Stockholm School Of Economics
1351 Master's thesis in Business Management
Academic year 2023–2025

Beyond Stars And Stories:
AI's Role In Shaping Consumer Behaviour Through Customer Reviews

Authors: Daniel Palmgren (50719) & Chris Saadeh (42565)

Abstract: Customer reviews are typically recognised as the digital version of word of mouth, and are widely regarded as the most genuine and cost-effective marketing method. Therefore, it is very common for e-commerce brands to include user-generated reviews on product pages. Several studies have already established that valence in reviews can increase purchase intention, product attitude and decision confidence. Review formats and ways of displaying reviews on product pages are a science in themselves, as small details can affect consumer behaviour. With the large mass of reviews some e-commerce brands have garnered, it becomes tricky for consumers to navigate through these many different opinions, which has been shown to have a negative effect on cognitive load. A new solution, AI-generated review summaries, has become more prevalent in the e-commerce space, with Amazon being at the forefront. We delve deeper into AI generated review summaries through social proof, cognitive load theory and persuasion knowledge. We employ a two-step experimental study. The pre-study established two suitable products for the main experimental study, which were gender neutral, equal in price and had different levels of complexity. In the main experiment, 408 participants were exposed to one of four stimuli with different product complexity and the presence or absence of an AI-generated review summary.

Results show that in the scope of social proof, AI-generated reviews have no effect on perceived authenticity and credibility. Neither does it have an effect on persuasion resistance, at least when user-generated reviews are also present. However, AI-generated review summaries reduce cognitive load, which improves decision confidence. To add to this, AI-generated review summaries also contributed to higher purchase intent and better product attitude. The moderating variable, product complexity, seems to have little to no effect on the variables within the scope of this study.

Keywords: Social Proof, Customer Reviews, Artificial Intelligence, Cognitive Load Theory, Persuasion Knowledge Model

Spuervisor: Jonas Colliander
Date Examined: 27/05/2025
Discussants: Ebba Langemar, Astrid Albertsson
Examiners: Magnus Söderlund, Constanze Eib

Table Of Contents

Glossary

PK (Persuasion Knowledge): Refers to consumers' awareness and understanding of marketers' persuasive tactics, which they use to create an active choice to engage and divert from the message.

PKM (Persuasion Knowledge Model): Explains how individuals use their knowledge about persuasion to evaluate and respond to persuasive messages based on their beliefs.

Social Proof: A psychological phenomenon where people look to others' behaviours to create mental shortcuts that guide their own decisions, especially in uncertain situations.

CLT (Cognitive Load Theory): Shows that individuals have a limited capacity in working memory, and decision making is hindered when too much information needs to be processed simultaneously.

Working Memory: Also known as short-term memory, it is responsible for temporarily capturing information needed for complex tasks and decision making.

E-Wom: Online consumer-generated content, such as reviews and recommendations.

PC (Product Complexity): The perceived difficulty of understanding, using and evaluating a product. Depends on the features, functionality and knowledge required to assess it.

PDP: Product Display Page

1.0 Introduction

Word of mouth is often regarded as the most cost-effective marketing tool in the scope of customer acquisition. This phenomenon has been widely studied and is regarded as "social proof" in behavioural economics. It entails that humans tend to latch onto others' behaviour and will use it to create shortcuts for their own behaviour. Social proof can be observed everywhere, it is the foundation of culture and religion, but also of our economy. This simple psychological phenomenon impacts many of the decisions we make on a daily basis. Anyhow, it also attracts new customers for businesses and is a very effective marketing tool if customers share their positive experiences. However, in the past decade, we have observed a digital shift of social proof in the scope of service and retail. Nowadays, more than 90% of consumers will actively search for reviews about a business before making a decision to commit or engage with the brand or business (Chen et al., 2022). Customer reviews have become the cornerstone for bridging the knowledge gap between the business and the customer, and are essential. Reviews act as a proxy, similar to word of mouth, guiding and helping customers make purchase decisions in a complex marketplace or environment. However, not all reviews are the same and certainly not equally persuasive. To add to this, in an era of constant digital development, it can be difficult for consumers to follow this rapid advancement, but one thing is certain: consumers are more informed than ever. The rise of e-commerce has contributed to this digital shift with widespread access to customer reviews, ratings and other stories of the customer journey and the overall experience. Concurrently, e-commerce businesses are leveraging the likes of artificial intelligence (AI) to further improve the customer experience by consolidating reviews and ratings to streamline decision-making for consumers. Businesses utilise AI to capture trends among customer reviews to reduce information overload, save time and create useful tools to gather data on product insights.

Beyond bridging the knowledge gap between the customer and the business, customer reviews help create a better overview of the product and its features. Not only does this increase sales and create trust, but it also reduces return rates and can aid in cost saving for E-commerce retailers (Sun et al., 2021). To add to this, for higher-priced products, reviews are exceptionally important as consumers need a stronger sense of confidence to make a correct purchase decision, as the monetary risk is much greater than a lower-priced item (Askalidis et al., 2016). The volume and valence of reviews are also significant, as the higher the number of reviews, the more confidence one can have to make a reasonable purchase decision, which goes hand in hand with the concept of social proof (Askalidis et al., 2016). As previously highlighted, humans depend on human behaviour and interaction to make behavioural shortcuts, which is a result of increased decision confidence when reviews are present.

The confidence that reviews constitute is deeply rooted in evolutionary psychology, where following the masses increases the chance of survival. Moreover, the same principle actually applies in commerce, as following the crowd makes it more likely that you make a better decision rather than a bad one. If individuals are presented with positive reviews and testimonials, we assume that the product must be good. That's why the valence of reviews is actually more impactful than star ratings alone, because they are more relatable and more human (Ketelaar et al., 2015).

To add to this, several well-known retail firms have adopted an AI summary approach to their product pages, with the aim of reducing cognitive overload from the hundreds of reviews. One example is Amazon, which was one of the first to adopt generative AI to create a more seamless customer experience (Amazon, 2024). Amazon highlights that the main aim of this adoption is to create improved purchase confidence and to create a tool for product improvement downstream. Moreover, Amazon has adopted AI review summaries to decrease time spent on product pages, dwelling on specific product attributes, but with little knowledge of how the general perception is towards AI use for user-generated reviews.

With the proliferation of user-generated reviews, commerce businesses have also been presented with an abundance of review volume, rustling in overwhelming potential customers. This cognitive onus presents itself as cognitive load theory, which highlights that humans can only process a limited amount of information before irrational decision making becomes more prevalent (Sweller., 1988). The basis for AI summaries aims to solve this problem by compiling excessive information into easily and readily available summaries to reduce the cognitive load and simplify the decision-making process. However, this presents an interesting problem. How do consumers perceive AI-generated review summaries compared to user-generated reviews? Despite the reduction of cognitive load, there is limited research on AI's effect on decision confidence, authenticity and trust in the scope of social proof. Our research aims to investigate this, namely the intersection between social proof and CLT in a digital commerce setting. Moreover, by employing an approach that evaluates the psychological effect of AI review summaries, we create a foundation for further research that can establish a broader depiction of potential managerial implications in this field.

1.1 Problem statement

It is yet unclear whether AI has the same effect on human behaviour in the scope of social proof as human-to-human interaction has. The presence of AI review summaries on E-commerce platforms may not create the same decision confidence as a real customer review does. Moreover, there is a general consensus that AI may not provide accurate and reliable information. Therefore, despite creating an easier experience for customers in the scope of researching a specific product, AI may hinder assurance and decision confidence due to its non-human nature. There is a salient gap in understanding how consumer behaviour is affected by the presence of AI, and more specifically, in the replacement or addition of a human-made narrative.

As previously highlighted, valent reviews have a large impact on buying decisions, and more than 90% of consumers make a buying decision with reviews in mind (Chen et al., 2022). To add to this, previous research in behavioural economics highlights that reviews heavily impact consumer behaviour as they create confidence in the scope of social proof. On the contrary, this raises a question whether AI-generated review summaries seem as authentic and credible as human-made reviews and if there are psychological barriers affecting commitment. AI also lacks human judgment, emotional intelligence and contextual distinction, which may lead to summaries lacking human sentiment and personality. Despite this, humans do acknowledge that machine learning actually has several benefits over humans, namely being precise and objective (Logg et al, 2019). It is also worth noting that this is solely dependent on the task type. The lack of previous research on AI's effect on consumer behaviour replacing human narratives creates a sense of uncertainty in regards to effective reviews consolidation. The uncertainty around this topic makes it difficult for E-commerce brands and other service businesses to navigate this space without harming their brand and product perception.

1.2 Purpose And Research Questions

This research paper seeks to delve deeper into the phenomenon of social proof, cognitive load theory and persuasion knowledge in the digital context of AI review summaries. We aim to explore how consumers perceive Artificial Intelligence in digital commerce and if consumers perceive AI to be just as credible, authentic and reliable as humans. Specifically, we aspire to investigate whether AI summarised reviews have an impact on consumer trust, decision confidence and authenticity compared to customer-generated reviews. Through an experimental approach, we will test factors such as authenticity, bias, purchase intent, product attitude and decision confidence to determine how the presence of an AI review summary affects these variables. We aspire to create a holistic view of the behavioural and psychological effects of AI review summaries from different angles of consumer behaviour, through a behavioural economic approach. Moreover, we also aim to compare the differences in these variables with another interaction variable, namely, product complexity. To accomplish this purpose, we will pursue the following research questions:

1. How do AI review summaries affect consumer behaviour and attitudes compared to regular user-generated reviews?
2. To what extent does product complexity and involvement impact these perceptions?

1.3 Expected Contribution

The current literature highlights psychological reasoning as to why humans will create behavioural shortcuts based on other people's actions, which is explained as social proof. Several studies have delineated that reviews are such a cornerstone for E-commerce businesses simply because of this psychological phenomenon, namely social proof. Moreover, current literature has also evaluated how trust and perceived bias for artificial intelligence depend on context. In other words, humans will trust others to make the right decisions and use these actions as heuristics to turn complex decisions into elementary ones. However, with the mixed perception of AI and its reliability, bias and authenticity in different contextual settings, an undiscovered dynamic exists. To add to this, research has shown that AI review summaries actually increase conversion and are more useful for search products (Wang & Wang., 2025). On the contrary, this novel research lacks a holistic view on consumer behaviour as conversion rates from Amazon product pages were used as the basis for AI summary success, without the inclusion of other variables. We aim to add to this research by creating a more comprehensive view by garnering the psychological factors affecting conversion and by limiting external factors that may affect purchase decisions and confidence. Also, to create more exhaustive results, we aim to deep-dive into product complexity and the effect of AI review summaries on the variables previously highlighted. Through this research, we will add a new revelation that not only scrutinises the conversion rate and sales. We apply a brand approach that considers the psychological factors in the scope of social proof and cognitive load theory (CLT). The findings of this research will thus contribute to the limited research on social proof in the context of AI and its role in digital commerce. Moreover, it creates a foundation for the exploration of research within this novel topic by introducing product complexity and involvement as a dependent variable.

2.0 Conceptual Background

The subsequent section builds a strong foundation for the empirical research of this specific topic. Given the novelty of the research topic, the aim of this segment is to tie together relevant literature on various topics which further abet the findings of this paper. Likewise, the topics that will be covered consist of previous literature on consumer trust in digital commerce, product complexity and involvement, perception of AI and the well-renowned behavioural economics concept, social proof. Nevertheless, the conceptual background will also provide a coherent and interesting discussion of the results that are drawn from the empirical study conducted in this particular line of research.

2.1 User-Generated Reviews

In recent times, user-generated reviews have become a cornerstone for e-commerce businesses to convey a message and increase purchase confidence and intention. Studies have shown that user-generated reviews are one of the most impactful factors in the scope of purchasing decision making (Zhong-Gang et al., 2015; Ruiz-Mafe et al., 2018). These studies have highlighted that there are several factors affecting the convincingness of user-generated reviews, such as level of detail, agreement and perceived credibility (Archak et al., 2011). For example, Archak et al imply that individuals will read and acknowledge product reviews depending on the readability, subjectiveness, informality and also correctness in terms of grammar and linguistic accuracy. Pleasant reviews with a positive attitude are also more likely to garner more decision-confident individuals, which is closely linked to higher purchase intention (Guo et al., 2020). Other studies have investigated other variables in regard to user-generated reviews, such as product satisfaction (Changchit & Klaus., 2020).

With this in mind, the general thread among previous literature is to evaluate consumer behaviour, such as purchase intention, with various interaction variables. However, there is a lack of research and understanding of the psychological mechanisms leading to specific behaviours such as cognition, emotions, and review helpfulness. The abundance of research on user-generated reviews typically focuses on the managerial implications from a sales approach and not from a brand approach. However, a clear link has already been established between emotional factors and purchase intention (Yuanyuan et al., 2009). For example, human-like emotion and distinct expressions have been shown to have a more positive effect on user-generated reviews and the overall perception of them (Yuanyuan et al., 2009). On the contrary, negative comments or reviews tend to have the opposite effect and result in lower emotional trust and perceived review usefulness (Archak et al., 2010). To add to this, consumers also believe that negative reviews are more useful when making a purchase decision, and they are more impactful at creating deviation and reluctance to specific products, compared to a five-star rating, garnering trust and high decision confidence (Ahluwalia et al., 2000).

While the literature on AI review summaries and their effect on consumer behaviour is limited, it is worth mentioning that there is related research in this area that aids in hypotheses formation of this specific research. For example, previous research has shown that the perceived effort that goes into review creation directly impacts purchase intention. This logic also demonstrates why influencer-generated reviews are more impactful than user-generated reviews, as they require more effort to craft (Puspitasari & Aruan., 2022). Furthermore, this conclusion goes hand in hand with social proof, as specific peers that individuals look up to can have a larger impact on purchase decisions. Anyhow, novel research on AI authorship has disclosed that emotional texts in any form are perceived as less authentic and credible when written by AI. Individuals are more likely to exhibit moral disgust and reluctance when generative AI is used for authorship unless the contents of the text are factual or a summary (Kirk & Givi., 2025). In the scope of user-generated reviews, the findings of this research are relevant as they correlate with Yuanyuan's research on the benefits of emotions in user-generated reviews and their impact on trust and credibility (Yuanyuan et al., 2009).

2.2 Collective Influence on Individual Choice

As previously highlighted, social proof relates to the phenomenon whereby individuals create cognitive shortcuts based on the behaviour of others. This behavioural economic phenomenon was originally probed by world-renowned psychologists Robert Cialdini (Cialdini., 2001). A great example of this is that individuals are more likely to buy a product if they have, in any medium, perceived other users as happy using that specific product (Cialdini, 2001). Moreover, individuals who are exposed to excess information and are uncertain about a specific choice are more likely to resort to this mental shortcut, namely social proof (Smith et al., 2007). However, when enough information is provided and decision confidence is assured, individuals typically do not resort to "following the crowd" and thus rely on their own personal preference and choice (Smith et al., 2007). Hence, individuals copy the behaviour of others solely when information is limited or convoluted.

Social proof is noticeable in all types of environments, including digital commerce. When making a purchase decision we more than often resort to the opinion of others to make a purchase decision. Previous research by Nielsen (2014) found that around 92% of individuals will trust recommendations from friends and family, and thus make a more swift and confident purchase decision. Interestingly, 70% of the participants in this research also highlighted that reviews and recommendations on social media platforms are the second most trusted source of information (Nielsen., 2014). This further exhibits that social proof is not only the foundation of human behaviour, but also the most cost-effective marketing tool. Moreover, it has also been found that electronic word of mouth (e-WOM), along with reviews and ratings, is the most efficient way to share information with an individual who has not yet made a purchase decision (Amblee & Bui, 2012). The reason why individuals feel more confident making a purchase decision based on the behaviour of others arises from this psychological phenomenon, namely, social proof. The psychological barrier dissipates as a result of heightened levels of perceived trust, authenticity and credibility, compared to other mediums of communication and information sharing (Liang et al., 2011).

Trust is previously defined by Mayer et al as; "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995). Due to the nature of digital commerce, the perceived risk of making a purchase decision is much higher as individuals are unable to touch, feel and experience the product before making a decision (Dachyar & Banjarnahor., 2017). Therefore, e-commerce businesses utilise user-generated reviews to reduce this perceived risk, which is typically achieved through user-generated reviews, with both a rating and qualitative information (Nazlan et al., 2018). However, the main variables that affect the convincingness of a user-generated review are the perceived authenticity and credibility (Boer, 2021; Thomas et al., 2019)

The effects of social proof already have an established root in digital commerce, which is what we experience as testimonials or customer reviews. User-generated reviews bolster the knowledge gaps that appear between the business and the consumer, thus increasing decision confidence as a result of user-generated testimonials (Isaac & Grayson, 2017). However, as highlighted in Cialdini's early work on social proof, false cues that are evident to individuals can trigger persuasion knowledge (Cialdini., 2001), which in an online setting leads to lost credibility and may result in opposite effects (Friestad & Wright, 1994). To add to this, the stimulation of persuasion knowledge impacts the effectiveness of social proof in commerce settings (Fenko et al., 2017). However, only a handful of studies have investigated the relationship between social proof and artificial intelligence.

