Monoculture in Matching Markets

Kenny Peng and Nikhil Garg
Cornell Tech

arXiv:2312.09841

Algorithmic Monoculture
What happens when many decision-makers use the same algorithm to evaluate applicants?

- Two firms using different evaluations hire better applicants than when using the same evaluation (Kleinberg and Raghavan, 2021)

- Reliance on the same (or similar) algorithms can result in more <u>systemic exclusion</u> of applicants (Creel and Hellman 2022, Bommasani et al. 2022, Jain et al. 2024)

Our setup

- Colleges share true preferences over students, but rank students noisily
- Two cases:
  - Monoculture: Same evaluation
  - Polyculture: Independent evaluations
- Study this in a matching markets model

Model

- Continuum of students with **true values** v dist. according to η
- College $c \in C$ ranks student using **estimated value** $v + X_c$ $\quad (X_c \sim \mathscr{D})$
    - **Monoculture:** $X_1 = X_2 = \cdots = X_C$
    - **Polyculture:** $X_1, X_2, \cdots, X_C$ i.i.d.
- Symmetric colleges (students uniformly random preferences, colleges equal capacities)

In stable matching, does student with true value v match? To whom?

Stable Matching

"Cutoff characterization" of stable matching (Azevedo and Leshno, 2016):

- Student v can **afford** college c iff est. value $v + X_c$ exceeds **cutoff** P_c
- Student matches to favorite college they can afford
- Cutoffs $P_1, P_2, \cdots, P_C$ **market clearing** iff college capacities properly filled
- Lemma: For market-clearing cutoffs, corresponding matching is stable

Key fact

For $P_1, P_2, \cdots, P_C$ market clearing, a student with true value v matches iff $v + X_c > P_c$ for some $c \in C$.

(i.e., the student can afford at least one college.)

Equal cutoffs

Symmetry implies equal cutoffs: $P_1 = P_2 = \cdots = P_C$

Let P_{mono} denote shared cutoff under monoculture

Let P_{poly} denote shared cutoff under polyculture

Main results

Theorem 1: Exactly top students match under polyculture

Theorem 2: Everyone more likely to match to top choice under monoculture

Theorem 3: Monoculture more robust to “differential application access”

Definition 1

A noise distribution $\mathscr{D}$ is **max-concentrating** iff

$$\lim_{n\to\infty} \mathrm{Var}\left[\max\{X_1, X_2, \cdots, X_n\}\right] = 0,$$

for $X_1, X_2, \cdots, X_n \sim \mathscr{D}$.

