Learning Personalized Decision Support Policies

Umang Bhatt[1,2*], Valerie Chen[3*], Katherine M. Collins[4], Parameswaran Kamalaruban[2], Emma Kallina[2,4], Adrian Weller[2,4], Ameet Talwalkar[3]

[1]New York University
[2]The Alan Turing Institute
[3]Carnegie Mellon University
[4]University of Cambridge
umangbhatt@nyu.edu, valeriechen@cmu.edu

Abstract

Individual human decision-makers may benefit from different forms of support to improve decision outcomes, but when will each form of support yield better outcomes? In this work, we posit that personalizing access to decision support tools can be an effective mechanism for instantiating the appropriate use of AI assistance. Specifically, we propose the general problem of learning a *decision support policy* that, for a given input, chooses which form of support to provide to decision-makers for whom we initially have no prior information. We develop `Modiste`, an interactive tool to learn personalized decision support policies. `Modiste` leverages stochastic contextual bandit techniques to personalize a decision support policy for each decision-maker. In our computational experiments, we characterize the expertise profiles of decision-makers for whom personalized policies will outperform offline policies, including population-wide baselines. Our experiments include realistic forms of support (e.g., expert consensus and predictions from a large language model) on vision and language tasks. Our human subject experiments add nuance to and bolster our computational experiments, demonstrating the practical utility of personalized policies when real users benefit from accessing support across tasks.

1 Introduction

Human decision-makers use various forms of support to inform their opinions before making a final decision (Keen 1980). Decision-makers with differing expertise may benefit from different forms of support on a given input (Yu et al. 2024). For example, one radiologist may provide a better diagnosis of a chest X-ray by leveraging model predictions (Kahn Jr 1994) while another may perform better after viewing suggestions from senior radiologists (Briggs et al. 2008): see Figure 1. In this paper, we study how to improve decision outcomes by *personalizing* which form of support we provide to a decision-maker on a case-by-case basis.

Since artificial intelligence (AI) is increasingly used as a form of decision support (Lai et al. 2023), even moving towards systems that could act as "thought partners" (Collins et al. 2024b), responsible machine learning (ML) model deployment requires clarity on who should have access to model outputs and when model outputs can be safely exposed to decision-makers (Amodei et al. 2016). Regulation increasingly calls for the "effective and appropriate use" of AI (Biden 2023), requiring careful consideration of when models ought to be accessible to decision-makers. In this paper, we formalize learning a *decision support policy* that dictates for each individual decision-maker when additional support (e.g., LLM output) should be viewed and used for a given input.

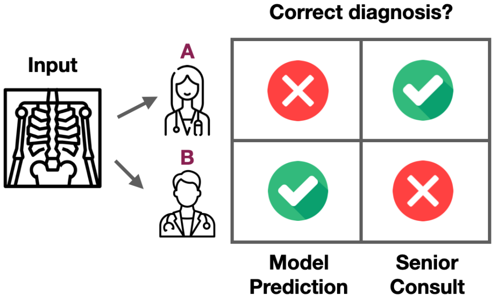


Figure 1: Depending on the input, decision-makers need different forms of decision support to make correct decisions. `Modiste` personalizes access to the right form of support at the right time for the right decision-maker online. Here, Alice would not benefit from model access, while Bob would not benefit from a senior consult.

While prior work has assumed access to offline human decisions under support (Laidlaw and Russell 2021; Charusaie et al. 2022) or oracle queries of human behavior (De-Arteaga, Dubrawski, and Chouldechova 2018; Mozannar and Sontag 2020) to learn decision support policies, we argue that this data is unrealistic to obtain in practice across all available forms of support *for a new decision-maker*. Thus, for individuals for whom we have no prior information initially, we propose learning how to personalize support *online*. We develop `Modiste`,[1] an interactive tool that leverages off-the-shelf stochastic contextual bandit algorithms (Li et al. 2010) to learn decision support policies by modeling human

[*]Both authors contributed equally: order decided by a coin flip.

[1] While a "modiste" usually refers to someone who tailors clothing and makes dresses/hats, we use the term to capture our tool's ability to alter a policy to a decision-maker.

prediction error under varying forms of support.

Our computational experiments explore the utility of personalization across multiple expertise profiles. Based on these experiments, we characterize decision-maker expertise into profiles where personalized policies outperform offline ones, such as population-wide majority vote. We demonstrate that if there is no benefit of personalization, `Modiste` recovers the same performance as the best form of support.

To validate `Modiste` on real users ($N = 80$), we conduct human subject experiments, where we explore forms of support that include expert consensus, outputs from an LLM, or predictions from a classification model. In contrast to prior work that only tests offline policies or evaluates in simulation, we demonstrate how `Modiste` can be used to learn personalized decision support policies online on both vision and language tasks. We emphasize our main contributions:

1. Formalizing decision support policies. We propose a formulation for learning a personalized decision support policy that selects the form of support that maximizes a given decision-maker's performance. We introduce `Modiste`, a tool to instantiate our formulation using existing methods from stochastic contextual bandits to model human prediction error under different forms of support. We open-source `Modiste` as a tool to encourage the adoption of personalized decision support policies.

2. Evaluating personalized policies in realistic settings. We use `Modiste` to learn personalized policies for new decision-makers through both computational and human subject experiments on vision and language tasks. We characterize under which settings we would expect personalized policies to improve performance. Our human subject experiments, where real users interact with `Modiste`, nuance our findings from the computational experiments on synthetic decision-makers, demonstrating the appropriate use of decision support has benefits in practice.

2 Related Work

Regulating AI Use. Our study of decision support policies has implications for safely deploying ML models to interact with users. This topic is of heightened importance, particularly in light of recent calls for the "effective and appropriate use" of AI in US President Biden's Executive Order (Biden 2023) and for clarity on when to "decide not to use [an] AI system" per the EU AI Act (EUA 2023). The disuse of AI can caution downstream misuse of models for assistive decision-making (Brundage et al. 2018). The refusal to use AI assistance can be strategic to empower decision-makers, thus preventing their overreliance on models and encouraging their agency on the task at hand (Gordon and Mugar 2020; Barabas 2022). Each decision-maker may require a different level of use to promote effective use of AI assistance in their decision-making (Kirk et al. 2024); for instance, experts and novices may prefer LLM access in different settings on theorem proving tasks (Collins et al. 2024a). Our experiments engage with such settings by learning when to provide LLM support for language-based tasks via a personalized policy.

