# A Model of ChatGPT

Tom Cunningham

2025-06-14

This paper develops a simple model of human and AI ability to answer questions. Each question $\boldsymbol{q}$ is a high-dimensional vector, with a true scalar answer $a$. An agent's estimate of the answer is an interpolation based on previously-seen questions and answers $(\boldsymbol{q}^i, a^i)_{i=1,\dots,n}$. This framework extends an earlier model developed for a different purpose.

The model yields several implications:

1. **The quality of an answer to a new question depends on its distance from the training set.** For a new question $\boldsymbol{q}$, the expected error is a function of the distance between $\boldsymbol{q}$ and the training set $\boldsymbol{Q}$.
2. **The quality of answers increases with the size of the training set.** The expected error decreases linearly with the number of linearly-independent examples in the training set.
3. **The value of advice from another agent depends on the distance between their training sets.**

This framework can be interpreted as a model of an agent, "the user," who must provide an estimate for the answer to a question $\boldsymbol{q}$ and can choose whether to consult an AI model like ChatGPT. The key components of the model are:

1. **The dimensionality of the question ($p$).** A higher-dimensional problem may be more costly to enter into the AI, but it also increases the potential benefit.
2. **The public information set.** These are the training questions that the AI has observed, which we can conceptualize as the corpus of public knowledge (e.g., the internet).
3. **The private information set.** These are the questions that the user has personally encountered and for which they have observed the true answer.

A user will consult the AI if and only if the expected improvement in their answer exceeds the associated cost. The model predicts that an AI will be most useful for questions with components that are novel to the user but contained within the AI's public training data.

This leads to several corollaries:

1. An AI will not be used for questions the user has encountered before.
2. An AI is more likely to be used for domains with higher *latent* dimensionality ($p$).
3. An AI is more likely to be used for domains with lower *surface* dimensionality, as this reduces the cost of specifying the question.
4. An AI is more likely to be used by humans with less experience in a domain (i.e., smaller $n_{\text{private}}$).

We can make some conjectures about adoption by occupation:
We can make some conjectures about adoption by task:
Additional things to add:

1. **High-dimensional answers.** Our model assumes *scalar* answers. In fact ChatGPT gives high-dimensional outputs. I think we can say some nice things here.
2. **Tacit knowledge.** ChatGPT will be more likely to be used for domains where humans have tacit knowledge.

Table 1: Conjectures about adoption by occupation

| **Occupation** | **Predicted ChatGPT use** | **Reason** |
|---|---|---|
| software engineer | high | many novel discrete problems, similar to those on |
| software engineer - idiosyncratic language | low | many novel discrete problems, not similar to thos |
| physician | high | many novel discrete problems, similar to those on |
| contact center worker | low | novel problems, but not similar to those on the i |
| architect | low | novel problems, not discrete, not text-based |
| manual worker | low | not not text-based |

Table 2: Conjectures about adoption by task

| **Task** | **Predicted ChatGPT use** | **Reason** |
|---|---|---|
| Intellectual curiosity | high | novel discrete problem, simil |
| Diagnosing medical problems | high | novel discrete problem, simil |
| Problems with widely-adopted systems (car, house, computer) | high | novel discrete problem, simil |
| Problems with idiosyncratic systems (custom setups) | low | novel discrete problem, *not* |

# 1 Model

**The State of the World and Questions.** The state of the world is defined by a vector of $p$ unobserved parameters, $\boldsymbol{w} \in \mathbb{R}^p$. A question is a vector of $p$ binary features, $\boldsymbol{q} \in \{-1, 1\}^p$. The true answer to a question $\boldsymbol{q}$ is a scalar $a$ determined by the linear relationship:

$$a = \boldsymbol{q}'\boldsymbol{w} = \sum_{k=1}^{p} q_k w_k$$

**Agents and Information.** There is a set of agents, indexed by $i \in \mathcal{I}$. Each agent $i$ possesses an information set $\mathcal{D}_i$, which consists of $n_i$ questions they have previously encountered, along with their true answers. We can represent this information as a pair $(\boldsymbol{Q}_i, \boldsymbol{a}_i)$:

