Paper ID APEN-MIT-2019_223
Applied Energy Symposium: MIT A+B
May 22-24, 2019 • Boston, USA

Optimising user engagement in highly automated virtual assistants to improve energy management

Fernando Galdon
Department of Global Innovation Design,
School of Design.
Royal College of Art
London, UK
fernando.galdon@network.rca.ac.uk1

Stephen Jia Wang
School of Design
Hong Kong Polytechnic University
Hong Kong, SAR
stephen.j.wang@polyu.edu.hk

Abstract— This paper presents a multi-dimensional taxonomy of levels of automation and reparation specifically adapted to Virtual Assistants (VAs) in the context of *Human-Human-Interaction* (HHI). Building from this framework, the main output of this study provides a method of calculation which helps to generate a trust rating by which this score can be used to optimise users' engagement. This tool may be critical for the optimisation of energy management and consumption. Based on the research findings, the relevance of contextual events and dynamism in trust could be enhanced, such as trust formation as a dynamic process that starts before a user's first contact with the system and continues long thereafter. Furthermore, following the continuously evolving of the system, factor-affecting trust during user interactions change together with the system and over time; thus, systems need to be able to adapt and evolve as well. Present work is being dedicated to further understanding of how contexts and its derivative unintended consequences affect trust in highly automated VAs in the area of energy consumption.

Keywords—trust, energy management, engagement, system design, calibration system,

I. Introduction

With around 120 million smart speakers circulating in the USA alone, raising 78% from the previous year [1], between 21% and 32% of the population now owns a smart speaker (depending on the study [1-3]). It is up from 16% at the end of the 2017 holiday season, and more than 50% of those people own two or more [1]. Virtual assistants are expected to dominate interactions in the near future, and will play a fundamental role on energy management and consumption at home via specific applications such as Google's Nest or Alexa home.

In this scenario, Virtual Assistants are transitioning from automation to autonomy. A recent demo presented by Google called Duplex presented an extraordinary level of fluency and autonomy never seen before. Therefore, design must focus attention to a new class of technology: highly autonomous systems. [4]. In this emerging Machine-Human-Interaction (MHI) paradigm is the technology who holds the initiative of the interaction [5]. This approach places highly autonomous systems at the centre and position trust as the fundamental element to design.

In this paradigm, the system will have the information and initiative to regulate human behaviour to optimise the impact of energy consumption. In this context, trust will be capital for the adoption of new strategies in energy management. However, as this persuasive approach will be fundamentally unsupervised, it may generate unintended consequences. If the system's failure rate goes beyond 30%, the user stops using it [6]. The main reasons are; high expectation of automation performance and unexpected errors [7-10]. Traditionally, complex autonomous systems required the human operator to appropriately calibrate their trust in the automation in order to achieve performance and safety goals. In this context, literature has focused on the Human-Machine-Interaction (HMI) and Human-Human-Interaction (HHI) trust paradigms to precisely define and measure trust in automation. In this article, the author minds the warning and propose a human-centred approach in the context of HHI directly aimed at ensuring that emerging highly autonomous systems interactions remain focused on the user's needs and preferences.

II. Trust dimensions

Research in this area have been complicated due to a lack of clear distinction amongst the factors that constitute trust, trust itself and the outcomes of trust. The main model from which all contemporary research underpins is Mayer's dimensional model. Who after an extensive revision on the topic, proposed a generic typology consisting fundamentally on three dimensions; ability, benevolence, and integrity [11].

These dimensions are conceptually distinct since they address different elements of cognitive and affective abstraction of trust. However collectively, they represent a comprehensive multi-dimensional space for trust. Their multidimensional model is one of the most widely accepted [12-13]. In Mayer's model three dimensions underpin the process of trust [14]:

- *Ability* - this area refers to "the trustor's perception of trustee's competencies and knowledge salient to the expected behaviour". They can be based on "prior (first-hand or second-hand) experience or institutional endorsements"
- *Integrity* – this area refers to the perception a trustor will follow a set of principles or rules

- *Benevolence* – this area refers to the intentionality and behaviour of the trustee. It is the intend of doing good to the trustor, beyond its own profit motives

These dimensions have embodied the model used in the Human-Machine-Interaction paradigm (HMI). However, these dimensions have been mutating due to the grown in independence of these emerging systems through unsupervised reinforcement learning and the ways in which they interact with users.