2.3 Attitude Toward Artificial Intelligence

Artificial intelligence has become prevalent in all sorts of industries, and is used in at least one business function for 71% of companies worldwide (Mckinsey, 2024). It is mainly used to redesign workflows and other strategic approaches to eliminate unnecessary costs. Moreover, it is also used to improve service operations, which is the core use for Amazon in regards to their adoption of generative AI review summaries. Despite the usefulness and the broad understanding of the capabilities, there is still some reluctance among users about the authenticity and credibility of generative AI. This scepticism is rooted in the perceived bias and lack of accuracy in some generative AI models. Research has shown that while generative AI may make users more efficient in decision-making processes and in task completion, there is still a general impediment in terms of reliability (Balbaa & Abdurashidova., 2024). Moreover, generative AI do make mistakes and in some cases provides false information, also known as AI "hallucinations". To add to this, the rate of AI hallucinations varies drastically across different platforms. For example, GPT-3.5 exhibits a hallucination rate of around 40%, and GPT-4 slightly lower than this (Chelli et al., 2024). However, some generative AI tools displayed hallucination rates above 90% (Chelli et al., 2024), which further goes to show that the trustworthiness and reliability of generative AI may be challenged.

In the context of AI review summaries, individuals may appreciate the convenience and the reduction of cognitive load to focus on important product attributes and other valuable information. On the contrary, users may become wary when AI-generated content lacks identifiable human input. This further aligns with research on AI authorship, essentially meaning the creation of texts, articles, reviews, etc, which shows that identifiable AI-generated content results in lower perceived credibility (Jia et al., 2024). The effectiveness of AI influencing a purchase decision thus depends on the perceived authenticity and credibility, where the user expects no bias to be present in the generated content.

While AI summaries are not a new phenomenon, it is a fairly novel use case in the scope of review consolidation. However, AI summaries are widely used internally by businesses to streamline processes, such as recruiting. However, research has found that for decisions with higher impact and higher involvement, AI summaries are merely used as a supplementary tool for effective decision making (Bakshi., 2024). The same can be assumed for other decision-making processes, which further highlights that high involvement and complex decision-making require individual evaluation of the choices at hand. In this case, AI tools are simply used as a "screening" to filter out unnecessary information, however, it does not act as a medium for decision making.

2.4 Product Complexity and Involvement

While the term product complexity has no clear red thread of definitions among the existing literature, it has been described as having many dimensions, components and modules (Jacobs, 2013). Moreover, based on the literature consolidation by Trattner et al on the subject, it is evident that products such as electronics are more complex than apparel per se, due to many aspects such as a number of variants, components, familiarity, variety and product platforms (Trattner et al., 2019).

Product complexity and involvement are pivotal factors that can have a huge influence on decision making, decision confidence and purchase intent. Firstly, product complexity can be divided into two different elements, namely, feature heterogeneity and feature interrelatedness. Feature heterogeneity refers to the wide array of functions that a product offers, whereas feature interrelatedness pertains to how these different features are connected and interlinked (Fürst et al., 2023). Moreover, Fürst et al also found that user-generated reviews are typically used as a tool to create a better understanding of products with high feature heterogeneity, as it builds "trust". Furthermore, products that are more complex, e.g having a wider array of features, higher price and longer descriptions, also require more involvement and thus individuals are more likely to read reviews more thoroughly (Chen et al., 2022). Fürst et al also highlight that an increase in feature heterogeneity lowers expected product usability, which may result in a lack of motivation for product adoption and usage. On the contrary, high feature interrelatedness improves the perceived capability of the products, and thus, more easily understood product attributes can make the product more attractive to customers (Fürst et al., 2023).
In addition to the findings by Fürst et al, novel research has also delved into the impact on consumer behaviour depending on the time spent reading product descriptions and learning about product attributes (Necula., 2023). By using machine learning, Necula analysed clickthrough data and determined that products with more complex attributes required more easily processable information to reduce exit rates, for example (Necula., 2023).

2.5 Product Attitude and Purchase Intention

Product attitude is typically referred to as a consumers opinion and judgment of a specific product, which typically comes from beliefs, emotions, feeling and behavioural intentions (Mitchell & Olson, 1981). In the context of e-commerce, product attitude is a antecedent of purchase intention, which is determined by the contextual information such as product reviews, product descioptions, images and ease of use (Cheung & Thadani, 2012). As highlighted in previous literature, the overall feeling toward a specific product also depends on the valance of information, richness and the perceived credibility of the source (Guo et al., 2020). When consumers are given relevant information about a specific product, with informational richness and from a credible source, they are more likely to have a positive attitude toward the product. Moreover, this goes hand in hand with other relevant theoroeis explored within this research, namely cognitive load. Another framework not widely used within this research, the elaboration likelihood model, highlights that consumers who seek infromtaion through a peripheral route are more likely to have positive attitudes (Petty & Cacioppo, 1986).

Product attitude and purchase intention have been widely studied and are closely linked. When prodcuta attitude is positive, a consumer is more likely to purchase the product. Purchase intention is the likelihood of a consumer purchasing a specific product (Ping et al., 2022). Prior literature shows that positive product attitude, perceived usefulness of information and emotional resoanance can have a direct positive impact on purchase intention (Yuanyuan et al., 2009). Customer reviews also have a large impact, as social proof is also a key antecedent of trust and credibility, which ultimately results in higher purchase intention (Mou et al., 2020). However, as previously highlighted, there is little research on how existing attitudes on artificial intelligence may affect purchase intention, especially in the form of review summaries.

2.6 Persuasion Knowledge Model (PKM)

Around 25 years ago the persuasion knowledge model was brought to life by Friestad & Wright (1994). The models highlight individuals as “targets” who are being persuaded by “agents” of any type of medium, such as an advertisement, testimonial or presentation, to skew the individual's initial beliefs and values in another direction. The model presents itself as a threefold, namely, agent knowledge, topic knowledge, and persuasion knowledge. PK is developed over time, in parallel with increased cognition and thus a well-developed understanding of PK is established around adultery, but becomes prevalent from the age of 6 (Wellman, 1990). To add to this, more educated and older people tend to have higher PK than other individuals, which also makes them more sceptical and reluctant to persuasive attempts, which can create a negative demeanour. A simple but great example of this is covert advertising, which may result in negative brand perception when individuals realise the true intention of the advertisement (Mohr & Kühl, 2021).

In the context of customer reviews and testimonials, the user-generated reviews act as a proxy, creating a persuasive message of product attributes while tapping into social proof. As highlighted by Cialdini, adopting the principles of social proof is an effective way to create persuasion. However, resistance to persuasion is also common among individuals (Fransen et al., 2015). Fransen et al present resistance to persuasion as threefold: threat to freedom, reluctance to change and concern for deception. To add to this, the general trust of AI is widespread depending on several variables, however, studies have found that AI has the power to deceive individuals on a systematic basis (Park et al., 2024), which in the framework of resistance to persuasion would lead to intransigence.

In recent times, the focus has shifted towards human attitudes to AI and how the effects on persuasion knowledge vary. Individuals who recognise that chatbots on e-commerce platforms are operated by AI are more reluctant to PK, which essentially reduces effectiveness (Kaplan & Haenlein 2020). Moreover, in instances where chatbots (agents) are perceived as human, targets (individuals) are actually more susceptible to persuasive knowledge, which further highlights that it is not the nature of the information, but rather the characteristics of the agent (Araujo., 2018). In particular cases, the presence of AI creates a negative connotation due to the non-human characteristics, simply when individuals recognise the agent is not a real human (Araujo., 2018). On the contrary, for simple tasks when AI is perceived as fair and objective, it may actually efficiently increase PK (Davenport et al., 2019).

2.7 Cognitive Load Theory (CLT)

Cognitive load theory (CLT) was developed by psychologist John Sweller in the late 1980s and is commonly described as individuals' limited cognitive function in the circumstance of excess information or stimuli (Sweller, 1988). Sweller identifies three types of cognitive load, which are: intrinsic, germane and extraneous. Intrinsic cognitive load refers to task complexity, germane load is the formation of schemas through enhanced learning. Finally, extraneous load arises when there is excess stimuli or information, creating unnecessary information processing by the individual (Sweller, 1988). CLT is widely used in instructional design to create a seamless decision-making process for individuals. For instance, products with higher complexity and more attributes tax cognitive function more significantly than simple, low involvement products (Scheiter et al., 2020). The need to process a wide array of information hinders cognitive function and can result in decision fatigue or reluctance to decision making overall (Payne et al, 1993). E-commerce brands typically keep this in mind to reduce cognitive load and improve the decision-making process by keeping product descriptions short and snappy and providing summaries of testimonials, for example. Research shows that summarisation improves decision speed and efficiency, especially for products that require high involvement and are more complex (Lurie., 2004). In e-commerce and digital environments, there are already established tools to prevent disorientation and excess cognitive load, such as keeping track of website position, by highlighting products, links and other URLS that have already been visited in specific colours (typically purple on domain web pages). On the other hand, completion of more difficult tasks that require more cognitive load may result in lower user satisfaction and may hinder further visits to the webpage (Eveland & Dunwoody., 2001).

Related to this, previous research on the topic highlights that a search function on an e-commerce website made individuals more efficient in finding specific items they were seeking, as a result of lower cognitive load (Schmutz et al., 2009). Individuals who were able to utilise the search function rather than browsing product catalogues were able to directly find the information sought after, rather than learning the structure of each catalogue and tracking their position on the website. In the scope of usability research, Chevalier and Kicka examined the differences on cognitive load when recovering information from an ergonomic website and a non-ergonomic website. While it seems contradictory, the findings highlight that there are no differences in cognitive load between professional designers, experienced users and neophytes (Chevalier & Kicka., 2006). The authors also displayed that while the users and neophytes used the website in different ways, the outcome of information retrieval was not statistically different in regard to the detail of analysis of these outcomes. However, it was also found that cognitive load was larger for novices on the ergonomic website compared to the users, which entails that users are more task driven on the ergonomic website compared to the non-ergonomic alternative (Chevalier & Kicka., 2006). The same principle can be applied to customer reviews and testimonials. As featured by Mirhoseini et al in their research on online shopping determinants, excess information on product pages leads to higher chances of abandoned carts and a shorter time spent in the environment. They also present a solution to reduce cognitive load and free up working memory, which is to keep relevant information short and summarised (Mirhoseini et al., 2021). On the other hand, despite product complexity creating a hindrance for decision making as a result of cognitive overload, novel information has the opposite effect; it induces motivation to stay on the website (Huang., 2000).

In the context of e-commerce, decision-making also depends on the structure and presentation of information. For example, previous research by Lurie and Mason showed that consumers who read summarised content in the form of product reviews felt more confident making a purchase decision Lurie & Mason (2007). This aligns with CLT, as individuals free up real estate in their working memory to carefully consider specific attributes, rather than filtering through a wide assortment of information. Furthermore, research has shown that Amazon's previous approach to summarized reviews through product attribute categorisation actually reduces decision time and improves trust as a result of task prioritisation and reduced cognitive stress (Mudambi & Schuff, 2010). However, it is important to keep in mind that this feature only allows for categorisation of product attributes depending on individual preferences. It is also worth mentioning that there are several other variables that affect task efficiency in the scope of CLT, such as website layout and design. Websites with an intuitive layout and a well-organised navigation system are more likely to create more efficient shopping experiences (van Schaik & Ling, 2008). Not to mention, other stimuli that can negatively affect attention span and create fatigue are pop-up windows, call to action windows and dense text (Beldad et al., 2010). On the contrary, previous research has established that high-quality product information has no direct positive effect on purchase intention, but readability and effectiveness of information retrieval do (Mou et al., 2020).

2.8 Discussion of Hypotheses

Based on the thorough literature review and conceptual background, the hypotheses are formulated after a short summarisation of the most relevant literature.

CLT highlights that decision making may be impaired when individuals are presented with an abundance of information exceeding the capacity of the working memory (Sweller, 1988). AI-created review summaries minimise extraneous cognitive load by filtering and summarising user-created content so that consumers can concentrate on relevant product features without going through unnecessary details. This is in accordance with Lurie and Mason (2007), who demonstrated that summary content improves choice confidence and speed. In addition, research confirms that presented information is rendered more usable and decreases mental exhaustion through streamlined presentation, e.g., concise endorsements or attribute-based summaries (Mudambi & Schuff, 2010; Mirhoseini et al., 2021). To add to this, summarisation of complex information has been shown to reduce cognitive load, which increases conversion and positive attitude towards a brand or a product (Mirhoseini et al., 2021). In the scope of CLT, product complexity is another dimension of uncertainty, along with excess information. Therefore, summaries will ease up working memory more effectively as a result of a higher degree of information processing (Mirhoseini et al., 2021). As previous research has shown, summaries reduce cognitive load and thus we expect AI generated review summaries to have the same effect.

H1: AI review summaries will reduce cognitive load

High cognitive load becomes prevalent when individuals must process inordinate amounts of information, leading to uncertainty, frustration and fatigue (Sweller., 1988; Payne et al., 1993). To add to this, decision confidence is improved when individuals are presented with more structured information that is more easily processed, such as the consolidation of information into summaries (Smith et al., 2007). Usability research has backed this up through a simulated approach, which shows that digital tools have already been established in e-commerce and general domain web pages to prevent disorientation, which improves decision speed and confidence (Schmutz et al., 2009); Chevalier & Kicka, 2006). Previous literature on CLT also emphasises that summaries have a positive impact on decision confidence as a result of lower cognitive load (Mirhoseini et al., 2021). Moreover, as we hypothesize that AI generated review summaries will reduce cognitive load, we can therefore alos hypothesize that decision speed and confidence will improve, as highlighted by previous research. Ultimately, CLT suggests that that individuals are able to make more confident and faster decisions when working memory is relieved. Therefore, H2 builds directly on H1 by testing the downstream impact of cognitive load on decision speed and confidence.

H2: Lower cognitive load will have a positive effect on decision speed and confidence

While some research concludes that AI subjectivity may impact attitude (Jia et al., 2024), other studies have found that reduced cognitive load will have a positive impact on purchase intention (Mou et al., 2020). Moreover, within our experiment, emotional appeal still remains within the AI-generated review summary, and the user-generated reviews are also present on the PDP, which aligns with the fact that emotional appeal improves purchase intention (Yuanyuan et al., 2009). Furthermore, previous research has highlighted that positive and pleasant reviews directly impact purchase intention (Guo et al., 2020). With lower cognitive load and emotional appeal still present, we expect purchase intention and product attitude to be higher when AI review summaries are utilised. To add to this, positive product attitude is closely linked to higher purchase intention and thus when consumers can obtain relevant information through a lower cognitive load with emotional abudannce, purchase intention will increase as a result of positive product attitude.

H3: AI review summaries will have a positive effect on purchase intent and product attitude

Social proof relates to the phenomenon whereby individuals create cognitive shortcuts based on the behaviour of others, specifically when individuals are faced with uncertainty (Cialdini., 2001). Moreover, AI-generated review summaries do not provide direct human endorsement, which may result in lower authenticity and credibility compared to regular user-generated reviews (Isaac & Greyson, 2017). Individuals are likely to create mental shortcuts when there is a sense of trust between the reviewee and the individual, which is not well established between individuals and generative AI as a result of response hallucinations (Chelli et al., 2024 ; Jia et al., 2024). In consequence, while useful, AI summary statements would potentially be given a lower ranking in perceived authenticity because no social endorsement indicators are given. This leads us to hypothesize that perceived authenticity will be lower for AI generated review summaries.

