Explainable artificial intelligence (XAI) and consumer products: an empirical assessment of the effects of explanations versus benefits in product information

European Journal of Marketing

463

Received 2 October 2023
Revised 14 November 2024
3 April 2025
9 September 2025
12 March 2026
Accepted 31 March 2026

Magnus Söderlund and Alona Natorina
Centre for Consumer Marketing, Stockholm School of Economics, Stockholm, Sweden

Abstract

Purpose – This study aims to examine one crucial assumption in the explainable artificial intelligence (XAI) literature, namely, that providing information with explanations of how an AI-based product works enhances understandability and trust. The study also extends the view of explanations in the XAI literature by examining the effects of another type of product information, product benefits and by including perceived product performance as a downstream variable.

Design/methodology/approach – A between-subjects experiment ($n = 480$) was used to manipulate pre-purchase information about an AI-based product in terms of both explanations about how the product works (absent vs present) and product benefits (absent vs present). Understandability, trust and perceived product performance were the measured response variables.

Findings – Explanations in the XAI sense boost understandability and trust, but they do more than that: such explanations also have a positive impact on perceived product performance (both directly and indirectly via trust). Information about product benefits produced a similar pattern of effects on understandability, trust and perceived product performance.

Research limitations/implications – One specific AI-based product was used as a stimulus in the present study (an add-on option for cars to provide safety for car drivers). It should be seen as a sample from a population of products to be used in situations in which there can be severe consequences for the user. AI, however, is also used for products with a lower potential for severe consequences, and it is possible that explanations of how they work would produce other reactions than those that were identified in the present study.

Practical implications – The results imply that marketers of AI-based products should provide consumers with explanations of how the product works and product benefit information.

Originality/value – This study offers novel extensions in relation to much existing research on XAI by examining a consumer setting, information about product benefits, perceived product performance as an overall assessment of the product subject to information and a pre-purchase situation.

Keywords Artificial intelligence (AI), Explainable AI (XAI), Explanations, Benefits, Understandability, Trust, Perceived product performance

Paper type Research paper

European Journal of Marketing
Vol. 60 No. 13, 2026
pp. 463-490
Emerald Publishing Limited
0309-0566
DOI 10.1108/EJM-10-2023-0736

1. Introduction

AI-based products are expected to become more prevalent in a broad range of industries, and such products are predicted to affect our work and everyday lives in profound ways (Blösser and Weihrauch, 2024; Kelly *et al.*, 2023; Rai, 2020). Indeed, such products can be expected to shift agency and control from humans to technology (Glikson and Woolley, 2020). For most people, however, it is difficult to understand how these products work, because many of the algorithms are inscrutable (Ali *et al.*, 2023; Glikson and Woolley, 2020; Laato *et al.*, 2022; Rai, 2020). This, in turn, can mitigate people's usage of AI-based products (Dwivedi *et al.*, 2021; Rai, 2020). From the point of view of firms that are developing AI-based products, then, algorithm inscrutability represents a potential threat to consumers' adoption of such products. Understanding how AI works is also needed from a societal point of view, given that AI has the potential to change many aspects of human lives – both for us who are around right now and for future generations (Dwivedi *et al.*, 2021).

One way to come to terms with this is to explain how AI-based products work. This is commonly referred to as explainable AI (XAI). Explainable AI provides visibility into how an AI-based system makes decisions, predictions and how it executes its actions (Rai, 2020). That is to say, XAI offers human-understandable justifications for procedures and output (Ali *et al.*, 2023; Shin, 2021). A main idea in the literature on XAI is that there will be an increasing demand for transparency, and that explanations are an important means to mitigate lack of transparency (Arrieta *et al.*, 2020; Chazette and Schneider, 2020; de Graaf and Malle, 2017; Hoffman *et al.*, 2018; Knapič *et al.*, 2021; Laato *et al.*, 2022; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Miller, 2019). There are also legal concerns in this area. For example, the European Union states in its GDPR framework that users have a right to an explanation concerning algorithm-created decisions based on personal information (Hoffman *et al.*, 2018; Knapič *et al.*, 2021).

Given that explanations are offered about how an AI system works, several consequences have been claimed. Two of them are particularly prevalent in the XAI literature: explanations can enhance understandability (Chazette and Schneider, 2020; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Shin, 2021; Wang *et al.*, 2016) and trust (Ali *et al.*, 2023; Chazette and Schneider, 2020; Glikson and Woolley, 2020; Grewal *et al.*, 2021; Knapič *et al.*, 2021; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Miller, 2019; Rai, 2020; Shin, 2021; Wang *et al.*, 2016; Yang *et al.*, 2020). Indeed, Dazeley *et al.* (2021) view the aims of XAI in terms of trust and understandability.

In empirical terms, however, and according to several authors (Cheng *et al.*, 2019; Kenny *et al.*, 2021; Leichtmann *et al.*, 2023; Shin, 2021), relatively few studies have been made of the effectiveness of explanations in an XAI setting. Some of the existing studies have also been criticized from a methodological point of view (Kenny *et al.*, 2021). For example, and with respect to both understandability and trust, some authors do not report which measurement items they used, use single-item measures with unknown reliability and use a mix of items reflecting different theoretical constructs, so that it is far from obvious what is measured. Some existing studies also expose participants for each of several treatments in experiments, use no control group for a no explanation condition and use a very small number of participants for each treatment. It may be seen as a genteelism, then, when Mueller *et al.* (2019) recommend that XAI researchers should be encouraged to include in their research reports fuller details on their empirical and experimental methods.

In any event, while some studies in an XAI context indicate that explanations can enhance understandability (Cheng *et al.*, 2019; Shin, 2021) and trust (e.g. Shin, 2021; Suen and Hung, 2023; Yang *et al.*, 2020; Wang *et al.*, 2016; Wang and Benbasat, 2007), other studies show that offering explanations has no impact on understandability (Kizilcec, 2016; Westphal *et al.*, 2023) and no impact on trust (e.g. Cheng *et al.*, 2019; Kenny *et al.*, 2021; Westphal *et al.*, 2023). Some studies also show that offering explanations *attenuates* both understandability and trust (Leichtmann *et al.*, 2023). In addition, it has been observed that users' need for explanations of intelligent interactive systems (Bunt *et al.*, 2012) and AI-based systems (Chazette and Schneider, 2020) is not always high. Given ambiguity regarding the effects of explanations, a first purpose of the present study is to reexamine the impact of product information comprising explanations about how an AI-based product works on consumers' understandability and trust. The end-user in the present study, then, is the consumer – not programmers, AI developers and other specialists that have been envisioned as users in much existing research on XAI (cf. Ali *et al.*, 2023).

From a traditional marketing point of view, however, another type of information than explanations of how a product works has been stressed for decades, namely, information about what benefits a product provides. Benefits are the desirable consequences of using a product (Gutman, 1982), such as "convenience" and "happiness". They represent reasons why people buy products and should be seen in contrast to the characteristics of the product *per se* (i.e. its attributes). In other words, people receive benefits, whereas products have attributes (Gutman, 1982). In the present study, we agree with Otnes *et al.* (2014) and Wu *et al.* (1988), who argue that incorporating consumers' benefit perceptions into preference models should provide better predictive power. Indeed, an emphasis on benefits can be seen as an alternative way of providing explanations in relation to the XAI literature; the XAI view is focused on explanations of the internal parts of a product, while benefits can be seen as explaining the potential consequences of using a product. A second purpose of the present study, then, is to examine whether information about the benefits of an AI-based product can influence consumers' understandability and trust in the same or different ways than XAI-based explanations.

Moreover, in contrast to many existing studies of the effects of explanations of AI systems in which understandability and/or trust are final dependent variables, our third purpose is to assess the impact of the two types of information about an AI-based product on the perceived performance of a product (the latter, then, is our downstream variable). That is to say, in the present study, understandability and trust are viewed as potential mediators (not as final dependent variables). Perceived product performance was selected as a downstream variable because it is one of the main antecedents to users' acceptance of new technology (Venkatesh *et al.*, 2003).

It should also be noted that basically all existing attempts to empirically assess the effects of explanations in the context of AI-based systems have been made in a setting in which explanations are provided *while* a system is used (or as post hoc explanations after usage). Typically, research along such lines means that the researcher – for the purpose of conducting a study – creates a setting in which participants are asked to use a system, and it is during this usage that participants are exposed to various types of explanations (e.g. explanation vs no explanation if it is an experiment). Examples are Alam and Mueller (2021) and Schraagen *et al.* (2020). This setting, however, is at odds with the marketing discipline's conceptualization of consumers' decision-making activities as a process with several stages (e.g. Batra and Keller, 2016). That is to say, particularly for products that users have to pay for, potential users of new products are in a pre-purchase phase before usage starts, and the outcomes of what happens in this phase determine if the potential user will become an actual user. Many AI-based products are indeed costly to develop and are unlikely to be made available free of charge, which calls on the firms that develop them to convince customers that the products should be purchased. Given a need for explanations of how these products work, we believe that this need arises already in the customers' pre-purchase phase.

Therefore, our fourth purpose is to examine the effects of providing information comprising explanations of how a product works (and providing information about benefits) in this particular phase.

In our empirical assessment, we employed an experimental approach in which pre-purchase information about an AI-based product was manipulated. The selected stimulus product, an add-on option for cars with the purpose of enhancing safety in traffic, is of the embedded AI type as opposed to AI with a visual representation – such as a robot or a virtual agent (Glikson and Woolley, 2020). The information about this product was manipulated in terms of both explanations about how the product works (absent vs present) and product benefits (absent vs present).

The present study is intended to contribute to existing literature in four ways. First, the user in the present study is the consumer, not a programmer or an AI developer as in the typical XAI study (e.g. Ali *et al.*, 2023). The main rationale behind our consumer focus is that it is unlikely that the consumer reacts to explanations in the same way as the developer and, in the case of AI-based products for consumers, that consumer reactions are likely to determine the success of AI-based products in the marketplace more than developers' reactions. Second, in addition to the effects of product information with explanations, the present study examines the effects of product information also in terms of benefits. Including the latter, we think, offers contributions in two ways. On the one hand, the XAI literature typically ignores benefit information, which the marketing literature has stressed as important for several decades (e.g. Gutman, 1982; Haley, 1968; Otnes *et al.*, 2014). On the other hand, despite the marketing literature's long tradition of emphasizing benefits, there have been relatively few attempts to examine the effects of benefit information in terms of the approach in the present study (i.e. manipulating such information with respect to its absence vs its presence). Third, in contrast to the typical XAI study, in which the focus is on the effects of explanations of how a product works on understandability and trust (e.g. Chazette and Schneider, 2020; Leichtmann *et al.*, 2023), the present study comprises perceived product performance as a downstream variable. This is a contribution in the sense that consumers' evaluations of products represent a more proximate cause than understandability and trust when it comes to their decisions to actually use (and pay for) products. Fourth, the present study addresses the impact of product information in the pre-purchase phase rather than in the usage phase. The motivation for our focus on the pre-purchase phase may be intuitively obvious from a marketing point of view, in the sense that the marketing literature typically acknowledges that consumers engage in various pre-purchase activities before products are purchased and used. In the XAI literature, however, the typical focus is on the user's reactions to product information in the usage phase (e.g. Cheng *et al.*, 2019; Shin, 2021). Since it has been argued that information processing is different in the pre-purchase phase compared to the usage phase (e.g. Hoch and Deighton, 1989; Hamilton and Thompson, 2007), we also believe that examinations of user reactions to product information restricted to comprise only the usage phase will constrain the development of theory about the effects of product information.

