CAREER: Behavior-Informed Machine Learning: Improving Robust Learning and Decision Support

1 Introduction

Machine learning (ML) is increasingly integrating into various aspects of everyday human life. Central to this evolution is ML's ability to *learn from humans*. By learning from human demonstrations, machines have been able to mimic, and even surpass, human capabilities across a range of domains [47, 132, 105, 15, 126]. Furthermore, as ML technology continues to advance, it offers immense potential to *augment human decision-making* [22, 12, 96, 21], fostering more informed choices and improved outcomes. While this two-way interaction between humans and ML holds transformative potential, it also introduces challenges to account for the role of humans in the design of ML systems.

Existing approaches to incorporating humans in ML often either treat humans as independent, stochastic data sources [27, 97, 24, 59, 131, 25, 137, 23] or assume humans are *rational* decision-makers [127, 19, 18, 46, 66, 6]. While these assumptions offer elegant and simple formulations, they deviate from real human behavior, as extensively documented in the psychology literature [122, 63, 123, 62, 65, 64]. Failing to account for these deviations in the design of ML systems can lead to unintended negative consequences. For instance, autonomous vehicles designed by learning from human driving behavior could adopt dangerous patterns from aggressive or unsafe drivers [30]. In healthcare, ML models trained on data generated by doctor annotations might recommend unnecessary treatments due to doctors' action bias in treating diseases even when the best course of action might be to wait and see [37]. Social media platforms using ML models might create echo chambers due to the confirmation bias in user behavior [10].

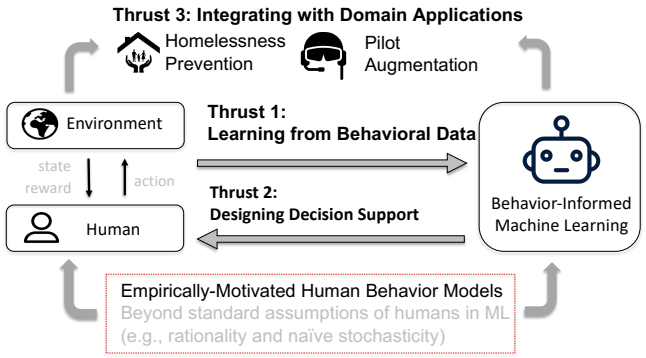


Figure 1: CAREER research plan. [CJ: to update]

This CAREER project proposes the development of a framework for *behavior-informed machine learning*, which examines and incorporates the impacts of human behavior into the design of machine learning systems. Specifically, we will focus on two key aspects of human behavior in the ML lifecycle: (1) the generation of data used for training machine learning models, and (2) human decision-making in tandem with machine assistance. The proposed research aims to devise ML systems resilient to biased training data, and capable of enhancing human decision-making for improved outcomes. In addition to theoretical contributions, through collaboration with domain experts, the research will be adapted to domain applications such as homelessness prevention and pilot augmentation. This ensures that the research findings will have practical relevance in domain applications, promoting their widespread adoption and impact. In more detail, as illustrated in Figure 1, we will investigate the following three research thrusts:

- **Thrust 1: Developing foundations for learning from human behavioral data.**
  Learning from human supervisions and demonstrations is widely studied in weakly supervised learning [13, 86, 103], truth inference in crowdsourcing [27, 97, 24, 59, 131, 25, 137, 23], and inverse reinforcement learning [87, 138, 94]. In these studies, humans are assumed to be either independent, stochastic data sources or to be rational decision makers. These assumptions enable inference techniques to uncover the ground truth from human demonstrations. However, these assumptions often fail to reflect actual human behavior, and failing to account for it could lead to suboptimal learning. This thrust aims to develop computationally practical, theoretically sound, and empirically grounded foundations for learning from behavioral data, explicitly accounting for human behavior in data generation.

- **Thrust 2: Designing assistive ML to improve human decision making.**
  As ML begins to outperform humans in certain areas, it is becoming more important than ever to explore whether and how ML can be leveraged to assist humans in making better decisions. In this thrust, we aim to develop assistive ML frameworks to enhance human decision making that take into account human biases and preferences. In particular, we will investigate approaches to determine when and what assistance ML should should provide through algorithmic, data-driven, and learning approaches. Furthermore, we will conduct behavioral experiments to account for human trust and reliance on ML advice that will in turn impact the design of the assistive ML framework.
- **Thrust 3: Integrating with domain applications.**
  While the main focus of this CAREER plan is to develop a general framework for behavior-informed ML, we will also collaborate with domain experts to tackle practical challenges in deploying this framework in domain applications. Specifically, the proposed research will be adapted for use in the domains of homelessness prevention (in collaboration with Prof. Patrick Fowler) and flight pilot augmentation (with Boeing). This approach ensures that our research findings are robust and practically applicable in domain applications, promoting their widespread adoption and potential for impact.

Long-term Goal. My career goal is to develop the foundations for humans and ML to collaborate together and solve problems neither can solve alone. This requires the advancements of machine learning, the understanding of humans, and the utilization of their interactions. This research proposal serves as the stepping stone to achieving this goal by developing behavioral-informed machine learning, designing learning algorithms that are robust to human behavior during data generation, and investigating how to design machine learning algorithms to assist humans in making better decisions.

Intellectual Merit. This proposed research will contribute to the empirical understanding of human behavior when interacting with ML and provide theoretical foundations for studying the human-machine interactions. 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.

PI Qualifications. The PI has extensive research experience in studying the interactions between humans and ML, using techniques drawn from machine learning, algorithmic economics, optimization, and online behavioral social science. From the perspective of learning from humans, the PI has explored the problem of eliciting and learning from noisy human-generated data [48, 51, 5, 55, 49, 115, 113, 33, 34] and designing incentives to encourage high-quality data [50, 52, 56, 76]. From the perspective of designing ML to assisting humans, the PI's recent works explored the design of when and what assistance to provide to humans using techniques from information design and environment design [134, 114, 40, 29] and investigating ethical considerations in leveraging ML in decision making [116, 118, 84, 85]. In addition to the theoretical and algorithmic studies, the PI has experiences in conducting large-scale online behavioral experiments to understand human behavior in computational environments [53, 115, 33, 34, 118, 134, 84, 85]. The PI is active in the research communities. The PI served as the Doctoral Consortium Co-Chair and Works-in-Progress and Demonstration Co-Chair of HCOMP (in 2022 and 2019, respectively), the premier conference in the study of human computation. The PI has also organized workshops at NeurIPS and HCOMP to explore the interactions between humans and machine learning, and served as the area chair, senior program committee, and program committee in major AI/ML conferences.

2 Background

We begin with a brief introduction to the classical decision making frameworks in ML, followed by a summary of well-known human behavioral models motivated from behavioral economics and psychology.

2.1 Decision Making Framework in Machine Learning

We first review the decision-making framework that serves as the foundation of the proposed work. Note that in this line of literature, the decision maker is assumed to the rational, and the goal of ML development is to either solve the optimal policy or infer the environment from (near-)optimal demonstrations.

Markov decision process (MDP) and reinforcement learning (RL). Markov decision process (MDP) is one of the most standard frameworks for modeling the sequential decision-making environment. An MDP can be characterized by the tuple $\langle S, A, T, R \rangle$, where

- State space S: characterizes the environment a sequential decision maker is interacting with.
- Action space A: actions the decision maker can chose from at each step.
- State transition function $T(s'|s,a)$: characterizes how decision maker's actions change the environment.
- Reward function $R_a(s,s')$: describes the benefits of taking each action.

The standard approach to solve the above MDP and obtain an optimal policy is through reinforcement learning (RL) [60, 112, 80, 81]. In the standard setup, the RL agent interacts with an unknown environment and attempts to maximize the total of its collected reward. At each time t, the agent in state $s_t \in \mathcal{S}$ takes an action $a_t \in A$, which returns a reward $R_{a_t}(s_t, s_{t+1})$, and leads to the next state $s_{t+1} \in S$ according to a transition probability kernel $T(s'|s,a)$, encoding the probability to state s' from s after taking action a. The goal of RL is to learn a policy $\pi(a|s)$ that maximizes the total time-discounted rewards $\mathbb{E}_\pi[\sum_t \gamma^t R_{a_t}(s_t, s_{t+1})|\pi]$, where $\gamma \in (0,1]$ is a discount factor ($\gamma = 1$ indicates an undiscounted MDP). RL has a long history of development, from the seminal Q-learning [129], to more recent deep learning aided approaches [73, 80, 81].

Inverse reinforcement learning (IRL). Inverse reinforcement learning tackles a challenging task of inferring the reward R from observing the sequence of (s_t, a_t)s. This problem has also been referred to as apprenticeship learning, or learning by watching, imitation learning etc. Ng et al. [87] is among the first to formalize this problem. They characterize the set of reward functions that would produce the same optimal policy as observed. The high-level idea is to find a feasible function $R(\cdot)$ such that a_t is the action maximizes the utility at s_t for all (s_t, a_t) pairs. Then the authors imposed smoothness constraints on each step's predicted policy to formulate a linear programming problem to solve. Follow-up works [4, 138, 94] have focused on variants of the optimization formulation. The common assumption in IRL is that the demonstrations (s_t, a_t)s are from unbiased and optimal decision makers.

