Trustworthy AI: How Ethicswashing Undermines Consumer Trust

Christian Peukert[1] and Simon Kloker[1]

[1] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, Karlsruhe, Germany {christian.peukert; simon.kloker}@kit.edu

Abstract. Ethicswashing is a neologism that has, due to the release of ethical guidelines for trustworthy Artificial Intelligence (AI) by the European Union, recently gained in popularity. Although the term is closely related to the concept of greenwashing, it is currently primarily used to describe companies' undertakings to keep ethical debates running in order to influence or avoid strict regulations. However, it is not clear yet whether ethicswashing has further implications similar to those already revealed for greenwashing or sharewashing. In an online survey with 94 participants, we find that perceived ethicswashing has a significant negative effect on consumer trust, whereby the effect is mediated by the perception of risk and consumer confusion (based on PLS SEM). With our results, we thus contribute a further flipside to the discussion of ethics in AI and provide a starting point for developing a comprehensive understanding of ethicswashing and its influence on trust.

Keywords: Ethicswashing, AI, ethics, consumer trust, online survey.

1 Introduction

In April 2019, the European Union (EU) published ethics guidelines for "trustworthy" Artificial Intelligence (AI) [1] to respond to the ever-increasing amount of systems that employ AI methods for value creation. The guidelines were developed by the so called High-Level Expert Group on Artificial Intelligence (AI HLEG). Thomas Metzinger, a famous German late philosopher, was part of this group and was involved in the development for nine months. With the outcome, he is, however, not satisfied and thus called the overall endeavor "Ethics washing made in Europe" [2], thereby coining the term *ethicswashing*. Ethicswashing is inspired by the much more popular term greenwashing [3–5], even though it has a slightly different meaning. Ethicswashing refers to overstating the role of ethics in a corporation's policy and culture and to (repeatedly) initiate ethical debates in order to (1) avoid or escape governmental regulations [5] and (2) to convince and reassure customers to keep with the company's products or services [1, 6].

While companies such as Amazon, Apple, Google, or Microsoft, which offer AI-powered services such as their voice assistant systems, excel each other by promoting their ethical principles, their integrity, or their confidentiality of handling user data,

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

https://doi.org/10.30844/wi_2020_j11-peukert

Brady [7] argues that a definition of "ethics" is actually hard to pin down and that "professional ethics" is basically not a topic for IT professionals. Ethics in the perception of consumers is often only "telling right from wrong" [7, p.5]. The current debate in the EU is, however, much more complicated. It is about "red lines," "fundamental human rights," interests, trade-offs, and also simple wording [2]. In a general sense, ethics can be defined as "doing the right thing even when nobody else is looking" [7, p.5]. Ethicswashing, anyhow, incorporates another perspective: Companies and organizations considerably advertise their engagement in ethical debates and their "code of ethics," leading to the question: "Whose ethics at all?" The perception of ethical behavior may strongly depend on one's role in the ecosystem and on whose interests one shares [7, 8], making the topic for consumers overall very confusing [9]. Especially recent press releases that human workers encode conversations with voice assistant systems raise doubts regarding the handling of data and with it the compliance with (claimed) ethical principles. Nevertheless, AI-based business models rely on enormous amounts of data to offer and continuously improve the desired services, which creates a field of tension.

The literature on green- and sharewashing has already proven that such "overstating" of principles and confusion related to ambiguous advertisements may sometimes do more harm than good [3, 10]. The money allocated to advertise a company's ethical culture and to keep the current debate running in order to avoid or influence regulations may have a flipside effect. It undermines consumer trust towards the company and subsequently to its products and services [3, 10]. Analogous to green- and sharewashing, the current debates and the company's advertisement of ethics may influence the consumer's willingness to make themselves dependent on such companies and their services (or ecosystems). If the concept of ethicswashing theoretically operates analogously to share- and greenwashing, it can be assumed that perceived ethicswashing, risk, and consumer confusion affect consumer trust and thereby consumers' expectations regarding the company's credibility, benevolence, and ability to really incorporate profound ethical principles in its corporate culture and services [11]. Trust, however, is considered as one of, if not the most important factor for companies operating with data and AI [1, 6], emphasizing that trust research is of utmost importance in this context. Our study thus focuses on the following research question: How does the perception of ethicswashing impact consumer trust in a company? Although we hypothesize to observe analogous effects as for green- and sharewashing, yet, several open questions make this research question interesting for the community. New is, that in ethicswashing, users do not pay with financial means, but with their privacy, potentially having an effect on the relations on consumer trust and the perception of risk. In addition, green- and sharewashing address issues *within* the goals of the user – while being treated ethically correct is perceived as a "right," which, in turn, may also effect the perception of risk and consumer trust.

In this paper, we provide first insights on how consumer perceptions of ethicswashing undermine consumer trust through the (partial) mediators consumer confusion and perception of risk by drawing on extant research on Corporate Social Responsibility, green-, and sharewashing. In doing so, we show that ethicswashing

https://doi.org/10.30844/wi_2020_j11-peukert

has a flipside and expand the debate on ethics in AI and in technology companies overall by the, hitherto, barely discussed effect of trust on the consumer side [6].