2. Theoretical framework and hypotheses

The foundation of the hypotheses in the present study is a theory about how commercial messages influence receivers. The underlying assumptions are not frequently articulated in the XAI literature, but they are prevalent in the marketing literature and particularly in advertising effectiveness research based on hierarchy of effects models, as well as in research making distinctions between indirect and direct experiences.

A main premise is that reactions to a stimulus message, such as product information, involve a response sequence in which each stage is causally related to the next (Barry and Howard, 1990; Eisend and Tarrahi, 2016; Kempf and Smith, 1998; Smith and Swinyard, 1982; Vakratsas and Ambler, 1999). Even in the complex contemporary marketplace, with new media, shifting media patterns and divided consumer attention, it is acknowledged that there are several steps that consumers go through before purchases are made (cf. Batra and Keller, 2016). Often, the understanding of a message (Eisend and Tarrahi, 2016; Goodwin and Etgar, 1980; Nantel and Sekhavat, 2008) and beliefs (in the present study, trust represents a belief variable) about the product are parts of the response sequence (Homer, 2006; Mitchell and Olson, 1981; Mittal, 1990; Orth and De Marchi, 2007). A typical application is also that in the downstream part of the response sequence, there is an overall evaluation variable, such as the attitude toward a product or a brand, which has the capability of influencing consumers' purchasing behavior (Eisend and Tarrahi, 2016; Kempf and Smith, 1998; Mitchell and Olson, 1981; Vakratsas and Ambler, 1999).

As indicated above, the specific communication context in the present study is pre-purchase information about an AI-based product intended for consumers. One main reason why there is a need for studying reactions in the pre-purchase phase is that consumers are likely to process information differently in the pre-purchase phase compared to the usage phase. In existing studies (outside the context of XAI), this is often conceptualized in terms of the consumer's indirect experience (e.g. reading a product description or seeing a product on display) versus direct experience (e.g. trying or using a product). Examples of studies in which this distinction is made are Daugherty *et al.* (2008), Hamilton and Thompson (2007), Mooy and Robben (2002) and Smith and Swinyard (1982). Several differences have indeed been discussed and verified empirically. An indirect experience produces less attention and is less easy to remember (Hoch and Deighton, 1989; Kempf and Smith, 1998); it results in a more abstract representation of the product (Hamilton and Thompson, 2007); it comprises lower involvement (Hoch and Deighton, 1989); and it can lead to less belief confidence, less trustworthiness and a lower level of message acceptance (Daugherty *et al.*, 2008; Hamilton and Thompson, 2007; Hoch and Deighton, 1989; Smith and Swinyard, 1982). Ultimately, in relation to a direct experience, an indirect experience can also result in lower evaluations of a product (Mooy and Robben, 2002). Hence, it is not surprising that several authors view the product itself (given that it can be tested and thus can generate a direct experience) as a decisive communication tool (Mooy and Robben, 2002). Indeed, direct experience gained from a trial is seen as a vivid, cogent and salient learning environment with a potentially powerful influence on evaluations of products and purchases (Kempf and Smith, 1998). In the light of this, then, one would expect that an examination of reactions in usage situations can exaggerate the effects of information about an AI product in relation to an examination of a pre-purchase situation.

In the present study, we hypothesize that if pre-purchase information comprises an explanation of how the product works, then it will enhance understandability compared to if no such explanation is offered. This is a common assumption in the XAI literature (e.g. Chazette and Schneider, 2020; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Shin, 2021). We also hypothesize that information comprising explanations of how the product works enhances trust, which represents another common assumption in the XAI literature (e.g. Chazette and Schneider, 2020; Glikson and Woolley, 2020; Leichtmann *et al.*, 2023). In contrast to the XAI literature, however, we hypothesize that product information comprising benefits would increase understandability and trust, too. Moreover, we hypothesize that understandability and trust can enhance perceived product performance. The expected net result is that information comprising explanations of how an AI-based product works and product benefits would boost the receiver's perceptions of product performance (see Figure 1).

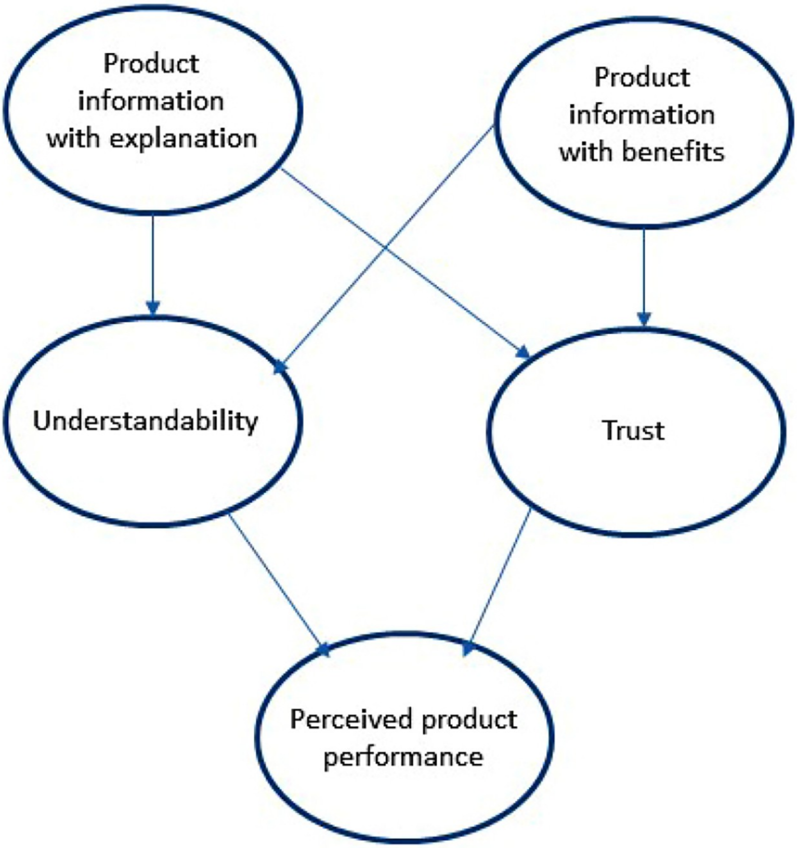

Figure 1. Overview of the theoretical framework
Source: Authors' own work

2.1 Explanations, benefits and understandability

The goal of explanations in XAI, it has been argued, is to answer questions such as "how does it work?" and "why did it just do that?" (Hoffman *et al.*, 2018). In the present study, our main interest is in the specific type of explanations that have been labeled "how-explanations;" such explanations inform a user about how a system works in terms of the processes and procedures it is using for its decisions (Wang and Benbasat, 2007).

As already indicated, a common assumption in the literature on XAI is that explanations along such lines would increase understandability in terms of the functioning of an AI-based product. That is to say, given that an explanation provides information about a system's processes and decisions, one would expect an increased understanding of the system (cf. Kraus *et al.*, 2020). One general reason is that we humans have a strong need to understand the world in terms of causes and effects (Keil, 2006), and if an explanation addresses why-questions, which is a common way to view the core of what an explanation is (e.g. Lombrozo, 2006; Keil, 2006), it would satisfy a fundamental need for understanding in causal terms. A more specific reason for an explanation–understanding link, in the case of new products, appears to be related to the resolution of incongruity. That is to say, a new product would almost by definition exhibit some degree of misfit with existing schemas used for sense-making (Rindova and Petkova, 2007), and an explanation can serve as a resolution to this. In any event, exactly what is meant by understandability varies between authors, but most of them conceptualize this in subjective terms. This is the case in the present study, too; here, understandability refers to the extent to which the receiver of an explanation believes that he or she understands the functioning of an AI-based product (Schoonderwoerd *et al.*, 2021). The following, then, is hypothesized for providing information about how an AI-based product works:

H1. Information about an AI product that comprises explanations of how the product works enhances understandability.

However, information about a product can also comprise benefits in terms of the desirable consequences of using a product (Gutman, 1982). In the marketing literature, the benefit construct has been used in several ways. First, it has been acknowledged that benefits offer a causal explanation of why people buy and use a specific product; the main idea is that desirable consequences – not product characteristics *per se* – represent the main reason for consumers' buying and usage behaviors (Botschen *et al.*, 1999; Gutman, 1982; Otnes *et al.*, 2014). Examples of specific benefits are "convenience," "good for health," "better economy" and "whiter teeth" (Haley, 1968; Vriens and Ter Hofstede, 2000). The marketing literature also stresses that such benefits can be seen as links to higher-level needs (e.g. "enjoyment" and "warm relations with others"), which represent a main underlying reason why one particular product is preferred (Gutman, 1982; Vriens and Ter Hofstede, 2000).

Second, many authors have suggested that benefits should be the basis for forming segments of customers that can be subject to subsequent marketing activities. A main argument is that the benefits that people are seeking in consuming a product are the basic reason for the existence of true market segments, because it is the benefits sought that would determine customers' behavior (Botschen *et al.*, 1999; Haley, 1968; Haley, 1984).

Third, explicit information in commercial messages about which benefits a product provides can facilitate receivers' information processing. This, then, means that marketing communications can be used to make connections between a product and the achievement of desirable consequences (Gutman, 1982; Otnes *et al.*, 2014). It has been argued that this is likely to be particularly useful for complex products (Vriens and Ter Hofstede, 2000) and for creating attention to a message (Haley, 1984). In any event, benefit-stressing messages are used frequently in firms' marketing communications activities (e.g. in printed ads and on websites). For products with AI content, one example is how Polestar describes the Polestar 3 SUV; it offers "safety" through technology that protects its occupants and prevents collisions. A second example is Tesla Y, which offers "maximum versatility," "flexible storage" and "quick and easy loading and unloading". The Super Cruise option for Cadillac Escalade-V is an additional example; its hands-free vehicle operation provides the driver with both "comfort" and "convenience". Indeed, in Hengstler *et al.*'s (2016) study of firms that develop and sell AI-related products, it is clear that communication about such products' benefits is a common strategy.

