Trustworthy Acceptance: A New Metric for Trustworthy Artificial Intelligence used in Decision Making in Food–Energy–Water Sectors

Suleyman Uslu, Davinder Kaur, Samuel J Rivera, Arjan Durresi, Mimoza Durresi and Meghna Babbar-Sebens

Abstract We propose, for the first time, a *trustworthy acceptance metric* and its measurement methodology to evaluate the trustworthiness of AI-based systems used in decision making in Food Energy Water (FEW) management. The proposed metric is a significant step forward in the standardization process of AI systems. It is essential to standardize the AI systems' trustworthiness, but until now, the standardization efforts remain at the level of high-level principles. The measurement methodology of the proposed includes human experts in the loop, and it is based on our trust management system. Our metric captures and quantifies the system's transparent evaluation by field experts on as many control points as desirable by the users. We illustrate the *trustworthy acceptance metric* and its measurement methodology using AI in decision-making scenarios of Food-Energy-Water sectors. However, the proposed metric and its methodology can be easily adapted to other fields of AI applications. We show that our metric successfully captures the aggregated acceptance of any number of experts, can be used to do multiple measurements on various points of the system, and provides confidence values for the measured acceptance.

Suleyman Uslu
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: suslu@iu.edu

Davinder Kaur
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: davikaur@iu.edu

Samuel J Rivera
Oregon State University, Corvallis, OR, USA e-mail: sammy.rivera@oregonstate.edu

Arjan Durresi
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: adurresi@iupui.edu

Mimoza Durresi
European University of Tirana e-mail: mimoza.durresi@uet.edu.al

Meghna Babbar-Sebens
Oregon State University, Corvallis, OR, USA e-mail: meghna@oregonstate.edu

1 Introduction

Despite so many advantages of AI systems and their uses, these systems sometimes directly or indirectly harm the users and society. It has become essential to make these systems safe, reliable, and trustworthy. Lately, trustworthy AI has been gaining increasing attention from governments, organizations, and scientific communities. So European Union (EU) has proposed ethnic guidelines and laws [50] for trustworthy AI to govern and facilitate the development and working of AI systems [13]. DARPA [15] also launched an XAI program, whose motive was to make these AI systems explainable and trustworthy. Garter estimates that 30% of all the digital products that use AI will require a trustworthy AI framework by 2025 [6], and 86% of users will trust and remain loyal to companies that use ethical AI principles [4]. So, developing AI systems using a trustworthy framework has become a necessity for today's society. Furthermore, there are various recent studies on developing trustworthy and explainable algorithms and AI [41, 29, 28, 49, 40, 19, 42, 55]. Kaur et al. [23] surveyed similar approaches that aimed to create trustworthy and explainable AI systems.

And as AI technologies mature, they have to follow the natural process that all established technologies have gone through, standardization. Among many advantages, standards allow manufacturers and users to speak the same language, enable the users to check the quality of products on the market, reduce the legal liability of manufacturers and providers, etc. Already ISO, an organization that deals with standardization, has presented different approaches to establish trust in AI systems using fairness, transparency, accountability, and controllability [20]. However, such works remain at the level of high-level principles. What is needed is to develop metrics and measurement procedures to establish standards for trustworthy AI [27]. We envision that in the future, various agencies will use such metrics to certify AI-based solutions, similarly as FDA certifies medications and treatments. For this reason, in this paper, for the first time, we propose a concrete metric, *trustworthy acceptance*, and its measurement methodology. Our metric captures and quantifies the system's transparent evaluation by field experts.

In this paper, we use the following definition: *Trustworthy AI is a framework to ensure that a system is worthy of being trusted concerning its stated requirements based on the evidence. It makes sure that the users' and stakeholders' expectations are met in a verifiable way* [20]. Furthermore, for the time being, AI lacks many aspects of human intelligence, including meaning, multidimensional data beyond the set used for algorithm training, meaningful causality, ethics, etc. Therefore, AI systems should complement and empower humans without replacing them. This is the essential requirement to make AI trustworthy. And for this reason, when developing the proposed metric and corresponding measurement procedures, we include human experts in the loop. Our contributions can be summarized as follows:

- We propose in Section 4 a *trustworthy acceptance* metric for the evaluation of the acceptance of AI-based systems by field experts.