Decision Support. While various forms of decision support have been proposed, such as expert consensus (Scheife et al. 2015) and changes to machine interfaces (Roda 2011), more recent forms of support focus on algorithmic tools where decision-makers are aided by machine learning (ML) models (Phillips-Wren 2012; Gao et al. 2021; Bastani, Bastani, and Sinchaisri 2022). In some prior work, the human does not always make the final decision, such as those that learn to defer decisions from a model to a single decision-maker (Madras, Pitassi, and Zemel 2018; Mozannar and Sontag 2020) or others that jointly learn an allocation function between a model and a pool of decision makers (Keswani, Lease, and Kenthapadi 2021; Hemmer et al. 2022). In our setting, the *human* is always the decision-maker, which includes settings where humans make the final decisions with support from ML models (Green and Chen 2019; Lai et al. 2023), as well as those where humans make decisions when provided with additional information beyond a model prediction, e.g., explanations (Bansal et al. 2021), uncertainty (Zhang, Liao, and Bellamy 2020), conformal sets (Babbar, Bhatt, and Weller 2022). While these studies *always* show a single form of support, recent works consider adapting when AI support is shown to users with a fixed policy. Ma et al. (2023) fit a decision tree to offline user's decisions to decide when to show AI support to users, and Buçinca et al. (2024) use offline reinforcement learning to estimate if AI support would be helpful, especially under time constraints (Swaroop et al. 2024). Our work considers general forms of support, beyond when to show AI support, and formalizes *learning* in which contexts each form of support should be provided to an unseen decision-maker online. An extensive comparison to prior work is in the Appendix.

Prior Assumptions About Decision-Maker Information. We briefly survey the assumptions made about the decision-maker when learning decision support policies. The model of the decision maker is either synthetic, thus lacking grounding in actual human behavioral data, or learned from a batch of *offline* annotations (Madras, Pitassi, and Zemel 2018; Okati, De, and Rodriguez 2021; Charusaie et al. 2022; Gao et al. 2023). For a new decision-maker or a new form of support, this set of data would not be available in practice. Instead, we propose to learn a decision support policy *online* to circumvent these limitations. Few works use some aspect of online learning for different decision-making settings or under strict theoretical conditions, as we describe in the Appendix.

3 Preliminaries

We consider a human decision-making process with different forms of decision support. In our setting, decision-makers may be shown support when selecting an outcome from a fixed set of labels (Seger and Peterson 2013; Lai et al. 2023).

General Problem Formulation. Decision-makers perform a classification task in observation/feature space $\mathcal{X} \subseteq \mathbb{R}^p$ and outcome/label space $\mathcal{Y} = [K]$. We operate in a stochastic setting where the data $(x, y) \in \mathcal{X} \times \mathcal{Y}$ are drawn iid from a fixed, unknown data generating distribution $\mathcal{P}$, an assumption that reflects typical decision-making settings (Bastani and Bayati 2020; Bastani, Bastani, and Sinchaisri 2022). Importantly, we consider an action set $\mathcal{A}$ corresponding to the forms of support available, which may consist of an individual piece of

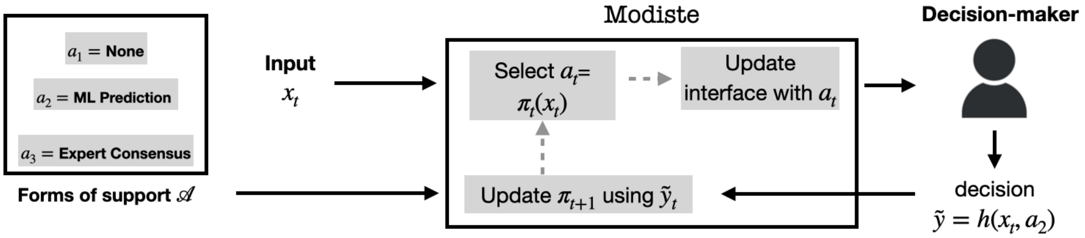

a_1 = None
a_2 = ML Prediction
a_3 = Expert Consensus
Forms of support $\mathscr{A}$
Input
x_t
Modiste
Select a_t= $\pi_t(x_t)$
Update interface with a_t
Update π_{t+1} using $\tilde{y}_t$
Decision-maker
decision
$\tilde{y} = h(x_t, a_2)$

Figure 2: We illustrate the process of learning a decision support policy π_t online to improve a decision-maker h's performance. Since assuming access to sufficient amounts of offline data is unreasonable in practice, our formulation learns a personalized policy *online*; each decision-maker's learned policy may differ from that of another decision-maker if they have decisions ($\tilde{y}$) and thus different expertise.

information (e.g., model prediction) or a particular combination of multiple pieces of information (e.g., model prediction and explanation). Given an observation $x \in \mathcal{X}$, the human attempts to predict the corresponding label $y \in \mathcal{Y}$ using the support prescribed by an action $a \in \mathcal{A}$. Note, we do not make assumptions on the specific forms of support a, but we provide multiple instantiations in our experiments. The quality of predictions is measured using a 0-1 loss function, where $\ell(y, y') = 1$ for $y \neq y'$ and $\ell(y, y') = 0$ for $y = y'$.

Decision-Making Protocol. A personalized decision support policy $\pi : \mathcal{X} \to \Delta(\mathcal{A})$ outputs a form of support for a given input. Let Π denote the class of all stochastic decision support policies. Let $\mathcal{A} = \{\mathrm{A}_1, \ldots, \mathrm{A}_k\}$, and $\pi(x)_{\mathrm{A}_i}$ denote $\mathbb{P}[\mathrm{A}_i \sim \pi(x)]$ for each $\mathrm{A}_i \in \mathcal{A}$. When the policy π prescribes the support A_i, the human decision-maker makes the prediction $\widetilde{y}$ based on the observation x and support A_i, i.e., the final prediction $\widetilde{y}$ is given by $\widetilde{y} = h(x, \mathrm{A}_i)$. The human decision-making process with different forms of support is described below. For $t = 1, 2, \ldots, T$:

1. A data point $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$ is drawn iid from $\mathcal{P}$.
2. A form of support $a_t \in \mathcal{A}$ is selected using a decision support policy $\pi_t : \mathcal{X} \to \Delta(\mathcal{A})$.
3. The human decision-maker makes the final prediction $\widetilde{y}_t = h(x_t, a_t)$ based on x_t and a_t.
4. The human decision-maker incurs a loss $\ell(y_t, \widetilde{y}_t) = 1$ if $y_t \neq \widetilde{y}_t$ and $\ell(y_t, \widetilde{y}_t) = 0$ otherwise.