2 Related Work

2.1 Ethics in AI

Ethics in AI is often discussed based on stories of AI weapons, autonomous drones in crisis regions, or vehicles killing civilians in our daily rush hour [12]. Russel et al. [12], however, illustrate that ethics in AI have several more dimensions: "liability and law," "machine ethics," "privacy," "professional ethics," and "policy questions." Research on ethics in AI focuses on how to embed ethics in a robust manner rather than addressing the more essential question of "what ethics to embed." Robustness is ensured by considering four basic questions [12]: "verification" (was the right system built?), "validity" (was the system built right?), "security" (is the system protected against unauthorized manipulation?), and "control" (can wrong behavior, decisions, etc. be fixed?). Answering these four questions is a crucial endeavor to ensure trust in AI besides adhering to the law [1]. However, many of the current dilemmas on "what ethics to embed" are currently solved by forwarding the decision to the user [12, 13]. Human users are expected to be capable of ethical decisions and to be liable to these decisions. In large scale implementations of AI, e.g. the voice assistants of companies like Amazon, Apple, Google, or Microsoft, the users' control is, however, somehow limited since the number of decisions and their complexity is often very high [14]. For instance, settings regarding privacy are limited by the minimum amount of data the service requires to run. Control regarding verification, validity, and the security of the data, need to be handed over to the company since individual users would not be able to assess them properly, exposing themselves to "trust" in the company or organization [15]. Therefore, Pieters [6] emphasizes the crucial point of explaining to the user what the technology is doing as one central aspect of ethics. The key to ethical and trustworthy AI is, therefore, to establish control and transparency in order to create and retain trust [1, 15].

2.2 Definition of Ethicswashing and Differentiation from other Concepts

The suffix "-washing" in the context of business models is usually used to indicate a gap between claim and reality. In greenwashing, this gap is between the actual environmental impact and the *communication* of environmental impact of the business and product [3, 16–18]. Companies are emitting misleading information regarding their ecological footprint or initiatives, building on an information asymmetry between company and customer. In contrast, the gap in sharewashing is between the actual business model of the company (or platform) and the *wording* that is used to advertise and describe it [10, 19–21]. An actual utilitarian value that is provided by the company is veiled in a sense of "community" and "sense of unity." Ethicswashing is similar to greenwashing with regard to information asymmetry (first dimension), since customers (and politicians) cannot proof the claims of companies

https://doi.org/10.30844/wi_2020_j11-peukert

regarding their "ethical" standards [6]. In addition, AI applications often represent black boxes, which are hard to explain to and to be understood by customers. In contrast to greenwashing, however, ethicswashing does not imply that companies advertise specific ethical initiatives. In this respect, ethicswashing is closer to sharewashing, as the wording used to describe the AI powered products and the respective company culture create a veil of ethics with terms scattered, e.g. in the company's description: "values," "equal opportunities," "freedom of expression." Beyond that, ethicswashing connotes one further dimension which is not shared by other "-washing" concepts. Ethicwashing pursues the goal of avoiding strict regulations – since regulations could, indeed, harm the overall business model – by actively engaging in public debates on sustainable and ethical use of AI and by trying to keep them running [2, 5]. While especially the latter described dimension is not targeted at the customers, both dimensions are perceived by the customer and can create a state of confusion and (intentional) disinformation or misinformation [6].

2.3 Corporate Social Responsibility, Consumer Confusion, and Risk Theory

Corporate Social Responsibility (CSR) includes business activities that report or advertise the social or environmental footprint of companies. If there is a misfit between actual reporting and (perceived) truth, consumers assume, for instance, greenwashing when there are unsubstantial environmental claims, or sharewashing in case of overstated socio-ecological claims through sharing of resources [3, 10]. To explain the discrepancy between companies' reporting narrative and the actual policies, one can refer to the benefits of reporting on corporate social responsibility [22]. Even though public institutions monitor and publicize the performance regarding environmental, social or ethical issues, gaps between reporting and actual behavior are still possible, whenever claims cannot be verified. Moreover, pressure from stakeholders or governments and a lack of legislation and regulation facilitate overly positive communication of CSR [23] and provide managers with incentives to report positive aspects while leaving out negative ones [24]. In many cases, consumers or monitoring entities have no choice but to trust that the claims by companies are objective – or to remain skeptical resulting in a lack of trust [25]. Greenwashing practices have already significantly reduced consumer trust in companies, leading to the fact that only one in five consumers trusts their environmental claims [26]. In ethicswashing, this effect may even be larger, as it is not really possible to publish (annual) "ethics reports." Today, applying ethics or owning an "ethical culture" is mere of a promise, which cannot be proven using common measures.

We base our research on the theoretical concept by Chen and Chang [3] and Hawlitschek et al. [10], who examine the impact of corporate greenwashing and sharewashing practices of peer-to-peer platforms on trust through the lens of consumers' perceptions of risk and consumer confusion. According to the perceived risk theory, minimizing risk is a dominant objective over maximizing expected utility in purchase decisions, especially since the majority of consumers is rather risk averse [e.g. 27]. We define *perceived risk* as a consumer's expectation of a company's inability to adhere to ethical standards and ethical data treatment. Wagner [5] and

https://doi.org/10.30844/wi_2020_j11-peukert

Rouillard & Giroux [9] state that companies' current emphasis on ethical principles lead to the (unintended) consequence of collective confusion. Similarly, we define *consumer confusion* as the failure of consumers to develop a correct interpretation of the ethical standard a company or organization proclaims and realizes in their culture and services.

