Intelligent Human Systems Integration (IHSI 2023), Vol. 69, 2023, 417–427
https://doi.org/10.54941/ahfe1002861

Affective Analysis of Explainable Artificial Intelligence in the Development of Trust in AI Systems

Ezekiel Bernardo and Rosemary Seva

Industrial and Systems Engineering Department, De La Salle University, Manila, Philippines

ABSTRACT

The rise of Explainable Artificial Intelligence (XAI) has been a game changer for the growth of Artificial Intelligence (AI) and the systems it powers. By providing human-level explanations, it systematically solves the most significant issue that AI faces: the black-box paradox realized from the complex hidden layers of its algorithm (i.e., machine and deep learning). Fundamentally, it allows users to learn how the AI operates and comes to decisions, thus enabling cognitive calibration of trust and subsequent reliance on the system. However, as human-computer interaction and social science studies suggest, relying on cognitive calibration might be limited as the affective processing component, which is also established from the interaction, was not yet considered. Considering the limited information regarding the affective route, this study aims to examine the effects of emotions associated with the interaction with XAI in adoption thru trust and reliance. One hundred and forty-three participants partake in the online experiment. The premise was that they were hired to classify different species of animals and plants, with an XAI-equipped image classification AI available to give them recommendations. Three key findings can be drawn from the results. First, if a person felt interestingly surprised emotions due to the XAI (e.g., interested, excited, surprised, pleased, and amazed), they would increase their trust in the AI and rely on its functionality. Second, only trust would improve for those who felt trusting emotions (e.g., happy, confident, secure, proud, and trusting). Third, fearfully dismayed (e.g., dismayed, afraid, fear, angry, and sad) or anxiously suspicious (e.g., suspicious, concerned, confused, nervous, and anxious) emotions do not translate to a significant change in trust and reliance on the AI.

Keywords: Explainable AI, Trust, Affective trust, Emotions, Reliance

INTRODUCTION

Artificial intelligence (AI) has radically changed technology's role in people's everyday life. By having – or sometimes exceeding – humans' capability in doing cognitive-based tasks, acceptance increased and consequently introduced the possibility of dependency (Glikson & Woolley, 2020). As such, AI can be seen powering different systems, augmenting humans at varying levels (Haenlein & Kaplan, 2019).

In recent years, however, society's adoption has become more challenging due to rising transparency concerns (Adadi & Berrada, 2018; Böckle et al., 2021; Rai, 2020). Fundamentally, this is when humans cannot understand AI's inner workings (e.g., how it operates and comes to decisions), thus appealing less trustworthy and subsequently affecting adoption. This is highly attributable to the complexity of the algorithms being used, which is virtually difficult to comprehend (i.e., hidden layers from machine learning and deep learning) (Chowdhary, 2020). This issue is becoming more pressing as AI is being more targeted for ordinary people who often possess low technical skills in understanding AI and highly critical tasks that demand more convoluted logic (Lewis et al., 2021).

To address the curtailing effects of transparency and subsequently alleviate trust, explainable AI (XAI) has been introduced. XAI is a human-level explanation aiming to provide insights into AI's purpose, process, and performance (Barredo Arrieta et al., 2020; Das & Rad, 2020; Gunning, 2017). This is created by drawing key information from the complex algorithm and is often presented in an interface thru the set of rules, a summary of features used, relative examples, or supplementary information (Das & Rad, 2020; Jin et al., 2021). The running hypothesis is that these explanations calibrate trust cognitively (Madsen & Gregor, 2000), allowing users to think and eventually create conclusions. This results in a research direction focused on improving the cognitive resource, by developing new techniques, for mental model building.

However, trust has long been known to work beyond cognition. Notably, many social sciences and human-computer interface (HCI) scholars have identified that trust from cognitive cues can also be developed via irrational factors like emotions (Lee & See, 2004; Madsen & Gregor, 2000; Riegelsberger et al., 2003). Previous studies with similar transparency utility such as for social robots (Gompei & Umemuro, 2018), warning alerts (Buck et al., 2018), intelligent personal assistance (Chen & Park, 2021), security seals (Bernardo & Tangsoc, 2021) had verified this, which profoundly changed how they are managed and used (e.g., focusing on design) to maximize its effectiveness in developing trust or reliance. Unfortunately, as of the time of writing, no study had confirmed the relationship for the case of XAI.

