Ruchira Ray EID: rr52486

Response 1: Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law

Concerns about artificial intelligence(AI) often arise from worries about AI's power being abused, mass job loss, violation of privacy, and a lack of regulation. The paper dives into the origins of AI mistrust and how, with the emergence of AI, technology has shifted gears from systems that used to clarify every step to low-transparency algorithms that automatically learn from existing data to estimate possible future outcomes.[1]

AI models are bound to contain flaws. The paper discusses an intriguing contradiction: allowing people to contribute feedback decreases trust, maybe by exposing people to the system's failures, complexities and edge cases[2] while also making them feel valued, heard, and respected.[3] So the question arises: how much should humans be in-the-loop? Studies show that confidence can significantly increase after poor first impressions if followed by constant and accurate user support.[4] Trust develops gradually based on the system's predictability and reliability.[4] It also shows that an increase in trust is influenced by the user's age and passion for/ease with technology.[4] Thus, even if it initially lowers trust, users and various stakeholders should be involved when making decisions about AI systems.

Interestingly, the paper further elaborates on invasions of privacy by social media platforms, e-commerce websites and so on, exponentially decreasing trust in AI-based industries. Organizations collect massive amounts of data by exploiting users' personal/social interactions. As society seeks to reap the benefits of personalization, the question arises of how firms may gather enormous amounts of data without jeopardizing people's trust. People are cautious about submitting information because they are unsure who will have access to it or how organizations collecting it will keep it secure.[5] One approach is to use differential privacy, which protects personally identifiable information while allowing valuable insights to be extracted by randomly changing some user data while guaranteeing that summary statistics remain true.[5,6]

Another intriguing point highlighted is that, despite recent initiatives, AI models remain Black Boxes, and that corporations tend to gatekeep their data collection, data processing, and models, leading to growing scepticism in AI. A possible solution is to employ blockchain to decentralize data, models, and analytics to boost consumer confidence by increasing AI traceability.[7]

The paper highlights the inadequacies of present AI systems and makes several important comments about how to move forward in establishing reliable AI systems. It concludes by noting that software to evaluate and rectify mistakes should be developed, even if costly, as it will serve as an investment in developing dependable AI systems. Such software would also make it easier to include user feedback and ratings from external monitors during evaluation.

REFERENCES

[1] Dwork, Cynthia, and Martha Minow. "Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law." *Daedalus* 151.2 (2022): 309-321.

[2] Donald R. Honeycutt, Mahsan Nourani, and Eric D. Ragan, “Soliciting Human-in-the- Loop User Feedback for Interactive Machine Learning Reduces User Trust and Impressions of Model Accuracy,” in Proceedings of the Eighth AAAI Conference on Human Computation and Crowdsourcing (Menlo Park, Calif.: Association for the Advancement of Artificial Intelligence, 2020), 63–72, https://ojs.aaai.org/index.php/HCOMP/article/view/7464.

[3] See E. Allan Lind and Tom R. Tyler, eds., The Social Psychology of Procedural Justice (New York: Springer, 1988).

[4] Tolmeijer, Suzanne, et al. "Second chance for a first impression? Trust development in intelligent system interaction." Proceedings of the 29th ACM Conference on user modeling, adaptation and personalization. 2021.

[5] Kaptchuk, G., Redmiles, E. M., & Cummings, R. (2020, October 21). People want data privacy but don't always know what they're getting. The Conversation. Retrieved August 24, 2022, from

https://theconversation.com/people-want-data-privacy-but-dont-always-know-what-theyre-getting-143782

[6] Fioretto, Ferdinando, et al. "Differential Privacy and Fairness in Decisions and Learning Tasks: A Survey." arXiv preprint arXiv:2202.08187 (2022).

[7] Gupta, D. (2021, December 20). Use of blockchain technology could increase human trust in AI. VentureBeat. Retrieved August 24, 2022, from https://venturebeat.com/datadecisionmakers/use-of-blockchain-technology-could-increase-human-trust-in-ai/