2.2 Empirically Motivated Human Behavioral Models

Existing approaches in modeling humans in ML frameworks mostly fall into two categories: (1) modeling humans as independent, stochastic data sources [27, 97, 24, 59, 131, 25, 137, 23], or (2) assuming humans are *rational* decision makers, taking actions that maximize their expected utility [127, 19, 18, 46, 66, 6]. While these models provide elegant and simple formulations, they do not always capture human behavior empirically observed in the field. In this CAREER plan, we aim to incorporate empirically grounded human models into the design of ML. To make the discussion more concrete, we summarize two important classes of human behavioral models in the literature of economics and psychology and provide formulations.

Time-inconsistent planning. Humans often cannot reason about future rewards in a consistent manner. For example, humans are rational and might not be able to reason future rewards due to cognitive and information limitations. Humans might also inherit time-inconsistent reasoning behavior. For example, when choosing between earning 10 dollars 100 days from now or 12 dollars 101 days from now, most people will choose the second option. However, when being asked to choose between earning 10 dollars now or 12 dollars tomorrow, many people will change their decisions and choose 10 dollars now. This example illustrates *present bias* [92], describing humans' tendency to give stronger weights on immediate costs and benefits rather than balancing them against costs and benefits in the future. These biases in time-inconsistent reasoning can be modeled by introducing a *discounting function* $d(t)$ that captures humans'

behavior in weighing future rewards. Let R_t denote the expected reward at time t, human's perceptions of the long-term rewards can be modeled as $\sum_t d(t)R_t$. This notion characterizes many behavioral models related to time-inconsistent planning, with some illustrative examples below:

- Standard model: $d(t) = \gamma^t$
- Bounded rationality: $d(t) = \gamma^t$ for all $0 \leq t \leq \tau$, and $d(t) = 0$ for all $t > \tau$.
- Present bias: One common model is hyperbolic discounting: $d(t) = \frac{1}{1+kt}$ for some pre-specified $k > 0$.

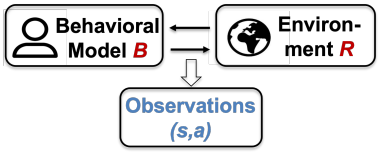
Biased reward evaluation. While it is commonly assumed that humans are rational, taking actions to maximize their expected utility (the expected utility theory [127]), humans are consistent observed to deviate from the assumption. For example, humans often over-estimate small probabilities and react more strongly to losses than gains. The most important theory that summarizes these systematic biases is the Nobel-winning *prospect theory* by Kahneman and Tversky [61]. Another commonly used theory, also Nobel-winning, is the discrete choice model [79, 108, 120], which accounts for the inherent randomness of human decision making by incorporating noises in the utility. These deviations from standard rational assumption can often be captured with humans' biased reward evaluations. Formally, let $(p_1, x_1, ..., p_K, x_K)$ be the *prospect* of an action, where p_k represents the probability of the outcome x_k happens after taking the action. Let $v(x_k)$ represent the utility of the outcome x_k. The above theories can be summarized below:

- Expected utility theory: it predicts that humans will take the action that maximizes $\sum_{k=1}^K p_k v(x_k)$.
- Prospect theory: it predicts that humans will take the action that maximizes $\sum_{k=1}^K \pi(p_k)u(v(x_k))$, where $\pi(\cdot)$ and $u(\cdot)$ models the humans' distorted interpretations on the probability and utility measure.
- Discrete choice model: It predicts that humans will take the action that maximizes $\sum_{k=1}^K p_k v(x_k) + \epsilon$, where ϵ is the additional noise term that incorporates the intrinsic randomness of human decision making.

3 Proposed Research

The proposed research aims to explore the human-machine partnership by focusing on integrating human behavior in the design of machine learning. We plan to develop theoretically sound and empirically grounded foundations for behavior-informed machine learning that both learns from humans and assists them. Moreover, through collaboration with domain experts, we will tackle the practical challenges associated with deploying behavior-informed ML in domain applications.

3.1 Thrust 1: Developing Foundations for Learning from Behavioral Data



This thrust aims to develop computationally practical, theoretically sound, and empirically grounded foundations that infer the latent environment parameters (e.g., MDP rewards or feature-label mappings) and human models from human behavioral data. Take RL for example, when humans are rational decision makers, they follow a policy π that maximizes $\mathbb{E}[\sum_t \gamma^t R_{a_t}(s_t, s_{t+1})|\pi]$. Conditional on this assumption, the standard learning framework (e.g., inverse reinforcement learning) aims to infer the underlying rewards R or recover the optimal policy π from observing the realized state-action pairs $\{(s_t, a_t)\}$.[1] However, humans often do not make decisions according to the optimal policy. Instead, they might follow some behavior models $\mathcal{B}$, e.g., $\mathcal{B}$ could include time-inconsistent planning $\mathcal{B}_T(t)$ and biased reward evaluation $\mathcal{B}_{R,a}(s, s')$, as reviewed in Section 2.2. Their goal is to maximize the *biased* expected rewards: $\mathbb{E}[\sum_t \mathcal{B}_T(t)\mathcal{B}_{R,a_t}(s_t, s_{t+1})|\pi]$. The aim of the algorithms that learns from behavioral data is to infer $\mathcal{B}_T, \mathcal{B}_R, R$ while only observing potentially biased (s_t, a_t)s.

More generally, in this learning setup, the observations (s_t, a_t) are generated from the *interactions* be-

[1] The discussion in this thrust also applies to (a simpler setting of) supervised learning, where we observe the feature-label pairs generated by humans $\{(x_n, y_n)\}_{n=1}^N$ and aim to undercover the latent mapping from features to labels.

tween humans $\mathcal{B}$ and the environment R.[2] This interactive nature raises challenges: Computationally, we have to grapple with a significantly larger space for the learning problem. Theoretically, there could exist scenarios where learning might prove unfeasible. For instance, if human decision-makers always choose actions with the highest empirical rewards, the resulted dataset is generated with *pure exploitation*, where humans overlook potentially high-utility actions that initially provide low rewards due to chance. Additionally, humans may attribute higher perceived utility to options that are chosen more frequently (e.g., herding bias [11, 101, 82, 107]), creating feedback loops that concentrate on popular choices. These scenarios might produce datasets that fail to offer a representative distribution to enabling learning. In fact, my previous work [113] has demonstrated that in certain simple stylized human models, learning can be infeasible even with infinitely many data observations. To tackle these challenges, this thrust will develop computationally efficient algorithms for jointly inferring $\mathcal{B}$ and R from behavioral data (**Task 1.1**). We will also construct theoretical foundations that characterize the feasibility of learning (**Task 1.2**). Lastly, in close collaboration with psychology researchers, we plan to conduct human-subject experiments to further our understanding of human behavior in the context of pervasive ML integration in everyday decision-making (**Task 1.3**).

Prior work. The proposed activities in this research thrust will be built on the PI's extensive prior work in crowdsourcing [48, 51, 52, 54, 55, 76, 115, 33, 34], where one key research theme is to infer ground truths from noisy human data. We will extend the standard models that assume humans exhibit some zero-mean noises to general human behavioral models. The PI's recent works on incorporating behavioral models motivated by psychology literature in the learning frameworks [113, 134, 40] and the experience in conducting human-subject experiments [54, 115, 33, 34, 85] will be the building blocks of the proposed research.

3.1.1 Task 1.1: Developing practical algorithms to learn from behavioral data

Our first task is to develop computationally practical algorithms to learn from human behavioral data. The proposed algorithms will be evaluated with both theoretical analysis (for a smaller set of general conditions) and simulations (for a wider range of conditions) to examine their accuracy (of estimating environment/behavioral parameters) and computational efficiency under different settings of human models $\mathcal{B}$ and environments R. Below we briefly describe our proposed approaches.

Two-stage learning. We will start with an easier setting: assume partial access to true environment parameters R_t (e.g., assuming some rewards are known in MDP), obtained through domain knowledge or historical information. With this assumption, we can leverage supervised learning techniques to first infer $\mathcal{B}$, utilizing provided R_t and observed $\{(s_t, a_T)\}$, and then infer R, by utilizing inferred $\mathcal{B}$ and $\{(s_t, a_t)\}$. This two-stage learning could significantly reduce the problem space and induce efficient learning.

Imposing constraints. The requirement of "ground truth" rewards is strong in practice. We will also investigate methods of learning $R, \mathcal{B}$ without accessing any of the true R_t. To address the computational issue, we will investigate methods of imposing proper constraints. For example, by leveraging the idea of my prior work in dealing with bandits with infinite search space [52], if we impose a mild condition that for two *similar* states s_t and $s_{t'}$, their rewards are also *similar*, we would be able to significantly reduce the search space and improve learning. In addition to this approach, we will also explore imposing constraints on the belief models and reward bias models, e.g., based on domain knowledge, to reduce the search space.

