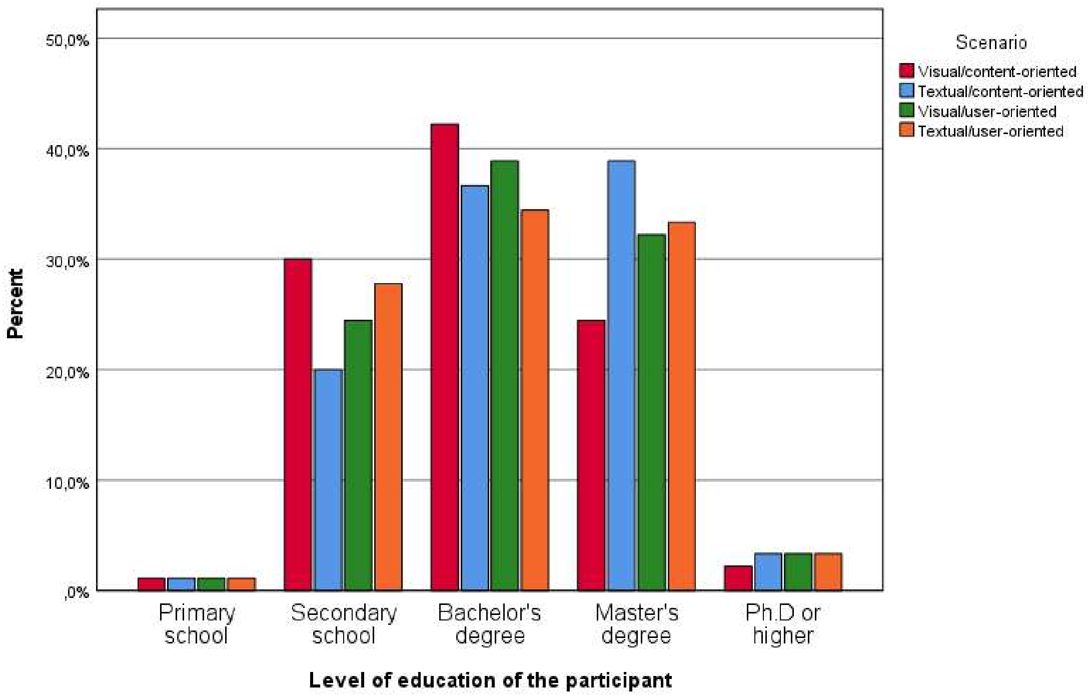


Figure 5: Bar chart of the distribution (in %) of education level of the samples by scenario

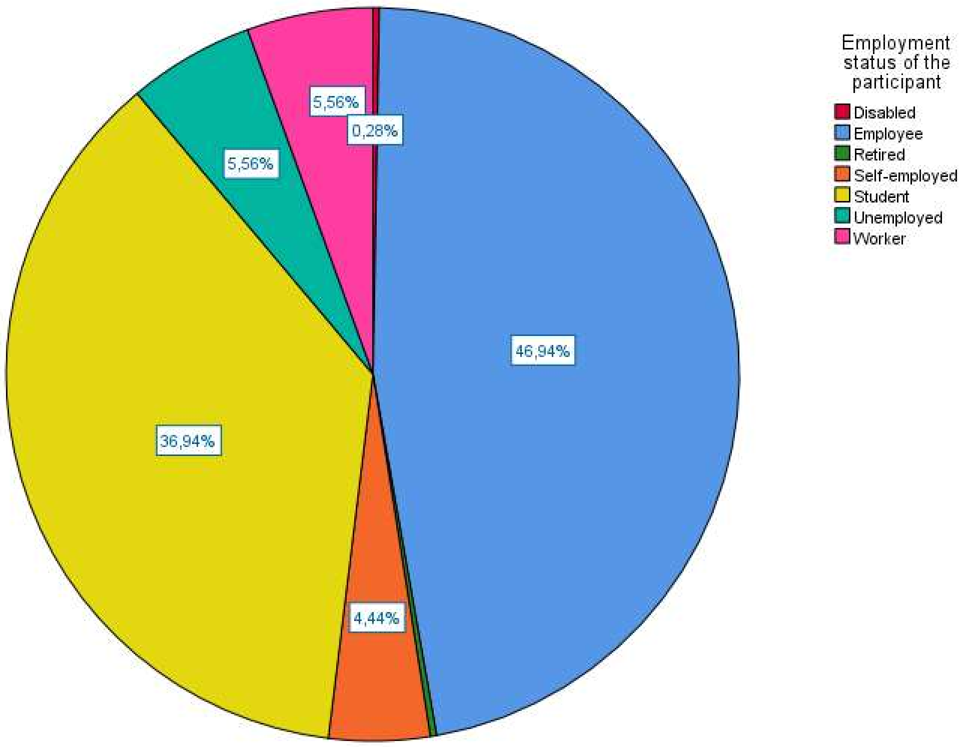


Figure 6: Pie chart of the distribution of employment status of the overall sample