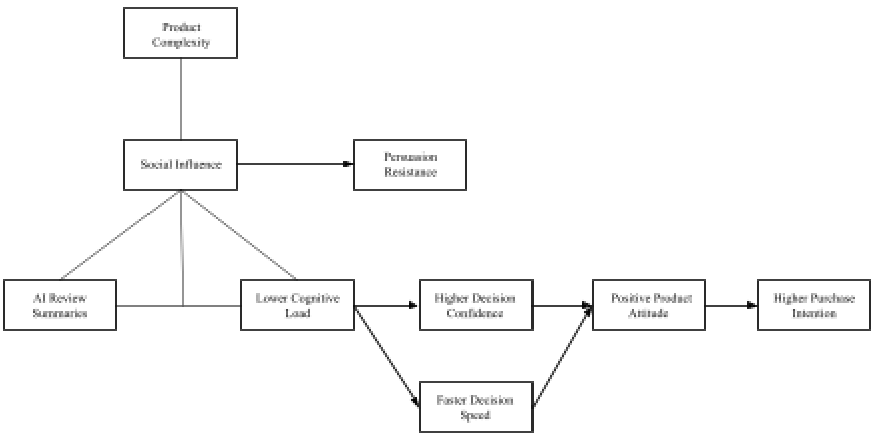
H4: AI review summaries are not considered human by the standards of social proof, resulting in lower perceived authenticity

Despite the adoption of digital tools to reduce cognitive load, persuasion resistance may arise, which results in reluctance and avoidance (Fransen et al., 2015). Reluctance may emanate from the perception of deception, which research has shown is more prevalent than expected in AI machine learning models (Park et al., 2024). Moreover, the non-human characteristics and lack of emotional appeal of AI may reduce the effectiveness of reviews as they are not perceived as authentic or credible. AI authorship has also been presented as problematic in previous research, as people perceive AI authorship as biased for typically human-made content (Jia et al., 2024). Adhering to the Persuasion Knowledge Model (PKM), consumers also build knowledge about persuasive techniques over a lifetime. When AI is openly flagged or recognized, consumers have access to such knowledge and can reject persuasion if a message is not of perceived believable or with emotional appeal (Araujo, 2018; Park et al., 2024). H5 thus hypothesizes that machine-composed summaries eradicate persuasive power.

H5: AI review summaries create persuasion resistance

2.8.1 Perceptual Map

The perceptual map highlights the four central theories that are used as the foundations of this research, as covered in the conceptual background. These four theories presented are Social Proof (SP), Cognitive Load Theory (CLT), Persuasion Knowledge Model (PKM) and Product Complexity (PC). By contrast, as highlighted in the literature review, purchase intention, credibility & authenticity, decision confidence, product attitude, and persuasion resistance are variables used within the frameworks of these theories. The map highlights the connections between the theories and what the hypotheses aim to measure.



3.0 Pre-Study

The following segment encapsulates the pre-study, which tests for product complexity and involvement in the scope of the chosen products.

3.1 Methodology

We adopt a two-step approach to answer the research questions presented earlier *(see problem statement and research questions)*. While the pre-study does not answer the research questions directly, it helps to create feasible interaction variables that support research question two. We conduct a pre-study to determine product complexity and product involvement for two products that have the same price category and are gender neutral. To add to this, before considering the two products to be used in the stimuli, we evaluated several product complexity indices (PCI) lists to determine feasible products to further investigate. Firstly, we conducted a questionnaire with two gender neutral products, namely a plain white T-shirt and a suitcase. However, we found that despite a small statistical difference between the two, it was not enough to establish a difference with confidence. Nevertheless, we conducted the same questionnaire between a plain white T-shirt and an electric toothbrush, which has a large difference in product complexity, as highlighted in multiple PCI lists (PCI Rankings, 2025). The questionnaire conducted had a relatively small sample *(N=50)*, but as previously mentioned, the results from the pre-study are not used to conclusively investigate the results of the hypotheses. We use a quantitative approach with ordinal Likert scale questions to establish the statistical difference and significance of our results. The questionnaire shows one of two stimuli to the participant, with an exposure time of a minimum of 7 seconds, to ensure proper observation of the product and the product description. Subsequently, participants are asked to rank their opinions on product complexity and involvement. The product description was included in the pre-study, as the same stimuli were used in the experimental study. However, summaries and AI-generated summaries were excluded from the pre-study questionnaire. (see Fig. 1 and Fig. 2 in the appendix for questionnaire stimuli).

For the sake of validity, a categorical (nominal) scale was also used to determine how complex each product is perceived to be. This scale was used to further test our hypotheses on product complexity below.

H0 = Product complexity and involvement are statistically the same across the two stimuli
H1 = Product complexity and involvement are statistically different across the two stimuli

3.2 Sampling

We used convenience sampling for the pre-study to quickly evaluate the possible products for the experiment. This allowed for easy access to an available pool of participants.

3.3 Results

We perform several statistical tests to ensure that the products used in the experimental study are solid and to test *H5 (see hypotheses)*. Moreover, we perform statistical tests using an Independent samples t-test for the Likert ordinal scale and Pearson's Chi-squared test to test the statistical significance of the categorical (nominal) scale. The results from these tests are presented below.

3.3.1 Descriptive Statistics

The following table highlights descriptive statistics of the results to show the mean for each question regarding complexity and involvement. Each column represents one of the questions regarding either complexity or involvement *(e.g. Complexity 1: 1st question on product complexity)* (see Fig. 4 & Fig. 5 in the appendix for actual questions). This is important as we compute a mean rank of the questions to create an aggregated score for each category to perform statistical testing.

Table 1.

Product complexity

Stimuli		**Complexity 1**	**Complexity 2**	**Complexity 3**
Electric Toothbrush				
	Mean	5.76	5.84	5.56
	SD	1.09	0.90	1.23
	N	25	25	25
T-Shirt				
	Mean	3.32	3.57	3.70
	SD	0.90	1.41	1.24
	N	25	23	25

Product Involvement

Stimuli		**Involvement 1**	**Involvement 2**	**Involvement 3**
Electric Toothbrush				
	Mean	5.72	5.20	6.32
	SD	1.20	1.55	1.15
	N	25	25	25
T-Shirt				
	Mean	3.80	3.42	4.08
	SD	0.82	1.14	1.63
	N	25	24	25

3.3.2 Independent Samples T-Test

Moreover, due to the nature of the sample, we perform an independent samples Mann-Whitney U test to test for the statistical difference between the two stimuli, namely the t-shirt and the electric toothbrush. The three different questions are recorded for this test to create a grouped variable of the total score for both complexity and involvement.

The following tables highlight the results of each test, along with the null hypothesis.

Table 2.
Product Complexity

Total N	50
Mann-Whitney U	28.0
Wilcoxon W	353
Test Statistic	28
Standard Error	51.3
Standardised Test Statistic	-5.55
Asymptotic sig. (2-sided test)	< 0.001

H0 = The distribution of product complexity is the same across the two stimuli

At the significance level of 5% (0.050), we can reject the null hypothesis as $p < 0.001$. To add to this, we can therefore assume that there is a statistical difference between the two stimuli, in regards to product complexity.

Table 3.
Product Involvement

Total N	50
Mann-Whitney U	55.5
Wilcoxon W	380.5
Test Statistic	55.5
Standard Error	51.23
Standardised Test Statistic	-5.02
Asymptotic sig. (2-sided test)	<0.001

H0 = The distribution of product involvement is the same across the two stimuli

At the significance level of 5% (0.050), we can reject the null hypothesis as $p < 0.001$. Therefore, we can assume that there is a statistical difference between the two stimuli in the scope of product involvement.

3.3.3 Chi-Squared test

As previously mentioned, for the sake of validity, a categorical (nominal) scale was also used to determine how complex each product is perceived to be. We utilised a Chi-squared test to determine the general attitude towards product complexity and involvement with simple yes or no questions.

The chi-squared test was run for two separate questions, namely difficulty of understanding the product's features and attributes (complexity) and time spent researching the product prior to making a purchasing decision (involvement).

Table 4.

Complexity	*Value*	*df*	*Asymptotic Sig.*	*Exact Sig.*
Pearson's Chi-Squared	28.78	2	<0.001	<0.001
Fisher's Exact Test	32.58			<0.001
Likelihood Ratio	36.31	2	<0.001	<0.001
Linear By Linear Association	28.2	1	<0.001	<0.001
N Valid Cases	50			

H0 = The distribution of product complexity is the same across the two stimuli
H1 = The distribution of product complexity is different across the two stimuli

Table 5.

Involvement	*Value*	*df*	*Asymptotic Significance*
Pearson's Chi-Squared	12.2	2	<0.001
Likelihood Ratio	13.4	2	<0.001
Linear By Linear Association	11.95	1	<0.001
N Valid Cases	50		

H0 = The distribution of product involvement is the same across the two stimuli
H1 = The distribution of product involvement is different across the two stimuli

Based on the results from the statistical analysis, we can reject the null hypotheses for both product complexity and involvement and accept the alternate hypotheses ($p < 0.001$), concluding that there is a statistical difference between the two stimuli in regards to complexity and involvement.

3.3.4 Overview Of Results

In conclusion, the results from the pre-study highlight that the two products that we use in the experimental study are significantly different in terms of complexity and required involvement. Moreover, the descriptive statistics show that the electric toothbrush has higher perceived complexity than the T-shirt, which goes hand in hand with the PCI list mentioned earlier. It is important to note that the findings of the pre-study are used to test product complexity as a mediator for the hypotheses previously presented.

4.0 Experimental Study

The following segment consists of the experimental study that we conducted to test our hypotheses.

4.1 Methodology

We use an experimental approach in our primary research that aims to answer our research questions and hypotheses linked to previous literature. We utilise a questionnaire with four different stimuli that are presented randomly, but with equal distribution. The core aim of the questionnaire is to simulate a real E-commerce web page to get closer to the real effects of how AI review summaries affect the dependent variables.

4.1.1 Questionnaire design & Stimuli

The four different stimuli are highlighted in the figure below for clarity.

Table 6.

Stimuli	*Product*	*Reviews*	*AI Review Summary*
1	T-Shirt	Present	Not Present
2	T-Shirt	Present	Present
3	Electric Toothbrush	Present	Not Present
4	Electric Toothbrush	Present	Present

We use a simple product page template to reduce attention to other areas of the product page. To add to this, we use plain images of the products with no brand or logo present, as well as no specific design feature that would indicate that the products are of a specific brand. Moreover, to ensure fairness across the different stimuli, the price of both the T-shirt and the electric toothbrush is the same, as well as the total review count and rating. Additionally, participants are presented with 9 different reviews, which have an aggregated rating that adds up to the total rating score of the product. Also, for stimuli two and four, the product page presents itself in the exact same way as the other stimuli, but with an AI-generated review summary of the nine reviews also present. Furthermore, the nine reviews on each product page have the same attitude and tone, with the only difference being the product features of the specific product. It is also worth noting that generative AI was used to create the AI review summary. Finally, the label "AI-generated review" as the title for the AI-generated review summary is the same as Amazon used during its introduction in 2024 (Amazon, 2024).

4.1.2 Questionnaire Structure

The questionnaire was created in a manner that would ensure that participants were not filling in the answers too quickly without thought. We utilise Likert scale questions where respondents are asked to rank statements on a scale from 1-7 concerning agreeableness. To add to this, we also utilise categorical questions to ensure that we provide a nuanced understanding of the participants' responses. The categorical questions were used for both data gathering on demographics and psychographics. Furthermore, categorical questions were also used to test for cognitive load and bias. This approach was adopted as it allows for comparisons and subgroup analysis to examine trends among different demographics. Additionally, a combination of question types reduces response bias as respondents may feel forced into a binary choice that doesn't fully capture their views or opinions for categorical questions only. Therefore, Likert scale questions allow for more nuanced opinions that reduce response bias.

The questionnaire contains 7 different blocks which measure different variables, based on the respondents' views and opinions from the viewed stimuli, whereby 6 were used for hypothesis testing.

Dependent variables

Credibility & Authenticity
Decision Confidence
Purchase Intention
Product Interest & Attitude
Persuasion Knowledge
Cognitive load

(Figures 6,7,8, and 9 in the appendix show each stimulus that was used in the experiment)

4.1.3 Procedure

As previously highlighted, respondents are shown one of the four stimuli randomly. The distribution of the stimuli is roughly equal, with each stimulus having an appearance rate of roughly one quarter of total viewings. Respondents were asked to observe the product page carefully as if they were purchasing the product for a minimum of 10 seconds, before the continue button appeared. Also, the total time spent on the product page was recorded for each participant to determine trends in time spent before making decisions on references and opinions. Followed by the product page, participants were asked to answer the questions presented (see Fig. 9), where all the questions were the same for each participant to ensure no skewing of results. Moreover, each participant was mostly presented with Likert-type ordinal questions, which allow for a better representation of participants' natural responses. To add to this, other question types, such as frequency scales and interval scales, were also used as they align with the intended statistical testing and allowed for question variation to reduce the risk of participants clicking through without close attention.

4.1.4 Questionnaire Sampling

We used convenience sampling for this experiment to achieve a large pool of available participants that would also allow for a larger sample size. The digital questionnaire (experiment) was shared on social media and other work and school-rated forums to achieve a diverse sample of different genders and ages. Moreover, the questionnaire tool used for this experiment was Qualtrics, which allowed for a strict following of GDPR guidelines to keep respondents' answers confidential and anonymous. Qualtrics also allows for a unique feature for our experiment, namely, analysis of the time spent on each stimulus. We also adopted control questions in the questionnaire to ensure that we could exclude data from respondents who had not followed the questionnaire in detail, which could skew findings. To add to this, participants were able to opt out at any moment, and informed consent was included on the opening page of the questionnaire.

4.1.5 Statistical Methods

In our research, we mainly utilize mean differences as an approach to assessing the impact of the experimental stimuli. This is mainly because, in an experimental design, comparing central tendencies across groups is suitable for determining whether observed discrepancies might be the result of stimulus manipulation or random fluctuations. According to Field (2013), mean differences represent an efficient standard approach in experimental studies since they directly measure the effect that independent variables have on quantifiable outcomes. In our analysis, when we have four varying categories with 408 respondents, one-way ANOVA is most suitable since it is suitable for comparison of mean values across more groups greater than two while keeping type I error in check. When comparing two groups, an independent sample t-test is used to ascertain whether the mean response is statistically significant across the two groups. These tests offer reliable evidence in terms of how stimuli impact outcome variables in differing experimental settings.

4.1.6 Limitations Of Methodology

One of the drawbacks of the experimental design and methodology relates to the lack of full ecological validity. While we aim to get as close to a naturalistic setting by creating a simulated e-commerce environment without external factors that may affect attitudes, the full complexities of actual consumer behaviour are hindered. Participants are aware that they are part of an experiment, which may affect their decision-making process and reduce validity due to social desirability bias. Additionally, to create a simulated e-commerce environment, we decided to use two different products, with some variation in the scope of complexity. This means that while our research gives insight into the psychological and behavioural effects of AI review summaries, it is difficult to assume the same results hold for other product categories. Moreover, the use of convenience sampling, while beneficial, may introduce potential biases in terms of representativeness, particularly if the sample is more tech-savvy, which is not measured in the experiment. Finally, the "AI-generated review" label on the stimuli may be presented in different ways to reduce negative pre-existing attitudes toward AI. However, we opted for the way Amazon presents theirs as a basis for research (Amazon,2024).

4.1.7 Participants

Based on the theoretical framework, we decided to only include demographic and psychographic variables that are related to the scope of the study. For example, persuasion knowledge improves over time based on the notion that processing ability develops over time (Wellman, 1990). To add to this, we included all ages up to 84 in the study to observe the difference across different ages. Moreover, despite other studies by Kirmani and Campbell (2004) setting 29 years as a limit to measure persuasion knowledge, it does not capture the effects on other age groups, which could be an important implication for business strategists. Also, previous research has shown that individuals with a higher level of education are more critical of information online (Metzger, 2004). Moreover, we decided to make the experiment available to all countries around the world, however, a large majority of the respondents were Swedish nationals. As the experiment is a simulation of an e-commerce environment, we decided to gather a diverse pool of respondents in the scope of age, gender and education level to validate the implications for businesses.

The approach we used resulted in a total of 432 responses across the four different stimuli presented at random to the participants. However, 8 participants opted out of the experiment as they did not give consent, 5 answered the control question wrong, and 11 exited the experiment before completion, which was highlighted in the experiment as a way of opting out in line with GDPR rules. The control question asked the participant which product they had been exposed to. This resulted in a total of 408 valid responses that were used for statistical analysis. The majority of the sample consisted of females (52.2%); the distribution of gender can be observed in Table 7. A chi-squared test was also conducted across the four stimuli, which revealed that gender was slightly different across the different experimental groups; $X^2 = 10.120,\ p = 0.018$). However, as the scope of the experimental design is randomised, the distribution of gender across the different stimuli is not possible.

All other demographics were kept consistent with the experimental design as described (Tables 8 and 9) highlight the distribution of age and education level, respectively. We conducted a Chi-squared test for both age also, which highlighted that there is an even distribution across the four different groups $X^2 = 21.289,\ p = 0.441$).

