Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Doctoral Consortium Track

Human-AI Collaboration in Recruitment and Selection

Neil Natarajan
University of Oxford
neil.natarajan@new.ox.ac.uk

Abstract

My research focuses on using algorithmic systems alongside human collaborators to outperform either humans or machines individually. I specialize in human-machine collaboration in recruiting and selecting talented groups of people.

1 Research Questions

My research focuses on using AI systems alongside human collaborators to perform tasks better than either humans or AIs can alone. I specialize in the design and evaluation of explainable AI systems and the use of human-machine collaboration in recruiting and selecting talented groups of people. My main research questions are:

RQ1 What attitudes do recruitment and selection domain experts hold towards use of and collaboration with intelligent systems?

RQ2 How do these attitudes and the unique nature of the domain affect human-AI cooperation?

(a) Are there tasks within the domain best performed by humans? by machines? by a concert?

(b) Supposing that there are tasks within the domain best performed by machines, what barriers exist to machine application to those tasks?

RQ3 Which properties of systems and architectures enable cooperation, and which properties hinder cooperation?

(a) What is the effect of explainability in AI systems on cooperation in this context? How does this differ from the effect of explainability elsewhere?

(b) What is the effect of algorithmic visualizations of diversity on cooperation?

RQ4 How can findings about attitudes and properties inform innovation in the recruitment and selection domains?

2 Work to Date

My work to date has focused primarily on two projects that are related to the main research questions above. The first project is a study of attitudes towards xAI for human-AI cooperation, and primarily helps answer RQ1 and RQ3. The second project is a study of the effect of algorithmic visualizations of diversity on human-AI cooperation, and primarily helps answer RQ2.

2.1 Attitudes Towards XAI

There is a growing body of work on trust in AI systems (such as [Jacovi *et al.*, 2021] and [Vereschak *et al.*, 2021]), and even on trust in AI explainers [Jacobs *et al.*, 2021]. A common finding is that presenting explanations of AI outputs increases trust in these systems [McCradden, 2021]. However, little work has been done examining why this effect exists or how this effect varies across domains. In particular, little attention has been paid to the selection and recruitment domain.

Thus, I began work on this project by arguing that trust in AI explainers, much like trust in AI systems themselves, is desirable exactly when it is warranted. Extending [Jacovi *et al.*, 2021]'s work on trust in AI systems, I argue in [Natarajan *et al.*, 2022] that AI explainers should only be trusted to do things they claim to do. In a paper currently under review, I conduct two quantitative surveys and present analysis demonstrating two tasks on which AI explainers, when used inappropriately, generate unwarranted *over-trust* in faulty models.

2.2 Algorithmic Visualizations of Diversity

The benefits of diversity are well known in the recruitment and selection domain [Page *et al.*, 2017]. However, unless the best performing applicants to a program or job form the most diverse cohort, cohort diversity and applicant performance are in tension [Page *et al.*, 2017]. Thus, it is important to measure and visualize diversity in the recruitment and selection domain. However, measuring cohort diversity is nontrivial (see [Budescu and Budescu, 2012]), finding the most diverse cohort is in-general NP-hard [Nemhauser *et al.*, 1978], and visualizing tradeoffs between cohort performance and diversity is yet-unsolved (see [Huppenkothen *et al.*, 2020] and [Schumann *et al.*, 2019]).

Thus, I began work on this project by conducting a literature review of the benefits of diversity in the recruitment and selection domain. I then developed a framework by which we view tradeoffs between applicant performance and cohort diversity as a *Selection Possibility Frontier* (analogous to a production possibility frontier), and devised an algorithm to approximate points along this frontier. Finally, I am currently collaborating with Rise[1], a scholarship program for brilliant young people, using data from their first year of applications to model how consideration of the Selection Possibility Frontier could improve the diversity of their cohorts. In a paper currently under review, I introduce this frontier, prove the correctness of the algorithm, and present results from the Rise case study.