3 Research Model and Hypotheses Development

Building on previous work on greenwashing [3] and sharewashing [10], we propose the research model illustrated in Fig. 1. The research model consists of four constructs: Ethicswashing (ET), consumer confusion (CF), perceived risk (RI), and consumer trust (TR). We model consumer trust as the consequence of the latter three constructs with ethicswashing having a direct effect and consumer confusion as well as perceived risk acting as (partial) mediators.

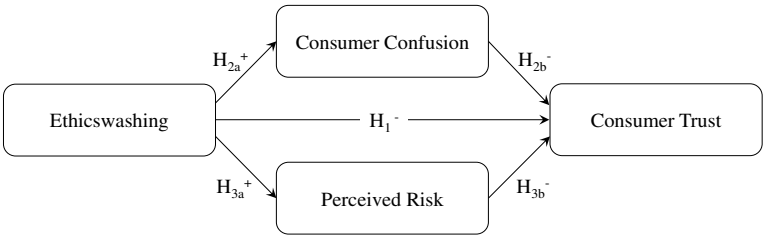

Figure 1. Research model ethicswashing

3.1 The Effect of Ethicswashing on Consumer Trust, Confusion, and Perceived Risk

In the area of products or services powered by AI, all kinds of data (probably also highly sensitive ones) are processed. Therefore, trust in the operating company is of highest concern to the consumer [6]. In the context of greenwashing, Chen and Chang [3] found that exaggerations or false statements about the eco-friendliness of products have a negative impact on the offering companies' trustworthiness. Similarly, Hawlitschek et al. [10] show in the realm of the sharing economy that platforms which falsely raise the claim to be part of the sharing movement, hence trying to benefit from the positive associations, are less trusted by the consumers. In the same vein, we argue that if users perceive that a company is only pretending to maintain compliance with ethical guidelines, this should have a negative impact on the trust placed in the company. We hence state: *The higher the degree of perceived ethicswashing, the lower the consumer trust.* **(H_1)**

Research in CSR repeatedly showed that overly emphasizing values, e.g. social values, environmental goals or further extraordinary engagements, may result in confusion among the consumers [3, 9, 28]. Wagner [5] even alleges companies to intentionally create confusion. Similarly, we argue that when companies overemphasize the compliance to ethical principles with regard to products powered

https://doi.org/10.30844/wi_2020_j11-peukert

by AI – whose functionalities are already difficult to understand – it can cause customer confusion. Thus: *Ethicswashing has a positive influence on consumer confusion.* **(H_{2a})**

Hawlitschek et al. [10] postulate that a prominent claiming of values may cause consumers to become especially aware of potential risks (e.g. legal and regulatory concerns or discrimination [29]). Similarly, consumers who have not previously associated the usage of AI-powered products or services with any risks are informed now by the ethicswashing activities. In general, all consumers might draw attention to the risks again through the ancillary publicity. In addition, the presence of ethical claims that seem insubstantial increases the perceived risk for consumers even further as there is no possibility to assess the risk [6]. Hence, we state: *Ethicswashing has a positive influence on perceived risk.* **(H_{3a})**

3.2 The Effect of Consumer Confusion and Perceived Risk on Consumer Trust

Consumer confusion can be caused by a variety of factors such as high product variety [30], great similarity between products, or ambiguous and misleading product descriptions [31]. It is widely argued [32] that consumer confusion negatively affects consumer trust. Chen & Chang [3] and Hawlitschek et al. [10] follow this argumentation for green- and sharewashing and find support for the relationship in their empirical analysis. Thus, we state: *Consumer confusion has a negative influence on consumer trust.* **(H_{2b})**

Consumers who consider the use of a product or brand to be highly risky would, in turn, not trust the product or brand either [33]. This also applies to services provided over the Internet. Dinev and Hart [34] argue that perceived Internet privacy risk negatively influences Internet trust. In line with Pieters [6], we also assume this relationship in the context of ethicswashing and the trustworthiness of companies offering AI-powered products and services. We therefore hypothesize: *Perceived risk has a negative influence on consumer trust.* **(H_{3b})**

4 Method

To evaluate the research model, we conducted an online survey. The data was collected in April 2019. The survey consisted of four parts: (1) An introduction explained the context of the study and framed participants to think especially of companies like Amazon, Apple, Google, and Microsoft and their voice assistant products (see Appendix B). (2) The second part asked for previous experience regarding AI-powered products and control variables (such as risk aversion, affinity to the internet, AI and privacy regulation). (3) The third part asked for the main constructs in five blocks (in random order). All items were distributed over the blocks and in random order within the blocks. (4) Finally, we collected some demographics and further control variables (such as age, gender, education, nationality). The questionnaire items were measured with 5-point Likert scales ranging from “totally

https://doi.org/10.30844/wi_2020_j11-peukert

disagree" (1) to "totally agree" (5). In addition, we implemented the marker variable technique to control for common method variance (CMV) by including the construct "Life satisfaction" by Diener et al. [35] as a theoretically unrelated construct [36]. Life satisfaction is suggested as a marker variable by Simmering et al. [37].

