

CAREER: Understanding and Accounting for Human Behavior in Machine-in-the-Loop Decision Making

Chien-Ju Ho, Washington University in St. Louis

Overview

Machine learning (ML) has been increasingly involved in decision making in our daily life. While fully automated decision making is the goal for tasks in some domains, such as autonomous driving, in many other tasks, we should not, or do not want to, delegate decision making entirely to ML. For example, when the stake is high and the objectives of the task are hard to be accurately specified, such as in government policy making or military applications, fully automatic decision making using ML may lead to suboptimal outcomes and has not achieved the level of being fully trusted. In these cases, humans are often brought into the loop to make the final call on what decisions to take. In addition, for tasks that involve human preferences or enjoyments, such as in deciding which restaurants to go to or which destinations to travel to, while ML might be able to provide useful information or recommendations, humans are still naturally the final decision makers in these tasks.

These considerations lead to a new paradigm of *decision making with machines in the loop*, where ML provides information to humans, who can then incorporate the information to make the final decisions. While this decision-making paradigm presents significant promises, there are also challenges. In particular, humans are known to exhibit behavioral biases in making decisions. How do we take into account of human behavior in designing ML algorithms? Moreover, there might be multiple objectives to balance and trade-off during decision making, how should we design ML that balance the objectives such that it aligns with human values? In this research proposal, I plan to investigate machine-in-the-loop decision making, with a focus on studying the *information exchange* between ML and humans. The goal is to utilize machine learning to assist humans in making better decisions while taking into account human behavior and complex objectives. More specifically, I will investigate research questions in the following three directions: (1) Developing optimization frameworks for information design in machine-in-the-loop decision making. (2) Understanding and modeling humans in machine-aided decision making. (3) Aligning the objectives of humans and machines by including humans in the loop.

Intellectual Merit

This proposed research will contribute to the empirical understanding of human behavior in ML-assisted decision making. It will also provide theoretical foundations for studying the interactions of humans and learning algorithms, through incorporating human models in learning frameworks. The results of the proposal will provide insights on developing human-centered machine learning algorithms and in combining humans and machines to solve problems neither can solve alone. This research is interdisciplinary in nature, combining ideas and techniques from machine learning, algorithmic economics, and online behavioral social science.

Broader Impacts

This research has a direct impact on the design of a broad range of online platforms with active human participation. Moreover, it also contributes to improve policy making for societal issues. In particular, the PI has existing collaborations with domain experts in the Brown School of Social Work and Medical School that apply computational approaches to practical problems such as allocation of scarce resources for homeless prevention and living donor kidney transplantation. The PI plans to continue and expand the collaborations through the Center for Collaborative Human-AI Learning and Operation (HALO) and the Division for Computational and Data Sciences (DCDS) at the Washington University in Saint Louis to apply the developed machine-in-the-loop decision making framework to address societal issues. This proposal also includes a comprehensive plan for enhancing the education and broadening the research outreach, including creating a course in human-AI collaborations, engaging undergraduate research, and hosting summer workshops for high-school students.