- $\boldsymbol{Q}_i$ is an $n_i \times p$ matrix where each row is a question vector. Let the $j$-th question for agent $i$ be $\boldsymbol{q}'_{i,j}$, so that:
$$\boldsymbol{Q}_i = \begin{bmatrix} \boldsymbol{q}'_{i,1} \\ \vdots \\ \boldsymbol{q}'_{i,n_i} \end{bmatrix} = \begin{bmatrix} q_{i,1,1} & \cdots & q_{i,1,p} \\ \vdots & \ddots & \vdots \\ q_{i,n_i,1} & \cdots & q_{i,n_i,p} \end{bmatrix}$$
- $\boldsymbol{a}_i$ is an $n_i \times 1$ vector of the corresponding answers. The answers are generated according to the true model:
$$\boldsymbol{a}_i = \boldsymbol{Q}_i \boldsymbol{w}$$

**Beliefs.** All agents share a common prior belief about the state of the world, assuming the weights $\boldsymbol{w}$ are drawn from a multivariate Gaussian distribution:

$$\boldsymbol{w} \sim N(\boldsymbol{0}, \Sigma)$$

where $\Sigma$ is a $p \times p$ positive-semidefinite covariance matrix. A common assumption we will use is an isotropic prior, where $\Sigma = \sigma^2 \boldsymbol{I}_p$ for some scalar $\sigma^2 > 0$. This implies that, a priori, the weights are uncorrelated and have equal variance.

Given their information set $\mathcal{D}_i$, agent $i$ forms a posterior belief about $\boldsymbol{w}$. When a new question $\boldsymbol{q}_{\text{new}}$ arises, the agent uses their posterior distribution to form an estimate of the answer, $\hat{a}_{\text{new}} = \boldsymbol{q}'_{\text{new}} \mathbb{E}[\boldsymbol{w} \mid \mathcal{D}_i]$.

## 2 Propositions

**Proposition 1** (Posterior over $\boldsymbol{w}$ given $\boldsymbol{Q}$ and $\boldsymbol{a}$)**.** *The agent's posterior mean and variance will be:*

$$\hat{\boldsymbol{w}} = \Sigma \boldsymbol{Q}^\top (\boldsymbol{Q}\Sigma\boldsymbol{Q}^\top)^{-1}\boldsymbol{a}$$
$$\Sigma_{|a} = \Sigma - \Sigma \boldsymbol{Q}^\top (\boldsymbol{Q}\Sigma\boldsymbol{Q}^\top)^{-1}\boldsymbol{Q}\Sigma.$$

*Proof.* The derivation follows from the standard formula for conditional Gaussian distributions. We begin by defining the joint distribution of the weights $\boldsymbol{w}$ and the answers $\boldsymbol{a}$. The weights and answers are jointly Gaussian:

$$\begin{pmatrix}\boldsymbol{w}\\ \boldsymbol{a}\end{pmatrix} \sim N\left(\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix}, \begin{pmatrix}\Sigma & \Sigma\boldsymbol{Q}'\\ \boldsymbol{Q}\Sigma & \boldsymbol{Q}\Sigma\boldsymbol{Q}'\end{pmatrix}\right)$$

where the covariance terms are derived as follows:

- $Cov(\boldsymbol{w}, \boldsymbol{w}) = \Sigma$ (prior covariance)
- $Cov(\boldsymbol{a}, \boldsymbol{a}) = Cov(\boldsymbol{Q}\boldsymbol{w}, \boldsymbol{Q}\boldsymbol{w}) = \boldsymbol{Q}Cov(\boldsymbol{w}, \boldsymbol{w})\boldsymbol{Q}' = \boldsymbol{Q}\Sigma\boldsymbol{Q}'$
- $Cov(\boldsymbol{w}, \boldsymbol{a}) = Cov(\boldsymbol{w}, \boldsymbol{Q}\boldsymbol{w}) = Cov(\boldsymbol{w}, \boldsymbol{w})\boldsymbol{Q}' = \Sigma\boldsymbol{Q}'$