The ways in which the benefit construct has been used in the marketing literature, and in marketing practice, indicate that providing information about benefits can be seen as an explanation, too – at least in the very general sense that an explanation comprises information that serves the purpose of clarification. There is also a causal component in providing benefit information, because such information typically involves (at least in an implicit sense) statements of the "if-then" type. Indeed, a typical marketing way to use arguments about benefits is to imply that if a product is used, then various benefits (e.g. convenience and happiness) would materialize for the user. Similarly, on the receiving end, it is likely that many, if not most, products and services are purchased because consumers infer a causal relationship; they believe, for example, that deodorants can improve their social life and that athletic shoes can enhance their performance (Folkes, 1988). Given that causal reasoning is a part of explanations in general (Hoffman *et al.*, 2018; Miller, 2019), providing information about benefits can be seen as an explanation attempt in the sense that it addresses causal aspects. Explanations of how a product works, in the XAI sense, and benefit information, however, have different foci when it comes to outcomes (XAI explanations focus on outcomes within a system, while benefit information focuses on outcomes for a user). In any event, there are several information-processing reasons why benefit information would increase the understanding of a product.

First, to make sense of a product in terms of its attributes can be challenging – particularly if there are several attributes that are related to each other. If the attributes are grouped in a limited number of higher-level benefit categories, however, the cognitive burden of the sense-making task can be reduced (Wang *et al.*, 2022). Second, the benefits stated by a sender of a commercial message are typically positive, which makes such information relatively easy to evaluate without much product knowledge (Graeff, 1997). Third, given that consumers comprehend product information by inferring personally relevant meanings (Graeff, 1997), stated benefits (e.g. "convenience" and "good for health") can be expected to be relatively easy to connect to what is self-relevant.

Given this, we expect that product information comprising benefits would boost understandability in relation to a product, and thus, the following is hypothesized for providing information about an AI-based product:

H2. Information about an AI product that comprises product benefits enhances understandability.

2.2 Explanations, benefits and trust

Trust in relation to an AI-based system can be defined as the belief that the system will help the individual to achieve his or her goals in a situation in which there is uncertainty and vulnerability (Kelly *et al.*, 2023; Kraus *et al.*, 2020; Lee and See, 2004; Lee, 2018). Typically, this belief comprises both confidence in the system (i.e. it is producing reliable and accurate results) and a willingness to follow its recommendations (Lee and See, 2004; Yang *et al.*, 2020). This mirrors a general view of customer trust as confidence in an exchange partner's reliability and integrity, that the exchange partner is seen as consistent and competent (Morgan and Hunt, 1994), and that this partner will perform actions with positive outcomes for the customer (Anderson and Narus, 1990). The main source of an individual's trust in a system, a vendor or a product is most likely direct experience (Anderson and Weitz, 1989; Anderson and Narus, 1990; Dwyer *et al.*, 1987; Lee and See, 2004; Swan and Nolan, 1985). In the case of an AI-based product, a high level of direct experience would be represented by continuous use of the product over time, as well as frequent communications with the vendor of the product. However, in the absence of direct experience, such as in a pre-purchase phase, message-related variables can influence trust, too. More specifically, in the present study, we hypothesize that both explanations about how a product works and its benefits can enhance trust.

With respect to explanations about how a product works, they can comprise characteristics that, in general, enhance trust. First, a sender of information who provides explanations is likely to be viewed as open and willing to share information, and these characteristics can enhance trust (Lee and See, 2004). Second, when a sender provides explanations, it can enhance perceptions of the sender's expertise and competence – and these two characteristics are also key determinants for trust (Lee and See, 2004; Moorman *et al.*, 1993; Swan and Nolan, 1985). In addition, communication involving meaningful information is a main precursor of trust (Anderson and Narus, 1990; Morgan and Hunt, 1994). Thus, given that an explanation increases perceived meaningfulness, this represents another way in which explanations can enhance trust. And for autonomous, intelligent systems, it has been argued that trust is enhanced if the user has accurate knowledge of the system's abilities (de Graaf and Malle, 2017). Therefore, if explanations comprise information about such abilities, one would expect that trust is enhanced. Empirical evidence for a positive explanation–trust association has been reported by Suen and Hung (2023). Based on this, the following is hypothesized for providing information about an AI-based product:

H3. Information about an AI product that comprises explanations of how the product works enhances trust.

We hypothesize that providing information about benefits can enhance trust, too. One reason is the presence of a goal component in trust; as noted above, trust in a system comprises the belief that the system will help the individual to achieve his or her goals (Kraus *et al.*, 2020; Lee and See, 2004; Lee, 2018). Typically, this goal component is addressed by benefit information, in the sense that benefits can be seen as means to achieve high-level needs with a goal character (Gutman, 1982; Vriens and Ter Hofstede, 2000). Another reason is that typical benefits (in commercial messages) are positively valenced and can thereby signal that the provider of such benefits is concerned with the well-being of the receiver. And in trust-related theory, such perceived concerns are either an antecedent to trust (Morgan and Hunt, 1994) or part of the trust construct *per se* (e.g. Anderson and Narus, 1990). Therefore, the following is hypothesized for providing information about an AI-based product:

H4. Information about an AI product that comprises product benefits enhances trust.

2.3 Implications for perceived product performance

The downstream variable in our conceptual framework is the perceived performance of an AI-based product. As already indicated, perceived product performance is one of the main antecedents to user acceptance of technology in the widely used UTAUT model (Venkatesh *et al.*, 2003). It also represents a common downstream variable in the literature on the effects of marketing communications (e.g. Sundar *et al.*, 2020; Usrey *et al.*, 2020), which basically shares the same main assumption as the XAI literature: information about a product has implications for the receiver's view of the product. More specifically, perceived product performance can have behavioral implications of interest for those who develop AI-based products, namely, a positive impact on variables related to buying and using a product (Mostafa and Kasamani, 2022; Usrey *et al.*, 2020). Performance in an AI context has been conceptualized as the degree of success of a system at effectively conducting the tasks for which it is designed (Hoffman *et al.*, 2018). In terms of empirical studies in an AI context, it has been used as an effect variable by Shin (2021). It may be noted that perceived product performance is essentially an evaluation (i.e. it can take on values on a bad–good continuum).

First, we hypothesize that understandability can have a positive impact on perceived product performance. One reason is that understanding is a main component of meaning in life (Martela and Steger, 2016), which makes it likely that understanding would be a positively charged state of mind for most people. Therefore, we expect that understanding can produce a positive affective reaction (i.e. it can enhance both pleasure and arousal). Then, in the next step, and according to the affect infusion model (Forgas, 1995), this positive affective reaction can have a valence-congruent impact on evaluations. In other words, the information processing is of this type: “If it makes me feel good, it must be good” (Pham, 2004). Moreover, understandability in relation to a product can enhance processing fluency; that is to say, the speed, efficiency and accuracy of stimulus processing (Winkielman *et al.*, 2006). Processing fluency typically generates positive affect (Winkielman *et al.*, 2006), because high processing fluency is associated with successful recognition of a stimulus, error-free processing and the availability of appropriate knowledge structures to interpret the stimulus (Reber *et al.*, 2004; Winkielman and Cacioppo, 2001). And, again according to the affect infusion model, positive affect generated in this way can have a valence-congruent impact on product evaluations. Indeed, high fluency is reliably associated with more positive evaluations from observers of a stimulus; observers are likely to interpret the positive affect elicited by a high level of fluency as their evaluative response to the stimulus (Reber *et al.*, 2004). Empirical evidence in support of an understandability-evaluation association for AI systems has been reported by Liu *et al.* (2022). Similarly, in an advertising setting, Goodwin and Etgar (1980) and Mick (1992) show that subjective ad comprehension is positively associated with the attitude toward the ad. In the context of information about an AI-based product that produces understandability, then, positive affect generated by (self-perceived) understanding of the product can carry over and inform the perceived performance of the product in a valence-congruent way. Hence, we hypothesize the following:

H5. Understandability is positively associated with perceived product performance.

Second, we hypothesize that trust has a positive impact on perceived product performance. One main reason is that trust is typically based on beliefs related to specific performance aspects, such as accuracy, reliability, competence, expertise and quality (Dwyer *et al.*, 1987; Gounaris, 2005; Moorman *et al.*, 1993; Mukherjee and Nath, 2007; Sichtmann, 2007; Swan and Nolan, 1985; Wang and Benbasat, 2007), and such beliefs can be expected to be the basis for on overall assessment of a product’s performance. Alternatively, trust is used for inferences about specific performance aspects. When this happens, trust is an “information surrogate” for less visible characteristics of an offer (Sichtmann, 2007). Moreover, it should be recalled that there is a goal component in trust; in a product context, trust can be seen as the belief that a product will help the individual to achieve his or her goals (Kelly *et al.*, 2023; Kraus *et al.*, 2020; Lee and See, 2004; Lee, 2018). Assuming that beliefs that one’s goals will be reached represent a positively charged state of mind, affect infusion in the same way as mentioned above (Forgas, 1995) can be expected: the positive affect generated by goal achievement colors the perceived performance of a product in a valence-congruent way.

In addition, trust is typically negatively associated with uncertainty (Morgan and Hunt, 1994) and perceived risk (Agag and El-Masry, 2017; Hengstler *et al.*, 2016; Mukherjee and Nath, 2007), and a product that is attributed with low levels of uncertainty and risk is likely to be rewarded with higher performance perceptions.

Empirically, and for AI systems, Mostafa and Kasamani (2022) and Shin (2021) have identified a positive association between trust and perceived performance, while Liu *et al.* (2022) report a positive correlation between trust and the perceived value of the system (i.e.

another type of evaluation). Indirect evidence for this association (given that perceived product performance is positively correlated with product acceptance) is also provided by numerous studies with a TAM framework, in which trust is typically influencing the acceptance of a new technology (Kelly *et al.*, 2023). Therefore, the following is hypothesized:

H6. Trust is positively associated with perceived product performance.

3. Research method

3.1 Research design, stimulus material and participants

A between-subjects experiment was employed to manipulate pre-purchase information about an AI-based product with respect to explanations about how it works (absent vs present) and the benefits it provides (absent vs present). This emphasis on presence versus absence in the manipulations is the same as in some previous studies of the effects of explanations in the XAI literature (Kenny *et al.*, 2021; Leichtmann *et al.*, 2023; Wang and Benbasat, 2007).

The specific stimulus product was an add-on option for cars, which we created for the purpose of this study. The idea was that it should be seen as "adjacent possible" (i.e. it is only a few steps away from what already exists). Using stimuli in an experiment that do not (yet) exist in the real world outside the experiment may perhaps be questioned. Nevertheless, the present study is influenced by a view stressing that it is less important how an experimental setting mirrors the real world outside the experiment in concrete terms; it is more important that the experimental setting has the same structural characteristics as settings existing outside the experiment. This view means that the main purpose of an experiment (in academic research) is not to create generalizations based on a specific stimulus in an experiment (here: an add-on option). The main purpose of an experiment is to test theoretically-based propositions; by definition, theory is general, and it is theory rather than reactions to a specific object that is the basis for generalizations (Mook, 1983).