Evaluation of π via Expected Loss. The quality of a policy π can be evaluated using the expected loss incurred by the decision-maker across the input space:

$$L_h(\pi) = \mathbb{E}_{(x,y)\sim\mathcal{P}}\big[\mathbb{E}_{\mathrm{A}_i\sim\pi(x)}[\ell(y, h(x, \mathrm{A}_i))]\big]. \quad (1)$$

We distinguish this metric from the more standard notion of regret, which is typically used to analyze policies in an online learning setting (Li et al. 2010); however, we cannot realize π^* for an unseen decision-maker in practical scenarios. Thus, we rely on $L_h(\cdot)$ as a proxy metric for evaluating the effectiveness of π.

4 `Modiste`: Learning Personalized Decision Support Policies

We introduce `Modiste`, a tool to translate our problem formulation into an interactive interface for learning personalized policies. The workflow, outlined in Figure 2, comprises a learning component to update the personalized policy and an interface to customize the appropriate form of support for each input and each decision-maker.

Learning Problem

To model the decision-making process of a human decision-maker *without access to their previous decisions*, we consider a stochastic contextual bandit set-up, where the forms of support are the arms, $\mathcal{X}$ is the context space, and the policy π can be learned online. In Algorithm 1, we detail an algorithm for learning a policy π^* that minimizes expected loss online. Our goal is to find an optimal decision support policy π^* that minimizes $L_h(\pi)$. We can rewrite Eq. 1 as follows:

$$L_h(\pi) = \mathbb{E}_x\big[\sum_{i=1}^{k} \pi(x)_{\mathrm{A}_i} \cdot r_{\mathrm{A}_i}(x; h)\big],$$

where $r_{\mathrm{A}_i}(x; h) = \mathbb{E}_{y|x}[\ell(y, h(x, \mathrm{A}_i))]$ is the human prediction error for input x and support A_i. Then, it can be shown that the optimal policy takes the form $\pi^*(x) = \arg\min_{\mathrm{A}_i\in\mathcal{A}} r_{\mathrm{A}_i}(x; h)$: see derivation in Appendix. For `Modiste` to run, we must maintain and update our estimate of human prediction error $r_{\mathrm{A}_i}(x; h)$ (Step 1 in Algorithm 1) and of our policy π (Step 2 in Algorithm 1).

To update the estimate of human prediction error (Step 1), `Modiste` implements two approaches to estimate $r_{\mathrm{A}_i}(x; h)$ for all $x \in \mathcal{X}$ and $\mathrm{A}_i \in \mathcal{A}$, but note that any online learning algorithm can be used to update the $\mathcal{U}_r$. We first consider **LinUCB** (Li et al. 2010), a common online learning algorithm that approximates the expected loss $r_{\mathrm{A}_i}(x; h)$ by a linear function $\widehat{r}_{\mathrm{A}_i}(x; h) := \langle\theta_{\mathrm{A}_i}, x\rangle$. Although the linearity assumption may not hold in general, we learn the parameters $\{\theta_{\mathrm{A}_i} : \mathrm{A}_i \in \mathcal{A}\}$ using LinUCB with the instantaneous reward function $R(x, y, \mathrm{A}_i; h) := -\ell(y, h(x, \mathrm{A}_i))$. We then normalize the resulting $\widehat{r}_{\mathrm{A}_i}(x; h)$ values to lie in the range $[0, 1]$. The second algorithm we use is an intuitive K-nearest neighbor (**KNN**) approach, which is a simplified variant of KNN-UCB (Guan and Jiang 2018). Here, we maintain an evolving data buffer $\mathcal{D}_t$, which accumulates a history of interactions with the decision-maker. For any new observation x, we estimate $\widehat{r}_{\mathrm{A}_i}(x; h)$ values by finding K-nearest neighbors in $\mathcal{D}_t$ and computing the average error of these neighbors.

Algorithm 1: Learning a decision support policy

```
Input: human decision-maker $h$
Initialization: data buffer $\mathcal{D}_0 = \{\}$; human error values
    $\{\widehat{r}_{\mathrm{A}_i,0}(x;h) = 0.5 : x \in \mathcal{X}, \mathrm{A}_i \in \mathcal{A}\}$; initial policy $\pi_1$
for $t = 1, 2, \ldots, T$ do
    data point $(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}$ is drawn iid from $\mathcal{P}$
    support $a_t \in \mathcal{A}$ is selected using policy $\pi_t$
    human makes the prediction $\widetilde{y}_t$ based on $x_t$ and $a_t$
    human incurs the loss $\ell(y_t, \widetilde{y}_t)$
    update the buffer $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(x_t, a_t, \ell(y_t, \widetilde{y}_t))\}$
    update the decision support policy:
        $\widehat{r}_{\mathrm{A}_i,t}(x;h) \leftarrow \mathcal{U}_r(\widehat{r}_{\mathrm{A}_i,t-1}(x;h), \mathcal{D}_t), \quad \forall \mathrm{A}_i \in \mathcal{A}$   (Step 1)
        $\pi_{t+1}(x) \leftarrow \mathcal{U}_\pi(\{\widehat{r}_{\mathrm{A}_i,t}\}_i)$   (Step 2)
end for
Output: policy $\pi^{\text{alg}}_\lambda \leftarrow \pi_{T+1}$
```

In practical settings where interactions are limited (like in our human subject experiments), the number of interactions T tends to be relatively small, which renders pure exploratory policies infeasible (Sutton and Barto 2018). Thus in Step 2, we guide exploration of the policy via:

$$\pi_{t+1}(x) = \arg\min_{\mathrm{A}_i \in \mathcal{A}} \widehat{r}_{\mathrm{A}_i,t}(x;h) + b_{\mathrm{A}_i,t}(x;h),$$

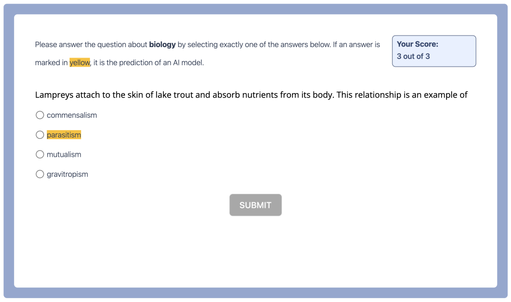
where $b_{\mathrm{A}_i,t}(x;h)$ corresponds to some exploration bonus. In the Appendix, we provide implementations of `Modiste` with LinUCB and with KNN.