3 Future Work

I plan to continue my work on the two projects described above. Additionally, I plan to begin work on a third project, which will focus on the effects of algorithmic interventions on selector confidence in their own intuitions and decision-making, which will primarily help answer RQ3. Finally, I plan to begin work on a fourth mini-project, the aim of which will be to synthesize results from the previous three projects and to answer RQ4.

3.1 Attitudes Towards XAI

After completing work on my paper under review, I intend to conduct a field study of attitudes towards xAI in the selection domain, using programs I am currently partnering with as case studies. (As the partnerships are not yet public, I have omitted the names of these organizations.) I will conduct interviews with selectors to determine whether the results of my survey study apply across the selection domain.

I also intend to implement a generative AI model designed to summarize applicant essay information in a way that is both accurate and explainable. I will then conduct a field study to determine whether this model (and explanations of this model) can be used to improve selector decision-making.

3.2 Algorithmic Visualizations of Diversity

After completing work on my paper under review, I intend to conduct a series of more in-depth case studies measuring the potential impacts of algorithmic visualizations of diversity on selector decision-making. I will use the Rise case study and will expand to other case studies with partner programs.

In tandem, I intend to conduct a co-design workshop with selectors to determine how the algorithm drawing the Selection Possibility Frontier could be tooled so as to be most useful, then to develop a tool for this purpose.

3.3 Algorithmic Interventions on Selector Confidence

A common thread with research into human-machine cooperation is careful attention to the human decision-maker's trust in the machine system's outputs (see [Vereschak *et al.*, 2021]). However, less-measured and equally important is the decision-maker's trust in their own decisions. In this project, I intend to conduct a series of surveys and field studies to determine whether algorithmic interventions affect selector confidence in their own intuitions and decision-making, the direction of said effect, and whether said effect can be used to calibrate decision-maker trust, or should be controlled for.

Though road-mapped, many of the later phases of this project are subject to change pending earlier results. Initially, I intend to run a mixed-methods survey analyzing the effect of different algorithmic tools on selector confidence.

[1] see risefortheworld.org

3.4 Informing Innovation

Following the completion of all three projects, I intend to spend time diagramming and extracting key findings, then using these findings to yield recommendations informing innovation in the recruitment and selection domain.

References

[Budescu and Budescu, 2012] David V. Budescu and Mia Budescu. How to measure diversity when you must. *Psychological Methods*, 17(2):215–227, 2012.

[Huppenkothen *et al.*, 2020] Daniela Huppenkothen, Brian McFee, and Laura Norén. Entrofy your cohort: A transparent method for diverse cohort selection. *PLoS ONE*, 15(7):e0231939, Jul 2020.

[Jacobs *et al.*, 2021] Maia Jacobs, Melanie F. Pradier, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational Psychiatry*, 11(1):1–9, Feb 2021.

[Jacovi *et al.*, 2021] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 624–635, New York, NY, USA, March 2021. Association for Computing Machinery.

[McCradden, 2021] Melissa D. McCradden. When is accuracy off-target? *Translational Psychiatry*, 11(1):369, Jun 2021.

[Natarajan *et al.*, 2022] Neil Natarajan, Reuben Binns, Jun Zhao, and Nigel Shadbolt. Trust explanations to do what they say. In *Workshop on Human-Centered AI at NeurIPS 2022*, New York, NY, USA, December 2022. Association for Computing Machinery.

[Nemhauser *et al.*, 1978] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, Dec 1978.

[Page *et al.*, 2017] Scott E. Page, Earl Lewis, and Nancy Cantor. *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy*. Princeton University Press, Sep 2017.

[Schumann *et al.*, 2019] Candice Schumann, Samsara N. Counts, Jeffrey S. Foster, and John P. Dickerson. The diverse cohort selection problem. *arXiv:1709.03441 [cs]*, Mar 2019. arXiv: 1709.03441.

[Vereschak *et al.*, 2021] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. How to evaluate trust in ai-assisted decision making? a survey of empirical methodologies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):327:1–327:39, Oct 2021.