In this context, specification, robustness and assurance have emerged as the dimensions to address the emerging Machine-Human-Interaction paradigm (MHI) [5].

- Specification problem arises when there is a mismatch between the ideal speciation and the revealed speciation, that is, when the AI system doesn't do what we would like it to do
- Robustness relates to the capability of the system to withstand perturbations which revolve around distributional shift, adversarial inputs, and unsafe exploration. Unsafe explorations are particularly difficult to address as they relate to "a system that seeks to maximise its performance and attain goals without having safety guarantees that will not be violated during exploration, as it learns and explores in its environment"
- Assurance involves monitoring and enforcing. The main problem is the incapability of an AI system for explaining its own decision and the difficulty of designing an off switch on the system to be able to turn off itself whenever necessary

However, recent research in the area of robustness in HAS shown 0% adversarial accuracy when evaluating a deep network against stronger adversaries [15-16]. In order to address this problem, they are using interval bound propagation to great success [17-19]. However, as the researcher acknowledge "no amount of testing can formally guarantee that a system will behave as we want. In large-scale models, enumerating all possible outputs for a given set of inputs…is intractable due to the astronomical number of choices for the input perturbation" [20]. In addition of levels of automation, papers in this area are also calling for the development of reparation strategies to address unintended consequences [21-23]. These strategies are becoming capital to address engagement and maintain trust in these systems. According to research in the area, Virtual Assistants need to generate less than 30% of errors, otherwise the user stop using them [24-26].

These elements position an intermediate Human-Human-Interaction (HHI) as a transitional paradigm to address trust in automated systems. In this context, the author presents three dimensions; autonomy, reparation and accountability to address the evolving and unpredictable nature of these systems.

- Autonomy - this area refers to the ability/robustness of the system. The competencies and knowledge of the system to perform according to expectations.
- Reparation - this area refers to benevolence/specification of the relationship. The predisposition of a trustee to integrate/develop reparation strategies to address unexpected behaviour. This element inserts a sense of balance in terms of vulnerability between users and developers enhancing trust in the interactive process.
- Accountability - this area refers to integrity/assurance of the relationship. The predisposition of a trustee to be accountable if something goes unintendedly. This element inserts a sense of balance in terms of integrity between users and developers enhancing trust in the interactive process.

III. User Engagement Challenges in Highly Automated Systems

A. Dependency

De Visser expects that the interaction with these highly automated systems will increase our emotional attachment and will be dominated by social and psychological factors [27]. In this scenario, recent investigations on Facebook's like button present the addictive implications of automated systems [28].

B. Asymmetries

One of the fundamental problems for preventing persuasion while designing trust in Highly Automated System such as Duplex is the level of asymmetry among the user and the system.

In a recent experiment Dylan Curran downloaded all his information from Goggle. The researcher presented evidences demonstrating that Goggle had stored 5.5 GB of information (around 3.000.000 million documents) [29]. Google knows where you have been, what you search, who are your friends, what do you like and dislike, your future plans, your preferences, the videos you watch on YouTube and trends you are interested. And we must point out that we do not know whether they are storing biometric data such as skin conductance, eyes tracking, pupil dilatation or face recognition through third parties. Clearly there are a range of data asymmetries between the system and the user in terms of data acquisition (personal, social, biometric and environmental), knowledge extraction capabilities (patterns, routines, trends, preferences), monitoring (sensors, cameras and microphones), and delivery (Information quality And Information usefulness)

C. Inferences

Inferences are assumptions/predictions about future behaviours enabled by data mining techniques. By using machine learning and deep learning algorithms, companies infer attributes such as sexual orientation, race, political opinions, imminent suicide attempts, eligibility for loans, political stances on abortion, susceptibility to depression, prediction of flu outbreaks, Alzheimer's disease, pregnancy by Target, assessment of users' satisfaction based on mouse tracking, or China's Social Credit Scoring system [30].

According to Wachter and Mittelstadt, the fundamental problem with inferences is that they cannot be verified at the time of decision making [30]. Furthermore, they impact our private lives, identity, reputation, and self-determination. And determine how we are viewed and evaluated by third parties. In this context, it is suggested that individuals must be protected against the inputs, and also, against the outputs of data processing. Unfortunately, as noted by Wachter and

Mitterlstant, no law and jurisprudence are providing it [30].