$\mathscr{D}$ is **max-concentrating** iff

$$\lim_{n \to \infty} \text{Var}\left[\max\{X_1, X_2, \cdots, X_n\}\right] = 0$$

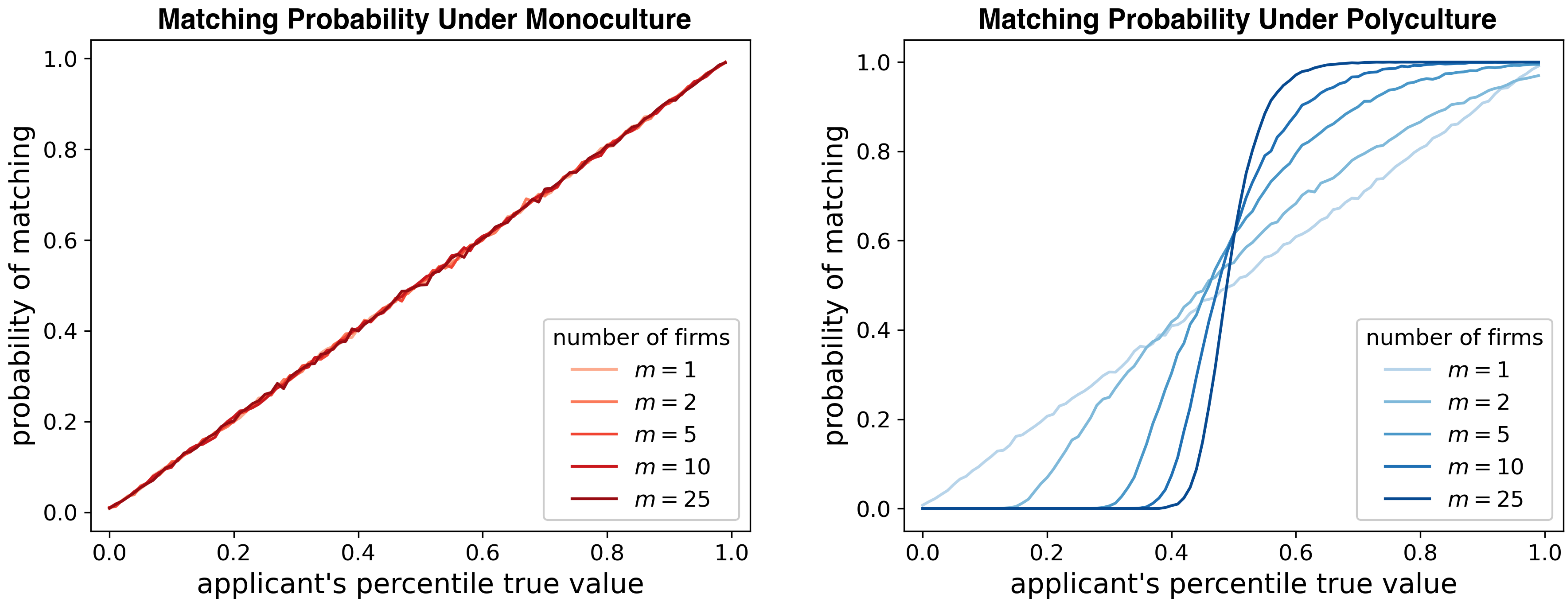
$$\sim \mathscr{D}$$

Theorem 1 (Informal)

If noise distribution $\mathscr{D}$ is max-concentrating:

$$\lim_{C \to \infty} \Pr\left[\text{student with true value } v \text{ is matched under polyculture}\right] = \begin{cases} 0 & v < v^* \\ 1 & v > v^* \end{cases}$$

.


Matching Probability Under Monoculture
probability of matching
applicant's percentile true value
number of firms
$m = 1$
$m = 2$
$m = 5$
$m = 10$
$m = 25$
0.0
0.2
0.4
0.6
0.8
1.0
Matching Probability Under Polyculture
probability of matching
applicant's percentile true value
number of firms
$m = 1$
$m = 2$
$m = 5$
$m = 10$
$m = 25$

Proof sketch.

P_{mono} or P_{poly}

$$v \text{ matches iff } v + X_c > \underbrace{P}_{\uparrow\, P_{\text{mono}} \text{ or } P_{\text{poly}}} \text{ for some } c \in C \iff v + \max_{c \in C} X_c > P$$

Monoculture: $\max_{c \in C} X_c \sim \mathcal{D}$

Polyculture: $\max_{c \in C} X_c$ vanishing variance

Intuition

Whether a student matches depends on <u>highest estimated value</u>

Theorem 1 holds for arbitrary student preferences, college capacities

Long-tailed noise results in “foolishness of crowds” effect

See Wisdom and Foolishness of Noisy Matching Markets (Peng and Garg, 2024)

Intuition

Whether a student matches depends on highest estimated value

Theorem 2 (Informal)

(i) For all v, $\Pr[v$ matched to top choice$]$ is at least as high under monoculture (strictly higher on set of positive η-measure)

(ii) For v on a set of positive η-measure, v is strictly more likely to match to top choice under monoculture, but strictly less likely to match overall

Proof sketch.

