Technological Forecasting & Social Change 105 (2016) 105–120

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices

Monika Hengstler *, Ellen Enkel, Selina Duelli

Zeppelin University, Dr. Manfred Bischoff Institute of Innovation Management of Airbus Group, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany

ARTICLE INFO

Article history:
Received 23 August 2015
Received in revised form 23 December 2015
Accepted 25 December 2015
Available online 16 February 2016

Keywords:
Artificial intelligence
Automation
Autonomous vehicle
Autonomous driving
Medical assistance
Trust

ABSTRACT

Automation with inherent artificial intelligence (AI) is increasingly emerging in diverse applications, for instance, autonomous vehicles and medical assistance devices. However, despite their growing use, there is still noticeable skepticism in society regarding these applications. Drawing an analogy from human social interaction, the concept of trust provides a valid foundation for describing the relationship between humans and automation. Accordingly, this paper explores how firms systematically foster trust regarding applied AI. Based on empirical analysis using nine case studies in the transportation and medical technology industries, our study illustrates the dichotomous constitution of trust in applied AI. Concretely, we emphasize the symbiosis of trust in the technology as well as in the innovating firm and its communication about the technology. In doing so, we provide tangible approaches to increase trust in the technology and illustrate the necessity of a democratic development process for applied AI.

1. Introduction

Autonomous driving and autonomous vehicles are currently among the most intensively researched and publicly followed technologies in the transportation domain (Beiker, 2012). Before realizing the vision of fully autonomous vehicles, many technical and legal challenges remain to be solved (ibid.). In addition to these technical and legal challenges, questions regarding user and societal acceptance come into play. For example, a survey of 1003 car buyers in Germany conducted by Puls (plus, 2015) in 2015 revealed that almost half of respondents were skeptical regarding the topic of autonomous driving. With respect to the actual use of such vehicles, approximately 49% of respondents preferred traditional driving and 43% preferred semi-autonomous driving. Only 5% of respondents preferred fully autonomous driving.

Autonomous vehicles involve the application of intelligent automation. In general, automation is defined as technology that actively selects data, transforms information, and makes decisions or controls processes (Lee and See, 2004). The decision-making process employed in the technology is based on inherent artificial intelligence (AI), hence the term "intelligent automation". The transportation industry is only one among many industries that are increasingly influenced by automation involving inherent AI. For instance, intelligent, personal robots have begun to noticeably appear in diverse application fields ranging from home automation to medical assistance devices.

In order to approach the topic of AI, the distinction between strong and weak notion of AI is essential. Strong AI implies a system with human or superhuman intelligence in all facets and is pure fiction today. Currently, only the weak notion of AI is of interest for commercial applications. This notion describes AI in terms of specific tasks that require single human capabilities, e.g., visual perception, understanding context, probabilistic reasoning and dealing with complexity (Russell and Norvig, 2010). In these domains, machines exceed human capabilities by far. However, intelligent technologies are not able to execute intelligent tasks such as ethical judgments, symbolic reasoning, managing social situations or ideation (Brynjolfsson and McAfee, 2014). This study will focus on applied AI in its weak notion, which is defined in terms of the tasks humans do rather than how humans think.

In human social interaction, trust is intuitively important (Slovic, 1993). Similarly, trust provides a valid foundation for describing the relationship between humans and automation (Lee and See, 2004). In academic literature, two different research streams on trust in the context of automation and innovation have emerged: First, there is prior work on trust in the technology (Ghazizadeh et al., 2012; Lee and Moray, 1992; Lee and See, 2004; Zuboff, 1988). Second, there is recent research on trust in the innovating firm and its communication (Brock, 1965; Chiesa and Frattini, 2011; Nienaber and Schewe, 2014; Ram, 1989; Ram and Sheth, 1989; Sternthal et al., 1978). In our study, we draw from both research areas to examine the mechanisms for building trust in applied AI.

* Corresponding author.
E-mail addresses: monika.hengstler@zu.de (M. Hengstler), ellen.enkel@zu.de (E. Enkel), s.duelli@zeppelin-university.net (S. Duelli).

http://dx.doi.org/10.1016/j.techfore.2015.12.014
0040-1625/

Thus, motivated by evidence of unusually high skepticism towards the concept, our aim is to develop an understanding of how firms have enhanced trust in applied AI. Hence, we examine the following research question: How is trust in applied AI fostered?

Our study makes five contributions to theory. First, we examine trust in applied AI via a discipline spanning approach and accordingly contribute to the cognitive engineering, innovation management, and sociology literature. Second, we consider technology acceptance based on the single construct of trust. Consequently, we diverge from earlier technology acceptance literature based on the technology acceptance model (TAM) (Davis, 1989). Third, our results illustrate the dichotomous constitution of trust in applied AI, which is formed by a symbiosis of trust in the technology as well as trust in the innovating firm and its communication. Fourth, our results provide tangible approaches that can be applied alongside the three bases of trust in automation identified by Lee and Moray (Lee and Moray, 1992). Accordingly, trust in the technology evolves alongside performance, process, and purpose information. The performance basis is primarily reliant on both operational and data security aspects; the process basis is determined by cognitive compatibility and usability of the application; and the purpose basis is founded on application context and design. Fifth, we illustrate the necessity of a democratic development process for applied AI (e.g., via stakeholder alignment, transparency in development, and early, proactive communication).