`Modiste` Interface

We provide an extendable interface for the study and deployment of decision-support policies. At each time step, `Modiste` sends each user's predictions to a server running Algorithm 1, which identifies the next form of support for the next input. `Modiste` then updates the interface accordingly to reflect the selected form of support. Our tool can be flexiblely linked to crowdsourcing platforms like Prolific (Palan and Schitter 2018). We implement three common forms of support: (1) HUMAN ALONE, where the human makes the decision solely based on the input, (2) MODEL PREDICTION, which shows decision-makers a model's prediction for the given input (Bastani, Bastani, and Sinchaisri 2022), and (3) EXPERT CONSENSUS, which presents the user with a distribution over labels from multiple annotators (Scheife et al. 2015). In Figure 3, we provide an example screenshot of the interface. Participants are informed of their own correctness after each trial and the correctness of the form of support (e.g., model prediction) if support was provided, so that participants can learn whether support ought to be relied upon.

5 Experimental Set-up

Before we evaluate `Modiste`, we overview the set-up of subsequent experiments, both computational and human subject. All other details are in the Appendix.



Please answer the question about **biology** by selecting exactly one of the answers below. If an answer is marked in yellow, it is the prediction of an AI model.

Your Score:
3 out of 3

Lampreys attach to the skin of lake trout and absorb nutrients from its body. This relationship is an example of

○ commensalism
○ parasitism
○ mutualism
○ gravitropism

Figure 3: Example of the `Modiste` interface for MMLU-2A where the human is provided responses from a LLM.

Decision-making Tasks

Following prior studies on human-AI interactions (Babbar, Bhatt, and Weller 2022; Mozannar et al. 2023; Lee et al. 2023), the decision-making tasks in our experiments center around the following vision and language datasets:

1. *CIFAR*-10 (Krizhevsky 2009), a 10-class image classification dataset;
2. *MMLU* (Hendrycks et al. 2020), a multi-task text-based benchmark that tests for knowledge and problem-solving ability across 57 topics in both the humanities and STEM.

In terms of the size of $|\mathcal{A}|$, we let kA denote when there are k forms of support for a task. We focus on $k = 2$ or 3, which captures a buffet of real-world scenarios in prior work where decision-makers have a few tools at their disposal. In particular, the two action setting covers practical use cases where a decision-maker has the option of using a model or not. The learned decision support policy then reflects appropriate use, as the model would be hidden when a decision-maker does not need it. While the forms of support we consider are common in practice (Lai et al. 2023), our choices of support are not intended to exhaustively demonstrate the diverse forms of support that `Modiste` can handle. We now describe our two main tasks, which are designed to be accessible to crowdworkers and will be featured in both the computational and human subject experiments.[2]

CIFAR-3A. In this task, we consider three forms of support: HUMAN ALONE, MODEL, CONSENSUS. Our goal is to construct a setup reflecting a realistic setting in which different forms of support result in different strengths and weaknesses for decision-makers. To instantiate this setting, we deliberately corrupt images of different classes to evoke performance differences – necessitating that a decision-maker appropriately calibrate when to rely on each form of support. We consider 5 of the animal classes in CIFAR-10; of these, we never corrupt images of Birds, do not corrupt images of Deers and Cats for the MODEL, and do not corrupt images of Horses and Frogs for the CONSENSUS.

[2] In the Appendix, we include computational experiments for two additional tasks (Synthetic-2A and CIFAR-2A), and experiments where we vary the size of k.

MMLU-$2A$. The two forms of support are Human Alone and LLM, where the human is provided responses generated from InstructGPT3.5, `text-davinci-003`, (Ouyang et al. 2022) using the same few-shot prompting scheme for MMLU as Hendrycks et al. (2020). We conducted pilot studies to select a subset of topics where the accuracy of the LLM and average human accuracy vary. We choose the following topics: Computer Science, US Foreign Policy, High School Biology, and Elementary Mathematics. The goal of this task is to evaluate whether we can learn personalized support *"in-the-wild,"* where we naturally expect people to excel at different topics, akin to real-world settings where decision-makers may have varying expertise.

Baselines and Other Parameters

Algorithms and Baselines. We compare personalized policies, learned using Algorithm 1 with LinUCB and with KNN reporting results as `Modiste`-LinUCB and `Modiste`-KNN respectively, against the following offline policies:

- *Human + Support*, where the decision-maker *always* receives the same form of support: $\pi(x) = A_i$ for all x. In CIFAR-$3A$, there are 3 fixed support baselines, corresponding to each form of support. In MMLU-$2A$, there are 2 fixed support baselines.
- *Population-wide,* where the decision-maker receives a form of support based on the majority vote from 10 learned policies (breaking ties at random). For this baseline, the form of support may vary across contexts but is not personalized to individual needs. This baseline is akin to recent offline policy learning (Ma et al. 2023; Buçinca et al. 2024).

Number of Interactions. While more interactions (higher T) provide more data points to estimate each r_{A_i}, we need to consider what a realistic value of T is given constraints of working with real humans (e.g., limited attention and cognitive load). In online learning, T is usually unreasonably large, on the order of thousands (Li et al. 2010; Guan and Jiang 2018). Via pilot studies, we found that 100 CIFAR images or 60 MMLU questions were a reasonable number of decisions to make within 20-40 minutes (a typical time limit for an online study), which we use throughout our experiments.

6 Computational Evaluation

Decision-makers may have different "expertise" (i.e., strengths and weaknesses) across input space $\mathcal{X}$ under each form of support. We evaluate `Modiste` using simulated human behavior to capture diverse decision-maker expertise.

Expertise Profiles

We can capture an individual h's expertise via an *expertise profile*, which is defined over the input space $\mathcal{X}$. We divide $\mathcal{X}$ into disjoint regions (i.e., $\mathcal{X} = \cup_{j \in [N]} \mathcal{X}_j$); these regions could be defined by class labels or by covariates, depending on the task.[3] We let $r_{\mathrm{A}_i}(\mathcal{X}_j; h)$ denote h's average prediction error under support A_i across region $\mathcal{X}_j$.