Appendix D – Factor analysis: Explanation perceived quality

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,538	64,828	64,828	4,538	64,828	64,828
2	,686	9,799	74,626			
3	,535	7,636	82,262			
4	,426	6,086	88,348			
5	,391	5,579	93,928			
6	,238	3,405	97,333			
7	,187	2,667	100,000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
Explanation understanding	,779
Explanation satisfaction	,851
Explanation detailed	,870
Explanation completeness	,851
Explanation usefulness	,784
Explanatuon accuracy	,780
Explanation trustworthiness	,708

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Reliability Statistics

Cronbach's Alpha	N of Items
,909	7

Appendix E - Factor analysis: Recommendation perceived trust

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,826	70,655	70,655	2,826	70,655	70,655
2	,558	13,953	84,608			
3	,356	8,905	93,513			
4	,259	6,487	100,000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component 1
Recommendation confidence	,861
Recommendation reliability	,841
Right recommendation	,892
Recommendation satisfaction	,762

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Reliability Statistics

Cronbach's Alpha	N of Items
,859	4

Appendix F – Factor analysis: Recommendation acceptance

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,290	71,505	71,505	4,290	71,505	71,505
2	,648	10,804	82,309			
3	,357	5,948	88,256			
4	,347	5,786	94,042			
5	,185	3,087	97,130			
6	,172	2,870	100,000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
I intend to use the recommendation system (AI) in the future	,886
I intend to use the recommendation system (AI) frequently	,860
I intend to recommend that other people use the recommendation system (AI)	,776
I am willing to use this recommendation system (AI) as an aid with my decision about which product to buy	,889
I am willing to let this recommendation system (AI) assist me in deciding which product to buy	,848
I am willing to use this recommendation system (AI) as a tool that suggests to me a number of products from which I can choose	,810

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Reliability Statistics

Cronbach's Alpha	N of Items
,920	6

Appendix G – Simple linear regression H1

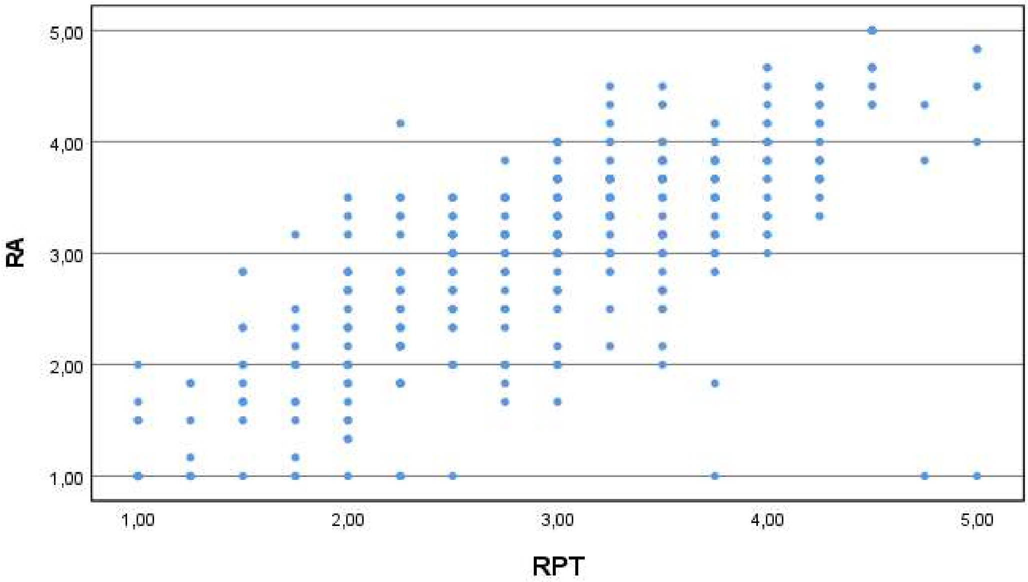
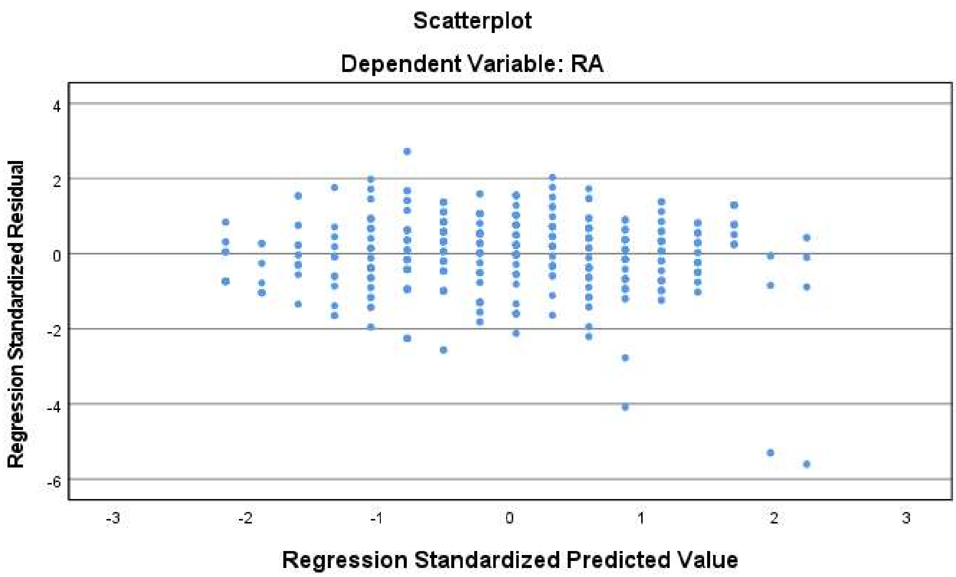