- The measurement procedure for the proposed metric is described in Section 4 and is based on the concept of a distance acceptance approach that is adaptable to a wide range of systems. In addition to the acceptance value, our metric provides the confidence of the acceptance.
- Our metric utilizes the trust of the experts in the given context, managed by our trust system, summarized in Section 3.
- Our metric can be measured in many points of the system in order to reach an assessment of the whole system, as discussed in Sections 4 and 5.
- Finally, in Section 5, we illustrate the application of our trustworthy acceptance metric and its measurement methodology using three systems for environmental decision making.

2 Related Work

As part of our daily lives, decision-making is also an essential element of processes of the most significant fields such as economics, finance, healthcare, and the environment. Kambiz [21] explained the importance of decision makings in such fields by giving examples of the crucial results caused by erroneous or dissatisfactory decisions such as the world economic crisis. He indicated that the difficulty of making such critical decisions relies on their complex nature and the involvement of multiple stakeholders who can have different expertise, background, perspectives, or maybe even competing for objectives.

As it is possible and common in some areas to have experts as stakeholders, there is a need to have a consensus mechanism for such decision makings. Dong and Xu [11] proposed approaches to minimize the modifications to the solutions that the experts propose at each round of the decision making. Furthermore, Hegselmann and Krause [17] investigated the consensus reaching mechanisms of decision makings involving agents with different behaviors in both mathematical modelings and computer simulations. Similarly, Babbar-Sebens and Minsker [3] proposed a design employing an algorithm for the utilization of expert feedback for an improved optimization in the environmental field. These studies clearly show a considerable need for methods and frameworks for advanced decision-making, especially for the ones where humans and machines need to collaborate for superior results.

Trust has long been a concept that is believed to be an essential part of the decision-making process especially involving multiple stakeholders. It has been shown that the utilization of trust contributed to the integrative behavior and helped disruptive activities to decline [14, 18, 25]. There has been multiple studies [52, 32, 31, 30] and surveys [53, 10, 33] discussing trust management and its frameworks. In [39], a trust management framework based on measurement theory is proposed for online social communities. There are several applications of this framework, such as stock market prediction with Twitter data [37]. Other examples include trust management in social networks [54], cloud computing [34, 35], internet of things [36, 38], healthcare [8, 9], emergency communications [12], and detection of crime [24] and fake users [22].

In the environmental field, Alfantoukh et al. [1, 2] proposed a model for water allocation problem and a more generic model for consensus reaching problem involving trust. We introduced a trust-based decision support system for natural resource sharing problems in Food-Energy-Water sectors [48]. Also, we presented the enhanced versions of our system utilizing discrete and precomputed solutions [43], game-theoretical approach [44], and scenarios with different evaluation criteria [45]. Furthermore, we presented the role of trust sensitivity of actors in environmental trust-based decision-making scenarios [47].

When computers make decisions, the liability of the decisions becomes a significant issue. Therefore, it requires a comprehensive testing process before deploying and utilizing such systems, which brings the acceptance of a system in the picture. Although there are studies to reach a desired level of acceptance, proposals could be concentrated on revealing and satisfying the needs of the users [26]. Also, there are studies to model users' desire to use such AI-based systems [16]. One of the most critical areas that we need the high acceptance of is healthcare, where trust between doctors and computerized systems is highly crucial [51]. In [24], it has been shown that in decision makings involving both humans and computers, there is no exact winner when considering all scenarios of crime detection, which opens the door for collaboration for more significant results. Similarly, in [22], AI-based fake user detection gave better results when it utilized the community's preferences.

3 Trust Management Framework

We summarize here our measurement theory-based trust management framework [39], because the new metric and its methodology are based on this framework. Trust defined in this framework has two main parameters: impression, m, and confidence, c. Impression represents the level of trust from one party to another, while confidence is the degree of certainty of the impression. Impression calculation is done by averaging the measurements, as shown in Eq. 1 whereas the confidence is calculated related to the standard error of the mean as shown in Eq. 2. In these formulas, $m^{A:B}$, $c^{A:B}$, and $r_i^{A:B}$ are the impression, confidence, and a measurement from A to B.