Exploring the possibility of affective calibration is significant, considering that it has been the primary tool for resolving the issue of transparency. If proven, research can be redirected to the affective properties of XAI and how to properly leverage it – opening new ways to use and improve XAI outside the cognitive norm. Considering the seeming gap and significance it may bring, this study is proposed to validate whether affect (i.e., emotions) will allow for trust to be calibrated and potentially increase adoption or reliance on AI.

BACKGROUND AND HYPOTHESIS

Trust and Reliance for XAI

The concept of trust has been studied across multiple dimensions. For HCI, contextualization centered on adoption through technology acceptance, with the majority highlighting its significance (Adadi & Berrada, 2018). Studies also identified trust as a predictor for reliance, varied as disuse (rejection), misuse (over-reliance), and abuse (detrimental use) (Dazeley et al., 2021).

Regarding development, two routes have been validated to generate it: cognitive and affective. The former uses information to calibrate mental models, while emotions that diffuse on judgments for the latter (Lee & See, 2004). However, the research stream on affective received lesser exploration between the two. A trend also reflected for XAI (Linardatos et al., 2020; Mohseni et al., 2020).

Affective trust studies for XAI have been limited on viability. For instance, the work of Guerdan et al. (2021) recognized emotion's potential bias in the decision-making process and perceptions. Because of that, they have analyzed humans' facial expressions and developed a feature to embed it to the XAI. Similarly, Kaptein et al. (2017b) created Emotion-aware XAI (EXAI) to leverage emotions on XAI's effectiveness. Another work by Kaptein et al. (2017a) shows the potential of using simulated emotions to generate explanations. Though progress has been made, current studies have yet to understand the total measurement of the affective route. Specifically, trust and reliance change relative to the different emotions developed from XAI.

XAI Emotion Set

In general affect studies, different taxonomies for emotions have been used to quantify different behavioral measures (e.g., trust and reliance). There is the PAD model (pleasure, arousal, dominance) (Mehrabian & Russell, 1974), the circumplex model (Russell, 1980), the structural model of affect (Plutchik, 1994), and the nine affects (Tomkins, 1992), to name a few. Varying results have been identified for different contexts of use. Still, the majority have been categorically consistent where individual control emotion and low certainty is significant to trust based on its valence (e.g., anger negatively affects trust, happiness positively affects trust, anxiety negatively affects trust).

For XAI, taxonomy has already been explored. The work of Bernardo & Seva (2022) has recognized emotions developed from the interaction with XAI. As summarized in their XAI emotion set (XES), there are four main groups: interestingly surprised (e.g., interested, excited, surprised, pleased, and amazed), trusting (e.g., happy, confident, secure, proud, and trusting), fearfully dismayed (e.g., dismayed, afraid, fear, angry, and sad), and anxiously suspicious (e.g., suspicious, concerned, confused, nervous, and anxious). However, XES has not yet been used to measure affective change for XAI, unlike the previously mentioned structure of emotions. Thus, there had been no precise measurement of trust change at multiple emotional dimensions.

Proposed Hypothesis

Considering the current delimitation in affect studies, findings from previous emotion taxonomies, and the viability of XES as a potential set to measure affect change, the following hypothesis is proposed. For better understanding, XAI affective trust and reliance (XATR) model presented in figure 1 is proposed.

***H1:** Interestingly surprised emotions positively affect (a) trust and (b) reliance*

***H2:** Trusting emotions positively affect (a) trust and (b) reliance*

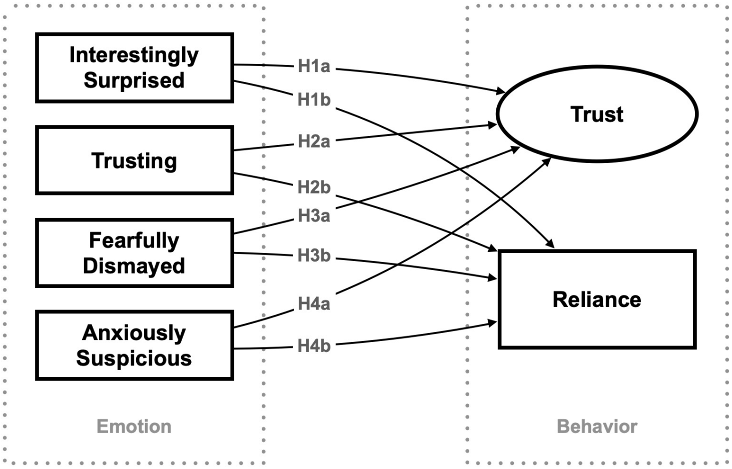

Figure 1: XAI Affective Trust and Reliance (XATR) calibration model with the corresponding proposed hypothesis.