The add-on option was described as based on a rear mirror camera inside a car; it continuously captures the driver's emotional state by facial recognition software, and it uses these data for decisions about informing the driver when he or she reaches critical emotional levels. A second idea was that the stimulus product should reflect the type of automation that is expected for many AI-based products. That is to say, it actively selects data, transforms information and makes decisions (cf. Lee and See, 2004). A third idea was to broaden the range of stimuli used in research on the effects of AI systems. Examples of stimuli employed in previous research is a computer algorithm to make decisions for university admission (Cheng *et al.*, 2019), autogenerated news content (Shin, 2021), an asynchronous video procedure for interviewing job applicants (Suen and Hung, 2023), a robot able to gather intelligence in a foreign town (Wang *et al.*, 2016) and a computer program that helps the user in classifying leafs (Yang *et al.*, 2020). The ambition was also to provide product information similar to what firms do when they communicate with potential and existing customers within their own Internet-based communication channels.

In our manipulations, all participants received the same baseline information about this add-on option (see the Appendix). For the manipulation of explanations of how a product works, and for the two presence-of-explanation-content conditions, an explanation part (see the Appendix) was added to the baseline information. This part was not included in the two absence-of-explanation-content conditions. It should be noted that there are several ways in which information about an AI-based product can comprise explanations. In the present study, one specific type of "how-explanations" (Wang and Benbasat, 2007) was chosen, namely what Cheng *et al.* (2019) label a "black-box approach". That is to say, this type of explanation comprises information about how inputs relate to outputs without showing the internal workings of the algorithms. For the manipulations of benefits, a benefit part (see the Appendix) was added to the presence-of-benefit-content conditions. This part stressed that the main benefit of the add-on option was to minimize the risk of traffic accidents and thus it would keep the driver safe in a traffic environment. For the absence-of-benefit-content conditions, this part was not added.

The participants were randomly allocated to one of the four versions of the product information. They were instructed to read the information and to answer the subsequent questions. We recruited the participants from Prolific, an online panel built for research purposes (Palan and Schitter, 2018). Four hundred 84 participants completed the study. However, four participants failed to respond correctly to the attention check item used in the study, and they were removed. The attention check item was formulated as follows: "In this study, you were asked questions about what?" followed by the response alternatives "An add-on option for boats," "An add-on option for cars" and "An add-on option for computers". The analysis was based on those that remained ($n = 480$, $M_{age} = 38.37$; 244 women, 235 men and 1 other; all were UK residents). Of these participants, 119 were exposed to the no explanation-no benefit condition, 122 to the explanation-no benefit condition, 117 to the no explanation–benefit condition, and 122 were exposed to the explanation-benefit condition.

3.2 Measurements

Multi-item measures were used for all variables in the hypotheses. Each item was scored on a 10-point scale, and Cronbach's alpha (CA), composite reliability (CR) and average variance extracted (AVE) for the measures, computed with Smart PLS 4.0, are presented below. A table with zero-order correlations is provided in the Appendix.

To assess the degree of the provided information's perceived *explanation content* regarding how the product works, the participants were given the following statements: "The description of the add-on option made it clear how this option works," "The description of the add-on option provided information that explains the function of the option" and "The description of the add-on option comprised facts about how it produces its output" (1 = Do not agree at all, 10 = Agree completely; CA = 0.85, CR = 0.86, AVE = 0.77; $M = 7.48$, SD = 1.76). The degree of perceived *benefit content* in the provided information was measured with the statements "The description of the add-on option provided information about the benefits of using it," "The description of the add-on option made it clear why this option would be helpful for car drivers," "The description of the add-on option clearly showed what the value would be for a driver who is using the option" and "The description of the add-on option articulated the advantages of using the option" (1 = Do not agree at all, 10 = Agree completely; CA = 0.94, CR = 0.96, AVE = 0.84; $M = 6.99$, SD = 2.23).

Self-reported *understandability* of an AI-based system has been assessed with single-item measures in several studies (e.g. Cheng *et al.*, 2019; Leichtmann *et al.*, 2023; Schoonderwoerd *et al.*, 2021). In the present study, however, a multi-item approach was used. The participants were asked the following four questions developed for the purpose of this study: "To what extent was it easy or difficult to understand how this AI-based option works?," "To what extent was it difficult or easy to understand how this AI-based option processes information?," "To what extent was it easy or difficult to understand how this AI-based option makes decision?" and "To what extent was it easy or difficult to understand how this AI-based option produces its output?" (1 = Difficult, 10 = Easy; CA = 0.93, CR = 0.95, AVE = 0.82; $M = 7.35$, SD = 1.87). The difficult versus easy scale endpoints were used so that the operationalization would be compatible with the classical literature on adoption of innovations in which understandability is often labeled complexity (defined as the degree to which an innovation is easy or difficult to understand and use; Rogers, 1995). Similar items, based on the extent to which a message is easy or difficult to understand, have been used in assessments of subjective comprehension of advertisements (e.g. Huhmann and Mott-Stenerson, 2008; Mick, 1992; Nantel and Sekhavat, 2008).

Single-item measures of *trust* in algorithms and AI-based systems have been used, for example, by Lee (2018), Leichtmann *et al.* (2023) and Schoonderwoerd *et al.* (2021). Again, however, in the present study, a multi-item approach was chosen and items adopted from Hoffman *et al.* (2018) and Sichtmann (2007) were used: "The output of the add-on option is trustworthy," "I am confident that this add-on option works well when it is used," "I feel that it would be safe to rely on this add-on option," "This option appears to be able to produce correct recommendations for users," "I believe that this add-on option will result in positive outcomes for the user" and "If I were driving a car with this add-on option, I would follow its advice" (1 = Do not agree at all, 10 = Agree completely; CA = 0.95, CR = 0.96, AVE = 0.80; M = 5.14, SD = 2.16).

Perceived product performance was measured with the question "What is your view of the performance of the add-on option?," which was followed by the items "Poor – Good," "Low performance level – High performance level," "It delivers below expectations – It delivers above expectations" and "Dissatisfying – Satisfying" (CA = 0.95, CR = 0.96, AVE = 0.87; M = 5.76, SD = 2.01). Similar items appear in a scale for product performance evaluations used by Usrey *et al.* (2020) and in Black and Kramer's (2009) scale for product performance expectations. To assess the nomological validity of this measure, and assuming that perceived product performance positively influences intentions to buy and use a product (e.g. Block and Kramer, 2009; Sundar *et al.*, 2020), an assessment of the willingness to use the product was included. The participants were asked, "To what extent would you be willing to use the add-on option while driving a car?" (1 = Not at all, 10 = Very much; M = 3.71, SD = 2.79). Then, we computed the zero-order correlation between the perceived product performance variable and the willingness-to-use variable. As the correlation was positive and significant (r = 0.67, p < 0.01), it can be concluded that the performance measure behaved as it was supposed to behave given one of its evidence-based consequences and thus that it exhibits nomological validity.

4. Analysis and results

4.1 Manipulation check

We used a 2 (no explanation vs explanation) × 2 (no benefit vs benefit) ANOVA to assess the effects of the explanation manipulation on perceptions about explanation content. This resulted in a significant main effect for the explanation factor [F (1, 479) = 23.64, p < 0.01, η^2 = 0.05]. The impact of the benefit factor was not significant [F (1, 479) = 3.70, p = 0.06 η^2 = 0.0008]. The perceived level of explanation content was lower when an explanation was not present (M = 7.08, SD = 1.93) compared to when an explanation was present (M = 7.85, SD = 1.49). Similarly, the same ANOVA was used to assess the effect of the benefit manipulation on the perceived benefit content. This resulted in a significant main effect of the benefit factor [F (1, 479) = 87.85, p < 0.01, η^2 = 0.16]. The explanation factor was not significant [F (1, 479) = 3.05, p = 0.08, η^2 = 0.006]. The level of perceived benefit content was lower in the conditions without benefits (M = 6.12, SD = 2.31) compared to the conditions with benefit content (M = 7.88, SD = 1.76). The interaction effects in these ANOVAs were not significant. It can be concluded, then, that both manipulations worked as intended.

4.2 Testing the hypotheses

The hypotheses were tested with a structural equation modeling approach (we used SmartPLS 4.0). The proposed model comprised the hypothesized associations *H1–H6* (see Figure 1). In the assessment of the proposed model, the perceived explanation content variable and the perceived benefit content variable were used as independent variables. In this proposed model, all item loadings were > 0.80. As for discriminant validity, SmartPLS recommends using the heterotrait-monotrait ratio of correlations (HTMT) for the assessment and 0.9 as a critical value (i.e. if the HTMT value is below 0.9, discriminant validity has been established between two reflectively measured constructs). In our case, all HTMT values were below 0.9 (see Table 1). Moreover, no critical levels of collinearity were at hand (all VIFs were < 3). The explained variance (i.e. R^2) in perceived product performance was 0.62.

The (standardized) path coefficients for *H1–H6* are reported in Table 2, which reveals that *H1–H6* were supported. In the subsequent parts, we examine these results more in detail with respect to each of the two types of product information (i.e. explanations and benefits).

4.3 Explanations a source of influence

H1 (information comprising explanations of how a product works enhances understandability) was supported, because there was significant positive association between these two variables ($b = 0.68$, $p < 0.01$). This outcome confirms the view of authors who argue that offering explanations of how an AI-based product works enhances understandability (e.g. Chazette and Schneider, 2020; Leichtmann *et al.*, 2023; Meske *et al.*, 2022). *H3* (information comprising explanations about how a product works enhances trust) was also supported ($b = 0.22$, $p < .01$), and this mirrors the XAI assumptions about the positive effects of such explanations on trust (e.g. Meske *et al.*, 2022; Miller, 2019; Rai, 2020; Shin, 2021; Wang *et al.*, 2016).

Table 1. Heterotrait–monotrait ratios (HTMT)

Variables	Benefits	Explanation	Performance	Trust
Benefits				
Explanation	0.563			
Performance	0.545	0.474		
Trust	0.522	0.455	0.827	
Understandability	0.466	0.815	0.447	0.481

Source(s): Authors' own work

Table 2. Path coefficients for the *H1–H6* associations

Hypotheses	b	SE	t	f^2	p
H1. Explanation – Understandability	0.68	0.035	19.51	0.73	< 0.01
H2. Benefits – Understandability	0.09	0.041	2.26	0.01	0.024
H3. Explanation – Trust	0.22	0.045	4.79	0.05	< 0.01
H4. Benefits – Trust	0.39	0.042	9.25	0.15	< 0.01
H5. Understandability – Perceived performance	0.08	0.037	2.14	0.01	0.033
H6. Trust – Perceived performance	0.75	0.026	29.47	1.20	< 0.01

Source(s): Authors' own work

Each of those two effects of providing information comprising explanations (i.e. understandability and trust) were hypothesized to influence perceived product performance, and these hypotheses were supported, too. That is to say, the association between understandability and perceived product performance was positive and significant ($b = 0.08$, $p = 0.03$), which provides support for *H5*. Similarly, the association between trust and perceived product performance was positive and significant ($b = 0.75$, $p < 0.01$); this provides support for *H6*.