Human-informed synthetic decision-makers. To construct expertise profiles with *realistic* values for each $r_{\mathrm{A}_i}(\mathcal{X}_j; h)$, we collect data on user decisions across different users and then calculate individual $r_{\mathrm{A}_i}(\mathcal{X}_j; h)$. The set of participant expertise profiles form a population of decision-makers that we refer to as *human-informed synthetic decision-makers*. From the estimated $r_{\mathrm{A}_i}(\mathcal{X}_j; h)$ of each human-informed synthetic decision-maker, we can simulate decision-maker behavior.

To construct human-informed synthetic decision-makers, we recruited 20 participants from Prolific (10 for CIFAR-$3A$ and 10 for MMLU-$2A$). We use the same recruitment scheme as the larger human subject experiment described in the Appendix. We define regions of expertise over class labels for CIFAR-$3A$ and over question topics for MMLU-$2A$, as we expect $r_{\mathrm{A}_i}(x; h)$ to be roughly constant for $x \in \mathcal{X}_j$ where $\mathcal{X}_j$ is defined by a class label or question topic. We showed each participant similar inputs with different forms of support to estimate $r_{\mathrm{A}_i}(\mathcal{X}_j; h)$ for each support A_i in each region $\mathcal{X}_j$. On each trial, each participant is randomly assigned a form of support; trials are approximately balanced by the type of support and grouping (i.e., topic or class). We compute participant accuracy averaged over all trials: 100 for CIFAR-$3A$, 60 for MMLU-$2A$. We denote expertise profiles as follows: if there were three regions in the input space, an individual's expertise profile under support A_i would be written as $r_{\mathrm{A}_i} = [0.7, 0.1, 0.7]$, meaning the individual incurs a loss of 0.7 on $\mathcal{X}_1$, 0.1 on $\mathcal{X}_2$, and 0.7 on $\mathcal{X}_3$.

Policies for each profile. Based our pilot study, we define the three expertise profiles and what kind of decision support policy we expect to be learned for each:

- **Approximately Invariant** expertise across all the regions under different forms of support, i.e., $r_{\mathrm{A}_1}(\mathcal{X}_j; h) \approx r_{\mathrm{A}_2}(\mathcal{X}_j; h) \approx \cdots \approx r_{\mathrm{A}_k}(\mathcal{X}_j; h), \forall j \in [N]$. Both a random decision support policy and a policy that selects a fixed form of support would suffice for such a profile.
- **Varying** expertise where a decision-maker excels in some areas but benefits from support in areas beyond their training (Schvaneveldt et al. 1985), i.e., $r_{\mathrm{A}_1}(\mathcal{X}_j; h) \leq r_{\mathrm{A}_2}(\mathcal{X}_j; h)$ and $r_{\mathrm{A}_2}(\mathcal{X}_k; h) \leq r_{\mathrm{A}_1}(\mathcal{X}_k; h)$, for some $j, k \in [N]$. For this expertise profile, we expect the decision support policy to select different forms of support in different regions. The quantity of $|r_{\mathrm{A}_1}$ - $r_{\mathrm{A}_2}|$ for a region will dictate how efficiently the policy can be learned.
- **Strictly Better** expertise (e.g., $\mathrm{A}_1 \succ \mathrm{A}_2 \succ \cdots \succ \mathrm{A}_k$) that is uniformly maintained across all the regions, i.e., $r_{\mathrm{A}_1}(\mathcal{X}_j; h) \leq r_{\mathrm{A}_2}(\mathcal{X}_j; h) \leq \cdots \leq r_{\mathrm{A}_k}(\mathcal{X}_j; h), \forall j \in [N]$. A decision support policy should learn the fixed form of support to use for all inputs.

Per the task design, we find participants generally only display varying expertise profiles on CIFAR-$3A$ while we find instances of all three expertise profiles on MMLU-$2A$.

When is Personalization Useful?

We investigate how personalized policies compare against offline baselines under each expertise profile (Table 1). We verify that learning decision support policies are not helpful

[3] While we instantiate decision-makers this way, `Modiste` does not take the expertise profiles or how they were constructed (e.g., the regions) as input.

Algorithm	Invariant	Strictly Better	Varying
H-ONLY	0.00 ± 0.01	0.09 ± 0.08	0.50 ± 0.06
H-MODEL	0.00 ± 0.01	0.22 ± 0.19	0.35 ± 0.05
H-CONSENSUS	0.00 ± 0.01	0.23 ± 0.13	0.27 ± 0.08
Population	0.00 ± 0.02	0.18 ± 0.08	0.15 ± 0.03
Modiste-LinUCB	0.00 ± 0.01	0.17 ± 0.05	0.19 ± 0.05
Modiste-KNN	0.00 ± 0.01	0.06 ± 0.01	$\mathbf{0.08 \pm 0.02}$

Algorithm	Invariant	Strictly Better	Varying
H-ONLY	0.01 ± 0.01	0.18 ± 0.17	0.22 ± 0.12
H-LLM	0.01 ± 0.01	0.18 ± 0.21	0.12 ± 0.17
Population	0.00 ± 0.02	0.19 ± 0.07	0.12 ± 0.09
Modiste-LinUCB	0.00 ± 0.01	0.12 ± 0.03	0.07 ± 0.04
Modiste-KNN	0.01 ± 0.01	0.05 ± 0.03	$\mathbf{0.05 \pm 0.03}$

Table 1: We evaluate Modiste across three expertise profiles. We compute the average excess loss $L_h(\pi) - L_h^{opt}$ (lower is better), and standard deviation across individuals in each expertise profile for both CIFAR-$3A$ (Left) and MMLU-$2A$ (Right). $L_h(\pi)$ is computed by averaging across the last 10 steps of 100 total time steps. We **bold** the variant with the lowest excess loss that is statistically significant from the other variants. Note that this only occurs in the "varying" expertise setting.

for decision-makers with "invariant" expertise profiles. For individuals who fall under the "varying" profiles, we find at least one personalized policy outperforms offline policies and learns a policy that is significantly closer to the decision-maker's optimal performance. This is because a personalized policy identifies *which* form of support is better in each context, compared to fixed offline policies which *always* show one form of support or to the population-wide variant, which may not provide the correct form of support to each individual. For individuals in the "strictly better" profile, while we do not find a statistically significant difference from the baselines due to the large variance in fixed policies (e.g., they work well for some decision-makers but poorly for others), we observe that much smaller variance with Modiste, particularly using KNN. For most individuals, the population-wide baseline performs poorly, emphasizing the need for personalization of decision support. Misalignment between the population policy and the optimal policy of the new decision-maker demonstrably leads to an ineffective use of decision support. We note that KNN generally outperforms LinUCB, the latter of which can be saddled by its implicit linearity assumption. We further study the effect of various parameters, e.g., exploration parameters, KNN parameters, embedding size, and the number of interactions in the Appendix.