$$m^{A:B} = \frac{\sum_{i=1}^{N} r_i^{A:B}}{N} \tag{1}$$

$$c^{A:B} = 1 - 2\sqrt{\frac{\sum_{i=1}^{N} (m^{A:B} - r_i^{A:B})^2}{N(N-1)}} \tag{2}$$

Another essential feature of this framework is to anticipate the trust even without communication between two parties. It supports the propagation methods, namely transition and aggregation, to assess trust between two entities that are connected through third-party nodes. Ruan et al. [39] proposed several trust propagation methods and provided their error propagation functions. For aggregation, we selected the averaging method, as shown in Eq. 3, and provided its error formula in Eq. 4. In these formulas, S represents the source, D is the destination, and Ts are the transitive nodes.

$$m_{T_1}^{SD} \oplus m_{T_2}^{SD} = \frac{m_{T_1}^{SD} + m_{T_2}^{SD}}{2} \tag{3}$$

$$e_{T_1}^{SD} \oplus e_{T_2}^{SD} = \sqrt{\frac{1}{2^2}((e_{T_1}^{SD})^2 + (e_{T_2}^{SD})^2)} \tag{4}$$

4 Trustworthy Acceptance Metric and its Methodology

This section describes the methodology of our trustworthy acceptance metric. We assume that an AI based system generates sets of solutions that a group of experts will evaluate. Each expert has its own set of solutions to evaluate. Part of each set of solution is a reference, optimal solution and some sub-optimal solutions, based on the trade-offs and the criteria applied by the system. Experts might chose the optimal solution or another sub-optimal solutions, based on their expertise, as shown in Fig. 1.

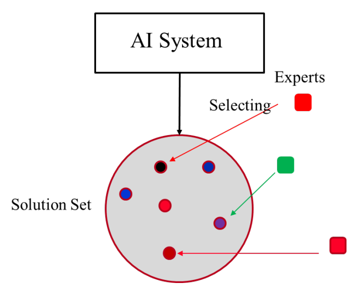

Fig. 1 Experts evaluate and select solutions from the set generated by the AI system.

Acceptance is based on and inversely related to the distance between the proposals of two parties. Distance is measured using the Euclidean distance formula, where each parameter of a proposed solution becomes a dimension. In other words, solutions can be considered as vectors when measuring the distance. After normalizing the each dimension, we also normalize the final distance where the maximum distance becomes 1. Also, distances are always non-negative. An example of the distance between n-dimensional solutions S and T, represented as d_S^T, is shown in Eq. 5 where S_i and T_i are the values of the dimension i of each solution.

$$d_S^T = \sqrt{\frac{\sum_{i=1}^{n} (S_i - T_i)^2}{n}} \tag{5}$$

Acceptance is measured as shown in Eq. 6 where A_e is the acceptance of the expert e, S is the solution selected by the expert, and T is the reference point, optimal solution. In this case, acceptance is also in the same range of distance which is [0,1]. Although the distance is bidirectionally the same, we usually stress the acceptance of the system evaluated by the expert. Therefore, if a system fails to provide appropriate solution alternatives close enough to the reference point, it is appropriate for experts to reject the system or rate it with lower acceptance.

$$A_e = 1 - d_S^T \tag{6}$$

We consider each acceptance as a trust assessment. Since it is possible to evaluate the acceptance of a system by multiple experts, we use our trust system, see Sec. 3, to aggregate the individual acceptances of experts weighted by their trust as shown in Eq. 7 where k is the number of experts. Similarly, we find the confidence of the acceptance measurements by calculating the population standard error of the mean as shown in Eqs. 8 and 9. We used the regular standard deviation formula instead of the weighted one for simplicity.

$$Tw_A = \frac{\sum_e A_e T_e}{k} \tag{7}$$

$$SE_{Tw_A} = \frac{\sqrt{\sum_e (Tw_A - A_e)^2}}{n} \tag{8}$$

$$c_{Tw_A} = 1 - 2(SE_{Tw_A}) \tag{9}$$

With these two values, the weighted mean acceptance and the confidence, we have our trustworthy acceptance metric, (Tw_A, c_{Tw_A}). It is also possible to have multiple sample measurements to evaluate a system's acceptance by the same group of experts. In this case, again we use our trust system for aggregation and calculate the aggregated mean acceptance of the system as shown in Eq. 10 where n represents the sample size. Furthermore, we update the standard error of the mean using Eq. 11 which is a generalized version of Eq. 4 that we use in our trust measurements explained in detail in [5, 39].