The study was conducted using the recruiting pool "Prolific" with participants from Australia, Canada, United Kingdom (UK), and the United States only. However, most participants reported being from the UK. 94 participants provided complete answers and passed the attention check. To ensure valid answers, we also excluded answers with total interview times below 4 min. A fixed payment of £ 0.7 was awarded for a complete answer. The median completion time was 7 min 51 sec (resulting in 5.35 £/h). The average age within the sample was 38.26 years (SD = 13.10) and 60.6% were female.

We operationalized all constructs in the research model in an online survey by adapting common scales from literature (Consumer Confusion from [3, 32, 38], Perceived Risk from [3, 39], Consumer Trust in Company from [3, 10, 40]). For the operationalization of ethicswashing, we stuck to the formulation of [3, 10], but adapted the items to fit our setting. We are aware that this procedure may result in an operationalization that only measures an aspect of ethicswashing as the definition is much broader than those of green- or sharewashing. However, in order to research the parallels of these concepts, we deemed this to be adequate for a first study. We refer to Appendix A for an overview of the applied items and definitons of the constructs.

5 Results

As a result of the exploratory research objective of our analysis, we use PLS SEM for data analysis [41]. In the analysis, we followed the two-stage approach by Hair et al. [42] to analyze and interpret the research model: We first analyzed the quality of the measurement model by assessing the internal consistency reliability, convergent validity, and discriminant validity. For all constructs, Cronbach's α and composite reliability (CR) are both well above the threshold value of 0.7 [42], thereby confirming internal consistency reliability [smallest Cronbach's α for RI (.760) and smallest CR for CF (.807)]. Next, we evaluate the convergent validity by examining each item's outer loading and the average variance extracted (AVE) for each construct. With respect to the latter, the values for all constructs are above the commonly applied cutoff value of 0.5 [42], except for the construct consumer confusion (AVE = .421). Regarding the outer loadings, the values of four items of CF, one item of RI and ET respectively, fall within the interval of 0.4 and 0.7 and hence do not exceed the threshold value of 0.7 (see Appendix A). According to Hair et al. [42], items with an outer loading between 0.4 and 0.7 shall only be considered for deletion when the deletion increases the AVE or measures for the internal consistency reliability above the cutoff value. We thus analyze for CF whether item deletion leads to an acceptable AVE value and find that deleting CF.1 and CF.4 results in the threshold value being exceeded (AVE = .51). Since we now meet the respective threshold values for all constructs, we decide to retain all other items with an outer

https://doi.org/10.30844/wi_2020_j11-peukert

loading below 0.7 in the model. For evaluating discriminant validity, we build upon three approaches: The consideration of cross-loadings, the Fornell-Larcker criterion [43], and the Heterotrait-Monotrait Ratio (HTMT). All three approaches provide evidence for the discriminant validity of the constructs. Table 1 shows the final values for the validation of the measurement model.

Table 1. Properties of measurement scales

							Correlations			
Construct	Mean	SD	Cron. α	CR	AVE	HTMT*	ET	CF	RI	TR
ET	3.45	0.69	.872	.908	.664	no	**.815**			
CF	3.66	0.61	.704	.800	.510	no	.658	**.714**		
RI	3.51	0.57	.760	.842	.524	no	.659	.635	**.724**	
TR	2.92	0.83	.931	.948	.784	no	-.641	-.583	-.592	**.886**

Note: Diagonal values indicate square root of AVE. *Indicates whether HTMT CI includes 1.

Second, after having confirmed the reliability of the measurement model, we next evaluate the structural model. Before we examine the results, we check for collinearity issues among predicting constructs and find that all VIF values are clearly below the threshold of 5 [42] (Inner VIF values of predicting constructs [ET=2.131; CF=2.018; RI=2.025] on TR), indicating that we do not encounter collinearity issues in our structural model. To control for common method variance, we applied Harman's single factor test [36] and the marker variable technique. For the prior, a single factor accounted for 38.6 percent of the variance (<50 percent) and for the latter, all latent variables' correlations with the marker variable are below .35. Fig. 2 shows the results for the PLS structural model [*p*-values for the path coefficients are based on bootstrapping procedure: two-tailed, 5000 samples, bias-corrected and accelerated (BCa) without sign change].

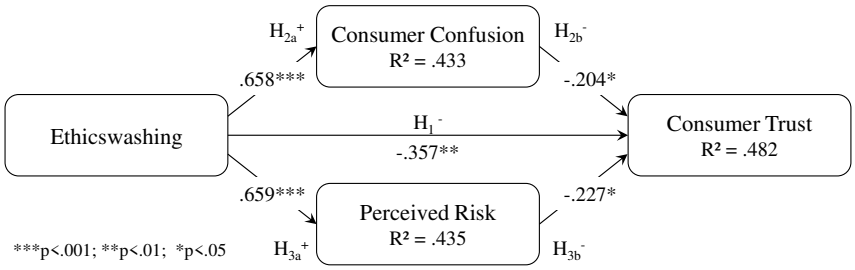

Figure 2. Results for the structural model

All relationships in the structural model are at least significant at a .05 level and point in the hypothesized directions. As hypothesized, a higher degree of perceived ethicswashing has a negative effect on the dependent variable consumer trust (H_1), whereas consumer confusion and perceived risk are positively influenced by ethicswashing (H_{2a}, H_{3a}). In line with H_{2b} and H_{3b}, consumer confusion and perceived risk, each, have a negative influence on the ultimate outcome variable consumer trust,