$$P_{\mathrm{mono}} < P_{\mathrm{poly}}$$

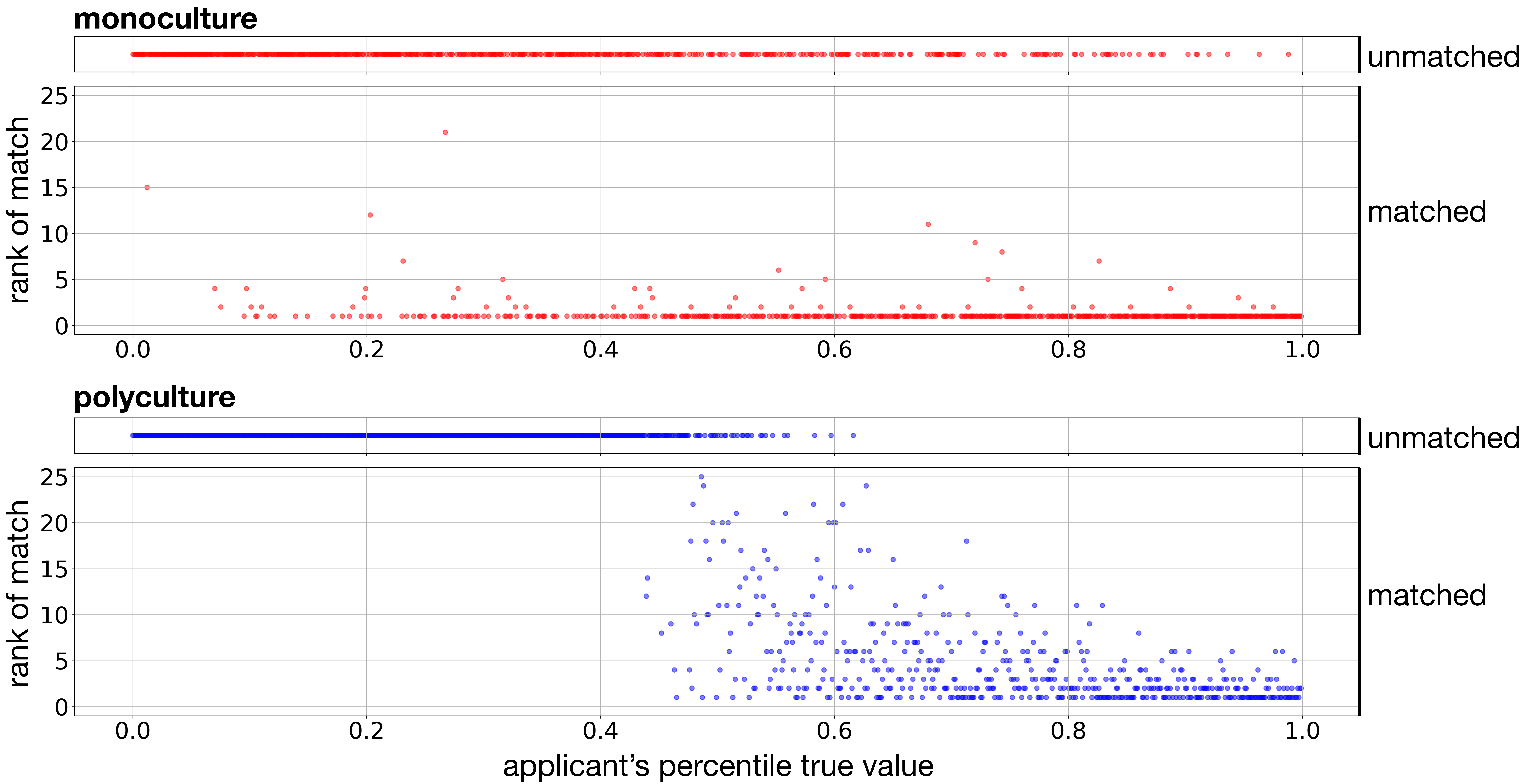
(More congestion under monoculture, so colleges each need to make more offers to clear market.)

Now suppose student's top choice is c. Then:

$$\Pr[v + X_c > P_{\mathrm{mono}}] \geq \Pr[v + X_c > P_{\mathrm{poly}}].$$

Individual Outcomes for 1000 Applicants

in a market with 25 colleges and 500 total openings



Theorem 3 (Informal)

Consider $v_1 > v_2$. Then by applying to more colleges, v_2 has a higher chance of being matched than v_1 under polyculture, but not under monoculture.

Main results

Theorem 1: Exactly top students match under polyculture

Theorem 2: Everyone more likely to match to top choice under monoculture

Theorem 3: Monoculture more robust to “differential application access”

ML Experiments

Colleges/firms use ML algorithm to predict binary outcome, rank according to predicted score

Monoculture: All use the same algorithm

Polyculture: All use different algorithms (trained on different features)


Monoculture v. Polyculture: ML-Based Evaluations (ACSIncome Texas)
monoculture
polyculture
Accuracy
0.64
0.62
0.60
0.58
Average Rank of Match
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
Subsets of Models


Monoculture v. Polyculture: ML-Based Evaluations (ACSIncome California)
monoculture
polyculture
Accuracy
Average Rank of Match
Subsets of Models

Emerging body of work

Modeling: (Ali et al., 2024, Castera et al., 2024, Peng and Garg, 2024)

Empirical Evaluations: (Bommasani et al., 2022, Toups et al., 2023)

Interventions: (Jain et al., 2024)