2. Theory

2.1. The essence of trust

In interpersonal relationships, the essence of trust is the willingness to be vulnerable to the actions of another person (Mayer et al., 1995). This behavior is founded in the expectation that the trustee performs a particular action that is important to the trustor, irrespective of the ability of the trustor to monitor or control the trustee. Trust is an evolving and fragile phenomenon and can be destroyed much more quickly and easily than it can be created. This asymmetry is due to the fact that negative, trust-destroying events are more visible and draw more attention than positive, trust-building events (Slovic, 1993).

2.2. Trust and perceived risk

Trust is essential to reducing perceived risk (Rousseau et al., 1998). Perceived risk has been formally defined as a combination of uncertainty plus the seriousness of the outcome involved (Bauer, 1967). In the innovation literature, perceived risk is mainly defined in terms of uncertainty about the possibility of the failure of a new product or the likelihood that the product will not work properly (Nienaber and Schewe, 2014). In the context of AI, perceived risk further stems from the delegation of control to a machine and its respective control mechanisms (Castelfranchi and Falcone, 2000). Accordingly, control is constitutive of trust. Interestingly, the original TAM (Davis, 1989) and its successors (Venkatesh and Davis, 2000; Venkatesh et al., 2003) do not mention the constructs of risk or trust. Trust has only recently been added to TAM, mainly in the context of high-tech innovations such as mobile banking or e-commerce (Luarn and Lin, 2005; Pavlou, 2003).

In the context of innovation, consideration of perceived risk is given high significance since it is a central determinant for adoption. Similarly, resistance to adoption is likely to arise due to the fact that any new product entails change, uncertainty, or risk (Nienaber and Schewe, 2014; Ram and Sheth, 1989). The perception of risk is influenced by the novelty of an innovation (Kleijnen et al., 2009). The more familiar users become with an innovation, for instance, through media coverage or experience, the more likely adoption is to occur (Wejnert, 2002). This is especially true for radical innovations where pre-purchase uncertainty regarding the benefits and consequences is higher (Chiesa and Frattini, 2011).

However, the adoption decision for a specific technology takes place not just at the individual level (Rogers, 2003). Whereas theory on new product development and marketing assumes that utility is a main determinant for adoption, sociological theory explains adoption and resistance in terms of interaction between the technology and its social context (MacVaugh and Schiavone, 2010). Since the perception of risk is a social process (Douglas and Wildavsky, 1983), technologies cannot be viewed as isolated from their social context and cultural values (Selwyn, 2003; Slovic, 1993). Whether a risk is accepted is a political choice that depends on values, beliefs, and alternatives (Fischhoff et al., 1978). Especially for radical innovations and automated technologies with partially unknown risks and consequences, social principles affect judgment regarding which types of risk are feared (Douglas and Wildavsky, 1983). Ultimately, there is a discrepancy between scientifically proven risks and their intuitive perception. “The risks that kill you are not necessarily the risks that anger and frighten you” (Sandman, 1987, p. 21). This gap between statistical safety and public perception can be attributed to a lack of trust (Slovic, 1993).

2.3. Trust in the technology

Drawing an analogy from interpersonal relationships, the interaction between humans and automation is mediated by trust (Ghazizadeh et al., 2012). Over time, trust in automation evolves alongside the three dimensions of predictability, dependability, and faith (Rempel et al., 1985). Accordingly, in early phases, trust is mainly driven by the predictability of the technology, which can be defined as the degree to which future behavior can be anticipated. McKnight et al. (McKnight et al., 2002) emphasized the importance of initial trust, especially in the case of radically new technologies where perceptions of risk must be overcome in order to create willingness to use the technologies. Over time, the driver of trust becomes dependability, which can be described as the degree to which an automation technology's behavior is consistent. Ultimately, the relationship shifts to faith, which is the condition in which the user relies on the technology.

Zuboff (1988) found that trust in a new technology depends on trial-and-error experience, followed by understanding of the technology's operation, and finally, faith, which are similar to Rogers' (2003) dimensions of experience, understandability, and observability. Understanding the motives of a machine creates trust that is more stable than trust based only on the reliability of performance (Lee and See, 2004). Lee and Moray (1992) identified three factors that are crucial for trust in automation: performance, process, and purpose. Although performance is a strong indicator of preference, it does not guarantee adoption, which is further influenced by factors such as resistance to change (Ghazizadeh et al., 2012). The process dimension refers to the understandability of the technology. When algorithms and functional logic are transparent, trust is likely to be reinforced (Lee and See, 2004). Purpose reflects faith in intentions. In the case of humans, this might represent motivations and responsibilities; with machines, purpose reflects the designer's intention in creating the system (Lee and Moray, 1992).

2.4. Trust in the innovating firm and its communication

High-tech companies often mistakenly believe that the quality of their technological innovations is sufficient to convince users and accordingly neglect marketing and commercialization of these products (Slater and Mohr, 2006). Several studies have revealed high failure rates, especially for radical innovations, indicating that the reasons for innovation resistance go beyond the technical characteristics of the products (Chiesa and Frattini, 2011; Heidenreich and Spieth, 2013; Ram, 1989; Ram and Sheth, 1989). Particularly, trust in the innovating firm and its communication influence the adoption decision (Sternthal et al., 1978).