These outcomes suggest an extended causal role for providing explanations of how a product works. More specifically, the supported *H1* and *H5* indicate that information comprising explanations of how a product works has an indirect impact on perceived product performance (via understandability), while the supported *H3* and *H6* indicate that explanations of how a product works is indirectly influencing perceived product performance (via trust). The outcomes, then, indicate the possibility of two routes of mediation:

(1) Explanation – Understandability – Product performance

(2) Explanation – Trust – Product performance

To assess this, the mediation analysis approach for structural equations modeling advocated by Nitzl *et al.* (2016) and Sarstedt *et al.* (2020) was used. This analysis approach comprises adding direct links to a proposed model to (a) control for the possibility that the impact of an independent variable can be direct rather than mediated and, (b) if mediation is at hand, to be able to assess the type of mediation. To examine the possibility of an indirect impact of providing explanations of how a product works on perceived product performance, then, a direct link between Explanation and Product performance was added to the proposed model in Figure 1.

For the first causal chain (1), the mediation assessment revealed that providing explanations did not have a significant indirect impact on product performance perceptions via understandability ($b = -0.001$, $t = 0.19$, $p = 0.85$). For the second causal chain (2), providing explanations had a significant indirect impact on product performance via trust ($b = 0.16$, $t = 4.73$, $p < 0.01$). Hence mediation was at hand with respect to trust. In addition, the direct Explanation – Product performance association was significant ($b = 0.13$, $t = 2.86$, $p < 0.01$), so the type of mediation in chain (2) should be seen as complementary (Zhao *et al.*, 2010). Taken together, then, and with respect to the effects of providing explanations of how an AI-based product works, the analysis revealed that explanations did more than just enhancing understandability and trust; they also contributed positively to the perceived performance of the product (both directly and indirectly via trust).

4.4 Benefits as a source of influence

Turning now to the other information type, product benefits, *H2* (i.e. benefits enhance understandability) was supported ($b = 0.09$, $p = 0.03$). In addition, benefits enhanced trust ($b = 0.39$, $p < 0.01$), which means that *H4* was supported.

Similar to above, the support for the hypotheses about the influence of understandability and trust on perceived product performance (*H5* and *H6*) suggests that providing information about benefits could have an indirect influence on perceived product performance. More specifically, and in analogy with the examination of the effects of information comprising explanations of how a product works, two possible routes of mediation call for a further examination:

(1) Benefits – Understandability – Product performance.

(2) Benefits – Trust – Product performance.

To assess these possibilities of mediation, we returned to the originally proposed model in Figure 1 and added a direct link between benefits and perceived product performance. For the causal chain (3), the mediation assessment including this direct link revealed that providing benefit information did not have a significant indirect impact on perceived product performance via understandability ($b = 0.003$, $t = 0.82$, $p = 0.41$). This, then, indicates that mediation was not at hand. For the causal chain (4), providing information about benefits had a significant indirect impact on perceived product performance via trust ($b = 0.27$, $t = 8.94$, $p < 0.01$). Thus, mediation could be supported for causal chain (4). The direct Benefit–Product performance association was significant ($b = 0.16$, $t = 4.06$, $p < 0.01$), which indicates complementary mediation (Zhao *et al.*, 2010) in the case of chain (4).

Hence, with respect to the effects of including information about the benefits an AI-based product would produce for a user, the analysis indicates that benefit information contributed not only to understandability and trust; it also contributed positively to perceptions of product performance (both directly and indirectly via trust). It should be noted that the effects of providing benefit information were similar to the effects of providing explanations. One exception, however, is how these two types of information influenced understandability; in our analysis, explanations had a relatively stronger effect ($b = 0.68$) than benefits ($b = 0.09$) on understandability (cf. Table 2). This difference in strength was significant ($Z = 7.21$, $p < 0.01$) according to the single sample test of associations between variables (p. 100) in Kleinbaum *et al.* (1988).

5. Discussion

5.1 Summary of main findings

The main findings in the present study are that both pre-purchase information about how an AI-based product works and information about its benefits enhance understandability and trust and boost perceived product performance. The results also show that the positive impact of these types of information on perceived product performance is both direct and indirect (mediated by trust).

5.2 Contributions

The present study offers several contributions to the literatures on explanations and benefits. With respect to explanations, and in general, they satisfy important needs for us humans, so it is not surprising that many authors stress that AI-based systems should comprise explanations of how such systems work. More specifically, several authors have argued that explanations of how AI-based products work would increase understandability (e.g. Chazette and Schneider, 2020; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Shin, 2021; Wang *et al.*, 2016). To date, however, only a limited number of empirical studies provide evidence for this, such as Cheng *et al.* (2019), Shin (2021) and Stange and Kopp (2020). The present study, then, contributes to the existing literature by showing additional evidence of an explanation–understandability association. Similarly, many authors have argued that providing explanations would enhance trust (Ali *et al.*, 2023; Chazette and Schneider, 2020; Glikson and Woolley, 2020; Leichtmann *et al.*, 2023; Meske *et al.*, 2022; Rai, 2020; Shin, 2021; Yang *et al.*, 2020), which has been empirically supported by Shin (2021), Yang *et al.* (2020) and Wang *et al.* (2016). The present study corroborates this and provides additional evidence for an explanation–trust link.

More importantly, however, most of the literature in the area of explainable AI has focused on the situation in which the AI system itself is the supplier of explanations while the system is used. This is a natural and important focus, because one can foresee many usage situations in which there is a need for AI systems to explain what they are doing while they are used. Before an AI-based product is actually used, however, there is a need to convince potential users that they should use it (and, in many cases, persuade them that this usage is worth paying for). Thus, there is a need to provide explanations also in a pre-purchase phase, and this is the situation that the present study has made attempts to address. As indicated earlier, a main reason behind the need to examine reactions in a pre-purchase phase is that consumers are likely to process information differently in this phase compared to when they are actually using a product (Daugherty *et al.*, 2008; Hamilton and Thompson, 2007; Mooy and Robben, 2002; Smith and Swinyard, 1982). This potential for differences, however, has to date not been acknowledged in the XAI literature. In the comprehensive literature review by Ali *et al.* (2023), for example, there is no explicit recognition of a pre-purchase situation versus a usage situation. In any event, our findings contribute to the existing literature by showing effects of providing explanations in a pre-purchase phase.

In addition, the present study shows that explanations of how an AI-based product works have a more extensive causal role than just influencing understandability and trust. That is to say, the present study contributes to the existing literature on XAI by including perceived product performance as a (downstream) evaluation variable. This should be seen in the light of the typical XAI study, in which overall evaluations of provided *explanations* is sometimes used as a dependent variable (e.g. Riveiro and Thill, 2021). The typical XAI study, however, rarely includes overall evaluations of the *product* or *system* of the type that is common in consumer-related research in marketing (e.g. customer satisfaction, perceived service quality, the overall attitude toward a product and perceived product performance). One argument for including such variables (in our case: perceived product performance) also in XAI research is that it would encourage XAI researchers to examine research outside their own specific field, which in turn is likely to stimulate the development of richer theories.

When it comes to information about product benefits, it should first be noted that such information is typically neglected in the XAI literature, which is focused on information about products *per se*. This represents a gap between the XAI literature and the marketing literature; marketing authors have over the years argued that a main reason behind customers' purchases is that products provide them with benefits and, consequently, that explicit pre-purchase information about product benefits is a useful communication approach (e.g. Gutman, 1982; Haley, 1968; Otnes *et al.*, 2014). Indeed, from a marketing point of view, it may be argued that relatively few consumers are likely to be helped by explanations of how AI works if they involve attempts to explain how AI actually works. That is to say, not everyone would find it useful with explanations comprising notions of neural networks, deep learning algorithms, posterior probabilities, feature vectors, lossless data compressors and overfitting. Nevertheless, relatively few studies in marketing have examined the effectiveness of providing explicit benefit information. One such study is Graeff (1997), who found that benefit information about a product boosts the attitude toward the product. Another study is Wu *et al.* (1988), showing that product benefit information predicts product preferences better than product attribute information. These studies, however, did not include mediators such as understandability and trust. The present study, then, contributes to the literature on marketing communications by showing that the presence of information about product benefits influences subsequent steps in the receiver's information processing activities – and that such information, ultimately, has a positive impact on perceptions of product performance.

Moreover, in our study, the impact of benefit-comprising information was mediated by trust (but not by understandability), which indicates that benefit information in a message may influence beliefs about the sender's concern for the receiver's goals and well-being. Thus, it may influence also other beliefs than those comprising how specific product attributes are means to reach goals. If this is so, it seems as if the commonly used means-end framework to handle the role of benefits (e.g. Botschen *et al.*, 1999; Gutman, 1982) would profit from taking also trust into account. Yet, it should be underlined that the present study shows that communicating explanations of the type stressed in the XAI literature (how a product works) indeed contributes to receivers' reactions to product information. This, then, suggests that a "benefit-only" approach may not be optimal – at least not for AI-based products. In any event, in the context of new products, the finding that communicating benefits can enhance perceptions of products should be seen in the light of a consistent pattern in the literature on adoption of innovations: customer-perceived product advantages can enhance adoption (e.g. Holak, 1988). However, the sources of perceived advantages have rarely been studied, so the present study contributes to the general literature on adoption of innovations by showing that providing benefit information in a product description is an influential activity.

5.3 Implications for decision makers

Our results show that both product information with XAI-based explanations (i.e. information about how a product works) and information about a product's benefits (i.e. consequences for the user) had direct and positive effects on perceived product performance – and both types of information had an indirect and positive impact on perceived product performance via trust. For designers, programmers, marketers and managers who are involved in communicating with customers about AI-based products in pre-purchase situations and who want customers to have a positive view of the product's performance, the main implication of the results from the present study is this: such communications should comprise both explanations and benefits.

Product information comprising explanations and benefits can be designed in many ways, so it is advisable to pre-test various versions. If the communication takes place online, it would be relatively easy to conduct such pre-testing in terms of randomized experiments (Thomke, 2020). In addition, some existing AI-based products have generated a vast number of users, such as ChatGPT and Google Translate, and the ways in which they provide explanations and benefits may be used as a source of inspiration for execution decisions.

Moreover, in a very broad sense, both explanations and benefits may be seen as disclosure activities. And disclosure, sometimes referred to as targeted transparency, has become ubiquitous for many products (Loewenstein *et al.*, 2014). From a legal point of view, and with respect to AI-based products, it seems as if explanations rather than benefits would be more likely to be subject to calls for mandatory disclosure (e.g. in the European Union; Hoffman *et al.*, 2018), so decision makers may need to be particularly mindful when it comes to providing information that explains how AI-based products work.