H3: Fearfully dismayed emotions negatively affect (a) trust and (b) reliance

H4: Anxiously suspicious emotions negatively affect (a) trust and (b) reliance.

METHODOLOGY

An asynchronous virtual experiment was designed to assess the proposed model. The main goal was to embed an XAI in a controlled AI system to viably measure both the independent (i.e., emotions developed upon exposure to XAI) and dependent variables (i.e., trust and reliance of the user towards the AI) for the hypothesis relationship testing.

Measurement Conceptualization

To contextualize the study, a pre-experiment survey was conducted with 52 current AI users. The objective was to determine the most used AI type and most acceptable design template for the XAI setup. Results showed that image classification AI was the optimal choice as almost all have experienced or understand its functionality. After the validation with a focus group discussion involving six AI application developers and six user experience (UX) experts, the Google Lens application was unanimously chosen to serve as the design environment's template for logic and composition.

Participants

Convenience sampling was the method followed for the data gathering. Initial leads were generated from social media advertisements, with qualifications specified as being able to communicate in English, being at least 18 years of age, having used any AI-powered system in the recent year, and having a normal or corrected-to-normal vision without any other sight problems. Facilitating requirements were also given for the device they will use and the experiment environment. A reward of 50 PHP (~1.00 USD) was also guaranteed upon their successful participation. For the sample size, the minimum number was set at 100 following the priori power computation of Westland (2010) and Cohen (1988) in detecting a significant effect for a model structure. Specifically, the anticipated effect size, desired statistical power level, and alpha level was set at 0.1, 0.8, and 0.05, respectively.

Tools

Two measurement tools were created for the study: an online questionnaire and XAI testbed. In terms of purpose, the former handled participant screening, consent confirmation, and acquisition of the subject's demographical information. The latter was where the main variables were measured. Integral emotion and latent trust construct were assessed using a seven-point unipolar slider (1 - strongly disagree, 7 - strongly agree). Emotion was quantified using the four types of emotion proposed by Bernardo & Seva (2022) in their XES (i.e., surprisingly happy, trusting, fearfully dismayed, and anxiously suspicious), while trust was explored using three questions developed for this study. As for reliance, it was measured binomially based on the behavioral dependency of the user on the recommendation of the AI. For the hosting, the questionnaire and XAI testbed was hosted using Google Forms and Quant-UX, respectively.

For assurance, both tools were pre-tested with 30 participants composed of AI users, language and grammar experts, AI programmers, and UX experts. Also, construct validity for the proposed trust questions was confirmed. All recommendations were incorporated and implemented before the main experiment.

Procedure

The experiment started with the participant accessing the online questionnaire link provided in the message sent to their social media account. They were initially prompted with the consent clause and screening questions upon opening. Those who agreed and passed were allowed to continue; otherwise, the questionnaire would terminate. Demographic information was then asked, along with the AI and XAI experience. After answering, the priming condition for using the XAI testbed was given: "An NGO hired you to recognize pictures of different species in the Philippines. To help you, an image recognition AI system is available for you to use. You may choose to agree with its recognition or provide your own". In the end, a link was provided to redirect the participants to the XAI testbed.

The XAI testbed started with general instructions on how to use the application and overall task. Each participant was required to recognize 25 random photos available in the application. In every trial, they were asked to rate their emotions on the XAI and trust in the AI. To limit the possible noise in the data, they were given three attempts to test the application and to feel how the sliders work. Once the trial run was up, participants proceeded with the actual experiment. After completing the 25 recognitions, the testbed ended with a question on the subject's availability for a post-experiment interview and instructions on how the rewards were to be distributed.

Data Recording and Analysis

The area under the curve (AUTC) was used to have representative data for all variables, following the findings of Yang et al. (2017). This was computed by averaging the data across all 25 recognitions. Since reliance was recorded from an observable action, scores above 0.5 were considered reliance on AI and vice versa.