https://doi.org/10.30844/wi_2020_j11-peukert

indicating that higher values in these variables lead to lower consumer trust. The model explains 48.2 percent (adj. R^2=.465) of the variance in consumer trust, with ethicswashing (f^2=.115; small effect, classification in accordance with [44]) contributing most, followed by perceived risk (f^2=.049; small effect) and consumer confusion (f^2 = .040; small effect). Similarly, the explained fraction of variance for consumer confusion (R^2=.433; adj. R^2=.427) and perceived risk (R^2=.435; adj. R^2=.429) are moderately high, whereby the effect of ethicswashing on consumer confusion ($f^2_{(ET \rightarrow CF)}$=.764) and on perceived risk ($f^2_{(ET \rightarrow RI)}$=.769) can be classified as large [44].

In order to further investigate the influence of the mediators on TR, we followed the approach by Hayes [45] and carried out a mediation analysis applying bootstrapping (percentile, 5000 samples). We created a parallel multiple mediation model (CF and RI), which allows to simultaneously investigate both mediators' influence within a single model (PROCESS Model 4 [46]). The direct effect of ET on TR (CI[-.734,- .213]) and the indirect effect mediated by RI (bootstrapped CI[-.380, -.018]) is significant, whereas the indirect effect through CF acting as mediator is not significant at a .05 level (bootstrapped CI[-.279, .019]; see Appendix C for further results). The result indicates that RI is the sole mediator in the model presented, building a complementary mediation to the significant direct effect [47]. This result contradicts the result obtained by the PLS SEM model, in which both constructs (CF and RI) act as mediators. Potential theoretical explanations for this finding are discussed next.

6 Discussion, Limitation & Future Research

Within this paper, we adapted an established model from the context of greenwashing [3] to the context of ethicswashing in order to evaluate whether the existing theory is transferable to the new phenomenon of ethicswashing. In Section 2.2, we argued that ethicswashing is similar to sharewashing especially against the background that the wording is used to frame the consumer. However, in addition to dimensions already covered by green- and sharewashing concepts, ethicswashing includes the dimension of trying to actively avoid regulation. In this paper, we focused on the user perspective, and therefore on trust in the company. The results indicated that the negative effect of perceived ethicswashing on trust is – according to results of the mediation analysis – primary mediated by perceived risk. This is in contrast to the findings for greenwashing [3] and sharewashing [10], in which the path over customer confusion was also acting as mediator. We interpret this deviation with another difference in the fit of the models to our context. While in green- and sharewashing the perceived risk by the consumer is carried by the society (more environmental impact or non-community usage), i.e., the society would be affected as a whole, in ethicswashing the risk is faced by the individual consumers who provide their data. In addition, the risk in green- and sharewashing is rather predictable (the consumer does not misspend more money than they pay for the product/service), in ethicswashing the

https://doi.org/10.30844/wi_2020_j11-peukert

consumer cannot assess the extent to which their data is misused, posing a further uncertainty.

In future research, therefore, further extensions of the model are conceivable to capture the phenomenon even more precisely. For instance, since trust has often proven to be a powerful predictor of actual system use (e.g. in e-commerce [48, 49]), an extension of the model could also give an indication of the behavioral intention to use AI-powered products or services. In addition, we referred to the dependent variable explicitly as *consumer trust in the company*. However, it could be promising to also investigate *consumer trust in the respective product/service,* since we argue that ethicswashing can be seen from two perspectives – first, with respect to specific products/services and, second, regarding the overall company (institution). Therefore, the scales need to be adapted to cover and differentiate between both perspectives. We also based our theoretical concept on the concepts used in green- and sharewashing [3, 10], especially with regard to the use of the concepts of perceived risk and consumer confusion. It is very likely that this concept has to be extended regarding further, potentially more contextualized constructs like privacy preferences, and others. However, to gain an initial insight into the phenomenon, it was helpful to build upon the existing models from green- and sharewashing, but future research could also consider a new model composition also questioning the current modelled relationships. In addition, the theory needs to be extended regarding diversity in consumer groups (culture), business and private contexts, and a time dimension.

Even though we admitted participants originating from different nationalities to the survey and also made sure that we activated the survey at different points in time to take the different time zones into account, 79.8 percent of participants currently live in the UK. In future research, we plan to further extend the sample to also be able to tackle cross-cultural questions as we expect differences in the answering behavior especially between European residents and people originating from North America in a similar fashion as Dinev et al. [50] who detected differences with respect to the privacy calculus model in e-commerce. We highlighted voice assistants as exemplary services which are powered by AI since it was assumed that these were products known to our participants (84 percent stated to have used one of these products at least once). This was necessary to make the abstract term "AI-powered products/services" more tangible. Nevertheless, we are aware that, depending on the product, consumers may also feel more or less bothered by the ethicswashing phenomenon as the nature and content of the data to be processed may have an influence (e.g. think of using a translator versus a movie recommendation system). Furthermore, as soon as the topic is becoming more widespread, it can also be ensured that the introductory text of future studies is written even in a more neutrally way to rule out any influences.