Furthermore, the nascent nature of these systems and the unavailability of them to conduct research prevents an adequate development of strategies.

IV. Engagement and trust

The ability of users to understand the system becomes more difficult when autonomous systems become more and more complex. Research illustrate that the higher the levels of automation, the lower the levels of trust [31]. In this context, reliability and predictability have been identified as a key factor influencing trust in automation [32]. Therefore, in order to address trust in highly automated systems, Trust must be appropriately calibrated to the actual system performance [33].

In the context of reliability, predictability has been identified as a fundamental quality for trust in automated systems. It is argued that prediction is necessary to mitigate potentially detrimental interaction behaviour to avoid unintended results that cannot be changed [34]. In this context, for the system to enhance reliability, the calibration system must enhance predictability. In Predictability, prior knowledge about potential automation failures reduces the level of uncertainty and risk [31]. Once reliability has been judged, the most important factor of trust in automation is predictability of performance over time [35].

Traditionally, in the context of automation, predictability is enhanced by implementing levels of automation (LoA). The notion of different levels of automation has been persistent in the automation literature since its introduction by Sheridan and Verplanck [36]. The idea of gradient-base models of approximation with positive, negative and neutral spectrums has been embodied through the concept of scales or Level of trust (LoT). Kaber points out that levels of automation (LoA) is a fundamental design characteristic that determines the ability of developers/designers to provide effective oversight and interaction with system autonomy [37]. Levels aim to improve transparency by simplifying interactions. In this context, transparency refers to the extent to which the actions of the automation are understandable and predictable by the user [38]. Automated systems which clarify their reasoning are more likely to be trusted [39-41]. Trust is an essential quality to build and maintain user engagement.

V. method

Scales addressing trust in automation range from one to ten points. The most common types are odd or uneven scales which allow the participant to record a neutral trust level. Recent studies using the scale presented excellent internal reliability (Cronbach's alpha = .93)[42]. Scales in automation functionality for measuring trust ranges from particular types of automation, such as autonomous vehicles [43], to robotics [44]. However, Bradshaw, Hoffman, Johnson, and Woods [45], argue that the notion of levels of automation are problematic because the Level of Autonomy is relative to the task, goals, and context. At the same time, literature points to LoA as a fundamental design characteristic that determines the ability of operators to provide effective oversight and interaction with system autonomy. In this context, LoA remains a central design decision associated with the design of automated and autonomous systems that must be addressed in system design.

VI. Discussion

In order to investigate these elements a preliminary investigation underpinned four highly sensitives areas where highly automated virtual assistants may impact significantly users; health and wellbeing, identity, economically related activities and social interactions.

Once the relevant contexts were identified, a workshop was conducted with 20 design students from the design department at the Royal college of Art to map unintended consequences in these highly sensitive areas. From this activity four main categories of unintended consequences emerged: unhappiness about the service, wrong predictions, losing something in the service and a service may unexpectedly end violently.

Then, a calibration system was designed by the lead author integrating dimensions, challenges, contexts and actions (Figure. 1). It was structured in four levels;

- Access - this area integrates a range of asymmetries related to data between the system and the user in terms of data acquisition (personal, social, biometric and environmental), and monitoring (sensors, GPS, cameras and microphones).
- Inferences - this area integrates the variables of knowledge extraction capabilities (patterns, routines, trends, preferences), and analysis (classification, labelling, probabilities and best option)
- Reasoning - this area integrates the scales of autonomy, reparation and accountability, as well as contexts (health and wellbeing, social interactions, emancipation and identity) and actions (unhappy services, wrong predictions, loses and unexpected violent endings)
- Calibration - this area integrates a matrix-based risk analysis tool.

After the system was designed, a survey was implemented to weight the impact of actions and contexts in highly automated VAs. From the areas aforementioned and based on demos, patents and prototypes, eight cases study were built to address different outcomes. Two cases were build to address each sensitive area ranging from low to high impact. The survey was structured around three sections addressing the three dimensions proposed; autonomy, reparation and accountability [41, 46-47].