As for analysis, covariance-based structural equation modeling (CB-SEM) was the principal method followed. This was considered given that it: can do simultaneous analysis for the hypothesized relationship, carries the multivariate techniques needed to confirm and validate the measurement tools, is insensitive to demanding parametric conditions, and can test, interpret, and compare contrasting models providing a more accurate estimate of the effects (Astrachan et al., 2014; Dash & Paul, 2021). IBM Statistical Package for the Social Sciences (SPSS) 23 and Analysis of a Moment Structure (AMOS) Graphics 23 were the primary tools used to do all the computations. For consistency, all statistical tests were measured at a 0.05 significance level.

RESULTS & ANALYSIS

The experiment ran for five days. Access mostly happened between 3:00 PM to 11:00 PM. On average, participants finished the whole experiment in 20 minutes, while 10 minutes for the post-experiment interview. No consent concern manifested, and the XAI testbed did not encounter any issues on all trials.

Data Screening

Overall, 165 participated in the experiment. Of this, 143 were determined usable after removing all those who failed the qualifications and did not finish the experimentation. Descriptively, the majority belongs to the millennial age group (24 to 30 – 31.47% and 31 to 39 – 24.48%), followed by generation X (40 to 47 – 15.38% and 48 to 55 – 6.99%), and generation Z (18 to 23 – 15.38%). Males (45.45%) were more compared to females (38.46%) and the undisclosed group (16.08%). In terms of educational attainment, the majority were college (70.63%) and post-studies (20.98%) graduates (Elementary - 0.81% and Highschooler - 4.88%). For AI-related info, almost all are innovators (More than 5 years - 76.42%, 4 to 5 years - 9.76%, 2 to 4 years - 11.38%, 1 to 2 years - 1.63%, less than one year – 0.81%) and virtually all have an encounter with an XAI (82.93%).

Trust Measurement

The introduced reflective trust questions showed outstanding measurement capacity as proven by the meritorious 0.833 Kaiser-Meyer-Olkin (KMO) measure and a significant Bartlett's test of sphericity ($p < 0.001$). Communalities also agree on this, given that all extraction was above 0.700. The three questions also exhibit high consistency and reliability, considering that all three proposed questions loaded on a single factor with a high 66.411% explained variance and a Cronbach's alpha of 0.739. Confirmatory factor

Table 1. Model fit measures and threshold.

Type	Indices	Estimate	Threshold	Reference
Absolute Fit	RMSEA	0.075	<0.08	(Westland, 2010)
	SRMR	0.039	<0.08	Hu & Bentler (1999)
Incremental Fit	CFI	0.984	>0.95	Schreiber et al. (2006)
	NFI	0.965	>0.95	Hu & Bentler (1999)
Parsimonious Fit	χ2/df	1.802	1 to 3	Hu & Bentler (1999)

Table 2. Sample human systems integration test parameters (Folds et al. 2008).

Hypothesis	From		To	Std. Est. (ß)	P-Value
H1a	Interestingly Surprised	→	Trust	0.229	***
H2a	Trusting			0.192	***
H3a	Fearfully Dismayed			−0.008	0.785
H4a	Anxiously Suspicious			0.013	0.653
H1b	Interestingly Surprised	→	Reliance	0.115	***
H2b	Trusting			0.043	0.165
H3b	Fearfully Dismayed			−0.011	0.73
H4b	Anxiously Suspicious			−0.012	0.707

*Notes: *** p-value < 0.01; ** p-value < 0.05; * p-value < 0.10.*

analysis also surfaced high validity with 0.763 and 0.53 composite reliability and average variance extracted (AVE) at an excellent fit (χ2/df - 1.395, RMSEA - 0.071, CFI - 0.964). With all these outstanding results, the three proposed questions were used to measure the latent variable of trust.

Structural Equation Modeling

A clean factored model was concluded from the 2000 bootstrapped SEM results. As presented in table 1, all the fit indices passed the threshold value, highlighting the model's validity in testing the relationship relative to the proposed hypothesis. More so, there were no modification indices, suggesting that no additional structural links or constraints were needed to increase the overall fit.

Of the eight proposed relationships, only three were determined to be significant ($p < 0.001$) – H1a and H2a for trust and H1b for reliance (see table 2). Notably, the interestingly surprised set relates significantly to trust and reliance, while the trusting set only has a significant link to trust.