7 Conclusion

Within this paper, we reported results for an online survey trying to shed light on the applicability of theories originating from greenwashing and sharewashing to the

https://doi.org/10.30844/wi_2020_j11-peukert

recent phenomenon of ethicswashing. Our results support the general applicability and demonstrate the negative influence of ethicswashing practices on consumer trust. Our results suggest that technology companies offering products or services powered by AI should be aware of the issues of ethicswashing and accordingly shall try to avoid ethicswashing in their marketing communication. In addition, ethicswashing strongly influences consumer confusion and perceived risk, which, in turn, affect consumer trust in the providing company (based on PLS SEM). However, we also showed where the model needs to be adapted in order to capture the phenomenon more precisely, especially emphasizing the differences in the risk component. Furthermore, we contribute to the ethics literature being the first to empirically test the effect of ethicswashing on trusting perceptions. We see our research as a first step towards understanding the complexity behind the concept of ethicswashing and encourage researchers and practitioners to further expand the direction of research since more and more AI applications will appear on the market in the near future.

References

1. AI HLEG: Ethics Guidelines for Trustworthy AI (By the High-Level Expert Group on Artificial Intelligence (AI HLEG) set up by the European Commission). (2019).
2. Metzinger, T.: Ethics Washing Made in Europe, https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html (Accessed: 06.08.2019), (2019).
3. Chen, Y.S., Chang, C.H.: Greenwash and Green Trust: The Mediation Effects of Green Consumer Confusion and Green Perceived Risk. J. Bus. Ethics. 114, 489–500 (2013).
4. Laufer, W.S.: Social Accounting and Corporate Greenwashing. J. Bus. Ethics. 43, 253–261 (2003).
5. Wagner, B.: Ethics as an Escape from Regulation: From "ethics-washing" to ethics-shopping? In: Hildebrandt, M. (ed.) Being Profiled, Cogitas Ergo Sum. pp. 84–90. Amsterdam University Press (2018).
6. Pieters, W.: Explanation and Trust: What to Tell the User in Security and AI? Ethics Inf. Technol. 13, 53–64 (2011).
7. Brady, D.: Ethics: IT Professional Pillar or Pillory. Mondo Digit. 13, 1–14 (2014).
8. Mingers, J., Walsham, G.: Toward Ethical Information Systems: The Contribution of Discourse Ethics. MIS Q. 34, 833–854 (2010).
9. Rouillard, C., Giroux, D.: Public Administration and the Managerialist Fervour For Values and Ethics: Of Collective Confusion in Control Societies. Adm. Theory Prax. 27, 330–357 (2005).
10. Hawlitschek, F., Stofberg, N., Teubner, T., Tu, P., Weinhardt, C.: How Corporate Sharewashing Practices Undermine Consumer Trust. Sustain. 10, 1–18 (2018).
11. Ganesan, S.: Determinants of Long-Term Orientation in Buyer-Seller Relationships. J. Mark. 58, 1–19 (1994).
12. Russell, S., Dewey, D., Tegmark, M.: Artificial Intelligence. AI Mag. 36, 105–114 (2015).
13. Johnson, D.G.: Technology with No Human Responsibility? J. Bus. Ethics. 127, 707–715 (2015).
14. Alaieri, F., Vellino, A.: Ethical Decision Making in Robots: Autonomy, Trust and

https://doi.org/10.30844/wi_2020_j11-peukert

Responsibility. In: Proceedings of the International Conference on Social Robotics 2016. pp. 159–168. Springer, Cham (2016).
15. Anderson, M., Anderson, S.L.: Machine Ethics: Creating an Ethical Intelligent Agent. AI Mag. 28, 15–26 (2007).
16. Hamann, R., Kapelus, P.: Corporate Social Responsibility in Mining in Southern Africa: Fair Accountability or just Greenwash? Development. 47, 85–92 (2004).
17. Lyon, T.P., Maxwell, J.W.: Greenwash: Corporate Environmental Disclosure Under Threat of Audit. J. Econ. Manag. Strateg. 20, 3–41 (2011).
18. Parguel, B., Benoît-Moreau, F., Larceneux, F.: How Sustainability Ratings Might Deter "Greenwashing": A Closer Look at Ethical Corporate Communication. J. Bus. Ethics. 102, 15–28 (2011).
19. Huang, L.-S.: #WeWashing: When "Sharing" Is Renting and "Community" Is a Commodity, http://www.huffingtonpost.com/leesean-huang/wewashing-when-sharing-is_b_6879018.html (Accessed: 06.08.2019), (2015).
20. Netter, S.: Exploring the Sharing Economy. Copenhagen Business School, PhD series, Frederiksberg (2016).
21. Light, A., Miskelly, C.: Sharing Economy vs Sharing Cultures? Designing for Social, Economic and Environmental Good. Interact. Des. Archit. 24, 49–62 (2015).
22. Lankoski, L.: Corporate Responsibility Activities and Economic Performance: A Theory of Why and How They are Connected. Bus. Strateg. Environ. 17, 536–547 (2008).
23. Pope, S., Wæraas, A.: CSR-Washing is Rare: A Conceptual Framework, Literature Review, and Critique. J. Bus. Ethics. 137, 173–193 (2016).
24. Hooghiemstra, R.: Corporate Communication and Impression Management - New Perspectives Why Companies Engage in Corporate Social Reporting. J. Bus. Ethics. 27, 55–68 (2000).
25. Niskanen, J., Nieminen, T.: The Objectivity of Corporate Environmental Reporting: A Study of Finnish Listed Firms' Environmental Disclosures. Bus. Strateg. Environ. 10, 29–37 (2001).
26. Ashley-Cantello, W.: Advertising Watchdog Receives Record Complaints Over Corporate "Greenwash," https://www.theguardian.com/environment/2008/may/01/corporatesocialresponsibility.ethicalliving (Accessed: 19.04.2019), (2008).
27. Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G.: Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. J. Eur. Econ. Assoc. 9, 522–550 (2011).
28. Pomering, A., Johnson, L.W.: Advertising Corporate Social Responsibility Initiatives to Communicate Corporate Image: Inhibiting Scepticism to Enhance Persuasion. Corp. Commun. An Int. J. 14, 101–118 (2009).
29. Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., Zwitter, A.: Will Democracy Survive Big Data and Artificial Intelligence? In: Helbing, D. (ed.) Towards Digital Enlightenment. pp. 73–98. Springer, Cham (2018).
30. Huffman, C., Kahn, B.E.: Variety for Sale: Mass Customization or Mass Confusion? J. Retail. 74, 491–513 (1998).
31. Mitchell, V.-W., Papavassiliou, V.: Marketing Causes and Implications of Consumer