7 Modiste with Real Users

To validate whether Modiste can improve decision-maker performance in practice, we run a series of human subject experiments (i.e., ethics-reviewed studies with real human participants). We first introduce the set-up of our user study; additional information can be found in the Appendix.

Recruitment. We recruit a total of 80 crowdsourced participants from Prolific to interact with Modiste ($N = 30$ and $N = 50$ for CIFAR-$3A$ and MMLU-$2A$, respectively). We recruit more participants for MMLU-$2A$, as we expect greater individual differences in regions where support is needed, e.g., some participants may be good at mathematics and struggle in biology, whereas others may excel in biology.

Each participant is assigned to only one task. Within a task, participants are randomly assigned to one algorithm variant; an equal number of participants are included per variant (i.e., 10 for MMLU and 5 for CIFAR). Participants are required to reside in the United States and speak English as a first language. Participants are paid at a base rate of \$9/hr, and are told they may be paid an optional bonus up to \$10/hr based on the number of correct responses. We allot 25-30 minutes for the CIFAR task and 30-40 for MMLU, as each MMLU question takes more effort. We applied the bonus to all participants in all studies. We run an ANOVA with Tukey HSD across the conditions for each task.

Modiste outperforms baselines for "varying" expertise profiles. By design, the CIFAR-$3A$ task compels "varying" profiles: Modiste's forte. This is reflected quantitatively in Figure 4 (left), where both Modiste variants have lower expected losses than any of the offline policies. In particular, KNN achieves statistically significantly lower expected loss over all variants ($p < 0.001$ across all pairs), including the population-wide policy learned from the pilot study. When visualizing the learned decision support policies, we observe that Modiste can reconstruct near-optimal policies, as depicted in the Appendix.

Modiste matches the best baseline for "strictly better" profiles. Polymaths are rare; we observe no different in our human subject experiments on MMLU, as most participants are better with LLM access, which places them in the "strictly better" category. While participants with both Modiste variants outperform the H-ONLY baseline ($p < 0.01$), we observe that on average Modiste settings are no different than H-LLM, the condition where people always have access to the LLM, as shown in Figure 4 (right). This confirms our hypothesis from the computational experiments. Further, while the average performance of Modiste variants is similar to that of the offline fixed or population-wide policies in Figure 4 (right), the variance is significantly smaller—particularly with KNN.

Modiste can facilitate appropriate use of decision support. Since the LLM only excels at three of the four topics on the MMLU task, Modiste learns in many cases to defer to human judgment on Mathematics questions, particularly when the human has strong expertise. In Figure 5, we visualize the learned decision support policies of various individuals in the study and illustrate how Modiste yields policies that provide support on different topics for different decision-makers.

Limitations. In this work, we consider the classification setting where we get immediate feedback (e.g., we can calculate the loss to update π). Future work can consider more complex decision-making tasks that may require extending to a delayed feedback setting or to a different cognitive task

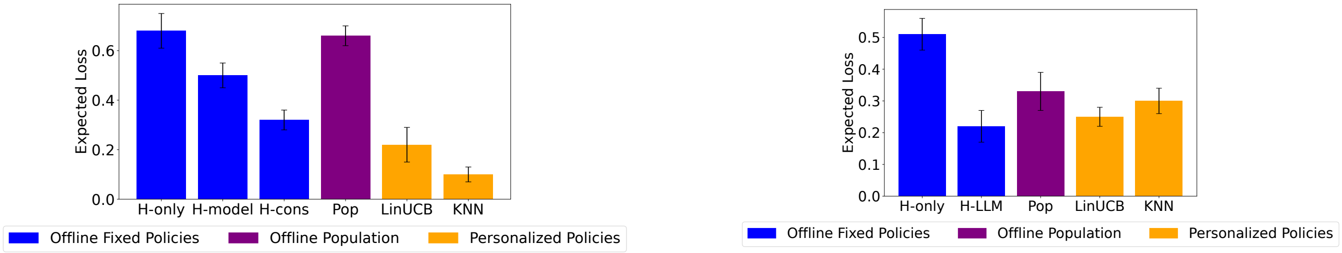


Figure 4: We report expected average loss $L_h(\pi)$ (lower is better) and standard error in the last 10 trials by Prolific participants for each algorithm, with CIFAR conditions on the left and MMLU conditions on the right. In the CIFAR setting, where individuals typically exhibit "varying" expertise profiles, we see significant benefits from using `Modiste`, particularly in the KNN setting. While we observe that most individuals in the MMLU condition exhibit "strictly better" expertise, which means personalized policies typically only perform as well as the best baseline, we still observe instances of deferred decisions to the human on a case-by-case basis—see Figure 5.