Then, based on all the variables a calibration matrix was designed to map the intend of the system. It was structured around the three dimensions proposed and organised in five levels: Low risk, Medium to low risk, Medium risk, Medium to high risk and high risk.

With this system is possible to obtain a trust rating illustrating the potential impact of an action/skill in context.

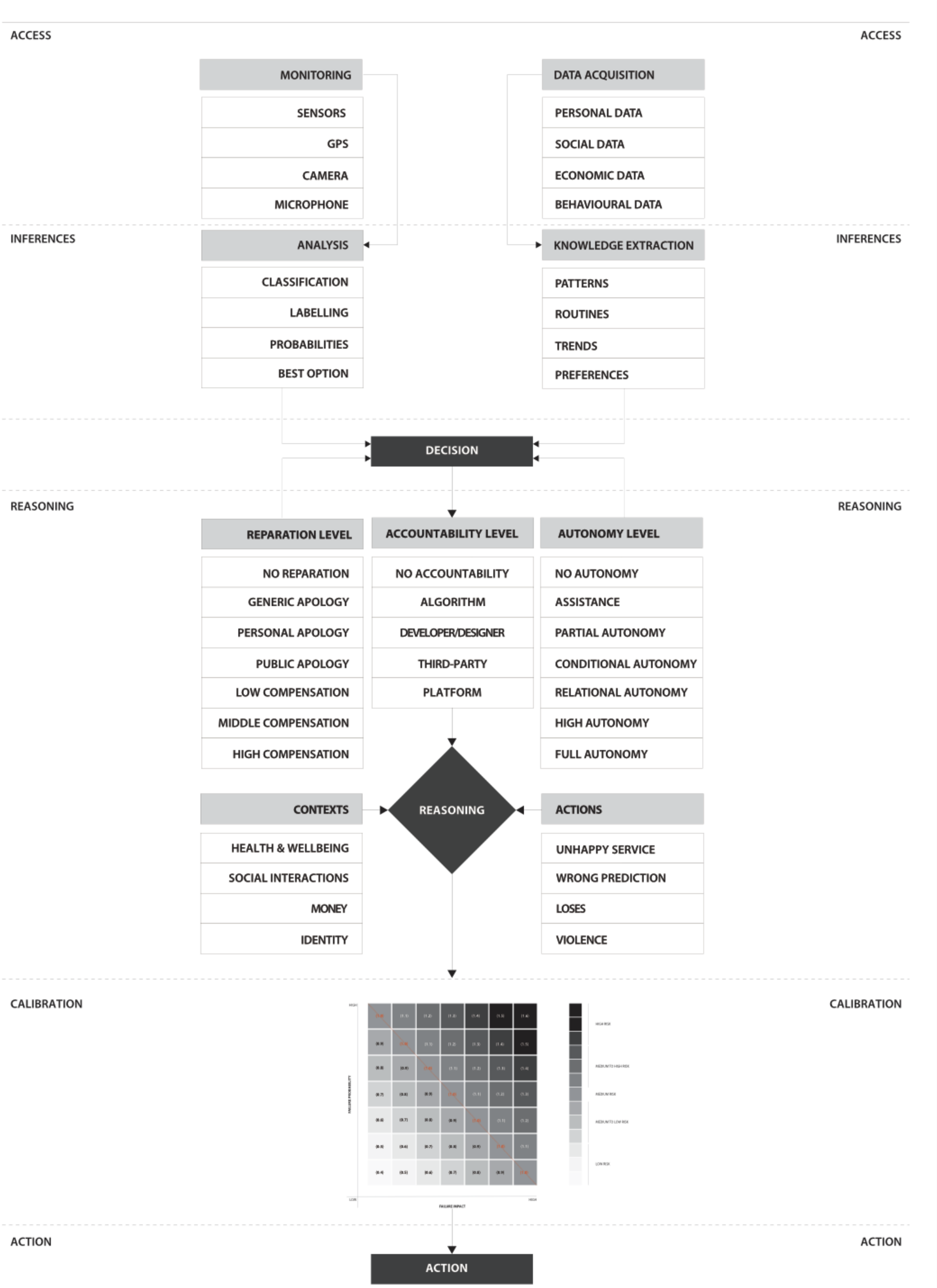

Figure 1. System design (Fernando Galdon)

A second workshop with 10 participants from the schools of Design and Architecture from the Royal College of Art was implemented to investigate energy management and consumption in the context of virtual assistants to understand future developments, and assess the weighing system.