$$System_{Tw_A} = \frac{\sum_n Tw_A}{n} \tag{10}$$

$$SE_{System_{Tw_A}} = \sqrt{\frac{1}{n^2} \sum_n (SE_{Tw_A})^2} \tag{11}$$

5 Simulated Measurement Results

To illustrate the evaluation of the proposed acceptance metric, we used the data set that consists of environmental solutions for over 200 fields in a region as explained in [47]. The genetic algorithmic system [7], generates near-optimal solutions while having constraints on the resource usage. The tests were run using three systems that generate solutions. System 1 is the default system where there is no constraint on the number of solutions generated for a field. In System 2 and 3, we reduced the number of available solutions to 7 and 4, respectively. Five agent experts were assigned to evaluate these systems by selecting the most appropriate solution for the specific field which could represent a neighborhood of farmers. Trust of such experts is assumed to be high because it makes sense that the organization which is testing the system deploy top experts. However, if needed, their trust can be dynamically adjusted by techniques described in our previous work [46, 45, 47]. Furthermore, a reference solution is also determined and assumed to be legitimate by the user. In our experiments, for simplicity, we used only the environmental protection values in the distance function.

An evaluation of a system starts by assigning each expert a field and presenting them the solutions prepared by the system for that specific field. After we completed the assignment, the experts selected the fittest solution based on their expertise. For an expert who didn't select the reference solution, a nonzero distance is measured between the expert and the reference solution, which is calculated using Eq. 5. Then, we calculated the individual acceptance rates using Eq. 6. After having individual acceptance rates, we averaged the acceptance rates weighted by their trust, as in Eq. 7, and also calculated the confidence of the acceptance, shown in Eq. 9. We call the pair of trust-weighted average acceptance and its confidence as our trustworthy acceptance metric. To increase the confidence in our metric, we repeated this process two more times for each system. We named the selected fieldsets A, B, and C. Then, we aggregated the trustworthy acceptance measurements for each fieldset using Eq. 10. After completing the evaluation for selected samples and generating the system's trustworthy acceptance, we performed the same tasks for System 2 and 3 and measured the acceptance of each system.

Fig. 2 shows the trustworthy acceptance of fields A, B, and C for Systems 1, 2, and 3. Compared to System 1, the acceptance declined in System 2 because of limiting the available solutions to 7 which eliminated some solutions closer to the optimal point. Similarly, a stricter constraint, having only four available solutions, resulted in even less acceptance of System 3. These results could be evident for a person who understands and can compare the systems. For example, such results could be used by USDA to certify only Systems 1 and 2 but not System 3. Also, the aggregation of samples' acceptances could be essential in real scenarios to reduce the bias of uneven sampling. As shown in Fig. 3, we aggregated the results of fields A, B, and C and presented the trustworthy acceptance of each system measured by our experts. As figures illustrate the acceptance of each system, Table 1 shows both the acceptance values and their confidence which together form our *trustworthy acceptance metric*.

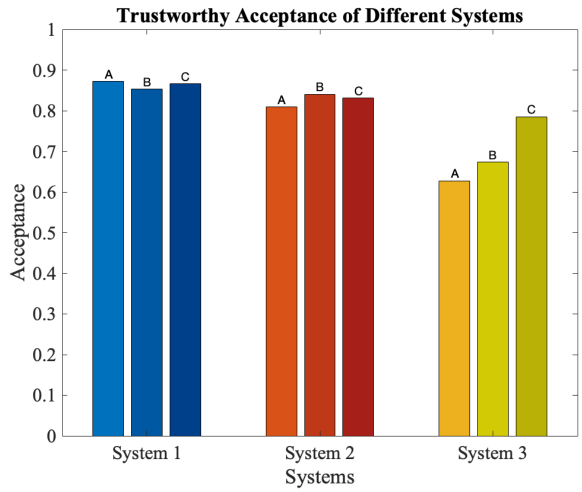

Fig. 2 Trustworthy acceptance of System 1, 2, and 3 measured over the sampled fields A, B, and C from the whole region is presented.

