Safe and Responsible AI for Citizen Science

Kobi Gal

Ben-Gurion University, University of Edinburgh

1   Research Focus

My research has made key contributions to the design of safe and responsible AI-based support for citizen science participants.

AI Generated Feedback: Together with colleagues from MSR and University of Oxford, we studied AI-driven approaches for personalizing motivational messages to individual participants [2, 4, 3, 5]. We addressed two key problems of incentive design. Finding a good representation of participants' interactions in the environment; and using the representation to inform a target intervention policy for generating messages to volunteers. Our approach automated the process of designing an intervention strategy for citizen science; it was fully deployed on Galaxy Zoo.

Applying the learned intervention policy significantly increased the contributions of thousands of volunteers (by up to 70%) while minimizing disruption to the volunteers, and maintaining the quality of their work. The breakthrough achieved in this work was an end-to-end and domain-independent method for automatically computing an intervention policy for human-in-the-loop setting. Our approach was subsequently embedded within the Galaxy Zoo infrastructure, providing adaptive support for thousands of volunteers.

Intelligent Recommendations: Finding the best citizen science project that suits the preferences and capabilities of participants has become a major challenge and is essential for keeping volunteers motivated and active contributors [6]. To address this challenge we adapted several recommendation algorithms to citizen science based on memory-based and model-based collaborative filtering approaches. The algorithms were trained on historical data of users' interactions in the SciStarter platform -as well as their contributions to different projects.

The trained algorithms were evaluated in SciStarter and involved hundreds of users who were provided with personalized recommendations for new projects they had not contributed to before. Using the new recommendation system led people to increased participation in new SciStarter projects when compared to groups that were recommended projects using non-personalized recommendation approaches, and compared to behavior before recommendations.

In subsequent work we studied whether combining project recommendations with explanations improves users' contribution levels and satisfaction [1]. We generated post-hoc explanations to users by learning from their past interactions as well as project content (e.g., location, topics). The vast majority of scistarter users highly preferred receiving explanations about why projects were recommended to them, and receiving such explanations did not impede on the contribution levels of users, when compared to other users who received project recommendations without explanations. Our approach is now fully integrated in SciStarter.

2 Potential Research Questions for Workshop:

The development of generative AI technologies and machine learning provides a real opportunity to transform how citizen science benefits society. I mention a few possible areas in this respect.

- How to incorporate collaborative agent design in existing workflows of citizen science projects?

- How to promote collaborative learning in which participants work together in tasks and AI is used to guide and assist them?

- How to infer participants skills and expertise using AI?

- How to ensure that AI based support is generated in a safe and responsible way, does not hinder participants and does not bias or discriminate?

References

[1] Daniel Ben Zaken, Avi Segal, Darlene Cavalier, Guy Shani, and Kobi Gal. Generating recommendations with post-hoc explanations for citizen science. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 69–78, 2022.

[2] Ya'akov (Kobi) Gal, Avi Segal, Ece Kamar, Eric Horvitz, Chris Lintott, and Mike Walmsley. A new workflow for human-ai collaboration in citizen science. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, GoodIT '22, page 89–95, New York, NY, USA, 2022. Association for Computing Machinery.

[3] Avi Segal, Kobi Gal, Ece Kamar, Eric Horvitz, and Grant Miller. Optimizing interventions via offline policy evaluation: Studies in citizen science. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[4] Avi Segal, Ya'akov Kobi Gal, Ece Kamar, Eric Horvitz, Alex Bowyer, and Grant Miller. Intervention Strategies for Increasing Engagement in Crowdsourcing - Platform, Predictions, and Experiments. *International Joint Conference of AI (IJCAI), year = 2016.*

[5] Avi Segal, Ya'akov Kobi Gal, Robert J Simpson, Victoria Victoria Homsy, Mark Hartswood, Kevin R Page, and Marina Jirotka. Improving productivity in citizen science through controlled intervention. In *Proceedings of the 24th International Conference on World Wide Web*, pages 331–337. ACM, 2015.

[6] Daniel Ben Zaken, Kobi Gal, Guy Shani, Avi Segal, and Darlene Cavalier. Intelligent recommendations for citizen science. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14693–14701, 2021.