First, participants mapped current skills/actions. Then, they projected them into the future using 'what if…' questions. After this task, participants were requested to conduct a consequential analysis, mapping desired and undesired consequences. Then, both groups confronted results and the unanticipated emerged for each group. This element presented participants with their own limitations and enhanced self-criticality. After this analysis, they mapped the prospective outcomes in terms of impact in contexts and impact in actions. They were presented with two quadrants to map the outcomes in highly sensitive areas in terms of contexts and actions. This analysis allowed them to understand context and action impact.

VII. CONCLUSION

Due to the ever-evolving nature of these systems, no amount of testing can formally guarantee that a system will behave as we want. In large-scale models such as VAs, enumerating all possible outputs for a given set of inputs, remains intractable due to the incredible number of choices for the input perturbation. This context demands the design of preventive *a priori* strategies and reparative *a posteriori* strategies to guarantee that emerging highly autonomous systems interactions remain focused on the user's needs and preferences.

In this context, trust has been identified as a fundamental variable to address. This paper presents a calibration system integrating dimensions, challenges, contexts and actions to obtain a trust rating illustrating the potential impact of an action/skill in context.

The specific workshop in energy management and consumption did not modify the weighting of contexts in the current state of VAs. However, the prospective workshop presented an energy management transition from information management to behaviour management. From concerns around monitoring information (privacy) to concerns around impact on health and wellbeing [48]. Consequently, it modified the weighting system in future developments: Health and wellbeing is the highly sensitive area concerning users the most. It is followed by social interactions, and identity. Finally, economically related activities are the least concerning highly sensitive area. In terms of actions, the same weighting system remains. Actions which may end violently causing death, harm or injury remain as the most concerning and penalised by the users in current and future developments. They are followed by losing something (specially money) and wrong prediction. Finally, unhappiness about an action/skill/service remains as the least impactful action.

The fundamental debate in future developments revolves around access. On one hand, the system needs to access data to tailor and optimise the service. On the other hand, the persuasive actions of the system impacting users' quality of life concern them [49]. The management of persuasion (dependency, asymmetries and inferences) remains capital for designers. The main tension revolves around intentionality: What should be the main priority, protecting the environment, protecting businesses or protecting the user?

In this context, the prospective nature of design revolving around preparedness, readiness and appropriateness contributes significantly to the development of unsupervised consequential systems by adjusting users and systems behaviour experience via the design of relational interventions [50].

In this scenario, by designing preventive strategies around the simulation of potential interactions and integrating reparation and accountability strategies to address unintended consequences, trust in the system can be build and maintained.

Future work is dedicated to further understanding how contexts, actions and its derivative unintended consequences affect trust in highly automated virtual assistants to build a self-calibrating algorithm in the context of what we call *synthetic consequential reasoning*. These systems designed to enhance trust aim to balance and accelerate the deployment of new concepts and technologies for managing social dynamics, mitigating unintended consequences and reduce environmental impact.