The conditional mean $E[\boldsymbol{w}|\boldsymbol{a}]$ is given by the formula:

$$E[\boldsymbol{w}|\boldsymbol{a}] = E[\boldsymbol{w}] + Cov(\boldsymbol{w}, \boldsymbol{a})Var(\boldsymbol{a})^{-1}(\boldsymbol{a} - E[\boldsymbol{a}])$$

Substituting the values from our model ($E[\boldsymbol{w}] = \boldsymbol{0}$, $E[\boldsymbol{a}] = \boldsymbol{0}$):

$$\hat{\boldsymbol{w}} = \boldsymbol{0} + (\Sigma\boldsymbol{Q}')(\boldsymbol{Q}\Sigma\boldsymbol{Q}')^{-1}(\boldsymbol{a} - \boldsymbol{0}) = \Sigma\boldsymbol{Q}'(\boldsymbol{Q}\Sigma\boldsymbol{Q}')^{-1}\boldsymbol{a}$$

This gives us the posterior mean of the weights. The posterior covariance is given by:

$$Var(\boldsymbol{w}|\boldsymbol{a}) = Var(\boldsymbol{w}) - Cov(\boldsymbol{w}, \boldsymbol{a})Var(\boldsymbol{a})^{-1}Cov(\boldsymbol{a}, \boldsymbol{w}) = \Sigma - \Sigma\boldsymbol{Q}'(\boldsymbol{Q}\Sigma\boldsymbol{Q}')^{-1}\boldsymbol{Q}\Sigma.$$

□

**Proposition 2** (Expected error for a given question)**.** *The expected squared error for a new question $\boldsymbol{q}$ is:*

$$\mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] = \boldsymbol{q}'\Sigma_{|a}\boldsymbol{q}$$

*For an isotropic prior where $\Sigma = \sigma^2\boldsymbol{I}$, the error is proportional to the squared distance of $\boldsymbol{q}$ from the subspace spanned by the previously seen questions in $\boldsymbol{Q}$:*

$$\mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] = \sigma^2\|(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Q}})\boldsymbol{q}\|^2$$

*where $\boldsymbol{P}_{\boldsymbol{Q}}$ is the projection matrix onto the row-span of $\boldsymbol{Q}$.*

*Proof.* The prediction error is $\boldsymbol{q}'\boldsymbol{w} - \boldsymbol{q}'\hat{\boldsymbol{w}} = \boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}})$. The expected squared error is the variance of this prediction error.

$$\begin{aligned}\mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] &= \mathbb{E}[\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}})(\boldsymbol{w} - \hat{\boldsymbol{w}})'\boldsymbol{q}]\\ &= \boldsymbol{q}'\mathbb{E}[(\boldsymbol{w} - \hat{\boldsymbol{w}})(\boldsymbol{w} - \hat{\boldsymbol{w}})']\boldsymbol{q}\\ &= \boldsymbol{q}'Var(\boldsymbol{w} \mid \boldsymbol{a})\boldsymbol{q} = \boldsymbol{q}'\Sigma_{|a}\boldsymbol{q}\end{aligned}$$

This proves the first part of the proposition. For the second part, we assume an isotropic prior $\Sigma = \sigma^2\boldsymbol{I}$. Substituting this into the expression for $\Sigma_{|a}$ from Proposition 1:

$$\begin{aligned}\Sigma_{|a} &= \sigma^2\boldsymbol{I} - (\sigma^2\boldsymbol{I})\boldsymbol{Q}'(\boldsymbol{Q}(\sigma^2\boldsymbol{I})\boldsymbol{Q}')^{-1}\boldsymbol{Q}(\sigma^2\boldsymbol{I})\\ &= \sigma^2\boldsymbol{I} - \sigma^4\boldsymbol{Q}'(\sigma^2\boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}\\ &= \sigma^2\boldsymbol{I} - \sigma^4(\sigma^2)^{-1}\boldsymbol{Q}'(\boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}\\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{Q}'(\boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q})\end{aligned}$$