Figure 1: Recommendation acceptance vs. Recommendation perceived trust



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,743[a]	,552	,551	,63580	1,933

a. Predictors: (Constant), RPT

b. Dependent Variable: RA

Figure 3: Durbin-Watson test

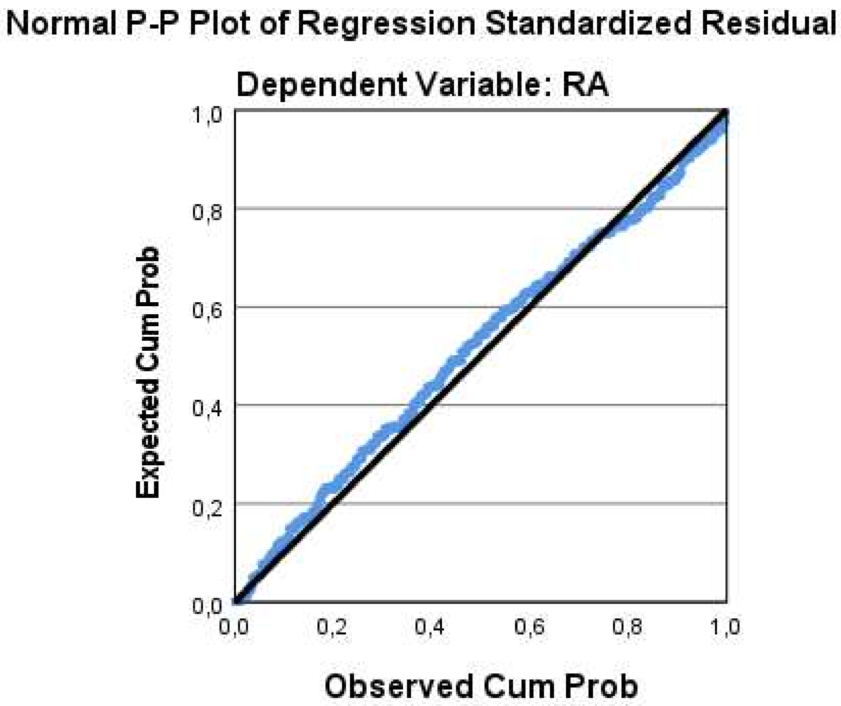


Figure 4: Gaussian plot of residuals

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	178,461	1	178,461	441,476	<,001[b]
	Residual	144,717	358	,404		
	Total	323,178	359			