The following aspect of our results should also be underlined: although both information about how a product works and information about benefits significantly enhanced trust, the two types of information content explained 27% of the variation in trust. Thus, more activities can be undertaken to boost trust – and such activities indeed seem to make sense, because trust had a relatively strong positive impact on perceived product performance in our study. Decision makers who want to capitalize on the causal potency of trust in a pre-purchase situation may therefore seek inspiration from general theory on trust-driving factors beyond those that are related to providing information comprising explanations and benefits. Given that a main source of trust is direct experience (Anderson and Weitz, 1989; Anderson and Narus, 1990; Dwyer *et al.*, 1987; Lee and See, 2004; Swan and Nolan, 1985), decision makers should consider making it possible to test a product before it is purchased. Decision makers should also consider offering testimonials from existing customers as a form of vicarious direct experience, because it has been shown that characteristics of users who make endorsements in commercial messages regarding high technology products can boost trust (Lafferty and Goldsmith, 2004).

The present study has implications also for another type of decision maker, namely, researchers – particularly those who examine the impact of product information on consumers' responses. More specifically, given the main finding (i.e. information about explanations and benefits in a message about a product influences receivers' evaluation of the product), researchers need to be mindful about these two types of product information in studies examining the effect of messages. This is especially relevant for experimental studies in which researchers create product information stimuli. Thus, researchers need to be mindful about the extent to which their manipulations comprise explanations of how a product works and product benefits, because the presence of such information can create confounding effects in research in which the focus is on the impact of other aspects of information.

5.4 Limitations and suggestions for further research

In the present study, one specific AI-based product was used as stimulus (an add-on option for cars to provide safety for car drivers). It should be seen as a sample from a population of products to be used in situations in which there can be severe consequences for the user. Examples of other such products are those that are used for medical diagnosis, recruitment, autonomous vehicles (Blösser and Weihrauch, 2024; Meske *et al.*, 2022) and for helping judges make decisions about defendants (Cheng *et al.*, 2019). AI, however, is used also for products with a lower potential for severe consequences, and it is possible that explanations of how they work would produce other reactions than those that we identified in the present study. Presumably, if the consequences do not matter much, an explanation would become a less potent causal variable. Therefore, further research is needed to examine the impact of information comprising how a product works and benefits for other usage situations.

Moreover, the specific stimulus product was of the embedded AI type (Glikson and Woolley, 2020) as opposed to AI-products with a visual representation (e.g. a chatbot and a service robot). Clearly, if an AI-based product has a distinct physical form, this form can contribute to consumers' understanding. This has been realized repeatedly over the years for other products. Thomas Edison, for example, deliberately created similarities between his electric light bulb and the kerosene lamp in an attempt to facilitate understanding of the new technology (Rindova and Petkova, 2007). Thus, further research is needed to examine the extent to which explanations of AI-based products with a visual representation would produce the same reaction patterns as in the present study.

It should also be noted that our manipulations were made in such a way that there was more text in the conditions in which explanations and benefit information were present versus absent. Given this, the mere volume of text may have influenced the results: it is possible that more text *per se* can have a positive impact on the receiver's reactions. The opposite, however, is also possible (i.e. the longer the text, the more taxing it is to read it, which can attenuate evaluations of what the text is about). We cannot test this with our data. But we can turn to existing studies on the effects of copy length in commercial messages for indications of how the length of a commercial message affects the receiver. The pattern in such studies is that copy length does not produce a significant effect on brand attitudes (e.g. Hanson, 2016). The same pattern is at hand in comparisons of the length of television commercials (e.g. Percy and Rossiter, 1992; Singh and Cole, 1993). Only one study that we have found points toward a potential advantage for longer copy; Soley (1986) show that industrial ads with more words generated a higher "started-to-read" score for "industrial prospects" than ads with fewer words (but there were no other dependent variables in that study). In any event, research of this type typically concludes that other aspects of a message than its length are more causally potent, so we believe that this is the case also for our stimulus material.

In addition, what is meant by an explanation in an XAI context is subject to variation between authors (Ali *et al.*, 2023; Dazeley *et al.*, 2021). The present study used "how-explanations" in terms of what Cheng *et al.* (2019) label a "black-box approach;" this type of explanation comprises information about how inputs relate to outputs without showing the internal workings of the algorithms. However, several other types of explanations exist (Dwivedi *et al.*, 2021; de Graaf and Malle, 2017; Preece, 2018). Further research, then, is needed to examine the effects of also other explanations than those that were examined in the present study. An important task is also to examine such explanations in relation to views of what an explanation is in the academic literature. An explanation in the latter sense, for example, can be seen as arguments demonstrating how what is being explained follows deductively from natural laws and empirical conditions (Lombrozo, 2006) or as "proofs in logic" (Keil, 2006). Explanations of this type, however, are unlikely to be helpful for users who want to understand either how a specific product works or what effects its usage has on the user.

Experimental researchers should also be mindful about how to develop stimuli for a condition with no explanations (and no benefits). If this condition becomes uninformative for participants, there is a possibility that it becomes negatively charged due to lack of information. And this could affect other variables negatively. In the present study, it can be noted that our baseline condition, with no explanations and no benefits, generated the lowest level of perceived product performance (M = 5.41, SD = 2.09) of all conditions. Yet, the level was close to the scale midpoint rather than its highly negative endpoint.

Another aspect to consider is that our mediation analyses resulted in two (non-hypothesized) significant direct associations; one between explanation content and perceived product performance, and one between benefit content and perceived product performance. This, then, indicates the possibility that other mediators (than understandability and trust) can be involved. Further studies are needed to examine this. For the explanation content route of influence, several potential mediators deserve attention. One possibility is that the influence of explanation content on perceived product performance is based on a perceived supplier effort factor: an explanation may signal effort from the firm, and such effort can be used as a sign of product quality by the customer (Kirmani, 1990). Then, in the next step, quality perceptions enhance perceived product performance. This possibility has been discussed by Buell (2019), who argues that operational transparency (showing how something is produced) can boost effort perceptions which, in the next step, can enhance perceptions of value and quality. It is also possible that explanations of how an AI-based system works would make a "machine heuristic" (Sundar and Kim, 2019) increasingly accessible to a receiver, and thereby also increase the accessibility to positively charged beliefs about machines (e.g. they are more objective, secure, precise and safe than humans), and this can have a positive impact on perceived product performance.

Potential additional mediators exist also for the benefit route of influence on perceived product performance. Presumably, a priming mechanism may be at hand (cf. Janiszewski and Wyer, 2014). More specifically, benefits mentioned in product information are typically positively charged (e.g. "safety" and "freedom") and they can make accessible material in cognitive categories with a positive charge, which in the next step carries over and informs a receiver, in a valence-congruent way, about the performance of a product. Another possibility is that claims about product benefits "become" product beliefs in the receiver's mind. Indeed, in general, people have a tendency to believe what is claimed in a message, because it is more convenient (and less costly from an information-processing point of view) to believe a claim than to question it (Gilbert, 1991; Gilbert *et al.*, 1993). And given that the receiver believes that a product offers a set of positively valued benefits, one would expect that such beliefs can positively influence the receiver's perceptions of product performance. It is also possible that explicit information about (positively charged) benefits can signal that the sender has good intentions *vis-à-vis* the receiver, which can boost the receiver's view of the sender's warmth (cf. Fiske *et al.*, 2007). Then, in the next step, warmth perceptions can influence performance perceptions positively.

There is also a possibility that the associations identified in the present study may be moderated. First, the present study comprised a pre-purchase situation, which may generate different strength in associations between variables compared to a situation involving actual use of a product (in which explanations can be offered in real time). Indeed, the mediators in the present study may change roles in a usage situation as opposed to a pre-purchase situation. Trust, for example, can evolve and change over time as usage occurs (Kraus *et al.*, 2020). Or, as stated by Hoffman *et al.* (2018), trust in an AI system will always be exploratory and can be subject to variation over time. Second, people are likely to react differently to explanations (and benefits) in an AI context depending on the level of their prior AI knowledge (Preece, 2018). In addition, several authors have suggested that individuals are subject to variation when it comes to their propensity for trust (Agag and El-Masry, 2017; Hoff and Bashir, 2015; Swan and Nolan, 1985; Wang and Benbasat, 2007), and such differences can influence the strength by which trust is affected by (and is affecting) other variables. Further research, then, is needed to examine if the results in the present study would be invariant or not in different situations (i.e. pre-purchase vs usage) and for individuals with different levels of AI knowledge and propensity to trust.

Funding

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

References

Agag, G.M. and El-Masry, A.A. (2017), "Why do consumers trust online travel websites? Drivers and outcomes of consumer trust toward online travel websites", *Journal of Travel Research*, Vol. 56 No. 3, pp. 347-369.

Alam, L. and Mueller, S. (2021), "Examining the effect of explanation on satisfaction and trust in AI diagnostic systems", *BMC Medical Informatics and Decision Making*, Vol. 21 No. 1, pp. 1-15.

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N. and Herrera, F. (2023), "Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence", *Information Fusion*, Vol. 99, p. 101805.

Anderson, E. and Weitz, B. (1989), "Determinants of continuity in conventional industrial channel dyads", *Marketing Science*, Vol. 8 No. 4, pp. 310-323.

Anderson, J.C. and Narus, J.A. (1990), "A model of distributor firm and manufacturer firm working partnerships", *Journal of Marketing*, Vol. 54 No. 1, pp. 42-58.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), "Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82-115.

Barry, T.E. and Howard, D.J. (1990), "A review and critique of the hierarchy of effects in advertising", *International Journal of Advertising*, Vol. 9 No. 2, pp. 121-135.

Batra, R. and Keller, K.L. (2016), "Integrating marketing communications: new findings, new lessons, and new ideas", *Journal of Marketing*, Vol. 80 No. 6, pp. 122-145.

Block, L. and Kramer, T. (2009), "The effect of superstitious beliefs on performance expectations", *Journal of the Academy of Marketing Science*, Vol. 37 No. 2, pp. 161-169.

Blösser, M. and Weihrauch, A. (2024), "A consumer perspective of AI certification–the current certification landscape, consumer approval and directions for future research", *European Journal of Marketing*, Vol. 58 No. 2, pp. 441-470.

Botschen, G., Thelen, E.M. and Pieters, R. (1999), "Using means-end structures for benefit segmentation: an application to services", *European Journal of Marketing*, Vol. 33 Nos 1-2, pp. 38-58.

Buell, R.W. (2019), "Operational transparency", *Harvard Business Review*, Vol. 97 No. 2, pp. 102-113.

Bunt, A., Lount, M., and Lauzon, C. (2012), "Are explanations always important? A study of deployed, low-cost intelligent interactive systems", in *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, *February 14-17*, Lisbon, Portugal, pp. 169-178.

Chazette, L. and Schneider, K. (2020), "Explainability as a non-functional requirement: challenges and recommendations", *Requirements Engineering*, Vol. 25 No. 4, pp. 493-514.

Cheng, H.F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F.M., and Zhu, H. (2019), "Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders", in *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, pp. 1-12.

Daugherty, T., Li, H. and Biocca, F. (2008), "Consumer learning and the effects of virtual experience relative to indirect and direct product experience", *Psychology and Marketing*, Vol. 25 No. 7, pp. 568-586.

Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S. and Cruz, F. (2021), "Levels of explainable artificial intelligence for human-aligned conversational explanations", *Artificial Intelligence*, Vol. 299, p. 103525.

de Graaf, M.M., and Malle, B.F. (2017), "How people explain action (and autonomous intelligent systems should too)", in *2017 AAAI Fall Symposium Series*.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T. and Williams, M.D. (2021), "Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy", *International Journal of Information Management*, Vol. 57, p. 101994.

Dwyer, F.R., Schurr, P.H. and Oh, S. (1987), "Developing buyer-seller relationships", *Journal of Marketing*, Vol. 51 No. 2, pp. 11-27.

Eisend, M. and Tarrahi, F. (2016), "The effectiveness of advertising: a meta-meta-analysis of advertising inputs and outcomes", *Journal of Advertising*, Vol. 45 No. 4, pp. 519-531.

Fiske, S.T., Cuddy, A.J. and Glick, P. (2007), "Universal dimensions of social cognition: warmth and competence", *Trends in Cognitive Sciences*, Vol. 11 No. 2, pp. 77-83.

Folkes, V.S. (1988), "Recent attribution research in consumer behavior: a review and new directions", *Journal of Consumer Research*, Vol. 14 No. 4, pp. 548-565.

Forgas, J.P. (1995), "Mood and judgment: the affect infusion model (AIM)", *Psychological Bulletin*, Vol. 117 No. 1, pp. 39-66.

Gilbert, D.T. (1991), "How mental systems believe", *American Psychologist*, Vol. 46 No. 2, pp. 107-119.

Gilbert, D.T., Tafarodi, R.W. and Malone, P.S. (1993), "You can't not believe everything you read", *Journal of Personality and Social Psychology*, Vol. 65 No. 2, pp. 221-233.

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: review of empirical research", *Academy of Management Annals*, Vol. 14 No. 2, pp. 627-660.

Goodwin, S. and Etgar, M. (1980), "An experimental investigation of comparative advertising: impact of message appeal, information load, and utility of product class", *Journal of Marketing Research*, Vol. 17 No. 2, pp. 187-202.

Gounaris, S.P. (2005), "Trust and commitment influences on customer retention: insights from business-to-business services", *Journal of Business Research*, Vol. 58 No. 2, pp. 126-140.

Graeff, T.R. (1997), "Comprehending product attributes and benefits: the role of product knowledge and means-end chain inferences", *Psychology and Marketing*, Vol. 14 No. 2, pp. 163-183.

Grewal, D., Guha, A., Satornino, C.B. and Schweiger, E.B. (2021), "Artificial intelligence: the light and the darkness", *Journal of Business Research*, Vol. 136, pp. 229-236.

Gutman, J. (1982), "A means-end chain model based on consumer categorization processes", *Journal of Marketing*, Vol. 46 No. 2, pp. 60-72.

Haley, R.I. (1968), "Benefit segmentation: a decision-oriented research tool", *Journal of Marketing*, Vol. 32 No. 3, pp. 30-35.

Haley, R.I. (1984), "Benefit segmentation–20 years later", *Journal of Consumer Marketing*, Vol. 1 No. 2, pp. 5-13.

Hamilton, R.W. and Thompson, D.V. (2007), "Is there a substitute for direct experience? Comparing consumers' preferences after direct and indirect product experiences", *Journal of Consumer Research*, Vol. 34 No. 4, pp. 546-555.

Hanson, C.B. (2016), "The effect of advertorial format and copy length on attitudes of female (target) and male (non-target) audiences", *Atlantic Marketing Journal*, Vol. 5 No. 2, pp. 41-54.

Hengstler, M., Enkel, E. and Duelli, S. (2016), "Applied artificial intelligence and trust–the case of autonomous vehicles and medical assistance devices", *Technological Forecasting and Social Change*, Vol. 105, pp. 105-120.

Hoch, S.J. and Deighton, J. (1989), "Managing what consumers learn from experience", *Journal of Marketing*, Vol. 53 No. 2, pp. 1-20.

Hoff, K.A. and Bashir, M. (2015), "Trust in automation: Integrating empirical evidence on factors that influence trust", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 57 No. 3, pp. 407-434.

Hoffman, R.R., Mueller, S.T., Klein, G. and Litman, J. (2018), "Metrics for explainable AI: challenges and prospects", *arXiv Preprint arXiv:1812.04608*.

Holak, S.L. (1988), "Determinants of innovative durables adoption: an empirical study with implications for early product screening", *Journal of Product Innovation Management*, Vol. 5 No. 1, pp. 50-69.

Homer, P.M. (2006), "Relationship among ad-induced affect, belief, and attitudes", *Journal of Advertising*, No. 1, pp. 35-51.

Huhmann, B.A. and Mott-Stenerson, B. (2008), "Controversial advertisement executions and involvement on elaborative processing and comprehension", *Journal of Marketing Communications*, Vol. 14 No. 4, pp. 293-313.

Janiszewski, C. and Wyer, R.S. Jr (2014), "Content and process priming: a review", *Journal of Consumer Psychology*, Vol. 24 No. 1, pp. 96-118.

Keil, F.C. (2006), "Explanation and understanding", *Annual Review of Psychology*, Vol. 57 No. 1, pp. 227-254.

Kelly, S., Kaye, S.A. and Oviedo-Trespalacios, O. (2023), "What factors contribute to acceptance of artificial intelligence? A systematic review", *Telematics and Informatics*, Vol. 77, p. 101925.

Kempf, D.S. and Smith, R.E. (1998), “Consumer processing of product trial and the influence of prior advertising: a structural modeling approach”, *Journal of Marketing Research*, Vol. 35 No. 3, pp. 325-338.

Kenny, E.M., Ford, C., Quinn, M. and Keane, M.T. (2021), “Explaining black-box classifiers using post-hoc explanations-by-example: the effect of explanations and error-rates in XAI user studies”, *Artificial Intelligence*, Vol. 294, p. 103459.

Kirmani, A. (1990), “The effect of perceived advertising costs on brand perceptions”, *Journal of Consumer Research*, Vol. 17 No. 2, pp. 160-171.

Kizilcec, R.F. (2016), “How much information? Effects of transparency on trust in an algorithmic interface”, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2390-2395.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. and Nizam, A. (1988), *Applied Regression Analysis and Other Multivariable Methods*, Duxbury Press, Belmont, CA.

Knapič, S., Malhi, A., Saluja, R. and Främling, K. (2021), “Explainable artificial intelligence for human decision support system in the medical domain”, *Machine Learning and Knowledge Extraction*, Vol. 3 No. 3, pp. 740-770.

Kraus, J., Scholz, D., Stiegemeier, D. and Baumann, M. (2020), “The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 62 No. 5, pp. 718-736.

Laato, S., Tiainen, M., Najmul Islam, A.K.M. and Mäntymäki, M. (2022), “How to explain AI systems to end users: a systematic literature review and research agenda”, *Internet Research*, Vol. 32 No. 7, pp. 1-31.

Lafferty, B.A. and Goldsmith, R.E. (2004), “How influential are corporate credibility and endorser attractiveness when innovators react to advertisements for a new high-technology product?”, *Corporate Reputation Review*, Vol. 7 No. 1, pp. 24-36.

Lee, J.D. and See, K.A. (2004), “Trust in automation: designing for appropriate reliance”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 46 No. 1, pp. 50-80.

Lee, M.K. (2018), “Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management”, *Big Data and Society*, Vol. 5 No. 1, pp. 1-16.

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M. and Mara, M. (2023), “Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task”, *Computers in Human Behavior*, Vol. 139, p. 107539.

Liu, C.F., Chen, Z.C., Kuo, S.C. and Lin, T.C. (2022), “Does AI explainability affect physicians’ intention to use AI?”, *International Journal of Medical Informatics*, Vol. 168, p. 104884.

Loewenstein, G., Sunstein, C.R. and Golman, R. (2014), “Disclosure: psychology changes everything”, *Annual Review of Economics*, Vol. 6 No. 1, pp. 391-419.

Lombrozo, T. (2006), “The structure and function of explanations”, *Trends in Cognitive Sciences*, Vol. 10 No. 10, pp. 464-470.

Martela, F. and Steger, M.F. (2016), “The three meanings of meaning in life: distinguishing coherence, purpose, and significance”, *The Journal of Positive Psychology*, Vol. 11 No. 5, pp. 531-545.

Meske, C., Bunde, E., Schneider, J. and Gersch, M. (2022), “Explainable artificial intelligence: objectives, stakeholders, and future research opportunities”, *Information Systems Management*, Vol. 39 No. 1, pp. 53-63.

Mick, D.G. (1992), “Levels of subjective comprehension in advertising processing and their relations to ad perceptions, attitudes, and memory”, *Journal of Consumer Research*, Vol. 18 No. 4, pp. 411-424.

Miller, T. (2019), “Explanation in artificial intelligence: insights from the social sciences”, *Artificial Intelligence*, Vol. 267, pp. 1-38.

Mitchell, A.A. and Olson, J.C. (1981), “Are product attribute beliefs the only mediator of advertising effects on brand attitude?”, *Journal of Marketing Research*, Vol. 18 No. 3, pp. 318-332.

Mittal, B. (1990), “The relative roles of brand beliefs and attitude toward the ad as mediators of brand attitude: a second look”, *Journal of Marketing Research*, Vol. 27 No. 2, pp. 209-219.

Mook, D.G. (1983), “In defense of external invalidity”, *American Psychologist*, Vol. 38 No. 4, pp. 379-387.

Moorman, C., Deshpande, R. and Zaltman, G. (1993), “Factors affecting trust in market research relationships”, *Journal of Marketing*, Vol. 57 No. 1, pp. 81-101.

Mooy, S.C. and Robben, H.S. (2002), “Managing consumers’ product evaluations through direct product experience”, *Journal of Product and Brand Management*, Vol. 11 No. 7, pp. 432-446.

Morgan, R.M. and Hunt, S.D. (1994), “The commitment-trust theory of relationship marketing”, *Journal of Marketing*, Vol. 58 No. 3, pp. 20-38.

Mostafa, R.B. and Kasamani, T. (2022), “Antecedents and consequences of chatbot initial trust”, *European Journal of Marketing*, Vol. 56 No. 6, pp. 1748-1771.

Mueller, S.T., Hoffman, R.R., Clancey, W., Emrey, A. and Klein, G. (2019), “Explanation in human-AI systems: a literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI”, *arXiv Preprint arXiv:1902.01876*.

Mukherjee, A. and Nath, P. (2007), “Role of electronic trust in online retailing: a re-examination of the commitment-trust theory”, *European Journal of Marketing*, Vol. 41 Nos 9-10, pp. 1173-1202.

Nantel, J. and Sekhavat, Y. (2008), “The impact of SMS advertising on members of a virtual community”, *Journal of Advertising Research*, Vol. 48 No. 3, pp. 363-374.