Let $\boldsymbol{P_Q} = \boldsymbol{Q}'(\boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}$, which is the projection matrix onto the row space of $\boldsymbol{Q}$. Then $\Sigma_{|a} = \sigma^2(\boldsymbol{I} - \boldsymbol{P_Q})$. The expected squared error is:

$$\mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] = \boldsymbol{q}'\sigma^2(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q} = \sigma^2\boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}$$

Since $\boldsymbol{I} - \boldsymbol{P_Q}$ is an idempotent projection matrix, $\boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q} = \boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q} = \|(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}\|^2$. Thus,

$$\mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] = \sigma^2\|(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}\|^2$$

□

**Proposition 3** (Error decreases with more independent questions)**.** *The average expected squared error over all possible new questions $\boldsymbol{q}$ decreases linearly with the number of linearly independent questions in the training set $\boldsymbol{Q}$. Specifically, with an isotropic prior $\Sigma = \sigma^2\boldsymbol{I}$, the average error is:*

$$\mathbb{E}_{\boldsymbol{q}}[\textit{error}(\boldsymbol{q})] = \sigma^2(p - \text{rank}(\boldsymbol{Q}))$$

*where the expectation is taken over new questions $\boldsymbol{q}$ with i.i.d. components drawn uniformly from $\{-1, 1\}$.*

*Proof.* The proof proceeds in two steps. First, we write the expression for the error for a given new question $\boldsymbol{q}$. Second, we average this error over the distribution of all possible questions.

1. **Predictive error for a fixed $\boldsymbol{q}$.** From Proposition 2, the expected squared error for a specific new question $\boldsymbol{q}$, given an isotropic prior $\Sigma = \sigma^2\boldsymbol{I}$, is:

   $$\text{error}(\boldsymbol{q}) = \mathbb{E}[(\boldsymbol{q}'(\boldsymbol{w} - \hat{\boldsymbol{w}}))^2] = \sigma^2\boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}$$

   where $\boldsymbol{P_Q} = \boldsymbol{Q}'(\boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}$ is the projection matrix onto the row-span of $\boldsymbol{Q}$.

2. **Average over random new questions.** We now take the expectation of this error over the distribution of new questions $\boldsymbol{q}$. The components of $\boldsymbol{q}$ are i.i.d. uniform on $\{-1, 1\}$, which implies that $\mathbb{E}[\boldsymbol{q}] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{q}\boldsymbol{q}'] = \boldsymbol{I}_p$. The average error is:

   $$\begin{aligned}
   \mathbb{E}_{\boldsymbol{q}}[\text{error}(\boldsymbol{q})] &= \mathbb{E}_{\boldsymbol{q}}[\sigma^2\boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}] \\
   &= \sigma^2\mathbb{E}_{\boldsymbol{q}}[\text{tr}(\boldsymbol{q}'(\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q})] \\
   &= \sigma^2\mathbb{E}_{\boldsymbol{q}}[\text{tr}((\boldsymbol{I} - \boldsymbol{P_Q})\boldsymbol{q}\boldsymbol{q}')] \\
   &= \sigma^2\,\text{tr}((\boldsymbol{I} - \boldsymbol{P_Q})\mathbb{E}_{\boldsymbol{q}}[\boldsymbol{q}\boldsymbol{q}']) \\
   &= \sigma^2\,\text{tr}(\boldsymbol{I} - \boldsymbol{P_Q}) \\
   &= \sigma^2(\text{tr}(\boldsymbol{I}) - \text{tr}(\boldsymbol{P_Q}))
   \end{aligned}$$