Table 7.

Gender	*Frequency*	*Percent*	*Cumulative Percentage*
Male	182	44.6	44.6
Female	213	52	96.8
Non-Binary	2	0.5	97.3
Undisclosed	11	2.7	100.0

Table 8.

Age	*Frequency*	*Percent*	*Cumulative Percentage*
<18	5	1.2	1.2
18-24	108	26.5	27.7
25-34	115	28.2	55.9
35-44	46	11.3	67.2
45-54	58	14.2	81.4
55-64	51	12.5	93.9
65-74	18	4.4	98.3
75-84	7	1.7	100.0

Table 9.

Education	*Frequency*	*Percent*	*Cumulative Perc.*
Less Than High School	5	1.2	1.2
High School	61	15.0	16.2
Some College	18	4.4	20.6
Bachelor's Degree	133	32.6	53.2
Master's Degree	181	44.4	97.5
Doctorate	10	2.5	100.0

4.1.8 Data Analysis Plan

The following part contains a description of how the data was used to conduct statistical testing in SPSS. Moreover, in the following segments, the products used in the experiment are referred to as Product 1 (Fig. 3) and Product 2 (Fig. 2). It is also important to note that Product 1 is the less complex product and Product 2 is the more complex product. For the hypotheses that we test in the results segment, we use an array of statistical tests to both test for statistical significance and effect sizes. To add to this, for each hypothesis that is tested, we create a twofold analysis. Firstly, we determine the statistical difference between AI review summaries and regular reviews on an aggregate level, which entails that the two different products are combined into two categories, namely, "AI review present" and "AI review not present" (see Table 10). Secondly, we test for the moderation variable, specifically product complexity. This allows for a clear understanding of whether the effect of AI-generated review summaries varies for different product complexities. Through this approach, we can capture the main effect of AI review summaries on consumer behaviour and attitudes while also investigating the effects of product complexity. This ensures that both direct and moderated effects are captured within the analysis.

The statistical analysis is conducted in SPSS. Depending on the nature of the dependent variable and hypothesis, we adopt relevant statistical testing such as T-tests, One-Way Anova, Kruskal-Wallis test and visualisation. Where appropriate, we also employ effect sizes such as Cohen's d and Eta-Squared to supplement the p-levels and create a clear image of the magnitude of the effect. Furthermore, assumptions such as homogeneity of variances for each test are checked before interpretation. Any violations of these assumptions, as clearly stated in the limitations or are addressed using adjusted statistics.

The distribution of the two different categorisations of groups is highlighted in Tables 10 and 11.

Table 10.

Stimuli Exposure	*Frequency*	*Percent*	*Cumulative Perc*
Product 1 With AI Review	101	24.8	24.8
Product 1 Without AI Review	96	23.5	48.3
Product 2 With AI Review	101	24.8	73.0
Product 2 Without AI Review	110	27.0	100

Table 11.

Stimuli Exposure (Aggregated)	*Frequency*	*Percent*	*Cumulative Perc*
AI Review Present	202	49.5	49.5
AI Review Not Present	206	50.5	100.0
(Total)	(408)		

4.2 Results

Before proceeding with the hypothesis testing, we conducted manipulation checks on the stimulus groups to ensure that participants saw that an AI-generated review summary was actually present on the PDP. Following these checks, we present the main results.

4.2.1 Manipulation Checks

Manipulation checks were done on two different variables to ensure that participants perceived the independent variable as intended. Firstly, as previously highlighted, a control question asking participants which product they were exposed to was used as a manipulation check for product complexity. Participants who did not answer this question correctly were excluded from the results, and thus, no statistical testing is required. All 408 participants who garnered valid results through the control question are used for analysis.

Secondly, to determine whether participants had seen the AI review summary, we included a yes or no question that was presented at the end of the experiment for the two groups exposed to an AI review summary. The question asked participants if AI was prevalent on the PDP that they had been exposed to. The question was strategically placed in the end to avoid social desirability bias. To test this, a One-way chi-squared test was undertaken. The One-way chi-squared test showed that the manipulation worked as a vast majority of participants exposed to the AI review summary saw that AI was present; $p < 0.001$ $X^2 = 23.229$. The expected and observed values are displayed in Table 12. Moreover, the manipulation surpasses the benchmark of 80%, which is typically the minimum a manipulation check should have (Oppenheimer et al., 2009).

Table 12.

One-Way Chi Squared Test	*Observed*	*Expected*	*Residual*
Yes	189	161.6	27.4
No	13	40.4	-27.4
Total	202		

Note that expected values were set to 80% for Yes and 20% for No to meet the benchmark.

4.3 Hypothesis Testing

The following segment contains a statistical analysis of all the hypotheses. Every hypothesis test includes a brief walkthrough of statistical methods, validity testing and actual results and interpretations of the results.

4.3.1 AI review summaries' effect on cognitive load

For the first hypothesis, we first conducted a one-way ANOVA to observe the differences between all the different stimuli. For this test, we created a mean of the three different variables concerning how difficult information was to process, both in terms of the PDP, reviews and the test altogether. We tested this mean value, indicating cognitive load on the PDP, for the four different stimulus groups. The results from ANOVA highlighted that there was a meaningful difference between at least two of the groups; $F = 4.293$, $p = 0.005$ (Table 13). Moreover, the Eta-squared of 0.065 entails that 6.5% of the variance in cognitive load can be explained through the stimuli. This further shows that the effect size can be considered medium-sized.

To add to this, the Tukey (HSD) also showed that both stimulus groups that were also presented with an AI-generated review summary had lower cognitive load, as highlighted in Table 13. Interestingly, the lowest cognitive load was observed for the more complex product with an AI review summary also presented ($M = 3.8218$).

Table 13.

ANOVA	*Sum Of Squares*	*df*	*Mean square*	*F*	*Sig.*
Between Groups	15.089	3	5.030	4.293	0.005
Within Groups	472.119	403	1.172		
Total	487.208	406			

Table 14.

Stimuli Exposure	*Frequency*	*Mean*	*Std. Deviation*
Product 1 With AI Review	101	3.7033	1.08339
Product 1 Without AI Review	96	3.3264	1.11132
Product 2 With AI Review	101	3.8218	1.00395
Product 2 Without AI Review	110	3.4636	1.12437

A higher mean in this test indicated lower cognitive load (reversely coded)

Anyhow, to test the four categories on an aggregated level, by consolidating the two groups presented with an AI review summary and the two without, we ran an independent samples T-test. The test showed that significant difference in cognitive load; t = 3.390, p = 0.001 (Table 15). We assumed unequal variances in this instance as Levene's test showed statistical significance of p = 0.009 (Table 16). This result shows that one of the groups experienced significantly lower cognitive load than the other, and as highlighted in Table 17, the participants presented with an AI-generated review summary experienced lower cognitive load and perceived the information to be easier to process than the other group. The effect size between the two groups is small to moderate according to Cohen's d; d = 336.

Table 15.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variance Not Assumed	3.3390	404	<0.001	<0.001	0.3318	0.10714

Table 16.

Levene's Test For Equality Of Variances	*F*	*Sig.*
Equal variances Assumed	6.830	0.009

Table 17.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present	202	3.7629	1.04330	0.07359
AI Review Not Present	206	3.3397	1.11769	0.07787

A higher mean indicates lower cognitive load (reverse-coded)

Finally, to further test the moderating variable, product complexity, we ran another T-test between the two different groups presented with an AI-generated review summary for product 1 and product 2. As Levene's statistic is not significant at p = 0.472, F =0.520, we assume equal variances. Furthermore, the test result shows that there is no statistical significance between the two different products in this instance; t = -0.804, p = 0.442 (Table 18). Moreover, the mean comparison can be observed in Table 19, which further highlights the very small difference between the two groups.

Table 18.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variances Assumed	-0.804	199	0.211	0.442	-0.11845	0.14731

Table 19.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Product 1 With AI	101	3.7033	1.08339	0.10834
Product 2 With AI	101	3.8218	1.00395	0.09900

With the tests for this hypothesis, we can conclude that AI-generated reviews reduce cognitive load, and individuals find it easier to process the content of reviews when an AI-generated summary is present along with the user-generated reviews. However, there is no difference between products between different product complexities in this regard; $p = 0.442$.

4.3.2 The effect of AI review summaries on decision speed and confidence

For the second hypothesis, we use two different variables to evaluate decision speed and decision confidence. Firstly, we recorded the time it took for individuals to proceed with the experiment after viewing the PDP The timer started the second the PDP was displayed, and the timer stopped when the "continue" button was pressed. Participants were asked to carefully evaluate the information on the PDP before proceeding to the questions. This variable is the decision speed. Moreover, similarly to the other dependent variables, we combined questions on decision confidence into a mean variable.

Decision speed

Firstly, we check the normality of residuals to determine if the distribution is normal between all stimulus groups. A Shapiro-Wilk test showed that the normality assumption is violated as $p < 0.001$ for all groups. Due to the nature of the data, we conduct a Kruskal-Wallis test to determine the difference in time taken to process the PDP for all the groups. Table 20 highlights the results from the test.

Table 20.

Kruskal-Wallis Test Summary	*Value*
Total	408
Test Statistic	1.897

Degrees Of Freedom	3
Asymptotic Sig.	0.594

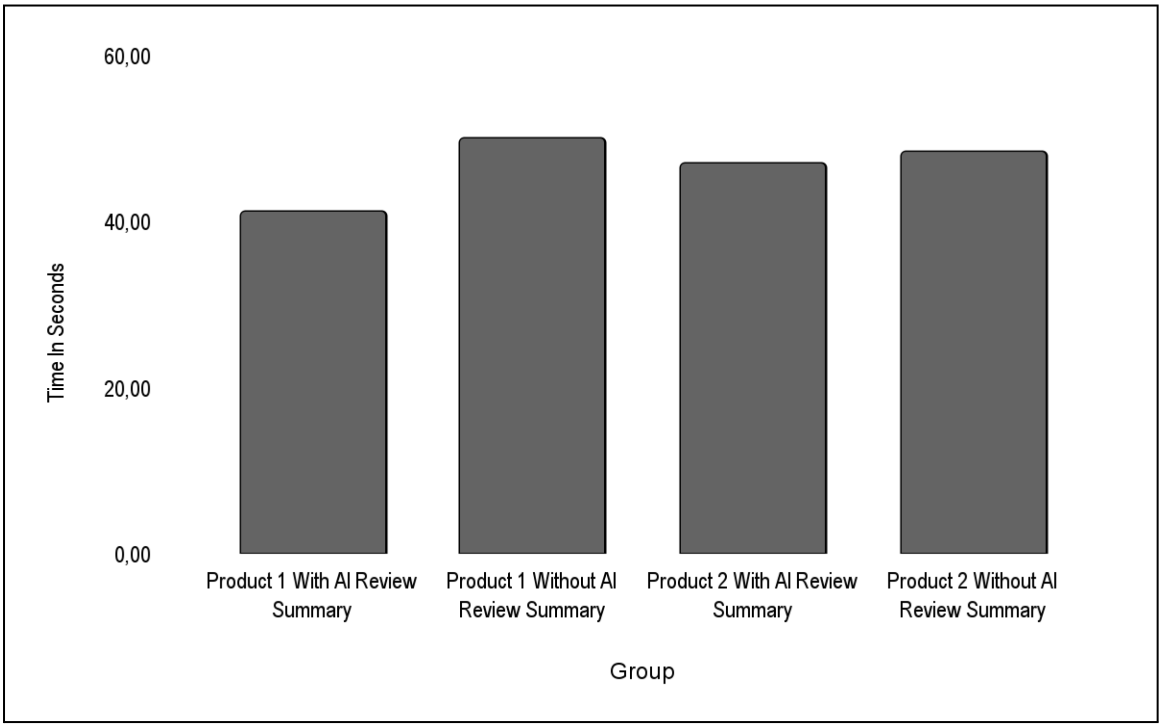
To further check the difference between the different groups in terms of decision speed, a pairwise comparison was also made to determine the significance between paired groups. The pairwise comparison showed that no pair was significantly different, with all p-values above 0.05, as highlighted in Table 21. We can thus reject the null hypothesis and conclude that there is no significant difference between the groups in a pairwise comparison.

Table 21.

Pairwise Comparison	*Test Statistic*	*Std.Error*	*Std.Test Statistic*	*Sig*
Product 1 With AI vs Product 2 With AI	-13.465	16.594	-0.811	0.417
Product 1 With AI vs Product 2 Without AI	-19.624	16.251	-1.208	0.227
Product 1 With AI vs Product 1 Without AI	-19.839	16.808	-1.180	0.238
Product 2 With AI vs Product 2 Without AI	-6.159	16.251	-0.379	0.705
Product 2 With AI vs Product 1 Without AI	6.374	16.808	0.379	0.705
Product 2 Without AI vs Product 1 Without AI	0.215	16.470	0.013	0.990

As shown both from the statistical testing and Figure 1, there is no significant difference in decision speed between the four different groups.

Figure 1. Histogram of Time Taken In Seconds To Proceed From PDP



This concludes that in terms of decision speed between the different groups, there is no meaningful difference.

Decision Confidence

To test whether decision confidence differs across the different groups, we first utilise Welch's ANOVA for all four groups. The test entails that there is a significant difference between at least one pair of the four different groups; $F = 5.602$, $p < 0.001$. Moreover, we conduct a robustness test for the equality of means, which further highlights that there is a significant difference between the groups, even when variances are unequal; $p = 0.001$. This is also shown in Table 22.

Table 22.

	Statistic	*df1*	*df2*	*sig*
Welch	5.582	3	223.388	0.001

To further evaluate the differences in decision confidence between the different groups, we ran a Games-Howell post hoc test to pinpoint exact differences. We utilise this specific test as equal variances are not assumed. Table 23 displays the results from this test.

Table 23.

Comparison (Games-Howell)	*Mean Difference*	*p*	*Interpretation*

Product 1 With AI vs Product 1 Without AI	+0.51093	0.015	Higher For Product 1 With AI
Product 2 Without AI vs Product 2 With AI	-0.50383	0.032	Higher For Product 2 With AI
Product 1 Without AI vs Product 2 Without AI	+0.04735	0.991	No Significance
Product 1 With AI vs Product 2 With AI	+0.05446	0.992	No Significance

As highlighted in Table 23, AI-generated review summaries do increase decision confidence; $p = 0.015$ & $p = 0.032$. However, the effect size is relatively small as the Eta-Squared is 0.040. This can be considered a small to medium-sized effect. For a robustness check, a T-test on an aggregated level was also performed, which evaluates the mean difference between all groups exposed to an AI review summary to the two other groups which were not. The T-test exhibits that there is a meaningful difference between the two groups; $t = 4.083$, $p < 0.001$. Table 24 further highlights the test statistics from the T-test.

Table 24.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variances Not Assumed	4.083	395.45	<0.001	<0.001	0.50899	0.12467

Equal variances not assumed as Levene's statistic = 0.016

Moreover, by looking at the mean comparison (Table 25), we can conclude that when AI review summaries are present on the PDP, individuals feel more confident in making the right purchasing decision.

Table 25.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present	202	4.9629	1.34556	0.09468
AI Review Not Present	206	4.4539	1.16412	0.08111

We can therefore conclude that AI review summaries increase decision confidence, but not decision speed. Also, product complexity has no significant effect on decision speed and confidence.

4.3.3 AI review summaries' effect on purchase intention and product attitude

Similarly to the other hypotheses, we first conduct a one-way ANOVA for all four stimulus groups to observe differences in both product attitude and purchase intention. Also, the questions regarding both purchase intention and product attitude were re-coded into a mean variable based on all questions in the scope of the specific variable, namely, product attitude and purchase intention. Following the one-way ANOVA, we also conduct a T-test on an aggregated level to test AI review-generated review summaries against user-generated review summaries with both products combined. We then run the same test for both dependent variables with the moderating variable, product complexity. Firstly, the tests on purchase intention are presented, followed by product attitude.

Purchase intention

The one-way ANOVA for purchase intention for all stimulus groups presents a small yet significant difference between at least one pair of the groups; $F = .593$, $p < 0.001$ (Table 26). Levene's test for homogeneity of variances entails that the assumption of equal variances is not valid in this instance ($p = 0.029$); however, based on the median and adjusted degrees of freedom, $p = 0.067$, which assumes variances are equal. Due to the large sample size and the ANOVA's robustness, we assume equal variances.

Table 26.