Sampling-based inference approach: Consider the Bayesian inference framework[16, 26, 124, 75]. Let θ be the parameters that represent $\mathcal{B}$, and $\mathcal{H}_t$ be the information set up to time t, we can formulate the inference problem by imposing a certain prior and Bayesian structure between R, θ and (s, a). We can then solve the inference problem: $\arg\max_{R,\theta} \log \mathbb{P}(\{(s_n, a_n)\})_{n=1}^{t} | R, \theta)$. The inference is often computationally heavy due to continuous and large parameter space. We will resort to sampling and variational approaches to solve this problem. For instance, we can adopt Gibbs sampling. More specifically, according to Bayes' theorem,

[2] We abuse the notation R in a broad sense to represent general environment parameters, not just rewards.

the conditional distribution of s_n, a_n satisfies $\mathbb{P}(s_n, a_n | R, \theta, \mathcal{H}_t) \propto \frac{\mathbb{P}(R,\theta|\mathcal{H}_{t+1})}{\mathbb{P}(R,\theta|\mathcal{H}_t)}$. Leveraging this, we generate samples via tracking the posterior distribution of the parameters $\mathbb{P}(R, \theta|\mathcal{H}_t)$, and compare the losses.

3.1.2 Task 1.2: Developing theories for the feasibility of learning from behavioral data

In learning from behavioral data, the dataset is generated according to human behavior. This creates concerns that the data distribution could be under-represented or biased, making learning challenging or even infeasible. My prior work [113] has proven that with certain human models, it's impossible to uncover the underlying latent environmental or human parameters even with an infinite number of data points. This observation highlights the need for a rigorous theoretical understanding for this learning setting. In this task, we aim to develop theoretical approaches to identify the conditions under which learning from behavioral data becomes feasible. Moreover, informed by these theoretical developments, we will investigate strategies (e.g., taking interventions or increasing diversity) during data generation to enable more efficient learning.

Investigate the feasibility of learning. To explore the feasibility of learning, we propose to utilize techniques from stochastic approximation [99, 43]. The main idea is to frame the state's realizations over time as a random variable, based on the interaction between the behavioral model and the environment. As an illustration intution of this approach: if a bijection exists between a pair $(\mathcal{B}, R)$ and the state at convergence (a weaker condition might suffice), we can infer $(\mathcal{B}, R)$ from converged state, and it implies the feasibility of learning. Therefore, by examining the convergence and convergence rate of the state trajectory, we can characterize the conditions for the feasibility and complexity of learning. My previous work [113] adopted this approach on a simpler setting with two specific human models and a bandit learning environment. The *state* was modeled as the empirical rewards received for each action. Humans were assumed to select the next action based on the state and their behavioral patterns. For one behavioral model, we demonstrated that the state converges to a fixed point, and we used the convergence rate to confirm the upper bound of learning efficiency. For another behavioral model, we showed that the state converges to a random variable with non-zero variance and used information theoretical arguments to prove that learning is unfeasible. In this task, we aim obtain general results beyond specific human models and environments. The goal is to characterize the conditions of human models and environments that make learning feasible.

Taking interventions and increasing diversity during data collection to improve learning. One of the main causes that leads to infeasible learning in our setting is that humans tend to perform *exploitation* during decision making, leading to potentially under-represented datasets that makes learning infeasible. Based on this observation, we plan to examine approaches during the data collection process to increase the amount of *explorations* (i.e., taking potential suboptimal actions to acquire information). The first approach is to take interventions, utilizing ideas from incentivizing exploration [42, 78] to provide incentives to motivate human decision makers to take explorations. As a first step, we will take an *epsilon-first* approach [121], assuming the data collection starts with random exploration then with user-guided exploitation. We'll quantify the minimum amount of explorations required to make learning feasible. We will then examine whether we can reduce the amount of explorations by designing when to explore. The second approach is to examine the connection between the diversity of humans in the data (e.g., in terms of their behavioral model $\mathcal{B}$) and the feasibility of learning. Some recent literature [93] has suggested that inherent diversity in the data could makes exploration unnecessary in bandit learning. We plan also to extend the results in our setting, examining whether and how much increasing diversity in our dataset could enable more efficient learning.

3.1.3 Task 1.3: Conducting experiments to understand human behavior in the ML age

This task focuses on examining real-world human behavior through human-subject experiments. In addition to validating human models and evaluating our approaches, the primary goal of this task is to enhance our understanding of human behavior in settings where ML is pervasively embedded in the decision-making process. In the previous two tasks, we start our investigation by leveraging existing behavioral models

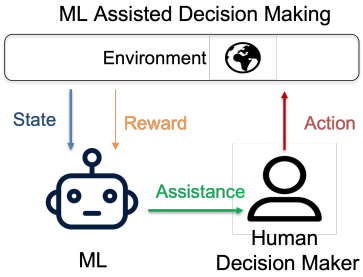
from the literature of psychology (e.g., see Section 2.2). Although these models are supported by extensive empirical evidence, they are primarily developed before ML becomes pervasive and widely recognized by the general population. Meanwhile, when people become aware that they are interacting with ML, their behavior might differ. My recent work [71] has demonstrated this phenomenon. We show that when people know their behavior will be used to train ML, they are willing to forgo rewards to ensure that the trained ML exhibits fair behavior. As ML continues to gain more societal attention, it becomes crucial to examine and understand the shifts in human behavior as ML becomes integrated into our daily lives.

Proposed research. We will conduct behavioral experiments to examine whether and how the presence of ML changes human behavior. The results will improve our understanding of human behavior with the presence of ML. It also serves as an improved foundation for addressing the tasks of learning from human behavioral data. To conduct the research, following the standard literature, we will start by utilizing social games, such as the ultimatum game [88], dictator game [39], prisoner's dilemma [8], to examine human behavior with the presence of ML. These social games provide succinct abstractions of human behavior and interactions in different contexts and are useful as the starting point towards a comprehensive understanding of humans. We will recruit participants from crowdsourcing platforms, e.g., Amazon Mechanical Turk or Prolific. In particular, we plan to vary the following independent variables in our experiment design and measure human responses as the dependent variables. Standard statistical tests (such as ANOVA and postdoc t-tests) will be conducted to examine the significance of the observations.

- Whether humans are explicitly interacting with ML. We hypothesis that humans are more likely to care more about ethics (e.g., being fair) when their partners in the game are other humans than ML.
- Whether human decisions will be used to train ML used to play with future players. We hypothesize humans are willing to sacrifice rewards to make the future ML behave in a more *ethical* manner.
- The context of the game, environment, and ML. For example, whether the trained ML will be playing with people they view favorably in the future. Whether the ML training mechanism is known to people.

For the research activities in this task, we will collaborate with Dr. Wouter Kool in the department of Psychology and Brain Sciences at WashU. Dr. Kool and I are currently co-advising a PhD student, Lauren Treiman, with whom we have generated the preliminary result [71] for this task (i.e., varying conditions on ML training in the ultimatum game). The proposed research will enable us to obtain a more comprehensive understanding of human behavior when ML is integrated in all aspects of decision making.

3.2 Thrust 2: Designing Behavior-Aware Assistive ML to Improve Human Decision-Making



Humans often make suboptimal decisions, especially in complex decision-making environments, and often need to engage in "on-the-job-training," i.e., learn to make better decisions while making these decisions [95, 109, 12]. On the other hand, the rapid development of ML suggests its potential of enhancing humans' performance and speeding up their learning in decision making settings with ML assistance. In this research thrust, our objective is to develop an algorithmic framework for ML-assistive decision making that takes into account human behavior. In this framework, ML provides recommendations to humans, who then make the final decisions. Here, the goal of machine learning is to *augment*, instead of *replacing*, humans in decision making.

We will initially focus on designing ML assistance in *low-complexity* environments where the optimal decision policy is algorithmically derivable using standard methods (such as value iteration). However, since the design of the ML assistance must consider the behavior of human decision maker, designing ML assistance leads to a more complex optimization problem. We will first presume known human models and propose algorithms to identify the optimal assistance policy (**Task 2.1**). We will then relax the low-complexity environment and known human model assumptions, proposing data-driven approaches for *high-*

complexity environments and developing algorithms to infer human models when unknown (**Task 2.2**). Finally, we will conduct human-subject experiments to understand the conditions for ML recommendation adoption, broadening our practical applicability (**Task 2.3**).

Prior work. The proposed activities in this research thrust are grounded in the PI's extensive prior work. Notably, the problem design aligns with a *Stackelberg game*, where ML initially determines the policy for providing assistance, and humans subsequently decide their course of action based on this assistance. The PI has explored the application of Stackelberg games across various domains, such as contract design [56], learning with strategic responses [119], Bayesian persuasion [29, 118, 40], and environment design [134]. In addition, the PI has substantial expertise in bandit learning [56, 76, 113, 118] and robust learning [119], which serve as technical foundations for addressing the problems of learning and robust design.