   The trace of the identity matrix is $p$. The trace of a projection matrix is the dimension of the subspace it projects onto, so $\text{tr}(\boldsymbol{P_Q}) = \text{rank}(\boldsymbol{Q})$. Thus, the average error is:

   $$\mathbb{E}_{\boldsymbol{q}}[\text{error}(\boldsymbol{q})] = \sigma^2(p - \text{rank}(\boldsymbol{Q}))$$

   Since the rank of $\boldsymbol{Q}$ increases with each linearly independent question added, the average error decreases linearly until $\text{rank}(\boldsymbol{Q}) = p$, at which point it becomes zero.

□

**Proposition 4** (Posterior in two-stage estimation)**.** *We consider a two-stage process. First, an agent (the "computer," C) with training data $(\boldsymbol{Q}_C, \boldsymbol{a}_C)$ forms an estimate for the answer to a new question $\boldsymbol{q}$. Second, another agent (the "human," H) with their own training data $(\boldsymbol{Q}_H, \boldsymbol{a}_H)$ observes the computer's estimate and updates their own belief.*

*The human has a prior over the weights $\boldsymbol{w} \sim N(\boldsymbol{0}, \Sigma)$. After observing their own data, the human's posterior for $\boldsymbol{w}$ is $N(\hat{\boldsymbol{w}}_H, \Sigma_H)$, where from Proposition 1:*

$$\begin{aligned}
\hat{\boldsymbol{w}}_H &= \Sigma\boldsymbol{Q}_H^\top(\boldsymbol{Q}_H\Sigma\boldsymbol{Q}_H^\top)^{-1}\boldsymbol{a}_H \\
\Sigma_H &= \Sigma - \Sigma\boldsymbol{Q}_H^\top(\boldsymbol{Q}_H\Sigma\boldsymbol{Q}_H^\top)^{-1}\boldsymbol{Q}_H\Sigma
\end{aligned}$$

*The human's initial estimate for the answer to a new question $\boldsymbol{q}$ is $\mu_H = \boldsymbol{q}'\hat{\boldsymbol{w}}_H$ with variance $\sigma_H^2 = \boldsymbol{q}'\Sigma_H\boldsymbol{q}$.*

*The computer has its own training data $(\boldsymbol{Q}_C, \boldsymbol{a}_C)$. It provides an estimate $\hat{a}_C = \boldsymbol{q}'\hat{\boldsymbol{w}}_C$ for the true answer $a = \boldsymbol{q}'\boldsymbol{w}$. The human observes $\hat{a}_C$ and updates their posterior for $a$. We assume the computer's observations may be noisy, such that $\boldsymbol{a}_C = \boldsymbol{Q}_C\boldsymbol{w} + \boldsymbol{\epsilon}_C$ with $\boldsymbol{\epsilon}_C \sim N(0, s_C^2\boldsymbol{I})$.*

*We analyze the human's final posterior for $a$ under different assumptions about what the human knows about the computer's process.*

**Proposition 4.1** (Updating with minimal information)**.** Assume the human has no knowledge of the computer's training set $\boldsymbol{Q}_C$ but believes the computer's estimate is unbiased with a known mean squared error $\tau^2$. That is, $\hat{a}_C = a + \eta$, where $\eta \sim N(0, \tau^2)$ and is independent of $\boldsymbol{w}$.

Upon observing $\hat{a}_C$, the human's posterior for $a$ is:

$$a \mid \hat{a}_C \sim N\left(\mu_H + \alpha(\hat{a}_C - \mu_H), (1-\alpha)\sigma_H^2\right)$$

where $\alpha = \frac{\sigma_H^2}{\sigma_H^2+\tau^2} \in [0,1]$. The human's new estimate is a weighted average of their own initial estimate and the computer's estimate. The weight $\alpha$ placed on the computer's estimate is higher when the computer is believed to be more accurate (smaller $\tau^2$) or when the human's own estimate is more uncertain (larger $\sigma_H^2$).