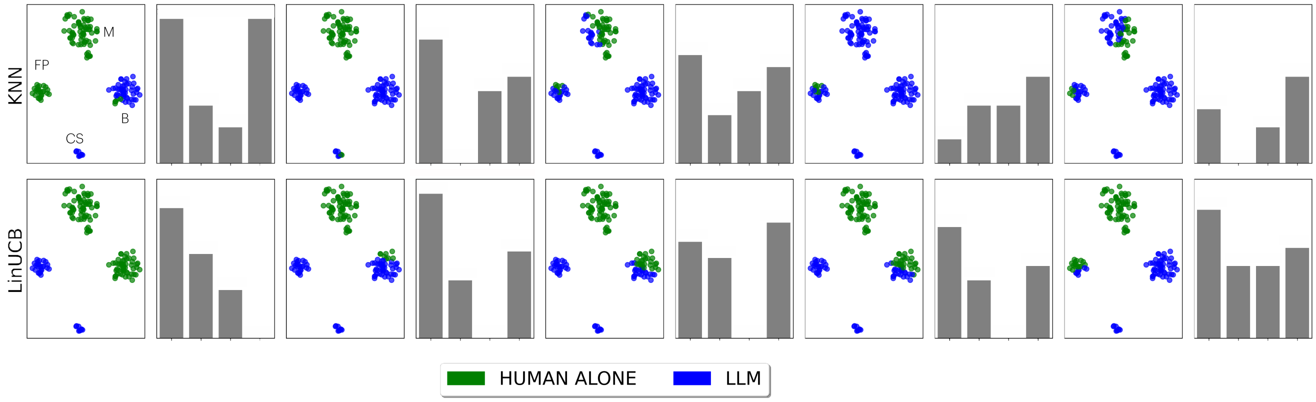


Figure 5: Snapshots of the learned decision support policies computed at the end of the study for 10 participants on the MMLU task. The forms of support are colored in t-SNE embedding space. All participants exhibit distinct policies across input space. The bar plot to the right of each scatter plot shows the relative performance of that decision-maker *alone* in each category, ordered from left to right as M=Mathematics, B=Biology, CS=Computer Science, FP=Foreign Policy per subplot. When a decision-maker performs well alone, `Modiste` learns policies to empower that decision-maker without LLM access. For example, the individual in the top left is highly competent at both Mathematics and Foreign Policy; the learned decision support policy reflects this.

(e.g., planning or perception). Though `Modiste` is promising, we note that significant issues can arise when decision-makers blindly rely on decision support (Buçinca et al. 2020; Chen et al. 2023), especially when the support is erroneous or ineffective; such over-reliance requires careful attention to prevent. Further, our problem definition hinges on domain experts defining the available forms of support (i.e., we need a clearly defined $\mathcal{A}$ to use `Modiste`). In practice, this may prove difficult, as one may not know how to define specific forms of support or decision-makers may have access to varying support sets.

8 Conclusion

A decision support policy captures when and which form of support should be provided to improve a decision-maker's performance. The selective use of AI-based decision support helps instantiate the "appropriate use" clauses in emerging regulation (Biden 2023), as we only provide AI assistance to decision-makers as and when it is beneficial to them. We introduce `Modiste`, an interactive tool for learning a decision support policy for each decision-maker using contextual bandits. To the best of our knowledge, we are the first to learn and validate such a policy online for unseen decision-makers. Within our `Modiste` interface, we instantiate two variants of Algorithm 1 using existing stochastic contextual bandit tools, namely LinUCB and online KNN. Our computational and human subject experiments highlight the importance—and feasibility—of personalizing decision support policies for individual decision-makers. Our human subject experiments show promise, as we personalize decision support policies in few iterations yet find nuances in decision-makers' need for support: some unskilled decision-makers uniformly benefit from LLM access, while others only need LLMs for some tasks. While encouraging rich cross-talk between domain experts and practitioners, future work integrating `Modiste` into existing decision-making workflows would pave a route towards responsible use of AI as decision support.

Acknowledgements

UB acknowledges support from DeepMind and the Leverhulme Trust via the Leverhulme Centre for the Future of Intelligence (CFI). AW acknowledges support from a Turing AI Fellowship under grant EP/V025279/1, EPSRC grants EP/V056522/1 and EP/V056883/1, and the Leverhulme Trust via CFI. KMC acknowledges support from a Marshall Scholarship and the Cambridge Trust. PK acknowledges support from The Alan Turing Institute. AT acknowledges support from the National Science Foundation grants IIS1705121, IIS1838017, IIS2046613, IIS2112471, an Amazon Web Services Award, a Facebook Faculty Research Award, funding from Booz Allen Hamilton Inc., and a Block Center Grant. This work is supported (in part) by ELSA - European Lighthouse on Secure and Safe AI funded by the European Union under grant agreement No. 101070617. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of any of these funding agencies, European Union or European Commission. We thank the Prolific participants and our volunteer pilot participants. We thank Alan Liu and Kartik Chandra for their invaluable input on designing our web-based server.

References

2023. “Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts” (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD).

Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; and Mané, D. 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.

Babbar, V.; Bhatt, U.; and Weller, A. 2022. On the Utility of Prediction Sets in Human-AI Teams. In Raedt, L. D., ed., *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 2457–2463. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Bansal, G.; Wu, T.; Zhou, J.; Fok, R.; Nushi, B.; Kamar, E.; Ribeiro, M. T.; and Weld, D. 2021. Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16.

Barabas, C. 2022. Refusal in Data Ethics: Re-Imagining the Code Beneath the Code of Computation in the Carceral State. *Engaging Science, Technology, and Society*, 8(2): 35–57.

Bastani, H.; Bastani, O.; and Sinchaisri, P. 2022. Improving Human Decision-Making with Machine Learning. In *Academy of Management Proceedings*, volume 2022, 17725. Academy of Management Briarcliff Manor, NY 10510.

Bastani, H.; and Bayati, M. 2020. Online decision making with high-dimensional covariates. *Operations Research*, 68(1): 276–294.

Biden, J. R. 2023. *Executive Order 14110, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. The White House.

Briggs, G. M.; Flynn, P. A.; Worthington, M.; Rennie, I.; and McKinstry, C. 2008. The role of specialist neuroradiology second opinion reporting: is there added value? *Clinical radiology*, 63(7): 791–795.

Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; et al. 2018. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

Buçinca, Z.; Lin, P.; Gajos, K. Z.; and Glassman, E. L. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 454–464.

Buçinca, Z.; Swaroop, S.; Paluch, A. E.; Murphy, S. A.; and Gajos, K. Z. 2024. Towards Optimizing Human-Centric Objectives in AI-Assisted Decision-Making With Offline Reinforcement Learning. *arXiv preprint arXiv:2403.05911*.

Charusaie, M.-A.; Mozannar, H.; Sontag, D.; and Samadi, S. 2022. Sample Efficient Learning of Predictors that Complement Humans. In *International Conference on Machine Learning*, 2972–3005. PMLR.

Chen, V.; Liao, Q. V.; Wortman Vaughan, J.; and Bansal, G. 2023. Understanding the role of human intuition on reliance in human-AI decision-making with explanations. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2): 1–32.

Collins, K. M.; Jiang, A. Q.; Frieder, S.; Wong, L.; Zilka, M.; Bhatt, U.; Lukasiewicz, T.; Wu, Y.; Tenenbaum, J. B.; Hart, W.; et al. 2024a. Evaluating language models for mathematics through interactions. *Proceedings of the National Academy of Sciences*, 121(24): e2318124121.