3.2.1 Task 2.1: Developing efficient algorithms for designing ML assistance

There have been recent works in leveraging ML to help humans make decisions [134, 12, 22], which only address settings with specific human models and settings. In this task, we aim to relax these assumptions and provide a general framework for designing ML assistance. In this task, we will focus on low-complexity environments and assuming known human models. Note that while solving the optimal policy in MDP is often feasible in low-complexity environments, my prior work [134] has proven that optimizing ML assistance while incorporating humans decisions is NP-hard in general. We aim to identify conditions that efficient algorithms are feasible, and propose the corresponding algorithms. We will investigate the relaxation of the assumptions of low-complexity environments and known human models in the next task.

The ML assistance framework. Suppose the human decision maker is solving a decision making problem formulated as an MDP, characterized by the tuple $\langle S, A, T, R\rangle$. Let the human decision-making policy be $\pi_{\mathcal{B}}(a|s)$, representing the probability for the human to choose action a at state s. The goal of ML is to maximize the total rewards derived from human actions by providing assistance. Let ML's assistance policy be $\rho(a|s)$, denoting the intervention ML makes at state s [3], and $\theta(s)$ be the *reliance policy* denoting whether the human adopt ML assistance at state s. We will start by assuming human reliance policy θ is known and given in task 2.1 and 2.2. We will examine θ in task 2.3. Now let $(\pi_{\mathcal{B}} \oplus \rho \oplus \theta)(a|s)$ be the final executed policy with ML assistance, e.g., $(\pi_{\mathcal{B}} \oplus \rho \oplus \theta)(a|s) = (1-\theta(s))\pi_{\mathcal{B}}(a|s) + \theta(s)\rho(a|s)$, the ML's assistance design problem is then to choose the assistance policy $\rho(a|s)$ to maximize the total expected reward within the pre-defined constraints of ML assistance policy. One natural example of the constraint would be to ensure ML does not intervene human decision-makers too much, i.e., the distance $D(\pi_{\mathcal{B}}, \rho)$ between human policy and ML policy is small for some distance measure D.

Designing the ML assistance policy $\rho(a|s)$ is useful in two perspectives. First, when deploying this assistance policy during human decision making, it augment human decision making and leads to better decision outcomes. Second, the difference between assistance policy and human policy represents the consequential mistakes humans often make in decision-making process, and therefore it has the potential to serve as the training materials and also has implication in education (See discussion in Section 4.3).

Proposed approaches. Consider the standard assumption that humans are rational, i.e., they choose the action with the highest expected utility with probability 1. When putting this decision function back to the optimization problem, the objective is non-continuous and the optimization is NP-hard to solve. On the other hand, when we consider the discrete choice model, i.e., the decision function is in the form of a continuous softmax function. With this human model, the objective of the optimization problem is continuously differentiable, and first-order optimization techniques might be applied. In fact, the PI's recent work [134] has demonstrated that it is indeed possible to derive computationally efficient algorithms with this particular human model. The above discussion highlights the need to understand how different human models impact the design of ML assistance. In this task we will address the following research questions:

[3] When $\rho(a|s) = \pi_{\mathcal{B}}(a|s)$ for all a, it means there is no ML assistance in state s.

- *Designing ML assistance with differentiable human models*: When the human decision model follows the discrete choice model or other models that lead to stochastic decision making, the optimization objective can usually be written as a continuous differentiable function. This enables the first-order optimization methods, such as gradient descent [100], to be applied. In this type of problems, we plan to characterize the computational complexity and convergence to the optimal solution with different human models. The key element is to quantify the *smoothness* of human models, i.e., how much human behavior changes with a small change of provided information. My prior work on the convergence rate of secure convex optimization [117] will serve as the technical foundation for this problem.
- *Designing ML assistance with non-differentiable human models*: When the human models are not differentiable (e.g, when the decision model follows the expected utility theory), the objective of the optimization problem will be non-differentiable, and standard first-order methods cannot be applied. We plan to utilize the techniques from algorithmic information design [35, 38, 9] (including the PI's own works [29, 40]) to characterize the solution. On a high-level, this approach involves utilizing the duality theory to characterize the properties of the optimal solution, which could help identify conditions for computationally feasible solutions to exist. We will also utilize techniques such as soft-max relaxations to derive approximation algorithms in the case that when the optimization is NP-hard.

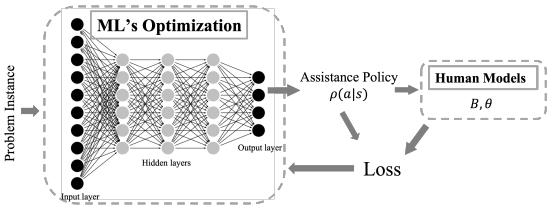
3.2.2 Task 2.2: Designing ML assistance with data-driven and learning approaches

In the above task, we focus on low-complexity environments and have assumed full knowledge of human models. While these assumptions could be approximately satisfied in simple domains and when we have access to an abundant amount of human behavior data, it is generally a strong assumption that might not hold in practice. In this task, we aim to move towards relaxing the above assumptions and take data-driven and learning approaches to optimize the ML assistance.

Background and challenges. To design ML assistance in high-complexity environments, one potential approach is to extend the idea of data-driven approaches of self-play [105, 28, 135], which learns the optimal decision policy of MDP through simulations. However, it is not trivial to incorporate human models in our setup to implement this approach. Moreover, when the human models are not know a priori, the problem becomes even more challenging as it is hard to estimate the policy performance using simulations. In this case, we will consider settings when the ML can interact with humans to obtain information over time and adaptively update policy based on what ML has learned. This naturally leads to the trade-off between exploration, taking potentially suboptimal actions to obtain information, and exploitation, taking optimal actions based on the available information, that need to be addressed.

Data-driven approaches to design ML assistance. We propose to leverage data-driven approaches to design ML assistance, assuming human models are known. In particular, we will extend the data-driven approach of self-play of optimizing decision policy in MDP to incorporate human models to optimize ML assistance. Specifically, we propose a neural-network-based structure that consists of two modules: the ML's optimization module and the human models. The ML's optimization follows the traditional neural-network-based structure, taking the details of an problem instance as input and outputs an assistance policy. The main difference compared with prior works is that we have incorporated human models, either in an analytical closed-form or a data-driven form in the optimization structure. The human models are treated as a black box for the ML's optimization module and is fixed before we begin training.

Given the assistance policy output by the ML's optimization module, we can compute the *loss* (the inverse of the reward of applying the assistance policy) by applying the assistance policy with the human models in the environment. For optimizing the assistance policy, we can follow the standard approach of using deep learning for



optimization [36, 90]: draw problem instances from a pre-specified distribution and perform stochastic gradient descent to minimize the loss function (applying soft-max approximation when the objective is not differentiable). We will examine the empirical performance of this neural-network based approach with different settings of human models, environments, instance distributions.

Bandit learning algorithms for unknown human models. While the above optimization architecture is general and powerful, it requires us to estimate the *loss* for the assistance policy in each iteration. In settings where the human models are unknown, we won't be able to infer the final human decisions to obtain an accurate loss estimation. To address this, we consider the setting in which the ML can sequentially interact with human decision makers in the environment, observe their responses, and adaptively update the assistance policy over time. This leads to an *online learning* setting in which we need to address the classical trade-off between exploitation (choosing policy with the highest estimated payoff) and exploration (choosing policy with uncertain payoff to obtain information), which can be formulated as a multi-armed bandit problem [70, 7, 20]. We plan to explore the usage of bandits in this setting. The main challenge is that the space of arms (i.e., the space of assistive policies) is large/infinite and could require too many explorations for bandit algorithms to be useful. To explore this challenge, we plan to adopt the technique in the PI's work on leveraging the similarities between arms in bandit learning [56]. The key intuition is that, if two policies lead to similar payoffs, they are considered "similar" arms, and we propagate the information we learned on one policy to other similar policies to achieve efficient learning. To quantify the arm similarity, we plan to leverage domain knowledge to characterize the problem structure (e.g., abstracting key properties of human models and environment states) to reduce the problem space. Our goal is to identify conditions for bandit approaches to work and develop corresponding algorithms.

3.2.3 Task 2.3: Understanding human reliance on ML assistance with human-subject experiments

This task aims to conduct investigations of humans' trust and reliance on ML recommendations in the setting with sequential decision making. In the previous two tasks, we have assumed that human reliance on ML assistance, $\theta(s)$, is known and given. However, in practice, appropriately formulating $\theta(s)$ is not trivial and not well understood. While there has been a growing line of literature in understanding humans' trust and reliance on ML [133, 77, 98, 136, 74, 125], including the PI's work [85], existing studies mostly focus on the one-shot decision making scenarios. Limited is known about how humans trust and rely on ML assistance in *sequential decision making settings*.