DISCUSSION

Three key findings can be drawn from the results. First, if a person felt interestingly surprised emotions due to the XAI (e.g., interested, excited, surprised, pleased, and amazed), they would increase their trust in the AI and rely on its functionality. Second, only trust would improve for those who felt trusting emotions (e.g., happy, confident, secure, proud, and trusting). Third, fearfully dismayed (e.g., dismayed, afraid, fear, angry, and sad) or anxiously suspicious (e.g., suspicious, concerned, confused, nervous, and anxious) emotions do not translate to a significant change in trust and reliance on the AI.

The findings show that for XAI, using valence as a marking for trust and reliance change is only ideal for positive emotions (i.e., interestingly surprised and trusting). This agrees with the general idea presented by Forgas (1995) in his Affect Infusion Model (AIM), where a person's integral emotions alter the appraisal of new stimuli parallel to the direction of emotion's valance. Also, from the insights generated in the interview, trust change from negative emotion was anchored by the level of reliability they feel. This follows the control theory discussed by Lerner and Keltner (2001) and Lerner et al. (2007), suggesting that high certainty or predictability does little to low change. The interview also highlighted that the reported emotions came from two sources: appealing via design and cognitive appreciation. The former took the composition of XAI as an indicator of trustworthiness, while the latter reflected the resource available. These resonate broadly on AIM, where affect can cause behavioral change via operating on the heuristic (appearance) and memory (mental representations) (Forgas, 1995). As for the design, principles of Kansei engineering were heavily observed as design factors in XAI (e.g., form of explanation, communication style) was pinpointed to cause affect change (Gan et al., 2021).

CONCLUSION

This study proved that integral emotions developed from XAI could calibrate trust and reliance on the AI system. Remarkably – as tested using the XATR calibration model – interestingly surprised emotion set positively affects trust and reliance, while the trusting set increases trust. Further, the negative emotion set of fearfully dismayed and anxiously suspicious does not significantly contribute to the calibration. These results challenge XAI's running cognitive calibration hypothesis, highlighting that explanations not only bring information to better understand how AI works but also diffuse emotions for behavioral change. Moving forward, this opens a new segment on how to improve XAI by leveraging emotions.

RECOMMENDATION

Three recommendations are proposed to improve the contribution of this paper. First, as suggested in the post-interview, the XAI design can be manipulated to verify the claims of affective change. Understanding this can position a better improvement plan to effectively leverage the benefits of using XAI. Second, the concept of time or usage experience with XAI can be further explored in terms of its comparative moderating effect. This can explain how exposure to XAI can change its impact over time. Lastly, studies can also check the relationship between dependent measures to understand the mediating properties of emotions.

ACKNOWLEDGMENT

The Department of Science and Technology funded this research under the Engineering Research and Development for Technology grant.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Astrachan, C. B., Patel, V. K., & Wanzenried, G. (2014). A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. Journal of Family Business Strategy, 5(1), 116–128. https://doi.org/10.1016/j.jfbs.2013.12.002

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bernardo, E., & Seva, R. (2022). Explainable Artificial Intelligence (XAI) Emotions Set. Manuscript Submitted for Publication.

Bernardo, E., & Tangsoc, J. (2021). Explanatory Modelling of Factors Influencing Loyalty on Smartphone Shopping Application: A Modification of the Shopping Application Adoption Model. In A. M. J. Gutierrez, R. S. Goonetilleke, & R. A. C. Robielos (Eds.), Convergence of Ergonomics and Design (Vol. 1298, pp. 181–191). Springer International Publishing. https://doi.org/10.1007/978-3-030-63335-6_19

Böckle, M., Yeboah-Antwi, K., & Kouris, I. (2021). Can You Trust the Black Box? The Effect of Personality Traits on Trust in AI-Enabled User Interfaces. In H. Degen & S. Ntoa (Eds.), Artificial Intelligence in HCI (Vol. 12797, pp. 3–20). Springer International Publishing. https://doi.org/10.1007/978-3-030-77772-2_1

Buck, R., Khan, M., Fagan, M., & Coman, E. (2018). The User Affective Experience Scale: A Measure of Emotions Anticipated in Response to Pop-Up Computer Warnings. International Journal of Human–Computer Interaction, 34(1), 25–34. https://doi.org/10.1080/10447318.2017.1314612

Chen, Q. Q., & Park, H. J. (2021). How anthropomorphism affects trust in intelligent personal assistants. Industrial Management & Data Systems.