REFERENCES

1. NPR and Edison research (2019). The smart audio report. https://www.nationalpublicmedia.com/wp-content/uploads/2019/01/Smart-Audio-Report-Winter-2018.pdf
2. IDC (2018). All Categories of Smart Home Devices Forecast to Deliver Double-Digit Growth Through 2022, Says IDC. https://www.idc.com/getdoc.jsp?containerId=prUS44361618
3. Kinsella, B. (2019). U.S. Smart Speaker Ownership Rises 40% in 2018 to 66.4 Million and Amazon Echo Maintains Market Share Lead Says New Report from Voicebot. https://voicebot.ai/2019/03/07/u-s-smart-speaker-ownership-rises-40-in-2018-to-66-4-million-and-amazon-echo-maintains-market-share-lead-says-new-report-from-voicebot/
4. Hancock, P. A. (2017). Imposing limits on autonomous systems. Ergonomics, 60(2), 284–291. DOI: 10.1080/00140139.2016.1190035
5. Ortega, B. P. A. (2018). Building safe artificial intelligence : specification , robustness , and assurance Specification : design the purpose of the system. Medium. Retrieved 12/12/2018 14:20 from https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f
6. Drnec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016). From trust in automation to decision neuroscience: Applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction. Frontiers in Human Neuroscience, 10. Doi: 10.3389/fnhum.2016.00290
7. Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operator's adaptation to automation. International Journal of Human-Computer Studies, 40, 153–184. https://doi.org/10.1006/ijhc.1994.1007
8. Parasuraman, R., and Riley, V. (1997). Humans and automation; use misuse, disuse, abuse. Hum. Factors 39, 230–253. https://doi.org/10.1518/001872097778543886
9. Moray, N., Inagaki, T., and Itoh, M. (2000). Adaptive automation, trust and self- confidence in fault management of time-critical tasks. J. Exp. Psychol. Appl. 6(1), 44–58. http://dx.doi.org/10.1037/1076-898X.6.1.44
10. Dzindolet, M., Peterson, S., Pomranky, T., and Beck, P. (2003). The role of trust in automation reliance. Int. J. Hum. Comput. Stud. 58, 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7
11. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. The Academy of Management Review. Vol. 20, No. 3, pp. 709-734. Doi: 10.2307/258792