Proposed research. In collaboration with Dr. Ming Yin, a leading expert in human trust and reliance on AI at Purdue University, we will conduct randomized human-subject studies to understand how human decision makers' reliance on ML are influenced by various factors under the sequential decision making setting. In particular, consistent with theoretical models previously proposed for human-automation interaction [57, 102], we expect humans' adoption of ML advice under sequential decision making settings can be influenced by factors related to *humans*, *ML*, and the *environment*. We will conduct experiments to understand:

- How factors related to ML, including the presentation format of ML recommendations, the provision of ML explanations, and the human-likeness of ML, influence humans' adoptions of ML advice?
- How factors related to the decision making environment, including the variability and complexity of the environment, influence humans' adoptions of ML advice?
- How factors related to humans, including their risk attitudes, their value similarity with ML, and their subjective perceptions of ML trustworthiness, influence humans' adoptions of ML advice?

General experimental designs. We plan to conduct randomized human subject experiments. For each human-subject experiment, we will start by designing experiment that only a single independent variable varies. That is, different experimental treatments will be created corresponding to different "levels" of the independent variable (e.g., timing of ML recommendations, type of ML explanations, human-likeness of

the ML policy). For dependent variables, we will record whether human subjects decide to rely on the ML's decision recommendations to estimate $\theta(s)$, as well as their final decision making performance. In addition, to align with the AI trust literature, we will ask human subjects to self-report their perceived trust level in the ML agent both at a fixed interval (e.g., after every 5 decisions are made) and at the end of the experiment. We can also have the human subjects complete a two-phase experiment, in which they make sequential decisions in the first phase with the assistance of the ML agent, while they make sequential decisions on their own in the second phase, and we can record their decision making performance in the second phase to understand if human decision makers can effectively learn from the ML agent in the first phase. After collecting the measurements on all the dependent variables, we can conduct statistical tests across treatments to examine if the independent variable varied in the experiment affects decision makers' adoption of ML advice and subjective trust on ML, sequential decision making performance, and learning outcome. Moreover, additional experiment can be carried out to vary multiple independent variables simultaneously, which will allow us to understand how they interact with one another to affect the dependent variables of interests.

3.3 Thrust 3: Integrating with Domain Applications

In research thrusts 1 and 2, our goal is to develop a framework for behavior-informed machine learning, incorporating human behavior in the design of ML systems. While the framework is intended to be general, deploying the framework in specific domain applications may introduce various domain-specific challenges. For instance, when allocating scarce societal resources for homelessness prevention, it is important not only to maximize the effectiveness of these resources but also to ensure that the allocation of resources is *fair and equitable* across different social groups. When designing decision support systems for airplane pilots, in addition to maintaining decision efficiency, *safety* is of the utmost importance.

In this thrust, we aim to collaborate with domain experts to tackle practical challenges when deploying this framework in domain applications. In particular, the proposed research will be tailored for use in the domains of homelessness prevention (with Prof. Patrick Fowler at the Brown School of Social Work) and flight pilot augmentation (with Boeing). In the long term, we plan to harness the interdisciplinary efforts at WashU to expand this research into other application domains, including the Division of Computational and Data Sciences (DCDS), the Center for Collaborative Human-AI Learning and Operation (HALO), and the Transdisciplinary Institute in Applied Data Sciences (TRIADS) at WashU that the PI is an active member in. These cross-disciplinary endeavors will help ensure that our research findings are practically applicable across various domains, thus promoting their adoption and potential for impact.

3.3.1 Task 3.1: Domain application: Data-driven decision support for homelessness prevention

This task extends our existing collaboration with Prof. Patrick Fowler on developing algorithmic solutions to homelessness prevention [32] to the scope of data-driven decision support for homelessness prevention. The problem of homelessness, a longstanding societal issue, presents significant personal and communal repercussions. Local systems dedicated to addressing homelessness often face a scarcity of resources, making it challenging to fulfill the demand for housing support. The current decision-making processes for distributing these limited resources are largely unexplored [17, 41, 104], leaving room for improvement in terms of both efficiency and equity. This opens up two important research directions that align with this CAREER plan: First, we can utilize historical data to understand the impacts of past resource allocation, thereby allowing us to derive insights to optimize future decisions. Secondly, by harnessing the power of ML, we can provide decision support for human decision makers in deciding the resource allocation.

Account for human behavior when learning from past data. There is a growing effort to use data-driven approaches to inform decision-making policies in homelessness prevention [44, 67, 69]. Specifically, Prof. Fowler has been involved in the St. Louis Regional Data Alliance [1], an initiative that aims to curate community data to improve community health, such as reducing homelessness. Building on this effort, Dr. Fowler and I have been co-advising a PhD student, Alex DiChristofano, in conducting preliminary analyses

of St. Louis regional data. We have identified two types of human behavior that could inject biases into the data. The first comes from the recipients of resources. In homelessness prevention, when people seek help, they are not immediately assigned resources due to the resource scarcity. Instead, they are placed on a waitlist and only receive resources when resources become available. This waiting process creates unequal *drop-out* rates across social groups, e.g., we found that females are more likely to leave the system before resources become available. Failure to account for this drop-out inequality could lead to biased predictions of resource efficacy. The second type of behavior that needs to be taken into account comes from the parties (e.g., social workers) that decide how to allocate resources. While there are general guidelines in the decision-making policy, the past data largely reflects the decision-makers' judgments. In this task, we aim to identify and incorporate this human behavior during the training of ML based on past data.

Designing decision support. In the decision-making process for allocating resources for homelessness, there isn't a clear right or wrong answer. Social workers often need to balance multiple ethical principles, such as prioritizing outcomes (reducing homelessness) or prioritizing the most vulnerable individuals [68]. When designing decision support systems, we must consider decision-makers' preferences and constraints. In this task, we will work with local homelessness service providers, the St. Louis Area Regional Commission on Homelessness (SLARCH) – a nonprofit organization that coordinates homeless service provision across the St. Louis region. By conducting qualitative surveys and interviews, we aim to gain better insights into their decision-making process, their objectives in decision-making, and the types of decision support needed to inform the design of our assistive ML. Furthermore, we will work with social workers, the decision-makers in the field, recruited through SLARCH, to evaluate and deploy our research.

3.3.2 Task 3.2: Domain application: Decision support for airplane pilots

This task aims to launch our newly initiated collaboration with Boeing in designing decision support for pilot decision-making. In this application domain, safety is of paramount importance, in addition to efficiency. To make the discussion more concrete, we will discuss the design of pilot augmentation to address runway incursions – a significant aspect of runway safety. Runway incursion [2] refers to an incident involving an incorrect presence of an aircraft, vehicle, or person on a runway designated for take-off or landing. In severe cases, runway incursions could lead to tragic events. Given the gravity of this problem, there has been research devoted to avoiding such incursions, including accident prediction [111, 106, 45] and system design to detect obstacles and alert pilots [58, 89, 91, 130]. Meanwhile, the Federal Aviation Administration (FAA) have reported that pilot behavior is involved in 65% of all runway incursions [3]. Therefore, in this task, aligning with this CAREER plan, we plan to adopt a behavior-informed approach in addressing the runway incursion problem. We will examine existing datasets and behavioral data from simulated platforms to identify pilot behavioral patterns in the context of runway incursions. Moreover, we will design decision support that provides interventions to prevent runway incursion events.

Proposed research. For the question of learning from behavioral data, we will leverage two data sources. The first is the public ASRS (Aviation Safety Reporting System) dataset, FAA's voluntary confidential reporting system that accepts confidential reports of near misses or close call events in the interest of improving aviation safety. This public dataset will enable us to identify generic characteristics for runway incursions. We will then leverage the flight simulator X-Plane, that WashU has acquired in the previous collaboration with Boeing, to collect individual behavioral data for identifying personalized behavioral patterns in runway safety. After identifying the behavioral patterns, we will design decision support systems that aim to maximize decision efficiency (e.g., time for departing/landing) while imposing safety constraints. The study will be initially conducted in an academic setting, recruiting general population (e.g., college students) in running the flight simulator. After developing the results, in collaboration with Boeing, the study will be extended to other contexts (e.g., inflight weather encounters, wake turbulence encounters), and the evaluations will conducted with domain experts and real pilots through simulations/surveys.

3.4 Evaluation Plan

The proposed research will span five years. The tasks in Thrust 1 and 2 have been organized in a way that can be performed in a sequential manner. We will perform the tasks in Thrust 3 after we have initial results for the first two thrusts. For the evaluation of the proposed research, there are three main components:

- Algorithm and theory: For task 1.1-1.2 and 2.1-2.2, we will develop new algorithms and theories. To evaluate our results, we will derive the performance guarantees (regret bounds or convergence rate) and analyze the computational complexity of the proposed algorithms. We will perform equilibrium analysis to characterize the human behavior in the equilibrium structure. Simulation will also be performed to evaluate the algorithm performance under the conditions both when users follow our proposed models and when users do not exactly follow to test for robustness of our proposed algorithms.
- Data collection: Task 1.3 and 2.3 involve collecting data using human-subject experiments. With collaborations with experts in psychology and HCI, we will follow the best practice in conducting the experiments, including pre-registering the hypothesis and performing appropriate statistical tests (e.g., ANOVA, post-hoc t-tests, mixed effects model). The collected data will be made publicly available to the research community. We believe the large-scale behavioral data would be of important research value.
- Deployment: For tasks in Thrust 3, we aim to deploy the proposed research in domain applications. In addition to the evaluations above, we will work with domain experts to develop our evaluation plan and solicit feedback of the proposed framework through interviews/surveys.