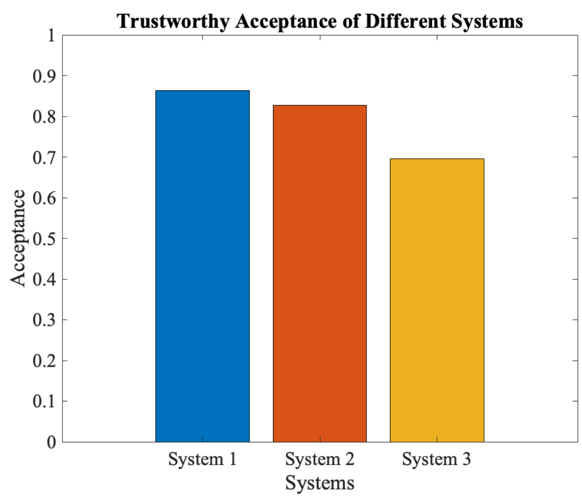

Fig. 3 Measured acceptances for fields A, B, and C are aggregated to find the trustworthy acceptance of System 1, 2, and 3.

Table 1 Trustworthy acceptances, which consist of acceptance and its confidence, are presented for fields A, B, and C and also their aggregation for System 1, 2, and 3.

System 1	Tw_Acceptance	Confidence
Field A	0.872	0.956
Field B	0.853	0.960
Field C	0.866	0.964
Tw_Acceptance_S1	0.863	0.977
System 2	**Tw_Acceptance**	**Confidence**
Field A	0.809	0.856
Field B	0.841	0.948
Field C	0.831	0.918
Tw_Acceptance_S2	0.827	0.942
System 3	**Tw_Acceptance**	**Confidence**
Field A	0.627	0.726
Field B	0.673	0.719
Field C	0.785	0.857
Tw_Acceptance_S3	0.695	0.861

6 Conclusions

We presented a new *trustworthy acceptance metric* and its measurement methodology for evaluating the approval of an AI system that generate solutions for environmental decisions for a region. Our metric can be used for standardization of Trustworthy AI. The human experts assigned for the evaluation selected the most appropriate solutions presented by each system. We measured the distance from these solutions to a optimal, reference point and calculated the trustworthy acceptance of a system using our trust framework. Furthermore, using our trust framework, we aggregated multiple measurements and provided the confidence of the acceptance using error propagation methods. Finally, we calculated and compared the trustworthy acceptance of each system measured by the assigned experts.

Our *trustworthy acceptance metric* can be applied to many AI applications that can use the concept of distance-based acceptance. Our approach to confidence measurement is based on our trust system that considers trust assessments as measurements. Lastly, the experts' inclusion of trust helped build a metric that can differentiate experts and is even more appropriate for scenarios where there is a possibility to engage different groups of experts.

Acknowledgements This work was partially supported by the National Science Foundation under Grant No.1547411 and by the U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) (Award Number 2017-67003-26057) via an interagency partnership between USDA-NIFA and the National Science Foundation (NSF) on the research program Innovations at the Nexus of Food, Energy and Water Systems.