Collins, K. M.; Sucholutsky, I.; Bhatt, U.; Chandra, K.; Wong, L.; Lee, M.; Zhang, C. E.; Zhi-Xuan, T.; Ho, M.; Mansinghka, V.; et al. 2024b. Building machines that learn and think with people. *Nature Human Behaviour*, 8(10): 1851–1863.

De-Arteaga, M.; Dubrawski, A.; and Chouldechova, A. 2018. Learning under selective labels in the presence of expert consistency. *arXiv preprint arXiv:1807.00905*.

Gao, R.; Saar-Tsechansky, M.; De-Arteaga, M.; Han, L.; Lee, M. K.; and Lease, M. 2021. Human-AI Collaboration with Bandit Feedback. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 1722–1728. International Joint Conferences on Artificial Intelligence Organization.

Gao, R.; Saar-Tsechansky, M.; De-Arteaga, M.; Han, L.; Sun, W.; Lee, M. K.; and Lease, M. 2023. Learning Complementary Policies for Human-AI Teams. *arXiv preprint arXiv:2302.02944*.

Gordon, E.; and Mugar, G. 2020. *Meaningful inefficiencies: designing for public value in an age of digital expediency*. New York, NY: Oxford University Press. ISBN 978-0-19-087017-1 978-0-19-087016-4 978-0-19-087015-7.

Green, B.; and Chen, Y. 2019. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–24.

Guan, M.; and Jiang, H. 2018. Nonparametric stochastic contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Hemmer, P.; Schellhammer, S.; Vössing, M.; Jakubik, J.; and Satzger, G. 2022. Forming Effective Human-AI Teams: Building Machine Learning Models that Complement the Capabilities of Multiple Experts. In *Proceedings of the Thirtieth International Conference on International Joint Conferences on Artificial Intelligence*.

Hendrycks, D.; Burns, C.; Basart, S.; Zou, A.; Mazeika, M.; Song, D.; and Steinhardt, J. 2020. Measuring Massive Multitask Language Understanding. In *International Conference on Learning Representations*.

Kahn Jr, C. E. 1994. Artificial intelligence in radiology: decision support systems. *Radiographics*, 14(4): 849–861.

Keen, P. G. 1980. Decision support systems: a research perspective. In *Decision support systems: Issues and challenges: Proceedings of an international task force meeting*, 23–44.

Keswani, V.; Lease, M.; and Kenthapadi, K. 2021. Towards unbiased and accurate deferral to multiple experts. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 154–165.

Kirk, H. R.; Whitefield, A.; Röttger, P.; Bean, A.; Margatina, K.; Ciro, J.; Mosquera, R.; Bartolo, M.; Williams, A.; He, H.; et al. 2024. The PRISM Alignment Project: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models. *arXiv preprint arXiv:2404.16019*.

Krizhevsky, A. 2009. Learning Multiple Layers of Features from Tiny Images. *Master's thesis, University of Toronto*.

Lai, V.; Chen, C.; Smith-Renner, A.; Liao, Q. V.; and Tan, C. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1369–1385.

Laidlaw, C.; and Russell, S. 2021. Uncertain Decisions Facilitate Better Preference Learning. *Advances in Neural Information Processing Systems*, 34: 15070–15083.

Lee, M.; Srivastava, M.; Hardy, A.; Thickstun, J.; Durmus, E.; Paranjape, A.; Gerard-Ursin, I.; Li, X. L.; Ladhak, F.; Rong, F.; Wang, R. E.; Kwon, M.; Park, J. S.; Cao, H.; Lee, T.; Bommasani, R.; Bernstein, M. S.; and Liang, P. 2023. Evaluating Human-Language Model Interaction. *Transactions on Machine Learning Research*.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, 661–670.

Ma, S.; Lei, Y.; Wang, X.; Zheng, C.; Shi, C.; Yin, M.; and Ma, X. 2023. Who should i trust: Ai or myself? leveraging human and ai correctness likelihood to promote appropriate trust in ai-assisted decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19.

Madras, D.; Pitassi, T.; and Zemel, R. 2018. Predict responsibly: improving fairness and accuracy by learning to defer. *Advances in Neural Information Processing Systems*, 31.

Mozannar, H.; Lee, J.; Wei, D.; Sattigeri, P.; Das, S.; and Sontag, D. 2023. Effective Human-AI Teams via Learned Natural Language Rules and Onboarding. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 30466–30498. Curran Associates, Inc.

Mozannar, H.; and Sontag, D. 2020. Consistent estimators for learning to defer to an expert. In *International Conference on Machine Learning*, 7076–7087. PMLR.

Okati, N.; De, A.; and Rodriguez, M. 2021. Differentiable learning under triage. *Advances in Neural Information Processing Systems*, 34: 9140–9151.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback.

Palan, S.; and Schitter, C. 2018. Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17: 22–27.

Phillips-Wren, G. 2012. AI tools in decision making support systems: a review. *International Journal on Artificial Intelligence Tools*, 21(02): 1240005.

Roda, C. E. 2011. *Human attention and its implications for human-computer interaction.* Cambridge University Press.

Scheife, R. T.; Hines, L. E.; Boyce, R. D.; Chung, S. P.; Momper, J. D.; Sommer, C. D.; Abernethy, D. R.; Horn, J. R.; Sklar, S. J.; Wong, S. K.; et al. 2015. Consensus recommendations for systematic evaluation of drug–drug interaction evidence for clinical decision support. *Drug safety*, 38(2): 197–206.

Schvaneveldt, R. W.; Durso, F. T.; Goldsmith, T. E.; Breen, T. J.; Cooke, N. M.; Tucker, R. G.; and De Maio, J. C. 1985. Measuring the structure of expertise. *International journal of man-machine studies*, 23(6): 699–728.

Seger, C. A.; and Peterson, E. J. 2013. Categorization= decision making+ generalization. *Neuroscience & Biobehavioral Reviews*, 37(7): 1187–1200.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction.*

Swaroop, S.; Buçinca, Z.; Gajos, K. Z.; and Doshi-Velez, F. 2024. Accuracy-Time Tradeoffs in AI-Assisted Decision Making under Time Pressure. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 138–154.

Yu, F.; Moehring, A.; Banerjee, O.; Salz, T.; Agarwal, N.; and Rajpurkar, P. 2024. Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nature Medicine*, 30(3): 837–849.

Zhang, Y.; Liao, Q. V.; and Bellamy, R. K. 2020. 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*, 295–305.