4 Education Plan

The PI aims to broaden research participation and develop education plans that integrate with the proposed research throughout the duration of the CAREER project. To maximize the impacts of the proposed activities, the PI will collaborate with several existing programs at WashU.

4.1 Broadening Research Participation

This project will invest efforts in broadening the participation in computing, including developing activities to expose high-school students in research, actively recruiting female and underrepresented minority students, and engaging undergraduate research participation.

Outreach to high-school teachers and students. The PI will work with the Institute for School Partnership (ISP) at WashU to design outreach activities. The goal is to provide professional developments for high-school teachers and broaden the dissemination of research ideas, and to cultivate next-generation scientists/engineers through exposing high-school students to academic research and stimulating their interests in computing. In particular, we will work with the ISP for the *Teacher-Researcher Partnership*, under which teachers work in the faculty's lab for 4-6 weeks in the summer, with the goal of learning and translating research ideas into lessons at grade level. We plan to host one teacher in each of the first two summers. Based on the partnership outcomes, we will participate in the *Hot Topic Series* at ISP and invite around 20 high-school teachers to disseminate the curriculum design to maximize the potential outreach.

The PI will also join force with existing efforts at the McKelvey School of Engineering, which has conducted a summer camp in Summer 2022 for local high school students of low-income backgrounds. This summer camp is planned to become an annual event. The PI will annually host a one-day summer workshop "Human-Centered Machine Learning" within this framework. The workshop will include a broad overview of machine learning and human behavior and engage students in group projects guided by Ph.D. students. We will prepare datasets and ML modules for students to explore the impact of human behavior in the design of ML, both for how human behavior leads to learning (how biased dataset leads to biased learning outcome) and how ML can assist humans in overcoming biases. For evaluations, the ISP will provide consultations for evaluation plans. In particular, we will conduct anonymous surveys to high-school

teachers/students before and after the event to evaluate their understanding of the topic and their aspirations in pursuing higher-education in STEM.

Engagements of traditionally underrepresented students. The PI is committed to recruiting female and underrepresented minority (URM) students. The PI is currently advising 5 PhD students, with whom one is female and another one is African American. The PI has also worked with 6 female and URM undergraduate/master students (out of 13 students that worked with the PI) at WashU so far. Among the 6 students, four have continued their graduate studies after graduation (at Stanford, Duke, Penn State, and Cornell), one went to the industry (at Google), and one is still in the undergraudate program. The PI will leverage the institutional effort for engaging female and URM students. In particular, WashU is committed to the goal of increasing the representation of women at the Ph.D. level. The CSE department, the McKelvey School of Engineering, and the Provost's Office of Diversity together fund a Platinum Sponsorship of Grace Hopper. Through WashU Summer Engineering Fellowship (WUSEF), which provides funds for students from backgrounds underrepresented in the STEM fields to perform summer research, the PI has advised one URM undergraduate student. In addition to working with WUSEF each summer, the PI will work with the Missouri Louis Stokes Alliance for Minority Participation (MOLSAMP), of which WashU is a participating institution, for offering summer research opportunities for minority participation.

Undergraduate research participation. Undergraduate students will be heavily engaged in the proposed research. The PI has been actively involved in the NSF REU site "Big Data Analytics" at WashU. The students the PI advised at the REU site have all continued their gradudate studies in the Computer Science field (at UT Austin, Duke, CMU, Yale, and Cornell) after graduation. The PI is committed to annually support REU/WUSEF research projects inspired by this proposal, such as understanding user behavior in computational systems through conducting behavioral experiments or analyzing existing datasets. The PI will also support undergraduate students on independent research projects during the academic year.

4.2 Course and Teaching Development

The research goal of the PI is to combine the strengths of both humans and machine learning (ML) to solve tasks neither can solve alone. To achieve this goal, we need to advance our understanding of ML, humans, and the interactions between them. Correspondingly, the education goal of the PI is to prepare students in these fronts. To achieve this education goal, the PI has been regularly teaching two courses: *Introduction to Machine Learning* and *Human-in-the-Loop Computation*. As part of this CAREER project, the PI plans to heavily revise the second course into a new course *Human-AI Interaction and Collaboration*. In addition to the general coverage of ML and human modeling (from behavioral economics, psychology, and HCI), there will be two main themes for the course topics. First. we will cover and discuss human-in-the-loop machine learning, addressing the techniques of incorporating humans in the learning process to advance machine learning. Second, we will discuss topics with a human-centered focus, including how humans process information from ML (such as interpretability, trustworthiness, and topics explores in this proposal) and how ML impacts human welfare (such as fairness, privacy, and ethical concerns). We will also include practical domain applications in social sciences and healthcare in the course materials (in the form of assignments, projects, or guest lectures) by leveraging the Division of Computational and Data Science (DCDS) and the Center for Collaborative Human-AI Learning and Operation (HALO) at WashU. The course materials will be made available online to enable self-study or to be used in other institutions.

Evaluation plan. The PI will work with the Center for Integrative Research on Cognition, Learning, and Education (CIRCLE) at WashU to develop evaluation plan for the proposed course. The evaluations will be conducted based on multiple metrics, including whether students obtain firm grasp of the subject (by constructing a knowledge inventory) and whether the course motivates students in applying the knowledge in different domains. The PI will coordinate with the Teaching Center at WashU to periodically videotape and assess the lectures from the course.

4.3 Long-term Vision: Towards Personalized Education

My long-term vision in education is to develop data-driven methods that enable personalized education. This vision aligns with this CAREER plan on designing ML that learns from behavioral data and assist human decision maker. As a starting point to realize this long-term vision, we have started to conduct research in the domain of Chess to develop personalized ML assistant. In collaboration with Kassa Korley [83], who holds the title of International Master and was the youngest African American to earn the title of National Master in the US, we have investigated the question of curriculum design, i.e., what moves should be provided to assist Chess players based on their skills, using data-driven approaches. In particular, leveraging the abundant amount of human play data in online Chess platforms (Lichess.org), we have developed ML models that can mimic human plays at different skill levels. We then leveraged both the idea of designing ML assistance in this proposal and curriculum learning [14, 128, 110] to design curriculum. Our preliminary results, showing that the approaches can identify curriculums that align with domain knowledge and improve human win rate, holds potential in designing personalized tool to improve human learning in Chess. In addition to creating assistive ML for chess learning, we plan to collaborate with Prof. Dennis Barbour, who has employed data-driven approaches to explore the connection between students' mathematical learning skills and general executive function skills, such as cognitive flexibility, working memory, and inhibitory/attentional control. The goal is to improve personalized education in the setting of enhancing mathematical skills.

5 Broader Impacts

The research of designing ML to learn from humans and assist humans creates a wide range of impacts. It impacts the design of a broad range of online platforms with active human participation, including recommendation systems, user-generated content platforms, social networking sites, in improving the way the platform interacts with users. Moreover, as algorithmic decision making gets deployed more widely in policy making, this research contributes to improving decision making for societal issues. In particular, the PI has existing collaborations with Prof. Patrick Fowler at Brown School of Social Work for homeless prevention [31] and with Dr. Jason Wellen at Medical School to apply computational approaches for living donor kidney transplantation [72]. The PI plans to continue and expand the collaborations through the Division for Computational and Data Sciences (DCDS) which brings together the Department of Computer Science & Engineering with the departments of Political Science and Psychological and Brain Sciences in Arts & Sciences and with the Brown School of Social Works. Moreover, the PI is a founding member of the Center for Collaborative Human-AI Learning and Operation (HALO), which provides additional collaboration opportunities with the medical school at Washington University on healthcare problems.

Dissemination of results. One of the main research effort in this proposed research is to collect human behavioral data through multiple sets of large-scale behavioral experiments. We plan to make the collected data publicly accessibly to the research community. To disseminate our research results to a broad audience, in addition to regular conference and journal publications, we will publicly release the software implementations of algorithms, simulation test-bed, and models developed in this project. Furthermore, we will disseminate results within the interdisciplinary DCDS program at Washington University through regular interaction with other faculty in the program, as well as its seminar series.

6 Results from Prior NSF Support

Dr. Ho is the co-PI on "FAI: FairGame: An Audit-Driven Game Theoretic Framework for Development and Certification of Fair AI" (IIS-1939677, $444,145, Jan 2020 to Dec 2023). *IM*: This project provides a general framework for fair decision making and auditing in stochastic, dynamic environments. PI Ho has published six publications in this project [84, 119, 33, 118, 34, 85]. *BI*: The work supports the training of graduate students and the development of new auditing algorithms that have impacts to AI and society.