a. Dependent Variable: RA

b. Predictors: (Constant), RPT

Figure 5 : Simple regression – Analysis of variance Table

Appendix H – Simple linear regression H2

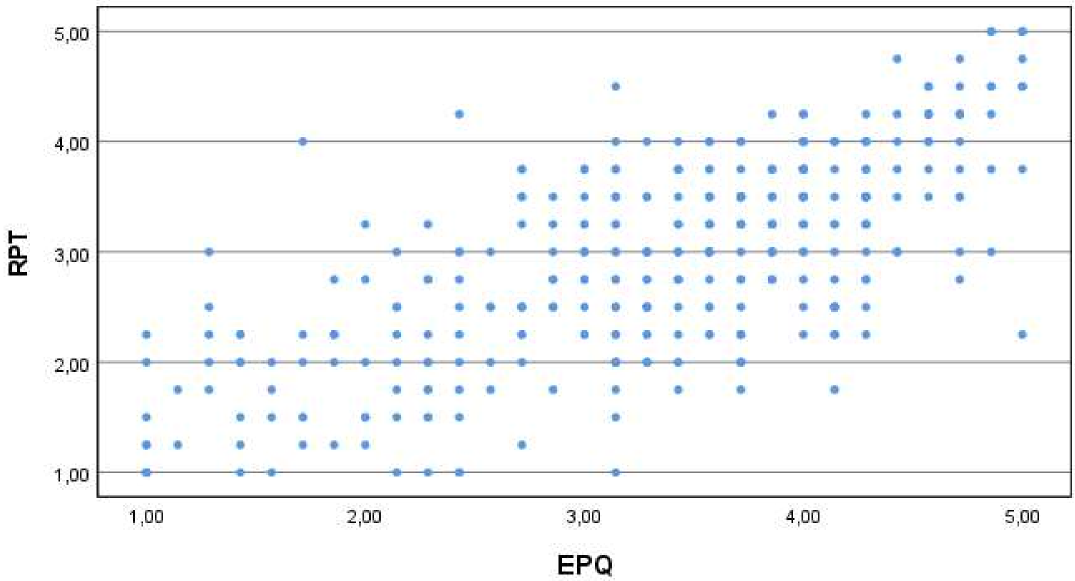


Figure 1: Recommendation perceived trust vs. Explanation perceived quality

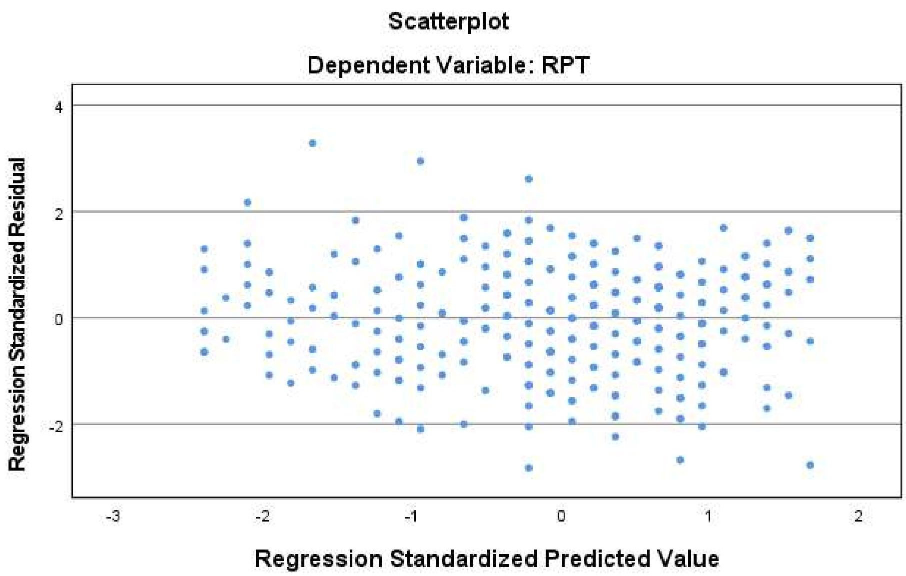


Figure 2: Regression standardized residuals vs. Regression standardized predicted value

