University of Pittsburgh | School of Computing and Information

Dissertation Defense

Doctor of Philosophy in Information Science

"**Countering Algorithmic Bias and Stereotypes with Human-centric Interactive Systems and Data-driven Approaches**" by **Yongsu Ahn**

Date: March 17, 2025
Time: 1:00 – 4:00 p.m.
Place: Room 828, Information Sciences Building, 135 N Bellefield Ave, Pittsburgh, PA 15260

Committee:

- (Chair) Dr. Yu-Ru Lin, Associate Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Peter Leonid Brusilovsky, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Rosta Farzan, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Dr. Adam Perer, Assistant Professor, School of Computer Science, Carnegie Mellon University

Abstract:

Artificial intelligence systems increasingly shape our daily lives through personalized recommendations and automated decisions. While these systems advance in capturing patterns from large-scale datasets, they often perpetuate bias and stereotypes, leading to various types of algorithmic effects, ranging from stereotype and discrimination against certain groups and individuals, to filter bubble or popularity bias in recommender systems. Despite growing concerns about the consequences of such algorithmic effects, existing literature still lacks systematic approaches to understand its nuanced impacts on users' algorithmic experiences and mitigate these effects in practice.

Through three interconnected studies, I propose a comprehensive approach for measuring, understanding, and mitigating algorithmic effects, focusing on recommender systems as an application of AI systems impacting many users' daily information use. First, I develop a unified quantitative framework that decomposes miscalibration into algorithmic bias and stereotype. This framework reveals how these effects are interconnected and disproportionately impact different user groups, particularly those with niche preferences or from underrepresented demographics. I introduce the notion of algorithmic stereotyping problems in recommender systems—an underexplored phenomenon of overgeneralizing individual user characteristics into typical preferences—highlighting its impacts on groups and individuals.

Second, I conduct a multi-staged qualitative study to investigate how algorithmic effects shape users' daily interactions with recommender systems. Using various types of activities such as interviews, mental model drawings, and card sorting exercises, I uncover users' folk theories about algorithmic effects, revealing a phenomenon of "algorithmic ambivalence" where algorithmic effects can be perceived as either benefit or harm depending on individual goals and contexts. Based on the findings, I develop RECalibrate, an add-on recommendation interface designed to enhance transparency in recommender systems. It allows users to see how recommended items are associated with certain algorithmic behaviors and effects (e.g., favoring the majority preference or popular items) and provide tools to adjust these factors, empowering users to personalize their experience and mitigate potential biases. This interface moves beyond simple item-level controls to allow users to steer their feeds towards their personal recommendation goals (e.g., personalization or diversity) and valuation over

algorithmic effects with visual summary and user controls at topic and feed level. The proposed interface demonstrates significant improvements in both usefulness and usability, for providing an intuitive summary of the feed and better aligning user perceptions with algorithmic outcomes. Compared to the baseline interface, it offers a more transparent user experience, allowing for better decision-making and user control. The study indicates that an AI interface should be designed to support alignment between human values and intent and algorithms' inner workings, based on deeper understanding of users' experiences over algorithmic platforms and content.

This dissertation contributes to the broader understanding of algorithmic effects in recommender systems while offering practical solutions for enhancing user agency in algorithmic systems. The findings emphasize the importance of considering both systematic patterns of bias and individual user perspectives in designing more equitable and user-centered recommender systems. Overall, this dissertation bridges the gap between the data-driven examination and human-centric understandings of algorithms' impacts, highlighting the importance of a more comprehensive and nuanced approach to analyzing these effects.