ANOVA	*Sum Of Squares*	*df*	*Mean square*	*F*	*Sig.*
Between Groups	36.264	4	12.088	5.593	<0.001
Within Groups	873.141	404	2.161		
Total	909.405	407			

Moreover, the post hoc test (Tukey HSD) showed that AI-generated review summaries increased purchase intention for both products. The largest difference between all four stimulus groups was for the more complex product accompanied by an AI-generated review summary, which had a significantly higher level of purchase intention compared to the groups not exposed to an AI-generated review summary. However, in the scope of product complexity, there was no significant difference. Table 27 showcases that the AI-generated review summaries increase purchase intention.

Table 27.

Comparison (Tukey)	*Mean Difference*	*p*	*Interpretation*
Product 1 With AI vs Product 1 Without AI	+0.54074	0.050	Higher for Product 1 With AI
Product 2 Without AI vs	-0.66157	0.012	Higher for Product 2 With AI

Product 2 With AI			
Product 1 Without AI vs Product 2 Without AI	-0.11989	0.937	No Significance
Product 1 With AI vs Product 2 With AI	-0.1947	0.783	No Significance

The aggregated groups also displayed the same effect, with AI-generated review summaries outperforming user-generated reviews in the scope of purchase intention; $t = 3.941$, $p < 0.001$ (Table 28). The mean difference between the groups exposed to an AI-generated review summary and the groups who were not was large enough by the standards of Cohen's d to be considered a medium to large effect size; $d = 0.574$. The mean differences can be observed in Table 29.

Table 28.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variances Not Assumed	3.941	392.6	<0.001	<0.001	0.57408	0.14567

Equal variances not assumed as Levene's statistic = 0.003

Table 29.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present	202	4.6485	1.58671	0.11164
AI Review Not Present	206	4.0744	1.34300	0.09357

Finally, to test the significance between the two groups exposed to AI-generated review summaries for both products, we ran a T-test. In line with the One-way ANOVA, there was no significant difference between the two different products; $t = -0.872$, $p = 0.384$, MD = -0.19472, Std. Error = 0.22341. Table 30 shows the mean difference between the two groups.

Table 30.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Product 1 With AI	101	4.5512	1.61963	0.16116
Product 2 With AI	101	4.7459	1.55503	0.15473

Equal variances assumed as Levene's statistic = 0.573

To conclude, purchase intention is higher when a PDP contains an AI-generated review summary, compared to only user-generated reviews. However, product complexity seems to have no effect in this regard.

Product Attitude

Similarly to purchase intention, we run the same statistical tests for product attitude. Firstly, a One-Way ANOVA was utilised to test the differences between all four groups. The test revealed a significant difference between at least one of the pairs; $F = 6.933$, $p < 0.001$ (Table 31). Levene's test for homogeneity of variances also shows that equal variances are assumed; $p = 0.020$. Furthermore, the effect size between the different groups is relatively small with an Eta-squared of 0.049, which entails that further statistical testing must be undertaken to further evaluate the mean differences for each group.

Table 31.

ANOVA	*Sum Of Squares*	*df*	*Mean square*	*F*	*Sig.*
Between Groups	33.678	3	11.226	6.933	<0.001
Within Groups	654.184	404	1.619		
Total	687.862	407			

To visualise the difference between each group, we also ran Tukey's HSD post hoc test, which revealed that the significant difference in groups is between AI-generated reviews and user-generated reviews and not between different product complexities. Table 32 highlights the mean differences and the significance between each relevant group.

Table 32.

Comparison (Tukey)	*Mean Difference*	*p*	*Interpretation*
Product 1 With AI vs Product 1 Without AI	+0.4778	0.043	Higher for Product 1 With AI
Product 2 Without AI vs Product 2 With AI	-0.62914	0.002	Higher for Product 2 With AI
Product 1 Without AI vs Product 2 Without AI	-0.06903	0.980	No Significance
Product 1 With AI vs	-0.22030	0.608	No Significance

Product 2 With AI

To add to this, we also ran T-tests on an aggregated level to firstly determine the difference between the groups exposed to an AI-generated review summary and the ones who were not in the scope of product attitude. The test revealed that there is a significant difference between the group that was exposed to an AI-generated review and the two groups that were not; $t = 4.370$, $p < 0.001$ (Table 33). This further entails that AI-generated review summaries improve product attitude, as further highlighted in Table 34.

Table 33.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variances Not Assumed	4.370	396.6	<0.001	<0.001	0.55116	0.12613

Equal variances not assumed as Levene's Statistic = 0.005

Table 34.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present	202	4.8849	1.36090	0.09575
AI Review Not Present	206	4.3337	1.17830	0.08210

Finally, we tested the statistical significance for product complexities through the same test, which revealed that there is no meaningful difference between the two different products; $p = 0.251$ (Table 35). The test does show that the more complex product scored a slightly better product attitude than the simple products, as displayed by Cohen's d; $d = -0.162$. However, this effect is not large enough to be considered statistically significant.

Table 35.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variances Assumed	-1.151	200	0.125	0.251	-0.2230	0.19135

Equal Variances assumed as Levene's statistic = 0.083

To summarise, similarly to purchase intention, there is a meaningful difference between the group that was exposed to an AI-generated review summary. This group scored higher in product attitude than the other groups, only exposed to user-generated reviews. However, product complexity does not have a statistically significant effect in terms of product attitude.

4.3.4 Perceived authenticity of AI in the scope of social influence

To test hypothesis four, we first conducted a One-way ANOVA between the four different experimental groups to determine if there was any observable difference in the scope of authenticity and credibility of the information from reviews in the PDP. The Anova revealed that there was no statistical difference between the four different groups; $p = 0.537$, Eta-Squared = 0,005, Table 36. As highlighted, the effect size was very small and can thus be interpreted as a non-effect. To further investigate the lack of significance, we conducted post-hoc analyses utilising Tukey HSD, which also controls for type 1 errors. At the significance of $p = 0.476$, we can conclude that the post-hoc test supports the ANOVA: there is no meaningful difference between the different groups (Table 37).

Table 36.

ANOVA	*Sum Of Squares*	*df*	*Mean square*	*F*	*Sig.*
Between Groups	2.905	3	0.968	0.725	0.537
Within Groups	539.514	404	1.335		
Total	542.419	407			

Table 37.

Tukey	*N*	*Subset For Alpha*
Product 1 With AI	96	4.7917
Product 1 Without AI	110	4.9159
Product 2 With AI	101	4.9678
Product 2 With AI	101	5.0248
Sig.		0.476

To further investigate this hypothesis, we conducted a T-test on an aggregated level, combining the products into one category, creating a more holistic view of AI review summaries and regular customer-generated reviews. As highlighted in Table 38, the mean comparison highlights that there is no difference between the two groups. The T-test garnered a result of sig. = 0.227, Mean Difference = 0.13828, Std. Error = 0.11435 (Table 39). These test results were based on the assumption that variances were equal, as shown by Levene's test prior to analysis (sig. = 0.172, Table 40). This entails that on an aggregated product level, there is no meaningful difference between AI-generated review summaries and user-generated summaries on an aggregate level. This was also shown by a very small effect size; Cohen's d = 0.12.

Table 38.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present	202	4.8863	1.11287	0.07830
AI Review Not Present	206	4.8580	1.19253	0.08309

Table 39.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variance Assumed	1.210	406	0.113	0.227	0.13828	0.11435

Table 40.

Levene's Test For Equality Of Variances	*F*	*Sig.*
Equal variances Assumed	1.875	0.172

Finally, to compare the effects of product complexity in the scope of AI review summaries, we ran a T-test for the stimulus groups that were presented product 1 and product 2 with an AI review summary. The test revealed; $p = 0.717$, $t = 0.363$ (Table 43). This entails that there is no significant difference in regards to product complexity for AI review summaries. We assumed that variances were equal for this test due to the results from Levene's test; $F = 1.107$, $p = 0.294$ (Table 42). Moreover, the effect size was very small as Cohen's d was a mere 0.051, which shows that there is practically no difference between the groups. The mean statistics are highlighted in Table 41.

Table 41.

Mean Comparison	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
AI Review Present (Product 1)	101	5.0248	1.01766	0.10126
AI Review Present (Product 2)	101	4.9678	1.20502	0.11990

Table 42.

Levene's Test For Equality Of Variances	*F*	*Sig.*
Equal variances Assumed	1.107	0.294

Table 43.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variance Assumed	0.363	300	0.359	0.717	0.05693	0.15694

The results for this hypothesis highlight that the different stimuli in this experiment had no effect on credibility and authenticity on both an aggregated level and for different product complexities. Therefore, we can conclude that hypothesis four has no effect and can thus be rejected.

4.3.5 Persuasion resistance and AI review summaries

For the final hypothesis, we test how AI review summaries affect persuasion resistance. Based on the questions that evaluated how persuasive respondents felt the information was in all the text that was available on the PDP, including user-generated reviews and the AI-generated review summary for the respective groups, we created a mean variable. Firstly, we determine the statistical significance between all four groups in the capacity of persuasion resistance using a One-way ANOVA. The results highlight that there is a difference among at least one pair of the four different groups; $F = 3.046$, $p = 0.029$ (Table 44). Moreover, the effect size between the different groups was relatively small; Eta-Squared = 0.052.

Table 44.

ANOVA	*Sum Of Squares*	*df*	*Mean square*	*F*	*Sig.*
Between Groups	11.260	3	3.753	3.046	0.029
Within Groups	497.730	404	1.232		
Total	508.990	407			

To further visualise the difference between the different groups, we ran a post hoc (Games-Howell) test. The test showed that there was no significant difference between the relevant groups within this study, as highlighted in Table 45. The only observable difference, which is also reflected in the ANOVA statistic of $p = 0.029$, is that individuals exposed to product 2 without an AI review summary perceived the information on the PDP as more persuasive than product 1 with the same layout.

Table 45.

Comparison (Games-Howell)	*Mean Difference*	*p*	*Interpretation*
Product 1 With AI vs Product 1 Without AI	+0.07740	0.956	No Significance

Product 2 Without AI vs Product 2 With AI	+0.04154	0.994	No Significance
Product 1 Without AI vs Product 2 Without AI	-0.38627	0.041	Higher for Product 2 With AI
Product 1 With AI vs Product 2 With AI	+0.26733	0.368	No Significance

Moreover, we also evaluated persuasion resistance on an aggregated level, comparing both groups exposed to AI-generated review summaries to the two that were not. We utilised a T-test for this, which further showed that there is no meaningful difference between the two groups; t = 0.043, p = 0.965 (Table 46).

Table 46.

T-Test	*t*	*df*	*One-Sided p*	*Two-Sided p*	*Mean Diff*	*Std. Error*
Equal Variance Assumed	0.043	406	0.483	0.965	0.00481	0.11087

Equal variance assumed as Levene's statistic = 0.203

Based on the One-way ANOVA and the t-test, we can conclude that there is no significant difference in persuasion resistance between the groups exposed to AI-generated review summaries and the ones who were not. However, there is an observable difference between the groups that were not shown an AI-generated review summary. This finding is not relevant for the hypothesis, but the nature of this result is covered in the limitations section of this research paper.

4.4 Overview Results

Hypothesis Conclusion	Conclusion
H1: AI review summaries will reduce cognitive load	**Accepted**
H2: Lower cognitive load will have a positive effect on decision speed and confidence	**Partially Accepted**
H3: AI review summaries will have a positive effect on purchase intent and product attitude	**Accepted**
H4: AI review summaries are not considered human by the standards of social proof resulting in lower perceived authenticity	**Rejected**
H5: AI review summaries create persuasion resistance	**Rejected**

5.0 Discussion

The objective of this research was to understand how AI-generated review summaries affect consumer behaviour and attitudes. After a comprehensive analysis of the dependent variables, this section provides a detailed discussion of the findings. Firstly, a detailed conclusion of the results is made, followed by theoretical and managerial implications.

5.1 AI review summaries' effect on cognitive load

The findings from the experiment show that participants exposed to an AI-generated review summary experienced the information on the PDP to be easier to process, despite the PDP containing more text than the groups who were not exposed to a summary. This outcome aligns well with the initial research on cognitive load theory, which entails that summaries can greatly reduce cognitive load (Sweller, 1988). Moreover, previous research on cognitive load in e-commerce has shown that keeping product information short and summarised greatly improves working memory, which supports the result of this hypothesis (Mirhoseini et al., 2021). Lurie & Mason's (Lurie & Mason, 2007) research on cognitive load and product reviews further demonstrated that consumers who read summarised review summaries are more likely to commit to a purchase. A similar result is reflected in our research, as the groups presented with the AI summary for both products perceived the information to be more easily processed, despite the abundance of information. Interestingly, processing an abundance of information can lead to cognitive stress and can often create negative attitudes and reluctance to decision making (Payne et al, 1993). With this in mind, our research posits the same conclusion as the majority of participants who felt that the information was difficult to process, were the groups were not exposed to an AI summary. It is also important to note that the participants exposed to an AI summary were shown more text than the groups which were not exposed to a summary. Despite this, the summary and perhaps the placement of the summary allowed participants to make a decision more swiftly.

In the scope of product complexity, previous research by Huang (Huang, 2000) revealed that higher product complexity has a negative effect on cognitive load, resulting in reluctance to decision making unless information is coherently displayed. Our research found that there is no meaningful difference between product complexity and cognitive load, especially when similar types of information are presented. On the contrary, Huang also highlights that novel information can create motivation to understand the product better, and in our research, an AI-generated review may be perceived as novel, resulting in higher motivation to acquire information.

From a managerial standpoint, including an AI-generated review summary can greatly reduce cognitive load and make information more easily processed. This can be particularly important for retailers with a large volume of user-generated reviews and for mobile on-the-go shopping experiences with limited screen real estate.

5.2 The effect of AI review summaries on decision speed and confidence

It has already been established in the previous hypothesis that AI summaries reduce cognitive load. Previous research in this space has shown that positive reviews are more likely to improve decision confidence (Guo et al., 2020). While the reviews the participants were exposed to were mainly positive and had the same star rating for all four groups, both on an aggregate level and review level, some reviews with negative comments were also included to make the stimuli more realistic. Despite this, decision confidence was much higher for the groups exposed to an AI review summary. This can also be explained through previous research on cognitive load theory, which suggests that summarised content results in higher decision confidence (Lurie & Mason, 2007).

Product complexity greatly influences decision confidence as products with higher levels of feature heterogeneity and feature interrelatedness require more consideration to commit to a decision (Fürst et al., 2023). However, we found that product complexity does not affect decision speed. This could also be because other variables are so similar between the two products, such as product information, price and layout. Also, consumers are more likely to spend a longer time reading reviews for products that have a wider array of features, long descriptions and a higher price (Chen et al., 2022). Out of these three variables, the complex product used in our scope of research had more features. Through our timing data, we found that there was no significant difference in time spent on the product page between all four groups. However, it is important to note that participants exposed to AI review summaries actually had more text to read, and yet still spent the same amount of time on the product page as the groups not exposed to an AI review summary. This could be an indication that participants exposed to the summary had higher decision speed, while our results show otherwise.

From a managerial view, our findings suggest that AI review summaries can streamline the purchase experience by reducing decision friction and cognitive load. For a large abundance of reviews, a summary allows consumers to swiftly filter through relevant information on the product.

5.3 AI review summaries' effect on purchase intention and product attitude

In the sphere of user-generated reviews and marketing, purchase intention and product attitude have been two of the most widely studied variables in this regard. We expected product attitude and purchase intention to increase when an AI-generated review was prevalent, which was a correct assumption based on previous research. Research has shown that customer reviews that are straight to the point positively improve product attitude (Changchit & Klaus, 2020). On the contrary, other research has shown that human-like factors are crucial for purchase intention, as user-generated reviews are testimonials containing emotions and attitudes (Yuanyuan et al., 2009). In our research, the AI-generated summary is merely a summarisation of the user-generated reviews also presented on the PDP. Thus, in the scope of cognitive load theory, the AI summary acts as a proxy to improve working memory and thus improves purchase intention (Sweller, 1988). Other research shows that valence in reviews is more important than star ratings alone (Ketelaar et al., 2015), and based on this research, we can conclude that AI review summaries should be present with the user-generated reviews, also, to display that product users have created the actual content of the reviews. Puspitasari & Aruan (Puspitasari & Aruan., 2022) also found that perceived effort that goes into user-generated reviews has a positive effect on purchase intention. In our experiment AI-generated review summaries may be perceived as efficient but potentially lacking in emotion and effort, which could moderate their effects on purchase intention in some contexts. Therefore, maintaining transparency and reviewing the origin is crucial in preserving the emotional appeal that drives purchase intention.