12. Rousseau, D.M., Sitkin, S.B., Burt, R.S. and Camerer, C. (1998) Not So Different after All: A Cross-Discipline View of Trust. Academy of Management Review, 23, 393-404. http://dx.doi.org/10.5465/AMR.1998.926617
13. Wasti, S. A.; TAN, Hwee Hoon; and Brower, H. H. (2007). Cross-Cultural Measurement of Supervisor Trustworthiness: An Assessment of Measurement Invariance across Three Cultures. Leadership Quarterly. 18, (5), 477-489. Research Collection Lee Kong Chian School Of Business. Doi; 10.1016/j.leaqua.2007.07.004
14. Bhattacherjee, A. (2002) Individual Trust in Online Firms: Scale Development and Initial Test, Journal of Management Information Systems, 19:1, 211-241, DOI: 10.1080/07421222.2002.11045715
15. Athalye, A., Carlini, N., Wagner, D. (2018). Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.
16. Uesato, J., O'Donoghue, B., van den Oord, A., and Kohli, P. (2018). Adversarial Risk and the Dangers of Evaluating Against Weak Attacks. Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.
17. Ehlers, R. (2017). Formal Verification of Piece-Wise Linear Feed-Forward Neural Networks. BT - Automated Technology for Verification and Analysis - 15th International Symposium, ATVA 2017, Pune, India, October 3-6, 2017, Proceedings
18. Katz, G., Barrett, C. W., Dill, D. L., Julian, K. And Kochenderfer, M. J. (2017). Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks. BT - Computer Aided Verification - 29th International Conference, CAV 2017, Heidelberg, Germany, July 24-28, 2017, Proceedings, Part I
19. Mirman, M., Gehr, T. & Vechev, M.. (2018). Differentiable Abstract Interpretation for Provably Robust Neural Networks. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:3578-3586
20. Kohli, P., Gowal, S., Dvijotham, K., and Uesato, J. (2019). Towards Robust and Verified AI: Specification Testing, Robust Training, and Formal Verification. Deepmind. Medium 28 March 2019. https://deepmind.com/blog/robust-and-verified-ai/. Accessed; 29/03/2019.
21. Bottom, W. P., Gibson, K., Daniels, S. E., & Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. Organization Science, 13(5), 497-513.
22. Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence-versus integrity-based trust violations. Journal of applied psychology, 89(1), 104.
23. Kohn, S. C., Quinn, D., Pak, R., de Visser, E. J., & Shaw, T. H. (2018). Trust Repair Strategies with Self-Driving Vehicles: An Exploratory Study. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 1108–1112. doi:10.1177/1541931218621254
24. Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. A Syst. Hum. 30, 286–297.
25. Wickens, C. D., and Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. Theor. Issues Ergon. Sci. 8, 201–212
26. Wang, L., Jamieson, G. A., and Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. Hum. Factors 51, 281–291.
27. de Visser, E. J., Pak, R., and Shaw, T. H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction, Ergonomics, 61:10, 1409-1427, DOI: 10.1080/00140139.2018.1457725
28. Osnos, E. (2018). Can Mark Zuckerberg Fix Facebook Before It Breaks Democracy?. The New Yorker. Retrieved 08/10/2018 21:10 from https://www.newyorker.com/magazine/2018/09/17/can-mark-zuckerberg-fix-facebook-before-it-breaks-democracy
29. Raphael, J. (2018). This Is Exactly How Much Information Google Has On You. Esquire. Retrieved 12/10/2018 19:10 from https://www.esquire.com/uk/latest-news/a19614546/this-exactly-how-much-information-google-has-on-you/
30. Wachter, S., & and Mittelstadt, B. (2018). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Business Law Review, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3248829
31. Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), Foundations of Trusted Autonomy (pp. 135–159). Australia: Springer Open. https://doi.org/10.1007/978-3-319-64816-3_8
32. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46, 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
33. Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics, 37, 1905–1922. https://doi.org/10.1080/00140139408964957
34. Drnec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016). From trust in automation to decision neuroscience: Applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction. Frontiers in Human Neuroscience, 10. Doi: 10.3389/fnhum.2016.00290
35. Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35, 1243–1270. https://doi.org/10.1080/00140139208967392
36. Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators: Fort Belvoir, VA: Defense Technical Information Center. https://doi.org/10.21236/ADA057655
37. Kaber, D. B. (2018). Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation. Journal of Cognitive Engineering and Decision Making, 12(1), 7–24. https://doi.org/10.1177%2F1555343417737203
38. Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. Human Factors: The Journal of the Human Factors and Ergonomics Society, 59, 5–27. https://doi.org/10.1177/0018720816681350
39. Simpson, A., Brander, G. N., & Portsdown, D. R. A. (1995). Seaworthy trust: Confidence in automated data fusion. In R. M. Taylor & J. Reising (Eds.) The Human-Electronic Crew: Can we Trust the Team, (pp 77–81). Hampshire, UK: Defence Research Academy. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a308589.pdf
40. Hoff, K. A., & 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, 57(3), 407–434. Doi: https://doi.org/10.1177/0018720814547570
41. Galdon, F., & Wang, S. J. (2019). Designing trust in highly automated virtual assistants: A taxonomy of levels of autonomy. International Conference on Industry 4.0 and Artificial Intelligence Technologies. Cambridge, UK. ISBN: 978-1-912532-07-0
42. Buckley, L., Kaye, S.-A., & Pradhan, A. K. (2018). Psychosocial factors associated with intended use of automated vehicles: A simulated driving study. Accident Analysis & Prevention, 115, 202–208. https://doi.org/10.1016/j.aap.2018.03.021
43. Garcia, D., Kreutzer, C., Badillo-Urquiola, K., & Mouloua, M. (2015). Measuring Trust of Autonomous Vehicles: A Development and Validation Study. In C. Stephanidis (Ed.), HCI International 2015 - Posters' Extended Abstracts (Vol. 529, pp. 610–615). Cham: Springer International Publishing.
44. Yagoda, R. E., & Gillan, D. J. (2012). You want me to trust a ROBOT? The development of a human–robot interaction trust scale. International Journal of Social Robotics, 4, 235–248.
45. Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013). The seven deadly myths of autonomous systems. IEEE Intelligent Systems, 28(3), 54–61. https://doi.org/10.1109/MIS.2013.70
46. Galdon, F., & Wang, S. J. (2019, August). From Apology to Compensation: A Multi-level Taxonomy of Trust Reparation for Highly Automated Virtual Assistants. In *International Conference on Human Interaction and Emerging Technologies* (pp. 42-46). Springer, Cham.
47. Galdon, F., & Wang, S. J. (2019, August). Addressing accountability in highly autonomous virtual assistants. In *International Conference on Human Interaction and Emerging Technologies* (pp. 10-14). Springer, Cham.
48. Wang S.J., Moriarty P. (2018) Urban Health and Well-Being Challenges. In: Big Data for Urban Sustainability. Springer, Cham.
49. Wang, S. J., & Moriarty, P. (2018). *Big Data for Urban Sustainability*. Springer.
50. Wang, S. J. (2013). *Fields Interaction Design (FID): The answer to ubiquitous computing supported environments in the post-information age*. Homa & Sekey Books.