References

1. Alfantoukh, L., Ruan, Y., Durresi, A.: Trust-based multi-stakeholder decision making in water allocation system. In: International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 314–327. Springer (2017)
2. Alfantoukh, L., Ruan, Y., Durresi, A.: Multi-stakeholder consensus decision-making framework based on trust: A generic framework. In: 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), pp. 472–479. IEEE (2018)
3. Babbar-Sebens, M., Minsker, B.S.: Interactive genetic algorithm with mixed initiative interaction for multi-criteria ground water monitoring design. Applied Soft Computing **12**(1), 182–195 (2012)
4. Barometer, E.T.: Edelman trust barometer global report. Edelman, available at: https://www. edelman. com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report_2. pdf (2019)
5. Berendsen, H.J.: A student's guide to data and error analysis. Cambridge University Press (2011)
6. Burke, B., Cearley, D., Jones, N., Smith, D., Chandrasekaran, A., Lu, C., Panetta, K.: Gartner top 10 strategic technology trends for 2020-smarter with gartner (2019)
7. Chipperfield, A., Fleming, P.: The matlab genetic algorithm toolbox (1995)
8. Chomphoosang, P., Durresi, A., Durresi, M., Barolli, L.: Trust management of social networks in health care. In: 2012 15th International Conference on Network-Based Information Systems, pp. 392–396. IEEE (2012)
9. Chomphoosang, P., Ruan, Y., Durresi, A., Durresi, M., Barolli, L.: Trust management of health care information in social networks. In: Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, pp. 228–235. IEEE (2013)
10. Chomphoosang, P., Zhang, P., Durresi, A., Barolli, L.: Survey of trust based communications in social networks. In: 2011 14th International Conference on Network-Based Information Systems, pp. 663–666. IEEE (2011)
11. Dong, Y., Xu, J.: Consensus Building in Group Decision Making. Springer (2016)
12. Durresi, A., Durresi, M., Paruchuri, V., Barolli, L.: Trust management in emergency networks. In: 2009 International Conference on Advanced Information Networking and Applications, pp. 167–174. IEEE (2009)
13. EC: Ethics guidelines for trustworthy ai (2018). URL https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
14. Gunia, B.C., Brett, J.M., Nandkeolyar, A.K., Kamdar, D.: Paying a price: Culture, trust, and negotiation consequences. Journal of applied psychology **96**(4), 774 (2011)
15. Gunning, D.: Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web **2**, 2 (2017)
16. Gursoy, D., Chi, O.H., Lu, L., Nunkoo, R.: Consumers acceptance of artificially intelligent (ai) device use in service delivery. International Journal of Information Management **49**, 157–169 (2019)
17. Hegselmann, R., Krause, U., et al.: Opinion dynamics and bounded confidence models, analysis, and simulation. Journal of artificial societies and social simulation **5**(3) (2002)
18. Hüffmeier, J., Freund, P.A., Zerres, A., Backhaus, K., Hertel, G.: Being tough or being nice? a meta-analysis on the impact of hard-and softline strategies in distributive negotiations. Journal of Management **40**(3), 866–892 (2014)
19. Hurlburt, G.: How much to trust artificial intelligence? IT Professional **19**(4), 7–11 (2017)
20. Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. Standard, International Organization for Standardization, Geneva, CH (2020)
21. Kambiz, M.: Multi-Stakeholder Decision Making for Complex Problems: A Systems Thinking Approach with Cases. World Scientific (2016)
22. Kaur, D., Uslu, S., Durresi, A.: Trust-based security mechanism for detecting clusters of fake users in social networks. In: Workshops of the International Conference on Advanced Information Networking and Applications, pp. 641–650. Springer (2019)