Similarly to the other hypotheses, no significant difference was found between the two different products in terms of product attitude and purchase intention. This contradicts previous research that suggests that more complex products require more information processing to make a purchase decision (Chen et al., 2022).
From a managerial standpoint, including an AI-generated review reduces cognitive load and improves information processing efficiency. However, to maintain the emotional appeal that drives purchase intention and positive product attitude, review transparency and origin must be maintained.

5.4 Perceived authenticity of AI in the scope of social influence

We hypothesised that perceived authenticity and credibility would diminish with an AI review summary present. Based on the mean comparison between the groups, we found that there was no significant effect on authenticity on a group or aggregate level. We based our hypothesis on Cialdini's framework of social proof, which entails that persuasive messages are only impactful through human endorsement (Cialdini, 2001). Moreover, with the established lack of trust for AI as a result of response hallucinations and its non-human characteristics, we expected authenticity to diminish (Chelli et al., 2024; Jia et al., 2024). As previously highlighted, perceived authenticity and credibility are the most important factors in terms of convincingness (Boer, 2021; Thomas et al., 2019). To add to this, 92% of consumers will trust recommendations from friends and family, and 70% will trust reviews from strangers (Nielsen., 2014). Based on this research and the association between how well consumers know the recommender, we assumed that AI will be perceived as non-human and foreign, thus reducing perceived authenticity. Adversely, we found that AI review summaries do not affect authenticity, at least when user-generated reviews are also present. It is also worth noting that the AI review summary in the PDP's contained content highlighting that the origin of the information was from previous users, similarly to the text provided on Amazon's reviews carousels (Amazon, 2024). Despite limited research on product complexity and authenticity, we stayed consistent with the moderation variable and found that product complexity does not have any effect on authenticity either. This is most likely because perceived authenticity is not derived from the product itself, but from the reviews and emotional appeal of the user-generated reviews.

The findings offer important insight into the concern that implementing AI may reduce the authenticity of a brand as a whole. AI is being utilised in content creation, customer support and other business areas, and we can conclude from our findings that if emotional appeal is apparent, there will be no negative effects on perceived trust, credibility and authenticity.

5.5 Persuasion resistance and AI review summaries

Finally, based on previous literature on persuasion knowledge and PKM, we hypothesised that there would be a higher degree of persuasion knowledge when an AI review summary is present. We depicted this from research indicating reluctance to AI due to its non-human characteristics and lack of emotion (Jia et al., 2024). Hallucination rates on current AI platforms such as GPT-3.5 are concerning for many individuals as they can be anywhere between 40%-90% percent depending on the language model (Chelli et al., 2024). Chelli et al also highlight how this negatively impacts trust for large language models such as GPT-3.5. In line with PKM, covert advertising or content that is not trustworthy leads to persuasion resistance (Mohr & Kühl, 2021). Other papers have also looked at the deception rate on a systematic basis, which in the framework of PKM leads to intransigence (Park et al., 2024). Moreover, previous research on AI through the scope of PKM and persuasion resistance has highlighted that AI is not recognised as a reliable agent, especially for complex tasks (Araujo., 2018). On the contrary, the same author recognised that for simple tasks, AI is perceived as fair and can actually reduce persuasion knowledge (Araujo., 2018). With the previous research in mind, we find, based on this experiment, that AI authorship in the context of customer reviews has little to no impact on persuasion knowledge. This contradicts previous research on the matter. However, one interesting takeaway from Araujo et al's research on the fairness of chatbots is that our experiment may be perceived as simple, thus no effect on persuasion knowledge is observed. Moreover, as the experiment includes products that participants may not be interested in, there is no persuasion resistance expressed in the findings. Persuasion knowledge in this context may depend on product interest.

Product complexity does not seem to have an impact on persuasion knowledge either. Looping back to the previous claim, it could be the case that the products included in the experiment lack interest from participants, and thus, persuasion knowledge is not as prevalent. Unfortunately, this goes beyond the scope of this research paper, but it would be an interesting moderation variable to explore further.

Despite concerns of AI authorship in previous literature, in the context of review summaries, there is no effect on persuasion knowledge. As discussed previously, AI authorship as a summary of user-generated reviews may depend on content origin and transparency. Therefore, brands may consider A/B testing AI review summaries for different products before a full-scale implementation.

6.0 Conclusions

How do AI review summaries affect consumer behaviour and attitudes compared to regular user-generated reviews?

Through our experimental approach that simulates a product page on an e-commerce website, we can conclude that AI review summaries affect consumer behaviour and attitude. Our results confirm that AI review summaries reduce cognitive load and make an abundance of reviews easier to process and garner important information. This leads to better product attitude, higher purchase intention and more decision confidence. However, through the scope of social proof and persuasion knowledge, AI review summaries do not affect perceived authenticity or persuasion knowledge, at least when user-generated summaries are also present on the product display page. While technology adds a new layer of mediation between the brand and the consumer, our research indicates that AI review summaries streamline decision-making if the summary is transparently applied.

To what extent does product complexity and involvement impact these perceptions?

Through a pre-study, we based our main experiment on two products that varied in product complexity, but maintained gender neutrality and price. Based on rigorous testing, we can conclude that product complexity did not have a significant effect on the variables measured in the scope of this study.

6.1 Research Limitations and Future Research

Just like other research, our experimental design has some flaws that are addressed in this segment. Firstly, our experiment is an artificial decision context where participants have to make a decision in a hypothetical context, for products they may not be interested in. This reduces external validity as money and personal preferences are not involved. Furthermore, while product complexity is broad, we used only two different products, which in our case did not generate any significant impact on the dependent variables. This also limits generalizability across other product categories, however, it was the only viable option to simulate a real-life PDP. Future research should observe the differences in consumer behaviour through A/B testing of different products and explore the differences in factors such as time spent on the PDP, return rates, conversion rate, cart abandonment and other relevant variables. This approach would also capture consumers in a real-life setting where external variables such as preferences and money are also prevalent.

Regarding our experimental design, we conducted a pre-study on product complexity to determine which two products to use for our main experiment. However, the participants of the main experiment on the product complexity of these two products could skew the moderation variable. Moreover, we used convenience sampling as our sampling method for the main experiment, and despite a large majority of the participants being between 18 and 34 (54,7%), a large proportion of the participants were above the age of 45. Older participants may be less tech-savvy and do not fully understand AI to the same extent as the younger generation. Moreover, a binary manipulation was used in the experiment. We compared the presence and absence of an AI review summary, but in reality, summaries can greatly vary in quality and style. Future research could potentially investigate the differences across products with high and low ratings to determine if AI review summaries impact consumer behaviour in this context. While on this topic, there is no longitudinal component that investigates long-term trust or review recall.

Additionally, the experimental design and stimulus of this study have some flaws. Most dependent variables were self-reported measures, which can be influenced by demand characteristics or social desirability bias. Other research on user-generated reviews has employed eye tracking, which is a more objective approach to data collection (Chen et al., 2022). Also, there is a somewhat limited scope of the dependent variables, as the foundation and scope of our research is based on behavioural economics. Actual behaviour, such as click-through rate, conversion and cart abandonment, would be a more objective way of measuring consumer behaviour. However, it is important to note that actual behavioural data does not capture attitudes and perceptions, but it would be a beneficial complement to the research.

Finally, there might be confounding variables where participants guessed the purpose of the experiment and expressed their pre-existing attitudes toward AI, regardless of the review content.

7.0 References

Ahluwalia, R., Burnkrant, R., and Unnava, H. (2000). Consumer response to negative publicity: the moderating role of commitment. J. Mark. Res. 37, 203–214. doi: 10.2307/1558500

Amazon. (2024) Vaughn Schermerhorn, Director. "How Amazon Continues to Improve the Customer Reviews Experience with Generative AI." Amazon Improves the Customer Reviews Experience with AI, US About Amazon, 14 Aug. 2024, www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-ai.

Amblee, N. and Bui. T. (2012) Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts, International Journal of Electronic Commerce, 16, 2, 91–113 (2012) doi:10.2307/23106395

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Computers in Human Behavior, 85, 183-189. https://doi.org/10.1016/j.chb.2018.03.051

Archak, N., Ghose, A., and Ipeirotis, P. (2011). "Deriving the pricing power of product features by mining consumer reviews." Management Science, vol. 57, no. 8, Aug. 2011, pp. 1485–1509,

Askalidis, Georgios, and Edward C. Malthouse. (2016) "The value of online customer reviews." Proceedings of the 10th ACM Conference on Recommender Systems, 7 Sept. 2016, pp. 155–158, https://doi.org/10.1145/2959100.2959181.

Balbaa, M. E., & Abdurashidova, M. S. (2024). The impact of artificial intelligence in decision making: A comprehensive review. EPRA International Journal of Economics, Business and Management Studies, 11(2), 27–38. https://doi.org/10.36713/epra15747

Bakshi, Amay. (2024). "How Do AI Candidate Summaries Enhance Recruiter Decision-Making? - Recruiting Resources and Hiring Solutions." Recooty HR Blog, 24 Sept. 2024, recooty.com/blog/ai-candidate-summaries-enhance-recruiter-decision-making/

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. Computers in Human Behavior, 26(5), 857-869. https://doi.org/10.1016/j.chb.2010.03.013

Boer, A. (2021). What you see is what you get: The effect of image style and typing errors of online hotel reviews on purchase intention, helpfulness and trust. University of Twente.

Campbell, M. L., Mohr, G. S., & Verlegh, P. W. (2013). Can disclosures lead consumers to resist covert persuasion? The important roles of disclosure timing and type of response. 54 Journal of Consumer Psychology, 23(4), 483–495. doi:10.1016/j.jcps.2012.10.012

Changchit, C., and Klaus, T. (2020). Determinants and impact of online reviews on product satisfaction. J. Internet Commer. 19, 82–102. doi: 10.1080/15332861.2019.1672135

Chevalier. A and Kicka. M. (2006) “Web designers and web users:influence of the ergonomic quality of the web site on the information search,” International Journal of Human Computer Studies, vol. 64, no. 10, pp. 1031–1048, 2006 https://doi.org/10.1016/j.ijhcs.2006.06.002

Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J. L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. Journal of Medical Internet Research, 26, e53164. https://doi.org/10.2196/53164

Chen, Tao, et al. (2022) “The impact of online reviews on consumers’ purchasing decisions: Evidence from an eye-tracking study.” Frontiers in Psychology, vol. 13, 8 June 2022, https://doi.org/10.3389/fpsyg.2022.865702.

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. Decision Support Systems, 54(1), 461–470. https://doi.org/10.1016/j.dss.2012.06.008

Cialdini, Robert B. (2001). “The Science of Persuasion.” Scientific American, vol. 284, no. 2, Feb. 2001, pp. 76–81, https://doi.org/10.1038/scientificamerican0201-76.

Dachyar, M., & Banjarnahor, L. S. (2017). Factors influencing purchase intention towards consumer-to-consumer e-commerce. Intangible Capital , 946-966. doi:10.3926/ic.1119

Davenport, Thomas, et al. **(2019).** “How artificial intelligence will change the future of marketing.” Journal of the Academy of Marketing Science, vol. 48, no. 1, 10 Oct. 2019, pp. 24–42, https://doi.org/10.1007/s11747-019-00696-0.

Eveland Jr. W. P. and Dunwoody. (2001). “User control and structural isomorphism or disorientation and cognitive load?”Communication Research, vol. 28, no. 1, pp. 48–78, 2001 doi:10.1177/009365001028001002

Fransen, M. L., Smit, E. G., & Verlegh, P. W. (2015). Strategies and motives for resistance to persuasion: an integrative framework. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01201

Fenko, A., Keizer, T., & Pruyn, A. (2017). Do social proof and scarcity work in the online context? 16th International Conferences on Research in Advertising.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (4th ed.). SAGE Publications. (Friestad & Wright, 1994) Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. Journal of Consumer Research, 21(1), 1. https://doi.org/10.1086/209380

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. Journal of Consumer Research, 21(1), 1. https://doi.org/10.1086/209380

Fürst, Andreas, et al. (2023). “How product complexity affects consumer adoption of new products: The role of feature heterogeneity and interrelatedness.” Journal of the Academy of Marketing Science, vol. 52, no. 2, 17 Apr. 2023, pp. 329–348, https://doi.org/10.1007/s11747-023-00933-7.
Guo, J., Wang, X., and Wu, Y. (2020). Positive emotion bias: role of emotional content from online customer reviews in purchase decisions. J. Retail. Consum. Serv. 52:101891. doi: 10.1016/j.jretconser.2019.101891

Huang M.H. (2000). Information load: Its relationship to online exploratory and shopping behavior, International Journal of Information Management, 20 (5) (2000), pp. 337-347, 10.1016/S0268-4012(00)00027-X

Isaac, M. S., & Grayson, K. (2017). Beyond skepticism: Can accessing persuasion knowledge bolster credibility? Journal of Consumer Research, 43(6), 895–912. https://doi.org/10.1093/jcr/ucw063

Jacobs, M.A. **(2013).** “Complexity: Toward an empirical measure”, Technovation, https://doi.org/10.1016/j.technovation.2013.01.001

Jia, H., Appelman, A., Wu, M., & Bien Aimé, S. (2024). News bylines and perceived AI authorship: Effects on source and message credibility. Computers in Human Behavior: Artificial Humans, 2(2), 100093. https://doi.org/10.1016/j.chbah.2024.100093

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Business Horizons, 63(1), 37-50. https://doi.org/10.1016/j.bushor.2020.01.003

Ketelaar, P. E., Willemsen, L. M., Sleven, L., & Kerkhof, P. (2015). The good, the bad, and the expert: How consumer expertise affects review valence effects on purchase intentions in online product reviews. Journal of Computer-Mediated Communication, 20(6), 649–666. https://doi.org/10.1111/jcc4.12139

Kirk Colleen P. & Julian Givi. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. Journal of Business Research, 186, 115026. https://doi.org/10.1016/j.jbusres.2023.115026

Kirmani, A., & Campbell, M. (2004). Goal seeker and persuasion sentry: How consumer Targets respond to interpersonal marketing persuasion. Journal of Consumer Research, 31(3), 573–582. https://doi.org/10.1086/425092

Liang, T. P., Ho Y. T., Y. W. Li & E. Turban. **(2011).** What drives social commerce: The role of social support and relationship quality. International Journal of Electronic Commerce, 16(2), 69-90. doi:10.2307/23106394

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organisational Behaviour and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lurie, N. H. **(2004)**. Decision making in information-rich environments: The role of information structure. Journal of Consumer Research, 30(4), 473-486. https://doi.org/10.1086/380283

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. Journal of Marketing, 71(1), 160-177. doi:10.1509/jmkg.71.1.160

Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organizational trust. Academy of Management Review, 709-734. https://doi.org/10.2307/258792

McKinsey & Company. (2024). The state of AI: 2024. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Metzger, Miriam J. (2006). "Privacy, Trust, and disclosure: Exploring barriers to electronic commerce." Journal of Computer-Mediated Communication, vol. 9, no. 4, 23 June 2006, pp. 00–00, https://doi.org/10.1111/j.1083-6101.2004.tb00292.x.

M. Mirhoseini, S.A. Pagé, P.M. Léger, S. Sénécal. (2021). What deters online grocery shopping? Investigating the effect of arithmetic complexity and product type on user satisfaction, Journal of Theoretical and Applied Electronic Commerce Research, 16 (4) (2021), pp. 1-18, 10.3390/JTAER16040047, 2021

Miao. Sun, et al. (2021). "The impact of online reviews in the presence of Customer Returns." International Journal of Production Economics, vol. 232, Feb. 2021, p. 107929, https://doi.org/10.1016/j.ijpe.2020.107929.

Mitchell, A. A., & Olson, J. C. (1981). Are product attribute beliefs the only mediator of advertising effects on brand attitude? Journal of Marketing Research, 18(3), 318–332. https://doi.org/10.2307/3150973

Mou, Jian, et al. (2020). "Impact of product description and involvement on purchase intention in cross-border e-commerce." Industrial Management & Data Systems, vol. 120, no. 3, 19 Dec. 2019, pp. 567–586. https://doi.org/10.1108/IMDS-05-2019-0280

Mudambi, and Schuff. (2010). "Research note: What makes a helpful online review? A study of customer reviews on Amazon.com." MIS Quarterly, vol. 34, no. 1, 2010, p. 185, https://doi.org/10.2307/20721420.

Nazlan, N. H., Tanford, S., & Montgomery, R. (2018). The effect of availability heuristics in online consumer reviews. Journal of Consumer Behaviour, 17(5), 449–460. https://doi.org/10.1002/cb.1731

Necula, Sabina-Cristiana. (2023). "Exploring the impact of time spent reading product information on e-commerce websites: A machine learning approach to analyze consumer behavior." Behavioral Sciences, vol. 13, no. 6, 23 May 2023, p. 439, https://doi.org/10.3390/bs13060439.