https://doi.org/10.30844/wi_2020_j11-peukert

Confusion. J. Prod. Brand Manag. 8, 319–342 (1999).
32. Walsh, G., Mitchell, V.-W.: The Effect of Consumer Confusion Proneness on Word of Mouth, Trust, and Customer Satisfaction. Eur. J. Mark. 44, 838–859 (2010).
33. Mitchell, V.-W.: Consumer Perceived Risk: Conceptualisations and Models. Eur. J. Mark. 33, 163–195 (1999).
34. Dinev, T., Hart, P.: An Extended Privacy Calculus Model for E-commerce Transactions. Inf. Syst. Res. 17, 61–80 (2006).
35. Diener, E., Emmons, R.A., Larsen, R.J., Griffin, S.: The Satisfaction With Life Scale. J. Personal. Assessment,. 49, 71–75 (1985).
36. Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., Podsakoff, N.P.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. J. Appl. Psychol. 88, 879–903 (2003).
37. Simmering, M.J., Fuller, C.M., Richardson, H.A., Ocal, Y., Atinc, G.M.: Marker Variable Choice, Reporting, and Interpretation in the Detection of Common Method Variance: A Review and Demonstration. Organ. Res. Methods. 18, 473–511 (2015).
38. Walsh, G., Hennig-Thurau, T., Mitchell, V.-W.: Consumer Confusion Proneness: Scale Development, Validation, and Application. J. Mark. Manag. 23, 697–721 (2007).
39. Chen, Y.S., Chang, C.H.: Enhance Green Purchase Intentions: The Roles of Green Perceived Value, Green Perceived Risk, and Green Trust. Manag. Decis. 50, 502–520 (2012).
40. Chen, Y.S.: The Drivers of Green Brand Equity: Green Brand Image, Green Satisfaction, and Green Trust. J. Bus. Ethics. 93, 307–319 (2010).
41. Gefen, D., Rigdon, E.E., Straub, D.: Editor's Comments: An Update and Extension to SEM Guidelines for Administrative and Social Science Research. MIS Q. 35, iii–xiv (2011).
42. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Los Angeles: Sage Publications (2016).
43. Fornell, C., Larcker, D.F.: Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. J. Mark. Res. 18, 382–388 (1981).
44. Henseler, J., Ringle, C.M., Sinkovics, R.R.: The Use of Partial Least Squares Path Modeling in International Marketing. Adv. Int. Mark. 20, 277–319 (2009).
45. Hayes, A.F.: Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. Commun. Monogr. 76, 408–420 (2009).
46. Hayes, A.F.: PROCESS: A Versatile Computational Tool for Observed Variable Moderation, Mediation, and Conditional Process Modeling. White Pap. Retrieved from http//www.afhayes.com/public/process2012.pdf. 1–39 (2012).
47. Zhao, X., Lynch, J.G., Chen, Q.: Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. J. Consum. Res. 37, 197–206 (2010).
48. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in Online Shopping: An Integrated Model. MIS Q. 27, 51–90 (2003).
49. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and Validating Trust Measures for e-Commerce: An Intergrative Typology. Inf. Syst. Res. 3, 334–359 (2002).
50. Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., Colautti, C.: Privacy Calculus Model in E-commerce - A Study of Italy and the United States. Eur. J. Inf. Syst. 15, 389–402 (2006).