Chowdhary, K. R. (2020). Fundamentals of artificial intelligence. Springer.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences.

Das, A., & Rad, P. (2020). Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. ArXiv:2006.11371 [Cs]. http://arxiv.org/abs/2006.11371

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. Technological Forecasting and Social Change, 173, 121092. https://doi.org/10.1016/j.techfore.2021.121092

Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S., & Cruz, F. (2021). Levels of explainable artificial intelligence for human-aligned conversational explanations. Artificial Intelligence, 299, 103525. https://doi.org/10.1016/j.artint.2021.103525

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). Psychological Bulletin, 117(1), 39–66. https://doi.org/10.1037/0033-2909.117.1.39

Gan, Y., Ji, Y., Jiang, S., Liu, X., Feng, Z., Li, Y., & Liu, Y. (2021). Integrating aesthetic and emotional preferences in social robot design: An affective design approach with Kansei Engineering and Deep Convolutional Generative Adversarial Network. International Journal of Industrial Ergonomics, 83, 103128. https://doi.org/10.1016/j.ergon.2021.103128

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gompei, T., & Umemuro, H. (2018). Factors and Development of Cognitive and Affective Trust on Social Robots. In S. S. Ge, J.-J. Cabibihan, M. A. Salichs, E. Broadbent, H. He, A. R. Wagner, & Á. Castro-González (Eds.), Social Robotics (Vol. 11357, pp. 45–54). Springer International Publishing. https://doi.org/10.1007/978-3-030-05204-1_5

Guerdan, L., Raymond, A., & Gunes, H. (2021). Toward Affective XAI: Facial Affect Analysis for Understanding Explainable Human-AI Interactions. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 3789–3798. https://doi.org/10.1109/ICCVW54120.2021.00423

Gunning, D. (2017). Explainable Artificial Intelligence (XAI). 36.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review, 61(4), 5–14. https://doi.org/10.1177/0008125619864925

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Jin, W., Fan, J., Gromala, D., Pasquier, P., & Hamarneh, G. (2021). EUCA: A Practical Prototyping Framework towards End-User-Centered Explainable Artificial Intelligence. 51.

Kaptein, F., Broekens, J., Hindriks, K., & Neerincx, M. (2017a). Self-explanations of a cognitive agent by citing goals and emotions. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 81–82. https://doi.org/10.1109/ACIIW.2017.8272592

Kaptein, F., Broekens, J., Hindriks, K., & Neerincx, M. (2017b). The role of emotion in self-explanations by cognitive agents. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 88–93. https://doi.org/10.1109/ACIIW.2017.8272595

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. Human Factors, 31.

Lerner, J. S., Han, S., & Keltner, D. (2007). Feelings and Consumer Decision Making: Extending the Appraisal-Tendency Framework. Journal of Consumer Psychology, 17(3), 181–187. https://doi.org/10.1016/S1057-7408(07)70027-X

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. Journal of Personality and Social Psychology, 81(1), 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Lewis, M., Li, H., & Sycara, K. (2021). Deep learning, transparency, and trust in human robot teamwork. In Trust in Human-Robot Interaction (pp. 321–352). Elsevier. https://doi.org/10.1016/B978-0-12-819472-0. 00014–9

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018

Madsen, M., & Gregor, S. (2000). Measuring Human-Computer Trust. 12.

Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. The Mit Press.

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

Plutchik, R. (1994). The psychology and biology of emotion (1st ed). HarperCollinsCollegePublishers.

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2003). The researcher's dilemma: Evaluating trust in computer-mediated communication. International Journal of Human-Computer Studies, 58(6), 759–781. https://doi.org/10.1016/S1071-5819(03)00042-9

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of Educational Research, 99(6), 323–338. https://doi.org/10.3200/JOER.99.6. 323–338

Tomkins, S. S. (1992). Affect, imagery, consciousness. 2: The negative affects (Repr). Springer.

Westland, C. (2010). Lower bounds on sample size in structural equation modeling. Electronic Commerce Research and Applications, 9(6), 476–487. https://doi.org/10.1016/j.elerap.2010.07.003

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating Effects of User Experience and System Transparency on Trust in Automation. Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, 408–416. https://doi.org/10.1145/2909824.3020230