Nielsen. (2014). E-Commerce: Evolution or Revolution in the Fast-Moving Consumer Goods World. Nielsen Global Survey for E-Commerce. http://www.nielsen.com/content/dam/nielsenglobal/apac/docs/reports/2014/Nielsen-Global-E-commerce-Report-August-2014.pdf

Oppenheimer, Daniel M., et al. (2009) “Instructional manipulation checks: Detecting satisficing to increase statistical power.” Journal of Experimental Social Psychology, vol. 45, no. 4, July 2009, pp. 867–872, https://doi.org/10.1016/j.jesp.2009.03.009.

Park et al. (2024) AI deception: A survey of examples, risks, and potential solutions, Park , P. Goldstein, S. O'Gara, A. Chen, M. Hendrycks, D. (2024) Patterns, Volume 5, Issue 5, 100988. 10.1016/j.patter.2024.100988

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker. Cambridge University Press. https://doi.org/10.1017/CBO9781139173933

PCI Rankings (HS92). (2025). The Observatory of Economic Complexity, oec.world/en/rankings/pci/hs4/hs92?tab=ranking.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. Advances in Experimental Social Psychology. https://doi.org/10.1016/S0065-2601(08)60214-2

Ping, C. Bang-jun, L. Bei (2022). Influence of streamer's social capital on purchase intention in live streaming e-commerce. 10.3389/fpsyg.2021.748172

Puspitasari, A. V., & Aruan, D. T. H. **(2022).** The effects of user-generated reviews versus influencer-generated reviews on consumer purchase intention. Journal Aplikasi Manajemen, 20(3), 528-540. https://doi.org/10.21776/ub.jam.2023.021.02.10

Ruiz-Mafe, C., Chatzipanagiotou, K., and Curras-Perez, R. (2018). The role of emotions and conflicting online reviews on consumers’ purchase intentions. J. Bus. Res. 89, 336–344. doi: 10.1016/j.jbusres.2018.01.027

Scheiter, K., Ackerman, R., & Hoogerheide, V. (2020). Looking at mental effort appraisals through a metacognitive lens: Are they biased? Educational Psychology Review, 32(3), 1003-1027. https://doi.org/10.1007/s10648-020-09555-9

Schmutz, Peter, et al. (2009) “Cognitive load in ecommerce applications—measurement and effects on user satisfaction.” Advances in Human-Computer Interaction, vol. 2009, no. 1, Jan. 2009, https://doi.org/10.1155/2009/121494.

Smith, J. R., Hogg, M. A., Martin, R., & Terry, D. J. (2007). Uncertainty and the influence of group norms in the attitude-behaviour relationship. British Journal of Social Psychology, 46(4), 769–792. https://doi.org/10.1348/014466606x164439

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Thomas, M.-J., Wirtz, B. W., & Weyerer, J. C. (2019). Determinants of Online Review Credibility and Its Impact on Consumers' Purchase Intention. Journal of Electronic Commerce Research, 20(1), 1–21.

Trattner, Alexandria, et al. (2019). "Product complexity and operational performance: A systematic literature review." CIRP Journal of Manufacturing Science and Technology, vol. 25, May 2019, pp. 69–83, https://doi.org/10.1016/j.cirpj.2019.02.001.

Van Schaik, P., & Ling, J. (2008). Modelling user experience with web sites: Usability, hedonic value, beauty, and goodness. Interacting with Computers, 20(3), 419-432. 10.1016/j.intcom.2008.03.001

Wang, H., & Wang, T. (2025). Does AI-generated review summarization affect consumer purchasing behavior?—An empirical study based on the Amazon platform. Proceedings of the 58th Hawaii International Conference on System Sciences, 4606–4615. https://hdl.handle.net/10125/109399

Wellman, H. M. (1990). The child's theory of mind. The MIT Press.

Yuanyuan, H., Peng, Z., Yijun, L., and Qiang, Y. J. M. R. (2009). An empirical study on the impact of online reviews sentimental orientation on sale based on movie panel data. Manag. Rev. 21, 95–103. doi: 10.1007/978-3-642-00205-2_9

Zhong-Gang, Y., Xiao-Ya, W. (2015). Economics, S. O. J. S. E. Research progress and future prospect on online reviews and consumer behavior. Soft Science. 6:20. doi: 10.3760/cma.j.cn112137-20200714-02111

7.1 AI Disclosure

In line with the AI disclosure guidelines at SSE, the following segment highlights how generative AI tools have been used throughout the research paper and other areas of relevance.

ChatGPT

While ChatGPT was not used extsnsively throughout our research, it was only used for two different purposes. Firstly, ChatGPT was used for proofreading the final draft of the thesis to ensure that there were no gaps or spelling mistakes that we had made. We also utilised ChatGPT to make sure that the research was following a red thread and to check if there were certain aspects of the research that needed more attention. We used commands such as: "Read through this thesis and give suggestions on how we can make the flow of the text easier for the reader and what are some aspects of the research that may not be as clear and concise as others" and "Are there any words that have been used too much that should be changed". Secondly, we used ChatGPT to create the AI generated review summaries that were used as stimuli in our experiment. We inserted the reviews for each product page into ChatGPT and asked for an review summary, similarly to how it would be done on an e-commer website.

8.0 Appendices

Fig 2.

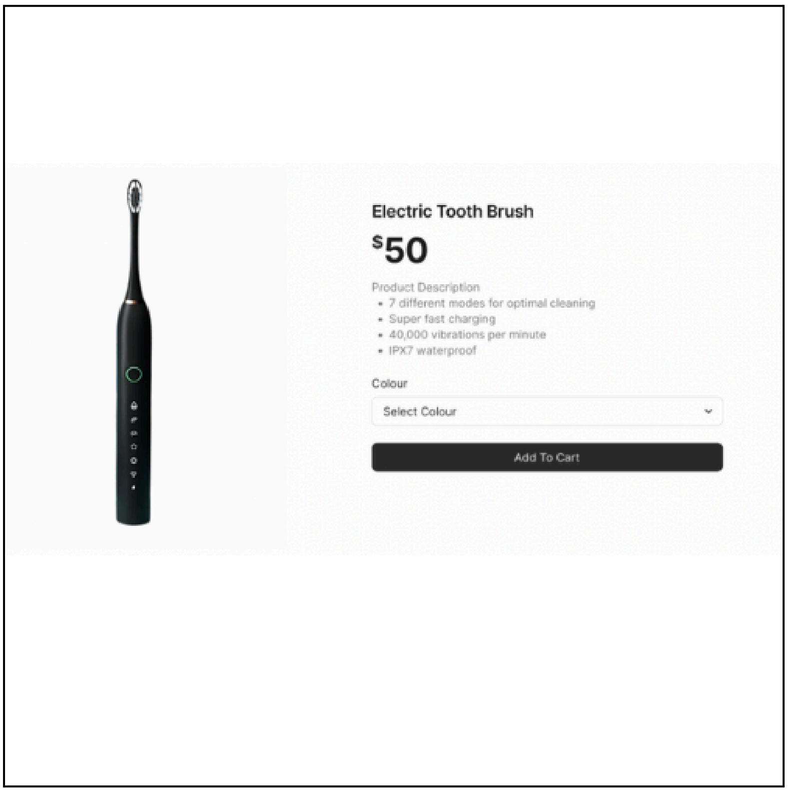
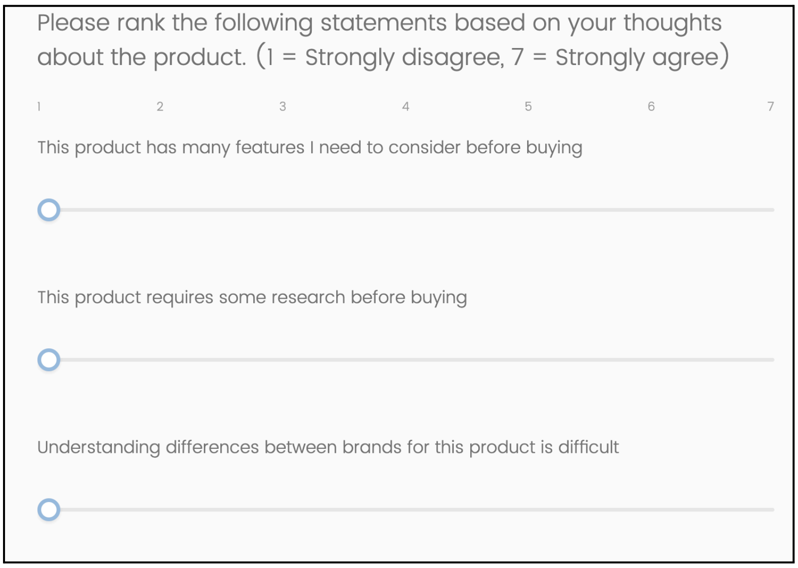


Fig 3.

Fig 4.


Please rank the following statements based on your thoughts about the product. (1 = Strongly disagree, 7 = Strongly agree)

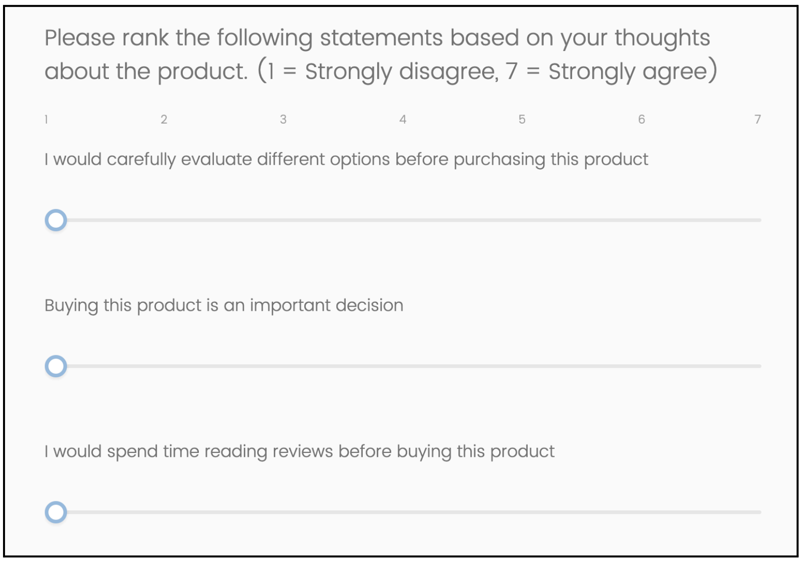
1 2 3 4 5 6 7

This product has many features I need to consider before buying

This product requires some research before buying

Understanding differences between brands for this product is difficult

Fig 5.


Please rank the following statements based on your thoughts about the product. (1 = Strongly disagree, 7 = Strongly agree)

1 2 3 4 5 6 7

I would carefully evaluate different options before purchasing this product

Buying this product is an important decision

I would spend time reading reviews before buying this product

Fig 6.

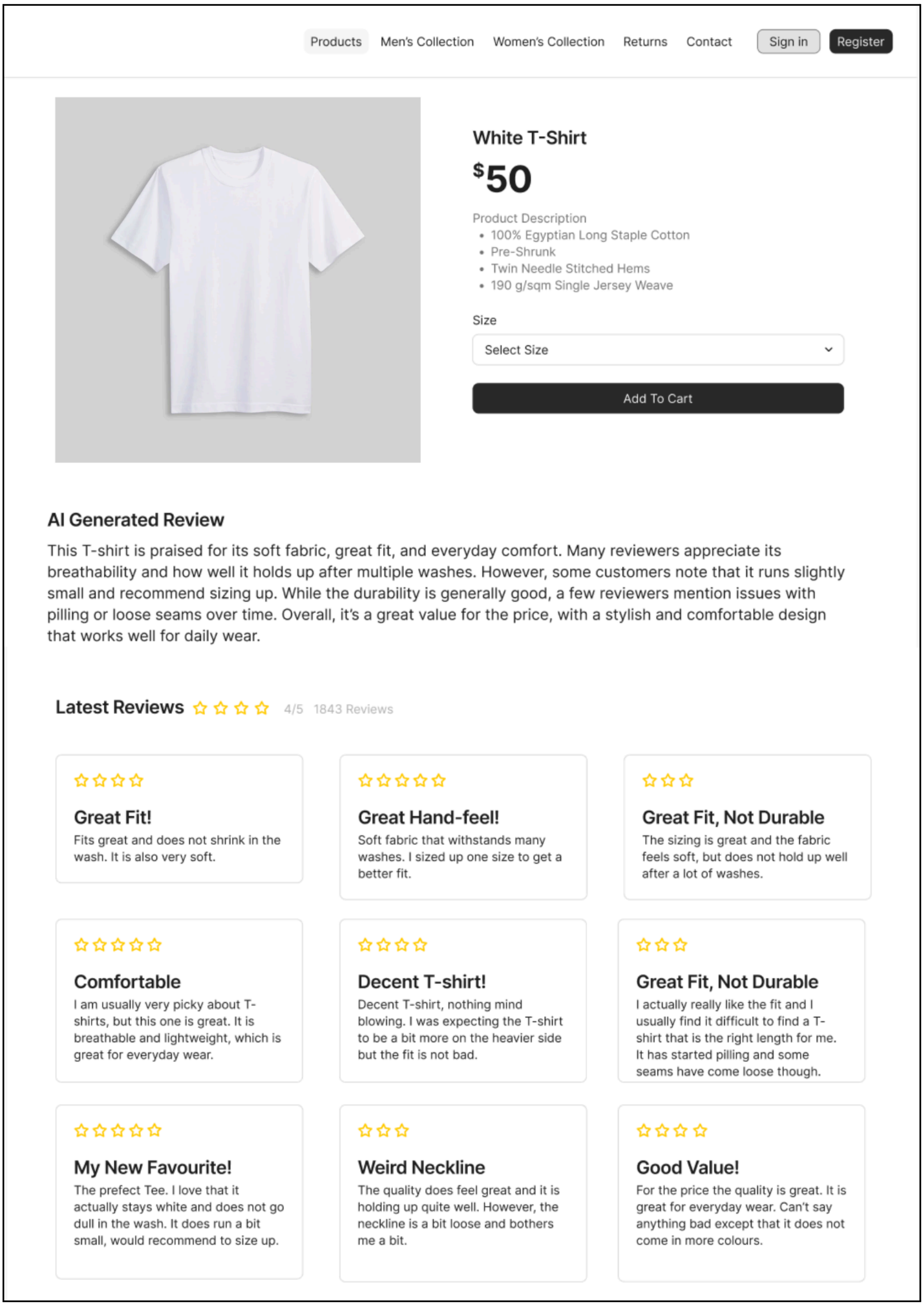

Products
Men's Collection
Women's Collection
Returns
Contact
Sign in
Register
White T-Shirt
$50
Product Description
• 100% Egyptian Long Staple Cotton
• Pre-Shrunk
• Twin Needle Stitched Hems
• 190 g/sqm Single Jersey Weave
Size
Select Size
Add To Cart
AI Generated Review
This T-shirt is praised for its soft fabric, great fit, and everyday comfort. Many reviewers appreciate its breathability and how well it holds up after multiple washes. However, some customers note that it runs slightly small and recommend sizing up. While the durability is generally good, a few reviewers mention issues with pilling or loose seams over time. Overall, it's a great value for the price, with a stylish and comfortable design that works well for daily wear.
Latest Reviews
4/5 1843 Reviews
Great Fit!
Fits great and does not shrink in the wash. It is also very soft.
Great Hand-feel!
Soft fabric that withstands many washes. I sized up one size to get a better fit.
Great Fit, Not Durable
The sizing is great and the fabric feels soft, but does not hold up well after a lot of washes.
Comfortable
I am usually very picky about T-shirts, but this one is great. It is breathable and lightweight, which is great for everyday wear.
Decent T-shirt!
Decent T-shirt, nothing mind blowing. I was expecting the T-shirt to be a bit more on the heavier side but the fit is not bad.
Great Fit, Not Durable
I actually really like the fit and I usually find it difficult to find a T-shirt that is the right length for me. It has started pilling and some seams have come loose though.
My New Favourite!
The prefect Tee. I love that it actually stays white and does not go dull in the wash. It does run a bit small, would recommend to size up.
Weird Neckline
The quality does feel great and it is holding up quite well. However, the neckline is a bit loose and bothers me a bit.
Good Value!
For the price the quality is great. It is great for everyday wear. Can't say anything bad except that it does not come in more colours.

Fig 7.