Nitzl, C., Roldan, J.L. and Cepeda, G. (2016), “Mediation analysis in partial least squares path modeling: Helping researchers discuss more sophisticated models”, *Industrial Management and Data Systems*, Vol. 116 No. 9, pp. 1849-1864.

Orth, U.R. and De Marchi, R. (2007), “Understanding the relationships between functional, symbolic, and experiential brand beliefs, product experiential attributes, and product schema: advertising-trial interactions revisited”, *Journal of Marketing Theory and Practice*, Vol. 15 No. 3, pp. 219-233.

Otnes, C.C., Ruth, J.A.M. and Crosby, E. (2014), “Product-agency benefits: consumer perspectives and strategic implications”, *European Journal of Marketing*, Vol. 48 Nos 5-6, pp. 878-898.

Palan, S. and Schitter, C. (2018), “Prolific. ac–a subject Pool for online experiments”, *Journal of Behavioral and Experimental Finance*, Vol. 17, pp. 22-27.

Percy, L. and Rossiter, J.R. (1992), “Advertising stimulus effects: a review”, *Journal of Current Issues and Research in Advertising*, Vol. 14 No. 1, pp. 75-90.

Pham, M.T. (2004), “The logic of feeling”, *Journal of Consumer Psychology*, Vol. 14 No. 4, pp. 360-369.

Preece, A. (2018), “Asking ‘why’ in AI: explainability of intelligent systems–perspectives and challenges”, intelligent systems in accounting”, *Finance and Management*, Vol. 25 No. 2, pp. 63-72.

Rai, A. (2020), “Explainable AI: from black box to glass box”, *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 137-141.

Reber, R., Schwarz, N. and Winkielman, P. (2004), “Processing fluency and aesthetic pleasure: is beauty in the perceiver’s processing exposure?”, *Personality and Social Psychology Review*, Vol. 8 No. 4, pp. 364-382.

Rindova, V.P. and Petkova, A.P. (2007), “When is a new thing a good thing? Technological change, product form design, and perceptions of value for product innovations”, *Organization Science*, Vol. 18 No. 2, pp. 217-232.

Riveiro, M. and Thill, S. (2021), “That’s (not) the output I expected! on the role of end user expectations in creating explanations of AI systems”, *Artificial Intelligence*, Vol. 298, p. 103507.

Rogers, E.M. (1995), *Diffusion of Innovations*, The Free Press, New York, NY.

Sarstedt, M., Hair, J.F., Jr, Nitzl, C., Ringle, C.M. and Howard, M.C. (2020), “Beyond a tandem analysis of SEM and PROCESS: use of PLS-SEM for mediation analyses!”, *International Journal of Market Research*, Vol. 62 No. 3, pp. 288-299.

Schoonderwoerd, T.A., Jorritsma, W., Neerincx, M.A. and Van Den Bosch, K. (2021), “Human-centered XAI: developing design patterns for explanations of clinical decision support systems”, *International Journal of Human-Computer Studies*, Vol. 154, p. 102684.

Schraagen, J.M., Elsasser, P., Fricke, H., Hof, M., and Ragalmuto, F. (2020), “Trusting the X in XAI: Effects of different types of explanations by a self-driving car on trust, explanation satisfaction and mental models”, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *SAGE Publications*, *Los Angeles*, *CA*, Vol. 64, No. 1, pp. 339-343.

Shin, D. (2021), “The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI”, *International Journal of Human-Computer Studies*, Vol. 146, p. 102551.

Sichtmann, C. (2007), “An analysis of antecedents and consequences of trust in a corporate brand”, *European Journal of Marketing*, Vol. 41 Nos 9-10, pp. 999-1015.

Singh, S.N. and Cole, C.A. (1993), “The effects of length, content, and repetition on television commercial effectiveness”, *Journal of Marketing Research*, Vol. 30 No. 1, pp. 91-104.

Smith, R.E. and Swinyard, W.R. (1982), “Information response models: an integrated approach”, *Journal of Marketing*, Vol. 46 No. 1, pp. 81-93.

Soley, L.C. (1986), “Copy length and industrial advertising readership”, *Industrial Marketing Management*, Vol. 15 No. 3, pp. 245-251.

Stange, S., and Kopp, S. (2020), Effects of a social robot’s self-explanations on how humans understand and evaluate its behavior”, in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 619-627.

Suen, H.Y. and Hung, K.E. (2023), “Building trust in automatic video interviews using various AI interfaces: tangibility, immediacy, and transparency”, *Computers in Human Behavior*, Vol. 143, p. 107713.

Sundar, A., Cao, E.S. and Machleit, K.A. (2020), “How product aesthetics cues efficacy beliefs of product performance”, *Psychology and Marketing*, Vol. 37 No. 9, pp. 1246-1262.

Sundar, S.S., and Kim, J. (2019), “Machine heuristic: when we trust computers more than humans with our personal information”, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 4-9*, Glasgow, Scotland, pp. 1-9.

Swan, J.E. and Nolan, J.J. (1985), “Gaining customer trust: a conceptual guide for the salesperson”, *Journal of Personal Selling and Sales Management*, Vol. 5 No. 2, pp. 39-48.

Thomke, S. (2020), “Building a culture of experimentation”, *Harvard Business Review*, Vol. 98 No. 2, pp. 40-47.

Usrey, B., Palihawadana, D., Saridakis, C. and Theotokis, A. (2020), “How downplaying product greenness affects performance evaluations: examining the effects of implicit and explicit green signals in advertising”, *Journal of Advertising*, Vol. 49 No. 2, pp. 125-140.

Vakratsas, D. and Ambler, T. (1999), “How advertising works: what do we really know?”, *Journal of Marketing*, Vol. 63 No. 1, pp. 26-43.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003), “User acceptance of information technology: toward a unified view”, *MIS Quarterly*, Vol. 27 No. 3, pp. 425-478.

Vriens, M. and Ter Hofstede, F. (2000), “Linking attributes, benefits, and consumer values”, *Marketing Research*, Vol. 12 No. 3, pp. 5-10.

Wang, N., Pynadath, D.V. and Hill, S.G. (2016), “Trust calibration within a human-robot team: Comparing automatically generated explanations”, in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 109-116.

Wang, W. and Benbasat, I. (2007), “Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs”, *Journal of Management Information Systems*, Vol. 23 No. 4, pp. 217-246.

Wang, X., He, J., Curry, D.J. and Ryoo, J.H. (2022), “Attribute embedding: learning hierarchical representations of product attributes from consumer reviews”, *Journal of Marketing*, Vol. 86 No. 6, pp. 155-175.

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G.B. and Rafaeli, A. (2023), “Decision control and explanations in human-AI collaboration: improving user perceptions and compliance”, *Computers in Human Behavior*, Vol. 144, p. 107714.

Winkielman, P. and Cacioppo, J.T. (2001), “Mind at ease puts a smile on the face: psycho physiological evidence that processing facilitation elicits positive affect”, *Journal of Personality and Social Psychology*, Vol. 81 No. 6, pp. 989-1000.

Winkielman, P., Halberstadt, J., Fazendeiro, T. and Catty, S. (2006), “Prototypes are attractive because they are easy on the mind”, *Psychological Science*, Vol. 17 No. 9, pp. 799-806.

Wu, T.W., Day, R.L. and MacKay, D.B. (1988), “Consumer benefits versus product attributes: an experimental test”, *Quarterly Journal of Business and Economics*, Vol. 27 No. 3, pp. 88-113.

Yang, F., Huang, Z., Scholtz, J., and Arendt, D.L. (2020), How do visual explanations foster end users’ appropriate trust in machine learning?”, in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 189-201.

Zhao, X., Lynch, J.G., Jr and Chen, Q. (2010), “Reconsidering Baron and Kenny: myths and truths about mediation analysis”, *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206.

Further reading

Cavique, L. (2024), “Implications of causality in artificial intelligence”, *Frontiers in Artificial Intelligence*, Vol. 7, p. 1439702.

Hoch, S.J. and Ha, Y.W. (1986), “Consumer learning: advertising and the ambiguity of product experience”, *Journal of Consumer Research*, Vol. 13 No. 2, pp. 221-233.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L. and Nass, C. (2015), “Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance”, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 9 No. 4, pp. 269-275.

Noor, N., Rao Hill, S. and Troshani, I. (2022), “Developing a service quality scale for artificial intelligence service agents”, *European Journal of Marketing*, Vol. 56 No. 5, pp. 1301-1336.

Sands, S., Campbell, C.L., Plangger, K. and Ferraro, C. (2022), “Unreal influence: leveraging AI in influencer marketing”, *European Journal of Marketing*, Vol. 56 No. 6, pp. 1721-1747.

Appendix

The baseline product information.

Here comes a description of an AI-based add-on option for cars:

This is an add-on option that can be selected when you buy a new car. It comprises a video camera attached to the rearview mirror inside a car. The camera is recording the driver while the car is used. The videos are processed by a software that is able to detect emotions in a human face. It can distinguish between seven emotions, and it calculates the intensity of each of these emotions 30 times per second. By doing this, the software can identify the emotional state of the driver at any point in time. The software is also collecting data about car behavior (location, speed, turning, what gear is used etc.). Based on its assessment of the driver’s emotional state and car behavior, it alerts the driver when critical emotion levels are reached. These alerts are voice messages (e.g. “Watch out, you are now angry”) followed by recommendations (e.g. “Please slow down”).

The explanation part

This is how it works: the AI-based software has been trained with videos of thousands of drivers' faces so that it has become highly accurate in detecting one specific driver's emotional state. It has also been trained with data from thousands of cars in traffic situations (e.g. the location of the car, the speed of the car, when and how it brakes, turns, accelerates and what gears are used). Moreover, it has learned to match human emotion data and car data so that it can find correlations between emotions and traffic accidents. Each unit (camera and software) is connected to the Internet and transmits facial data and car data from each car to a central database located at the car manufacturer's AI department. All data generated by the units are saved and stored and used for further training of the algorithms.

The benefit part

The idea is to make the driver aware of his or her emotional states so that traffic accidents can be avoided. Thus, this add-on option is designed to boost the driver's safety. So, for example, if the software detects that the driver is angry, it informs the driver about this so that the driver can avoid the type of aggressive driving that typically goes hand in hand with the emotional state of anger, which is not a safe state. That is to say, many accidents occur because of angry drivers. As another example, it can detect if the driver is in a highly happy state, which can mean that the driver fails to pay careful attention to what goes on in a traffic environment.

Table A1. Means, SDs and zero-order correlations for the variables used in the testing of hypotheses ($n = 480$)

Variable	M	SD	1	2	3	4
1. Explanation content	7.48	1.76				
2. Benefit content	6.99	2.23	0.50			
3. Understandability	7.35	1.87	0.73	0.43		
4. Trust	5.14	2.16	0.41	0.49	0.45	
5. Perceived performance	5.76	2.01	0.43	0.51	0.42	0.73

Note(s): All correlations are significant ($p < 0.01$)

Corresponding author
Magnus Söderlund can be contacted at: Magnus.Soderlund@hhs.se