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,708[a]	,501	,500	,64450	2,041

a. Predictors: (Constant), EPQ

b. Dependent Variable: RPT

Figure 3: Durbin-Watson test

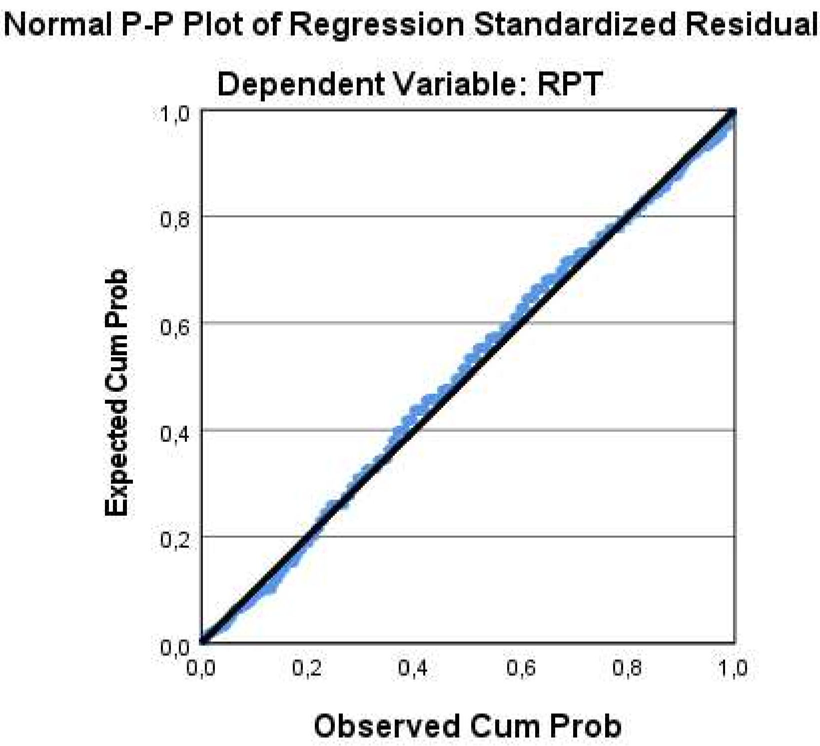


Figure 4: Gaussian plot of residuals

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	149,458	1	149,458	359,808	<,001[b]
	Residual	148,708	358	,415		
	Total	298,166	359			

a. Dependent Variable: RPT

b. Predictors: (Constant), EPQ

Figure 5 : Simple regression – Analysis of variance Table

Appendix I – Levene's Tests

Independent Samples Test

		Levene's Test for Equality of Variances			
		F	Sig.	t	df
Recommendation understanding	Equal variances assumed	33,092	<,001	-3,448	358
	Equal variances not assumed			-3,448	316,086
Explanation detailed	Equal variances assumed	20,308	<,001	-5,237	358
	Equal variances not assumed			-5,237	340,344