https://doi.org/10.30844/wi_2020_j11-peukert

Appendix

A. Applied measurement scales in the research model and outer loadings

Code	Items (adapted)	Outer loading
Perceived Ethicswashing (EW; reflective**):** User's perception that a technology company is deliberately misleading with regard to their statements about their ethical claims and principles. [3, 10] (Source Items: [3])		
EW1	Technology companies mislead with words in their ethical principles with respect to their products and services powered by AI.	.881
EW2	Technology companies mislead with visuals or graphics in their ethical principles with respect to their products and services powered by AI.	.818
EW3	Technology companies possess an ethical claim that is vague or seemingly unprovable with respect to their products and services powered by AI.	.821
EW4	Technology companies overstate or exaggerate how their ethical principles actually are with respect to their products and services powered by AI.	.692
EW5	Technology companies leave out or mask important information, making the ethical claim sound better than it is, with respect to their products and services powered by AI.	.851
Consumer Confusion (CF; reflective**):** User's failure to develop a correct interpretation of the ethical standards which are applied by a technology company when using their products or services powered by AI. [3, 10] (Source Items: [3, 32, 38])		
CF1	Due to the great similarity of technology companies with respect to their ethical claims regarding the application of AI, it is often difficult to choose the most ethical product or service.	.535†
CF2	It is difficult to recognize the differences between technology companies with respect to their ethical use of AI in their products and services.	.572
CF3	There are so many technology companies offering products and services powered by AI that you are really confused with respect to their ethical use of AI when using their services or products.	.546
CF4	There are so many products and services powered by AI that it is difficult to decide which one you should choose with respect to ethical standards the technology companies claim.	.471†
CF5	When using a product or service powered by AI, you rarely feel sufficiently informed with respect to ethical standards applied by the offering technology company.	.862
CF6	When using a product or service powered by AI, you feel uncertain about the ethical standards of the offering technology company.	.819
Perceived Risk (RI; reflective**):** User's expectation of experiencing negative consequences, especially with regard to non-compliance with the claimed ethical principles, associated with usage of products or services powered by AI. [3, 10] (Source Items: [3, 39])		
RI1	There is a chance that there will be something wrong with the compliance with the ethical principles of the company offering the product or service powered by AI.	.737

https://doi.org/10.30844/wi_2020_j11-peukert

RI2	There is a chance that the product or service powered by AI will not handle my data with the ethical principles the company claimed.	.807
RI3	There is a chance that you would suffer loss of data sovereignty if you used a product or service powered by AI.	.769
RI4	There is a chance that my data may be subject to unethical procedures when using a product or service powered by AI.	.797
RI5	Using a product or service powered by AI would damage your reputation or image with respect to your value of ethical principles.	.449
Consumer Trust in Company (TR; reflective**):** User's willingness to depend on a technology company based on the beliefs or expectation resulting from its credibility, benevolence, and ability associated with its stated ethical principles. [3, 10] (Source Items: [3, 10])		
TR1	I have faith that the ethical reputation of technology companies offering products and services powered by AI is generally reliable.	.909
TR2	I have the impression that compliance with ethical principles by technology companies offering products and services powered by AI is generally dependable.	.839
TR3	I have faith that ethical claims by technology companies offering products and services powered by AI are generally trustworthy.	.922
TR4	I have the impression that ethical concerns by technology companies offering products and services powered by AI meet my expectations.	.874
TR5	Technology companies offering products and services powered by AI generally keep their promises and commitments for compliance with ethical principles.	.882
Note: All items used a 5-point Likert scale (strongly disagree | strongly agree). † initial item loading for items which were removed in course of the evaluation of the measurement model.		

B. Survey Introduction

The focus of this study is on **your perception of companies that provide products and services powered by Artificial Intelligence (AI)**. AI is embodied in a growing number of products and services, such as smart speakers, e.g., Amazon Echo Dot, Google Home, or voice assistants like Microsoft's Cortana and Apple's Siri. Since services offered by these products **heavily rely on the available data** (mostly user data), some people, organizations, or governments are concerned whether the use of this data and also the **outcome adheres to ethical standards**. In products relying on AI, the user cannot fully reproduce, how the outcomes (e.g., interactions, answers, recommendations) were generated and has to fully trust that the stated outcome is **neither biased nor that the data provided will be used for purposes beyond the requested** – at the current or later point in time.

While the business models of these companies at least partly rely on **collecting as much user data as possible**, the users demand **fundamental rights to retain data sovereignty**. This leaves the companies in a field of tension. As a result, some companies heavily **advertise their responsible handling of data, their privacy settings and their general ethical principles** with respect to the implementation of AI in their products and services. Additionally, some companies also **actively engage in the public debate** on the ethical use of AI.

In this survey, we ask you to provide us with **your perception of this field of tension** against the background of technology companies like Amazon, Apple, Google, or Microsoft. If you do

https://doi.org/10.30844/wi_2020_j11-peukert

not have any experiences with products or services powered by AI from these companies, please answer the following questions from a hypothetical or a general perspective. In any case, please answer truthfully and intuitively. There are no "wrong" answers.

C. Results of parallel multiple mediation analysis

			Mediator CF			Mediator RI		
	total effect	direct effect	a1	b1	a1 x b1 [a]	a2	b2	a2 x b2 [a]
ET	-.766***	-.473***	.519***	-.175	-.091	.547***	-.370*	-.202*

Note: * $p < .05$, ** $p < .01$, *** $p < .001$; unstandardized coefficients; [a] 95% CI (bootstrapped, 5000 samples) for a1 x b1 (CI [-.279, .019]) and a2 x b2 (CI [-.380, -.018]).

https://doi.org/10.30844/wi_2020_j11-peukert