References

[1] St. louis regional data alliance. commnuity information exchange. URL https://stldata.org/project-community-information-exchange/.

[2] Runway incursions. URL https://www.faa.gov/airports/runway_safety/resources/runway_incursions.

[3] Runway incursion totals for fy 2014, 2014. URL http://www.faa.gov/airports/runway_safety/statistics/regional/?fy=2014.

[4] Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1. ACM, 2004.

[5] Jacob Abernethy, Yiling Chen, Chien-Ju Ho, and Bo Waggoner. Low-cost learning via active data procurement. In *16th ACM Conf. on Economics and Computation (EC)*, 2015.

[6] Tal Alon, Magdalen Dobson, Ariel Procaccia, Inbal Talgam-Cohen, and Jamie Tucker-Foltz. Multiagent evaluation mechanisms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[7] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002. Preliminary version in *15th ICML*, 1998.

[8] Robert Axelrod. Effective choice in the prisoner's dilemma. *Journal of conflict resolution*, 24(1): 3–25, 1980.

[9] Ashwinkumar Badanidiyuru, Kshipra Bhawalkar, and Haifeng Xu. Targeting and signaling in ad auctions. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2545–2563. SIAM, 2018.

[10] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[11] Abhijit V Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3): 797–817, 1992.

[12] Hamsa Bastani, Osbert Bastani, and Wichinpong Park Sinchaisri. Improving human decision-making with machine learning. *arXiv preprint arXiv:2108.08454*, 2021.

[13] Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *COLT*, volume 3, page 1, 2009.

[14] Yoshua Bengio, JÃ©rÃ´me Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, page 41â48, 2009.

[15] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Dębiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

[16] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, (just-accepted), 2017.

[17] Molly Brown, Camilla Cummings, Jennifer Lyons, Andrés Carrión, and Dennis P Watson. Reliability and validity of the vulnerability index-service prioritization decision assistance tool (vi-spdat) in real-world implementation. *Journal of Social Distress and the Homeless*, 27(2):110–117, 2018.

[18] Michael Brückner and Tobias Scheffer. Stackelberg games for adversarial prediction problems. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 547–555, 2011.

[19] Michael Brückner, Christian Kanzow, and Tobias Scheffer. Static prediction games for adversarial learning problems. *Journal of Machine Learning Research*, 13(Sep):2617–2654, 2012.

[20] Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

[21] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–21, 2021.

[22] Frederick Callaway, Yash Raj Jain, Bas van Opheusden, Priyam Das, Gabriela Iwama, Sayan Gul, Paul M Krueger, Frederic Becker, Thomas L Griffiths, and Falk Lieder. Leveraging artificial intelligence to improve peopleâs planning strategies. *Proceedings of the National Academy of Sciences*, 119(12):e2117432119, 2022.

[23] Yiling Chen, Chara Podimata, Ariel D. Procaccia, and Nisarg Shah. Strategyproof linear regression in high dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation (EC)*, 2018.

[24] Sharath R. Cholleti, Sally A. Goldman, Avrim Blum, David G. Politte, and Steven Don. Veritas: Combining expert opinions without labeled data. In *Proceedings 20th IEEE international Conference on Tools with Artificial intelligence (ICTAI)*, 2008.

[25] A. P. Dawid and A. M. Skene. Maximum likeihood estimation of observer error-rates using the EM algorithm. *Applied Statistics*, 28:20–28, 1979.

[26] Alexander Philip Dawid and Allan M Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied statistics*, pages 20–28, 1979.

[27] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

[28] Anthony DiGiovanni and Ethan C Zell. Survey of self-play in reinforcement learning. *arXiv preprint arXiv:2107.02850*, 2021.

[29] Bolin Ding, Yiding Feng, Chien-Ju Ho, Wei Tang, and Haifeng Xu. Competitive information design for pandora's box. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 353–381. SIAM, 2023.

[30] Vinayak V Dixit, Sai Chand, and Divya J Nair. Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one*, 11(12):e0168054, 2016.

[31] Zehao Dong, Sanmay Das, Patrick Fowler, and Chien-Ju Ho. Efficient nonmyopic online allocation of scarce resources. In *Proceedings of the 20th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2021.

[32] Zehao Dong, Sanmay Das, Patrick Fowler, and Chien-Ju Ho. Efficient nonmyopic online allocation of scarce reusable resources. In *AAMAS Conference proceedings*, 2021.

[33] Xiaoni Duan, Chien-Ju Ho, and Ming Yin. Does exposure to diverse perspectives mitigate biases in crowdwork? an explorative study. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, pages 155–158, 2020.

[34] Xiaoni Duan, Chien-Ju Ho, and Ming Yin. The influences of task design on crowdsourced judgement: A case study of recidivism risk evaluation. In *Proceedings of the ACM Web Conference 2022*, pages 1685–1696, 2022.

[35] Shaddin Dughmi and Haifeng Xu. Algorithmic bayesian persuasion. *SIAM Journal on Computing*, 2019.

[36] Paul Dütting, Zhe Feng, Harikrishna Narasimhan, David Parkes, and Sai Srivatsa Ravindranath. Optimal auctions through deep learning. In *International Conference on Machine Learning*, pages 1706–1715. PMLR, 2019.

[37] Glyn Elwyn, Adrian Edwards, Martin Eccles, and David Rovner. Decision analysis in patient care. *The Lancet*, 358(9281):571–574, 2001.

[38] Yuval Emek, Michal Feldman, Iftah Gamzu, Renato PaesLeme, and Moshe Tennenholtz. Signaling schemes for revenue maximization. *ACM Transactions on Economics and Computation (TEAC)*, 2 (2):1–19, 2014.

[39] Christoph Engel. Dictator games: A meta study. *Experimental economics*, 14:583–610, 2011.

[40] Yiding Feng, Chien-Ju Ho, and Wei Tang. Rationality-robust information design: Bayesian persuasion under quantal response. *arXiv preprint arXiv:2207.08253*, 2022.

[41] Patrick J Fowler, Peter S Hovmand, Katherine E Marcal, and Sanmay Das. Solving homelessness from a complex systems perspective: insights for prevention responses. *Annual review of public health*, 40:465–486, 2019.

[42] Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 5–22, 2014.

[43] Noufel Frikha, Stéphane Menozzi, et al. Concentration bounds for stochastic approximations. *Electronic Communications in Probability*, 17, 2012.

[44] Yuan Gao, Sanmay Das, and Patrick Fowler. Homelessness service provision: a data science perspective. In *Workshops at the thirty-first AAAI conference on artificial intelligence*, 2017.

[45] Jean-Baptiste Gotteland, Nicolas Durand, Jean-Marc Alliot, and Erwan Page. Aircraft ground traffic optimization. In *ATM 2001, 4th USA/Europe Air Traffic Management Research and Development Seminar*, pages pp–xxxx, 2001.

[46] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[48] Chien-Ju Ho and Jennifer Wortman Vaughan. Online task assignment in crowdsourcing markets. In *26th AAAI Conference on Artificial Intelligence (AAAI)*, 2012.

[49] Chien-Ju Ho and Ming Yin. Working in pairs: Understanding the effects of worker interactions in crowdwork. *arXiv preprint arXiv:1810.09634*, 2018.

[50] Chien-Ju Ho, Yu Zhang, Jennifer Wortman Vaughan, and Mihaela van der Schaar. Towards social norm design for crowdsourcing markets. In *4th Human Computation Workshop (HCOMP)*, 2012.

[51] Chien-Ju Ho, Shahin Jabbari, and Jennifer Wortman Vaughan. Adaptive task assignment for crowdsourced classification. In *30th Intl. Conf. on Machine Learning (ICML)*, 2013.

[52] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *15th ACM Conf. on Electronic Commerce (EC)*, 2014.

[53] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*, pages 419–429, 2015.

[54] Chien-Ju Ho, Aleksanrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. Incentivizing high quality crowdwork. In *24th Intl. World Wide Web Conf. (WWW)*, 2015.

[55] Chien-Ju Ho, Rafael Frongillo, and Yiling Chen. Eliciting categorical data for optimal aggregation. In *30th Advances in Neural Information Processing Systems (NIPS)*, 2016.

[56] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research*, 55:317 – 359, 2016.

[57] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3):407–434, 2015.

[58] Yasuo Ishihara and Steve Johnson. Aircraft systems and methods for managing runway awareness and advisory system (raas) callouts, February 12 2019. US Patent 10,204,523.

[59] Rong Jin and Zoubin Ghahramani. Learning with multiple labels. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[60] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[61] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decisions under risk. *Econometrica*, pages 263–291, 1979.

[62] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

[63] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[64] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 5(1):193–206, 1991.

[65] Niklas Karlsson, George Loewenstein, and Duane Seppi. The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, 38:95–115, 2009.

[66] Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 825–844, 2019.

[67] Amanda Kube, Sanmay Das, and Patrick J Fowler. Allocating interventions based on predicted outcomes: A case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 622–629, 2019.