23. Kaur, D., Uslu, S., Durresi, A.: Requirements for trustworthy artificial intelligence–a review. In: International Conference on Network-Based Information Systems, pp. 105–115. Springer (2020)
24. Kaur, D., Uslu, S., Durresi, A., Mohler, G., Carter, J.G.: Trust-based human-machine collaboration mechanism for predicting crimes. In: International Conference on Advanced Information Networking and Applications, pp. 603–616. Springer (2020)
25. Kimmel, M.J., Pruitt, D.G., Magenau, J.M., Konar-Goldband, E., Carnevale, P.J.: Effects of trust, aspiration, and gender on negotiation tactics. Journal of personality and social psychology **38**(1), 9 (1980)
26. Kocielnik, R., Amershi, S., Bennett, P.N.: Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2019)
27. Lakkaraju, S., Adebayo, J.: Neurips (2020) tutorial," in tutorial: (track2) explaining machine learning predictions: State-of-the-art, challenges, and opportunities (2020)
28. Mueller, S.T., Hoffman, R.R., Clancey, W., Emrey, A., Klein, G.: Explanation in human-ai systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable ai. arXiv preprint arXiv:1902.01876 (2019)
29. Rossi, F.: Building trust in artificial intelligence. Journal of international affairs **72**(1), 127–134 (2018)
30. Ruan, Y., Alfantoukh, L., Durresi, A.: Exploring stock market using twitter trust network. In: Advanced Information Networking and Applications (AINA), 2015 IEEE 29th International Conference on, pp. 428–433. IEEE (2015)
31. Ruan, Y., Alfantoukh, L., Fang, A., Durresi, A.: Exploring trust propagation behaviors in online communities. In: Network-Based Information Systems (NBiS), 2014 17th International Conference on, pp. 361–367. IEEE (2014)
32. Ruan, Y., Durresi, A.: Trust management for social networks. In: Proceedings of the 14th Annual Information Security Symposium, p. 24. CERIAS-Purdue University (2013)
33. Ruan, Y., Durresi, A.: A survey of trust management systems for online social communities–trust modeling, trust inference and attacks. Knowledge-Based Systems **106**, 150–163 (2016)
34. Ruan, Y., Durresi, A.: A trust management framework for cloud computing platforms. In: Advanced Information Networking and Applications (AINA), 2017 IEEE 31st International Conference on, pp. 1146–1153. IEEE (2017)
35. Ruan, Y., Durresi, A.: A trust management framework for clouds. Computer Communications **144**, 124–131 (2019)
36. Ruan, Y., Durresi, A., Alfantoukh, L.: Trust management framework for internet of things. In: Advanced Information Networking and Applications (AINA), 2016 IEEE 30th International Conference on, pp. 1013–1019. IEEE (2016)
37. Ruan, Y., Durresi, A., Alfantoukh, L.: Using twitter trust network for stock market analysis. Knowledge-Based Systems **145**, 207–218 (2018)
38. Ruan, Y., Durresi, A., Uslu, S.: Trust assessment for internet of things in multi-access edge computing. In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), pp. 1155–1161. IEEE (2018)
39. Ruan, Y., Zhang, P., Alfantoukh, L., Durresi, A.: Measurement theory-based trust management framework for online social communities. ACM Transactions on Internet Technology (TOIT) **17**(2), 16 (2017)
40. Smith, C.J.: Designing trustworthy ai: A human-machine teaming framework to guide development. arXiv preprint arXiv:1910.03515 (2019)
41. Smuha, N.A.: The eu approach to ethics guidelines for trustworthy artificial intelligence. CRi-Computer Law Review International (2019)
42. Sutrop, M.: Should we trust artificial intelligence? TRAMES: A Journal of the Humanities & Social Sciences **23**(4) (2019)
43. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Decision support system using trust planning among food-energy-water actors. In: International Conference on Advanced Information Networking and Applications, pp. 1169–1180. Springer (2019)

44. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Trust-based game-theoretical decision making for food-energy-water management. In: International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 125–136. Springer (2019)
45. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Trust-based decision making for food-energy-water actors. In: International Conference on Advanced Information Networking and Applications, pp. 591–602. Springer (2020)
46. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M., Tilt, J.H.: Control theoretical modeling of trust-based decision making in food-energy-water management. In: Conference on Complex, Intelligent, and Software Intensive Systems, pp. 97–107. Springer (2020)
47. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M., Tilt, J.H.: A trustworthy human–machine framework for collective decision making in food–energy–water management: The role of trust sensitivity. Knowledge-Based Systems **213**, 106,683 (2021)
48. Uslu, S., Ruan, Y., Durresi, A.: Trust-based decision support system for planning among food-energy-water actors. In: Conference on Complex, Intelligent, and Software Intensive Systems, pp. 440–451. Springer (2018)
49. Varshney, K.R.: Trustworthy machine learning and artificial intelligence. XRDS: Crossroads, The ACM Magazine for Students **25**(3), 26–29 (2019)
50. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: Automated decisions and the gdpr. Harv. JL & Tech. **31**, 841 (2017)
51. Wang, W., Siau, K.: Trusting artificial intelligence in healthcare (2018)
52. Zhang, P., Durresi, A.: Trust management framework for social networks. In: 2012 IEEE International Conference on Communications (ICC), pp. 1042–1047. IEEE (2012)
53. Zhang, P., Durresi, A., Barolli, L.: Survey of trust management on various networks. In: Complex, Intelligent and Software Intensive Systems (CISIS), 2011 International Conference on, pp. 219–226. IEEE (2011)
54. Zhang, P., Durresi, A., Ruan, Y., Durresi, M.: Trust based security mechanisms for social networks. In: Broadband, Wireless Computing, Communication and Applications (BWCCA), 2012 Seventh International Conference on, pp. 264–270. IEEE (2012)
55. Zhang, Y., Liao, Q.V., Bellamy, R.K.: Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. arXiv preprint arXiv:2001.02114 (2020)