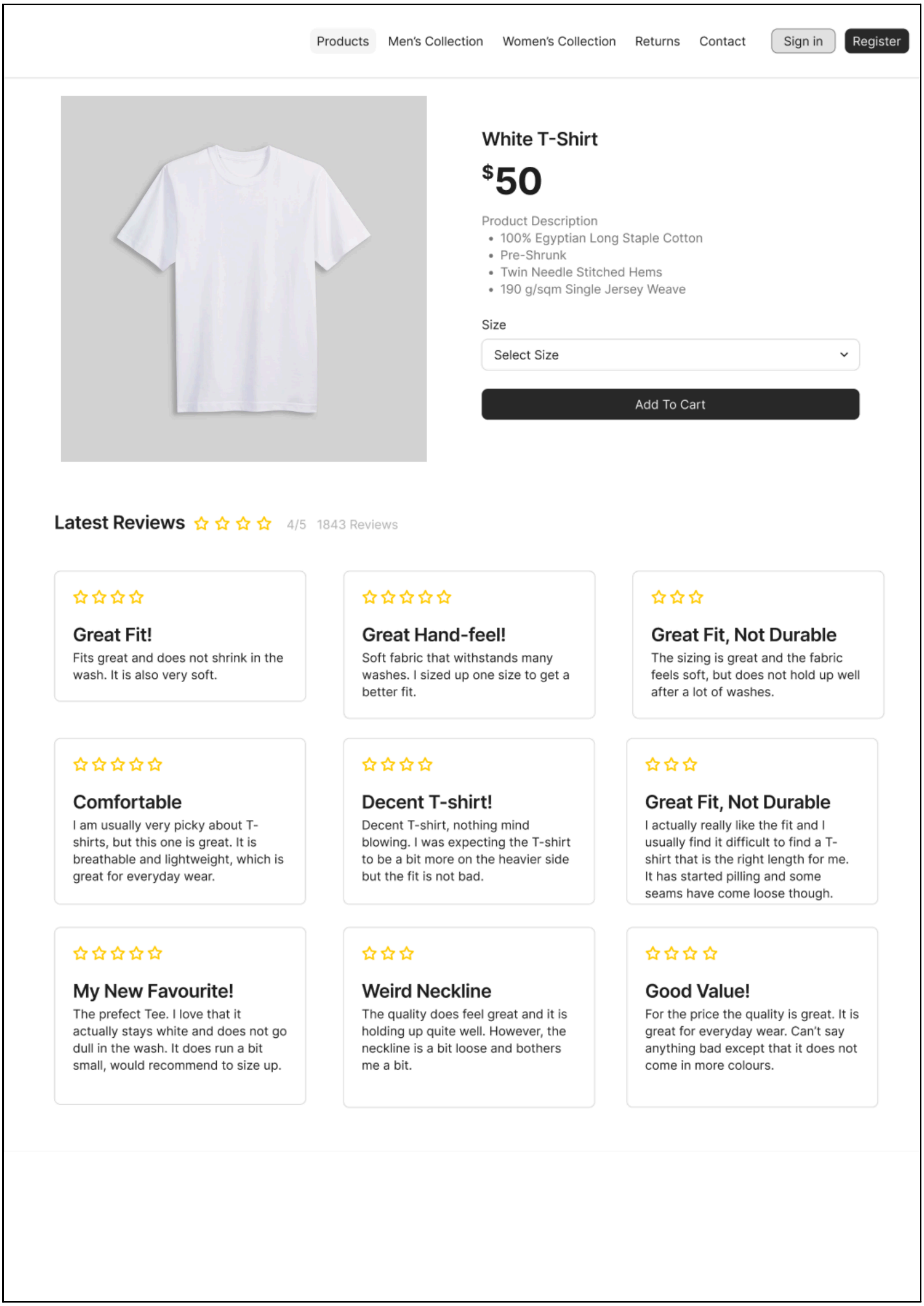
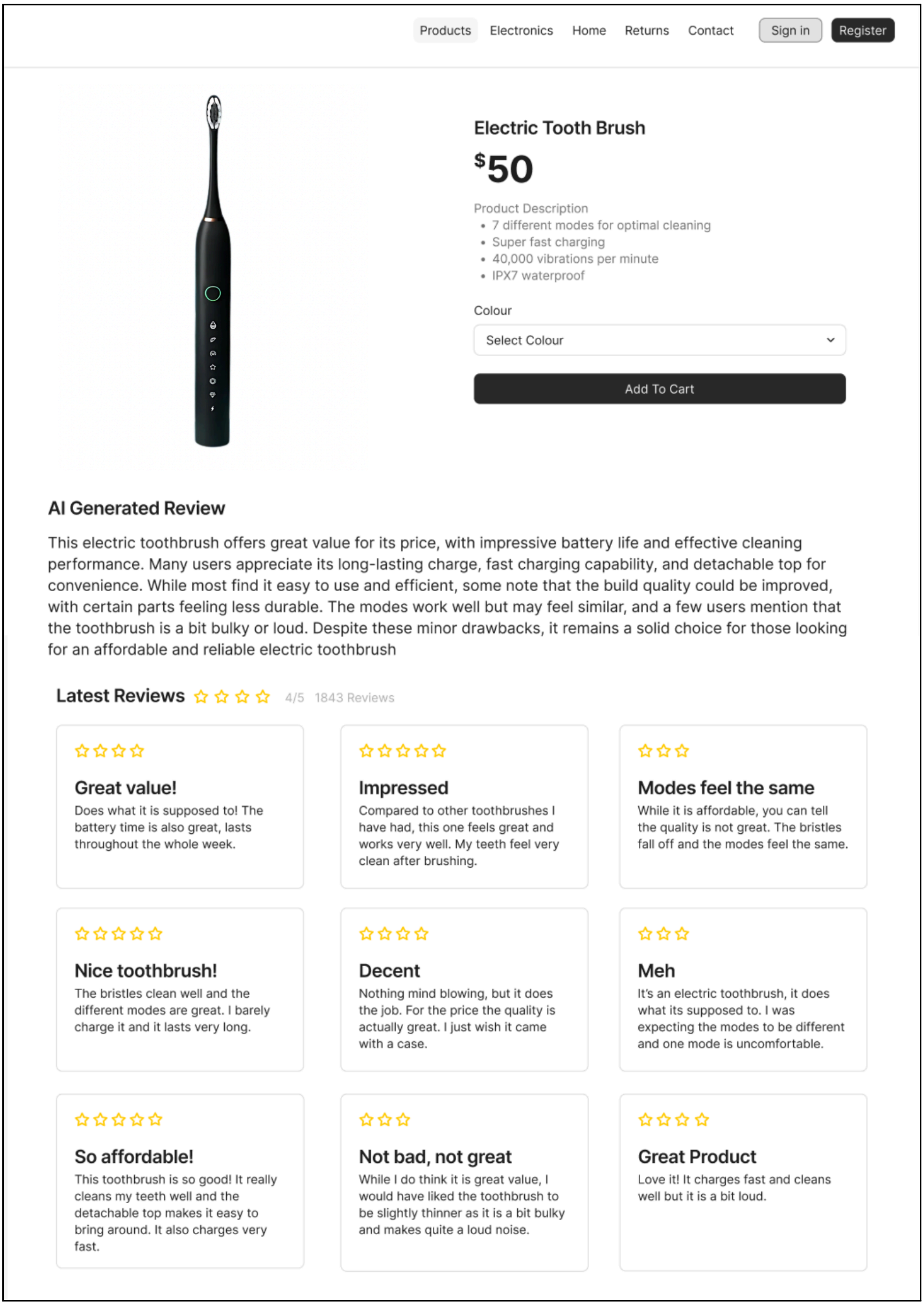

Products
Men's Collection
Women's Collection
Returns
Contact
Sign in
Register
White T-Shirt
$50
Product Description
• 100% Egyptian Long Staple Cotton
• Pre-Shrunk
• Twin Needle Stitched Hems
• 190 g/sqm Single Jersey Weave
Size
Select Size
Add To Cart
Latest Reviews
4/5 1843 Reviews
Great Fit!
Fits great and does not shrink in the wash. It is also very soft.
Great Hand-feel!
Soft fabric that withstands many washes. I sized up one size to get a better fit.
Great Fit, Not Durable
The sizing is great and the fabric feels soft, but does not hold up well after a lot of washes.
Comfortable
I am usually very picky about T-shirts, but this one is great. It is breathable and lightweight, which is great for everyday wear.
Decent T-shirt!
Decent T-shirt, nothing mind blowing. I was expecting the T-shirt to be a bit more on the heavier side but the fit is not bad.
Great Fit, Not Durable
I actually really like the fit and I usually find it difficult to find a T-shirt that is the right length for me. It has started pilling and some seams have come loose though.
My New Favourite!
The prefect Tee. I love that it actually stays white and does not go dull in the wash. It does run a bit small, would recommend to size up.
Weird Neckline
The quality does feel great and it is holding up quite well. However, the neckline is a bit loose and bothers me a bit.
Good Value!
For the price the quality is great. It is great for everyday wear. Can't say anything bad except that it does not come in more colours.

Fig 8.



Products
Electronics
Home
Returns
Contact
Sign in
Register
Electric Tooth Brush
$50
Product Description
• 7 different modes for optimal cleaning
• Super fast charging
• 40,000 vibrations per minute
• IPX7 waterproof
Colour
Select Colour
Add To Cart
AI Generated Review
This electric toothbrush offers great value for its price, with impressive battery life and effective cleaning performance. Many users appreciate its long-lasting charge, fast charging capability, and detachable top for convenience. While most find it easy to use and efficient, some note that the build quality could be improved, with certain parts feeling less durable. The modes work well but may feel similar, and a few users mention that the toothbrush is a bit bulky or loud. Despite these minor drawbacks, it remains a solid choice for those looking for an affordable and reliable electric toothbrush
Latest Reviews
4/5 1843 Reviews
Great value!
Does what it is supposed to! The battery time is also great, lasts throughout the whole week.
Impressed
Compared to other toothbrushes I have had, this one feels great and works very well. My teeth feel very clean after brushing.
Modes feel the same
While it is affordable, you can tell the quality is not great. The bristles fall off and the modes feel the same.
Nice toothbrush!
The bristles clean well and the different modes are great. I barely charge it and it lasts very long.
Decent
Nothing mind blowing, but it does the job. For the price the quality is actually great. I just wish it came with a case.
Meh
It's an electric toothbrush, it does what its supposed to. I was expecting the modes to be different and one mode is uncomfortable.
So affordable!
This toothbrush is so good! It really cleans my teeth well and the detachable top makes it easy to bring around. It also charges very fast.
Not bad, not great
While I do think it is great value, I would have liked the toothbrush to be slightly thinner as it is a bit bulky and makes quite a loud noise.
Great Product
Love it! It charges fast and cleans well but it is a bit loud.

Fig 9.

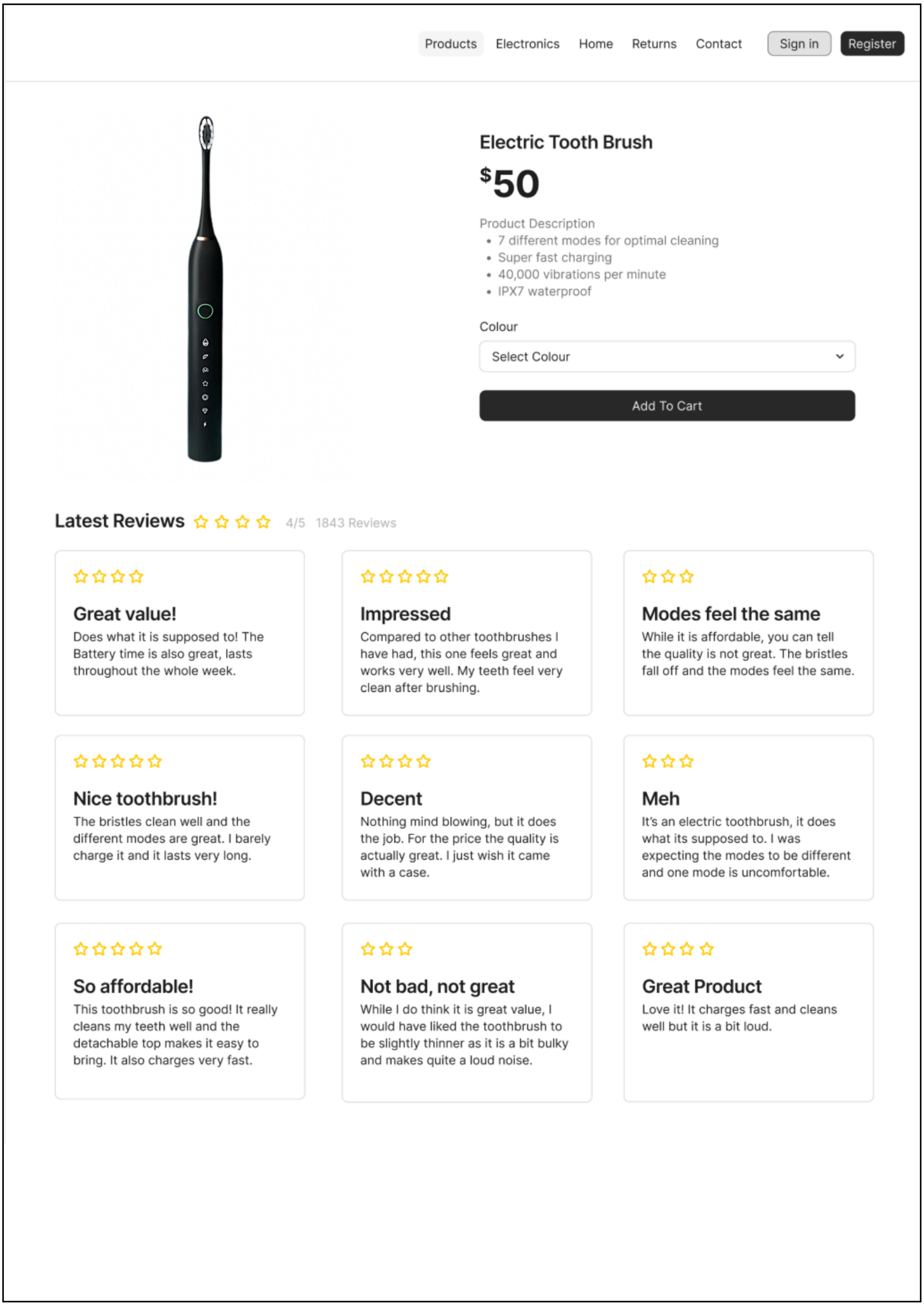


Fig 9. Questionnaire Layout

Variables	*Question*	*Scale from 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)*
Demographic	1. What is your age? 2. What is your gender? 3. How often do you shop online? 4. What is your highest level of education?	1. Under 18 to over 85 2. Male,Female,Undisclosed 3. Daily, Weekly, Monthly, Few times a year, rarely, never. 4. Less than school, High school graduate, Some college, Bachelor's degree, Masters degree, Doctorate.
Review Usefulness	1. The reviews for this product seem authentic. 2. The reviews for this product seem credible. 3. I trust the information provided in the reviews. 4. I feel confident in relying on the reviews to make a purchase decision.	From 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)
Decision confidence	1. The reviews helped me understand the product's strength and weaknesses 2. The reviews would save me time in making a decision 3. The reviews are clear and easy to read 4. I would rely on the reviews to make an informed decision	From 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)
Purchase intention	1. I feel confident in my ability to make a purchase decision based on the information provided 2. I would not hesitate to proceed buying this product 3. I seek additional information about this product	From 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)
Product interest & Attitude	1. I would consider buying this product 2. I would recommend this product to others 3. The likelihood of me buying this product is high 4. This product seems to be of good quality 5. I find this product appealing 6. I trust this product to perform as expected 7. I am interested in this product	From 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)
Bias & Persuasion	1. I believe the reviews provide an unbiased perspective on the	From 1-7, (1 = Strongly Disagree, 7 = Strongly Agree)

Knowledge	product 2. The reviews seem to favour the product too much 3. The reviews were made by real customer 4. I feel that the information was used to influence my decisions	
Cognitive Load	1. When viewing the product page how difficult or easy did you feel the information was to process? 2. How difficult or easy was it to understand the information on the product page? 3. Was it difficult or easy to process the reviews on the product page?	● Extremely Difficult ● Somewhat difficult ● Neither easy nor difficult ● Somewhat easy ● Extremely easy
Control & Manipulation	1. What product did you see on the product page? 2. Was there any AI generated text on the product page (only for the groups exposed to AI review summary)	1. Range of products (2 of which correct) 2. Yes or No