[68] Amanda Kube, Sanmay Das, Patrick J Fowler, and Yevgeniy Vorobeychik. Just resource allocation? how algorithmic predictions and human notions of justice interact. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 1184–1242, 2022.

[69] Amanda R Kube, Sanmay Das, and Patrick J Fowler. Community-and data-driven homelessness prevention and service delivery: optimizing for equity. *Journal of the American Medical Informatics Association*, 30(6):1032–1041, 2023.

[70] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocations rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[71] Wouter Kool Lauren Treiman, Chien-Ju Ho. Humans forgo reward to instill fairness into AI. Working paper, 2023.

[72] Zhuoshu Li, Kelsey Lieberman, William Macke, Sofia Carrillo, Chien-Ju Ho, Jason Wellen, and Sanmay Das. Incorporating compatible pairs in kidney exchange: A dynamic weighted matching model. In *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*, 2019.

[73] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[74] Han Liu, Vivian Lai, and Chenhao Tan. Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–45, 2021.

[75] Qiang Liu, Jian Peng, and Alexander Ihler. Variational inference for crowdsourcing. In *26th Advances in Neural Information Processing Systems (NIPS)*, 2012.

[76] Yang Liu and Chien-Ju Ho. Incentivizing high quality user contributions: New arm generation in bandit learning. In *32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[77] Zhuoran Lu and Ming Yin. Human reliance on machine learning models when performance feedback is limited: Heuristics and risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2021.

[78] Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. *arXiv preprint arXiv:1602.07570*, 2016.

[79] Daniel McFadden. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272, 1981.

[80] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[81] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[82] Lev Muchnik, Sinan Aral, and Sean J Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):641 – 651, 2013.

[83] Saumik Narayanan, Kassa Korley, Chien-Ju Ho, and Siddhartha Sen. Improving the strength of human-like models in chess. In *Human in the Loop Learning (HiLL) Workshop at NeurIPS*, 2022.

[84] Saumik Narayanan, Guanghui Yu, Wei Tang, Chien-Ju Ho, and Ming Yin. How does predictive information affect human ethical preferences? In *ACM Conference on AI, Ethics, and Society*, 2022.

[85] Saumik Narayanan, Guanghui Yu, Chien-Ju Ho, and Ming Yin. How does value similarity affect human reliance in ai-assisted ethical decision making? In *ACM Conference on AI, Ethics, and Society*, 2023.

[86] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. *Advances in neural information processing systems*, 26, 2013.

[87] Andrew Y Ng, Stuart J Russell, et al. Algorithms for inverse reinforcement learning. In *International Conference on Machine Learning*, 2000.

[88] Martin A Nowak, Karen M Page, and Karl Sigmund. Fairness versus reason in the ultimatum game. *Science*, 289(5485):1773–1775, 2000.

[89] Shutai Okamura, Takeshi Hatakeyama, Takahiro Yamaguchi, and Tsutomu Uenoyama. Radar detection system and radar detection method, January 19 2021. US Patent 10,895,638.

[90] Neehar Peri, Michael Curry, Samuel Dooley, and John Dickerson. Preferencenet: Encoding human preferences in auction design with deep learning. *Advances in Neural Information Processing Systems*, 34:17532–17542, 2021.

[91] Joseph T Pesik and David Matty. Determination of collision risks between a taxiing aircraft and objects external to the taxiing aircraft, February 4 2020. US Patent 10,553,123.

[92] Matthew Rabin and Ted O'Donoghue. Doing It Now or Later. *American Economic Review*, 1999.

[93] Manish Raghavan, Aleksandrs Slivkins, Jennifer Vaughan Wortman, and Zhiwei Steven Wu. The externalities of exploration and how data diversity helps exploitation. In *Conference on Learning Theory*, pages 1724–1738. PMLR, 2018.

[94] Deepak Ramachandran and Eyal Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

[95] Kamalini Ramdas, Khaled Saleh, Steven Stern, and Haiyan Liu. Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science*, 64(6):2590–2608, 2018.

[96] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R Varshney, Amit Dhurandhar, and Richard Tomsett. Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–22, 2022.

[97] Vikas Raykar, Shipeng Yu, Linda Zhao, Gerardo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from crowds. *Journal of Machine Learning Research*, 11:1297–1322, 2010.

[98] Amy Rechkemmer and Ming Yin. When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models. In *CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2022.

[99] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[100] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[101] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854 – 856, 2006.

[102] Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors*, 58(3):377–400, 2016.

[103] Clayton Scott. A rate of convergence for mixture proportion estimation, with application to learning from noisy labels. In *Artificial Intelligence and Statistics*, pages 838–846. PMLR, 2015.

[104] Marybeth Shinn, Andrew L Greer, Jay Bainbridge, Jonathan Kwon, and Sara Zuiderveen. Efficient targeting of homelessness prevention services for families. *American journal of public health*, 103 (S2):S324–S330, 2013.

[105] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[106] Abneesh Singla, Srinivas D Gonabal, Pradeep Huncha, Vedavyas Rallabandi, Jaibir Singh, Sunil Kumar KS, et al. System and method for monitoring compliance with air traffic control instructions, August 25 2020. US Patent 10,755,583.

[107] Ruben Sipos, Arpita Ghosh, and Thorsten Joachims. Was this review helpful to you?: It depends! context and voting patterns in online content. In *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, pages 337–348, 2014.

[108] Kenneth A Small. A discrete choice model for ordered alternatives. *Econometrica: Journal of the Econometric Society*, pages 409–424, 1987.

[109] Hummy Song, Anita L Tucker, Karen L Murrell, and David R Vinson. Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science*, 64(6):2628–2649, 2018.

[110] Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey. *International Journal of Computer Vision*, 130(6):1526–1565, 2022.

[111] Zhe Sun, Cheng Zhang, Pingbo Tang, Yuhao Wang, and Yongming Liu. Bayesian network modeling of airport runway incursion occurring processes for predictive accident control. In *Advances in Informatics and Computing in Civil and Construction Engineering: Proceedings of the 35th CIB W78 2018 Conference: IT in Design, Construction, and Management*, pages 669–676. Springer, 2019.

[112] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[113] Wei Tang and Chien-Ju Ho. Bandit learning with biased human feedback. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1324–1332, 2019.

[114] Wei Tang and Chien-Ju Ho. On the bayesian rational assumption in information design. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 9, pages 120–130, 2021.

[115] Wei Tang, Ming Yin, and Chien-Ju Ho. Leveraging peer communication to enhance crowdsourcing. In *The World Wide Web Conference*, pages 1794–1805. ACM, 2019.

[116] Wei Tang, Chien-Ju Ho, and Yang Liu. Differentially private contextual dynamic pricing. In *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2020.

[117] Wei Tang, Chien-Ju Ho, and Yang Liu. Optimal query complexity of secure stochastic convex optimization. In *34th Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[118] Wei Tang, Chien-Ju Ho, and Yang Liu. Bandit learning with delayed impact of actions. *Advances in Neural Information Processing Systems*, 34:26804–26817, 2021.

[119] Wei Tang, Chien-Ju Ho, and Yang Liu. Linear models are robust optimal under strategic behavior. In *International Conference on Artificial Intelligence and Statistics*, pages 2584–2592. PMLR, 2021.

[120] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[121] Long Tran-Thanh, Archie Chapman, Enrique Munoz De Cote, Alex Rogers, and Nicholas R Jennings. Epsilon–first policies for budget–limited multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1211–1216, 2010.

[122] Amos Tversky and Daniel Kahneman. Belief in the law of small numbers. *Psychological bulletin*, 76 (2):105, 1971.

[123] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[124] Dimitris G Tzikas, Aristidis C Likas, and Nikolaos P Galatsanos. The variational approximation for bayesian inference. *IEEE Signal Processing Magazine*, 25(6):131–146, 2008.

[125] 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):1–39, 2021.

[126] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[127] John von Neumann and Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[128] Xin Wang, Yudong Chen, and Wenwu Zhu. A survey on curriculum learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4555–4576, 2021.

[129] Christopher J.C.H. Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[130] Christabel Wayllace, Sunwoo Ha, Yuchen Han, Jiaming Hu, Shayan Monadjemi, William Yeoh, and Alvitta Ottley. Dragon-v: detection and recognition of airplane goals with navigational visualization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13642–13643, 2020.

[131] Jacob Whitehill, Ting fan Wu, Jacob Bergsma, Javier R. Movellan, and Paul L. Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[132] Wayne Xiong, Lingfeng Wu, Fil Alleva, Jasha Droppo, Xuedong Huang, and Andreas Stolcke. The microsoft 2017 conversational speech recognition system. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5934–5938. IEEE, 2018.

[133] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 279. ACM, 2019.

[134] Guanghui Yu and Chien-Ju Ho. Environment design for biased decision makers. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

[135] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pages 321–384, 2021.

[136] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 295–305, 2020.

[137] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. Truth inference in crowdsourcing: Is the problem solved? *Proceedings of the VLDB Endowment*, 10(5):541–552, 2017.

[138] Brian D Ziebart, Andrew Maas, J Andrew Bagnell, and Anind K Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2008.