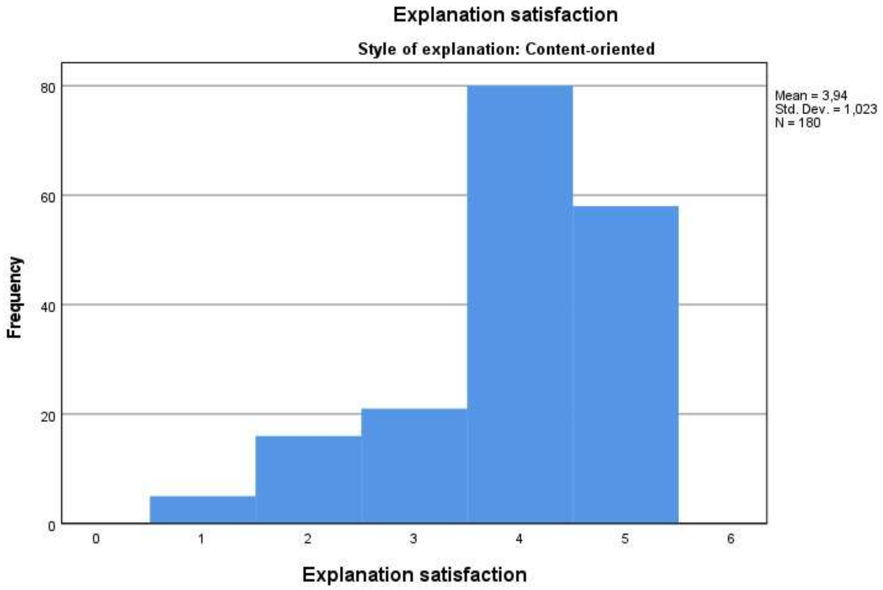
Independent Samples Test

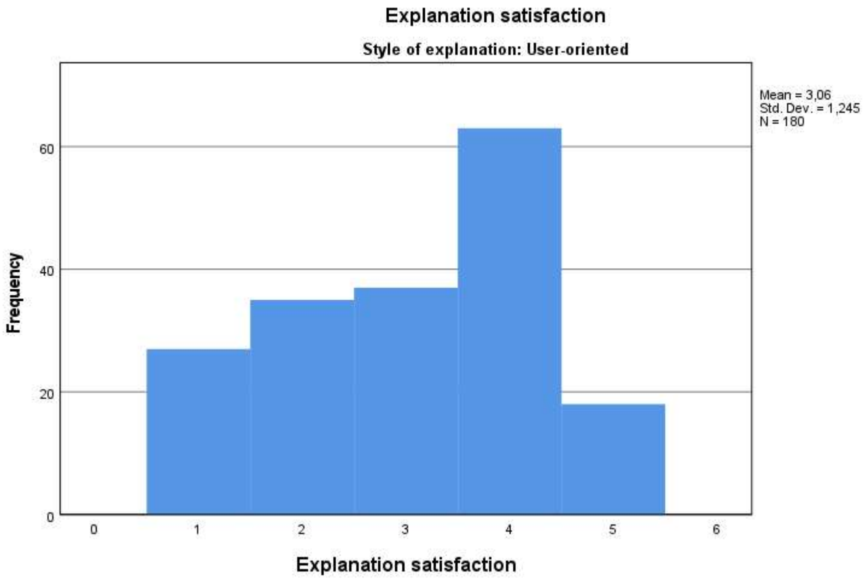
		Levene's Test for Equality of Variances			
		F	Sig.	t	df
Explanation satisfaction	Equal variances assumed	19,383	<,001	7,400	358
	Equal variances not assumed			7,400	345,067
Recommendation accuracy	Equal variances assumed	13,733	<,001	7,847	358
	Equal variances not assumed			7,847	348,383

Appendix J – Explanation Satisfaction: Variance

Explanation satisfaction

Style of explanation			Frequency	Percent	Valid Percent	Cumulative Percent
Content-oriented	Valid	I disagree strongly	5	2,8	2,8	2,8
		I disagree somewhat	16	8,9	8,9	11,7
		I am neutral about it	21	11,7	11,7	23,3
		I agree somewhat	80	44,4	44,4	67,8
		I agree strongly	58	32,2	32,2	100,0
		Total	180	100,0	100,0	
User-oriented	Valid	I disagree strongly	27	15,0	15,0	15,0
		I disagree somewhat	35	19,4	19,4	34,4
		I am neutral about it	37	20,6	20,6	55,0
		I agree somewhat	63	35,0	35,0	90,0
		I agree strongly	18	10,0	10,0	100,0
		Total	180	100,0	100,0	





Appendix K – Recommendation Accuracy: Frequencies

Recommendation accuracy

Scenario			Frequency	Percent	Valid Percent	Cumulative Percent
Visual/content-oriented	Valid	I disagree strongly	3	3,3	3,3	3,3
		I disagree somewhat	8	8,9	8,9	12,2
		I am neutral about it	18	20,0	20,0	32,2
		I agree somewhat	51	56,7	56,7	88,9
		I agree strongly	10	11,1	11,1	100,0
		Total	90	100,0	100,0	
Textual/content-oriented	Valid	I disagree strongly	4	4,4	4,4	4,4
		I disagree somewhat	14	15,6	15,6	20,0
		I am neutral about it	17	18,9	18,9	38,9
		I agree somewhat	33	36,7	36,7	75,6
		I agree strongly	22	24,4	24,4	100,0
		Total	90	100,0	100,0	
Visual/user-oriented	Valid	I disagree strongly	29	32,2	32,2	32,2
		I disagree somewhat	22	24,4	24,4	56,7
		I am neutral about it	18	20,0	20,0	76,7
		I agree somewhat	17	18,9	18,9	95,6
		I agree strongly	4	4,4	4,4	100,0
		Total	90	100,0	100,0	
Textual/user-oriented	Valid	I disagree strongly	13	14,4	14,4	14,4
		I disagree somewhat	16	17,8	17,8	32,2
		I am neutral about it	25	27,8	27,8	60,0
		I agree somewhat	32	35,6	35,6	95,6
		I agree strongly	4	4,4	4,4	100,0
		Total	90	100,0	100,0	