**Proposition 4.2** (Knowledge of computer's questions)**.** Assume the human knows the computer's training questions $\boldsymbol{Q}_C$ and its noise level $s_C^2$, but not the observed answers $\boldsymbol{a}_C$.

The human can model the computer's estimate as $\hat{a}_C = \boldsymbol{q}'\boldsymbol{P}\boldsymbol{w} + \boldsymbol{q}'\boldsymbol{\zeta}$, where $\boldsymbol{P} = \Sigma\boldsymbol{Q}_C'(\boldsymbol{Q}_C\Sigma\boldsymbol{Q}_C' + s_C^2\boldsymbol{I})^{-1}\boldsymbol{Q}_C$ and $\boldsymbol{\zeta} = \Sigma\boldsymbol{Q}_C'(\boldsymbol{Q}_C\Sigma\boldsymbol{Q}_C' + s_C^2\boldsymbol{I})^{-1}\boldsymbol{\epsilon}_C$.

The pair $(a, \hat{a}_C)$ is jointly Gaussian, conditional on the human's data. The posterior for $a$ is:

$$a \mid \hat{a}_C \sim N\left(\mu_H + \kappa(\hat{a}_C - \mu_C), \sigma_H^2 - \kappa\sigma_{HC}\right)$$

where:

- $\mu_C = \boldsymbol{q}'\boldsymbol{P}\hat{\boldsymbol{w}}_H$ (human's expectation of computer's estimate)
- $\sigma_{HC} = \boldsymbol{q}'\Sigma_H\boldsymbol{P}'\boldsymbol{q}$ (covariance)
- $\sigma_C^2 = \boldsymbol{q}'\boldsymbol{P}\Sigma_H\boldsymbol{P}'\boldsymbol{q} + \boldsymbol{q}'\Sigma_\zeta\boldsymbol{q}$ (variance of computer's estimate)
- $\Sigma_\zeta = s_C^2\boldsymbol{P}\Sigma^{-1}\boldsymbol{P}'$
- $\kappa = \frac{\sigma_{HC}}{\sigma_C^2}$ (the weight on the computer's prediction error)

The weight $\kappa$ depends on the covariance structure, which is influenced by the overlap between the subspaces spanned by $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_C$.

**Proposition 4.3** (Limiting cases)**.** The framework of Proposition 4.2 nests two extreme cases:

1. **Oracle Trust:** If the human believes the computer's estimate is perfect (e.g., $s_C^2 \to 0$ and $\boldsymbol{Q}_C$ spans the relevant subspace), then $\kappa \to \sigma_H^2/(\boldsymbol{q}'\boldsymbol{P}\Sigma_H\boldsymbol{P}'\boldsymbol{q})$, and the posterior variance collapses towards zero. In the simplified Kalman model, if $\tau^2 \to 0$, then $\alpha \to 1$, and the human adopts the computer's answer, $a \mid \hat{a}_C \to N(\hat{a}_C, 0)$.
2. **Total Skepticism:** If the human believes the computer provides no information (e.g., $\sigma_{HC} \to 0$ because $\boldsymbol{Q}_C$ is irrelevant to $\boldsymbol{q}$), then $\kappa \to 0$. In the Kalman model, if $\tau^2 \to \infty$, then $\alpha \to 0$. In both cases, the human ignores the computer's estimate and reverts to their original posterior, $a \mid \hat{a}_C \sim N(\mu_H, \sigma_H^2)$.

# 3 Related Literature

## 3.1 Agrawal et al. (2018) "Exploring the Impact of Artificial Intelligence: Prediction versus Judgment"

https://www.nber.org/system/files/working_papers/w24626/w24626.pdf

### 3.2 Kleinberg et al. (2017) "Human Decisions and Machine Predictions"

## 4 References