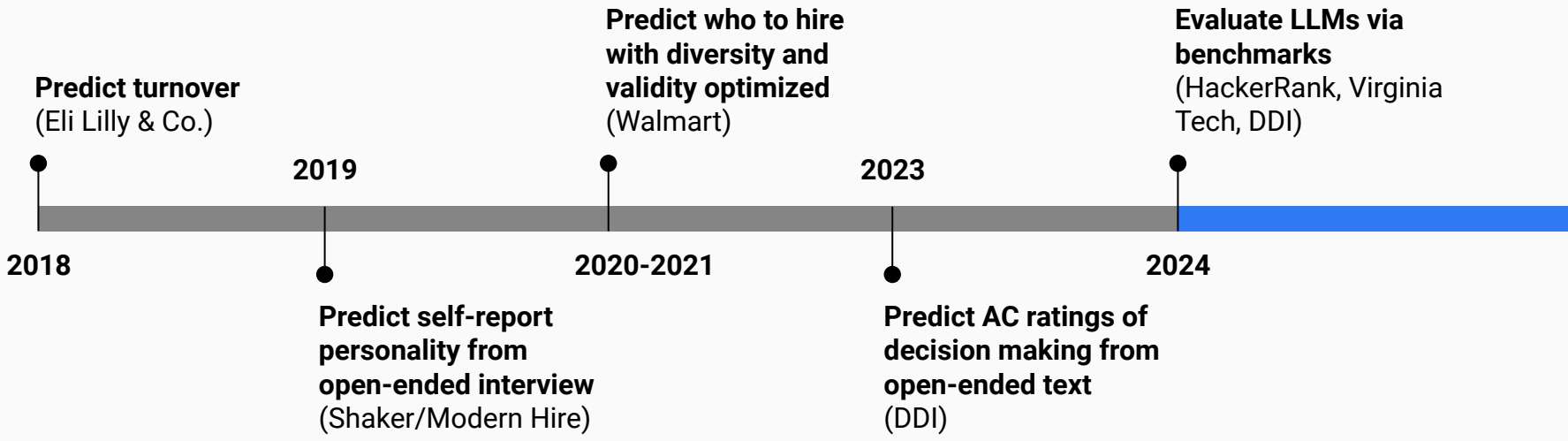
SIOP 2024 ML Competition

Sebastian Marin, Ivan Hernandez
Isaac Thompson, Georgi Yankov, Tyler Mirando

History of the Machine Learning Competition



The Machine Learning Competition

A data set is released with a problem statement (training set)

Community attempts to solve the problem statement, empirically

Scaled evaluation of approaches is accomplished online on a public leaderboard (dev phase)

Best generalizable solution wins as teams submit to a final private leaderboard (test phase)

Winners are decided based on the empirical quality of their work

The benchmark lives beyond the competition as these solutions are released to the public and new methods become available

This Year

Integration of LLMs and a detailed primer for up-skilling

Cash prizes and swag for winning teams from our generous sponsor HackerRank

Multiple data sponsors from Virginia Tech, DDI

Multiple benchmarks of diverse problem sets related to IO tasks

Stats

Participants: 30 teams (200 people)

Number of submissions: 800+

First ever competition? 20%

University vs Industry: 55% vs 45%

Education: 55% PhD, 45% MA

Programming Languages: 100% Python, 40% R too

Competition Challenges

Predicting Empathy (DDI): Classify whether empathy was demonstrated or not in a job candidate's simulated response (metric: accuracy)

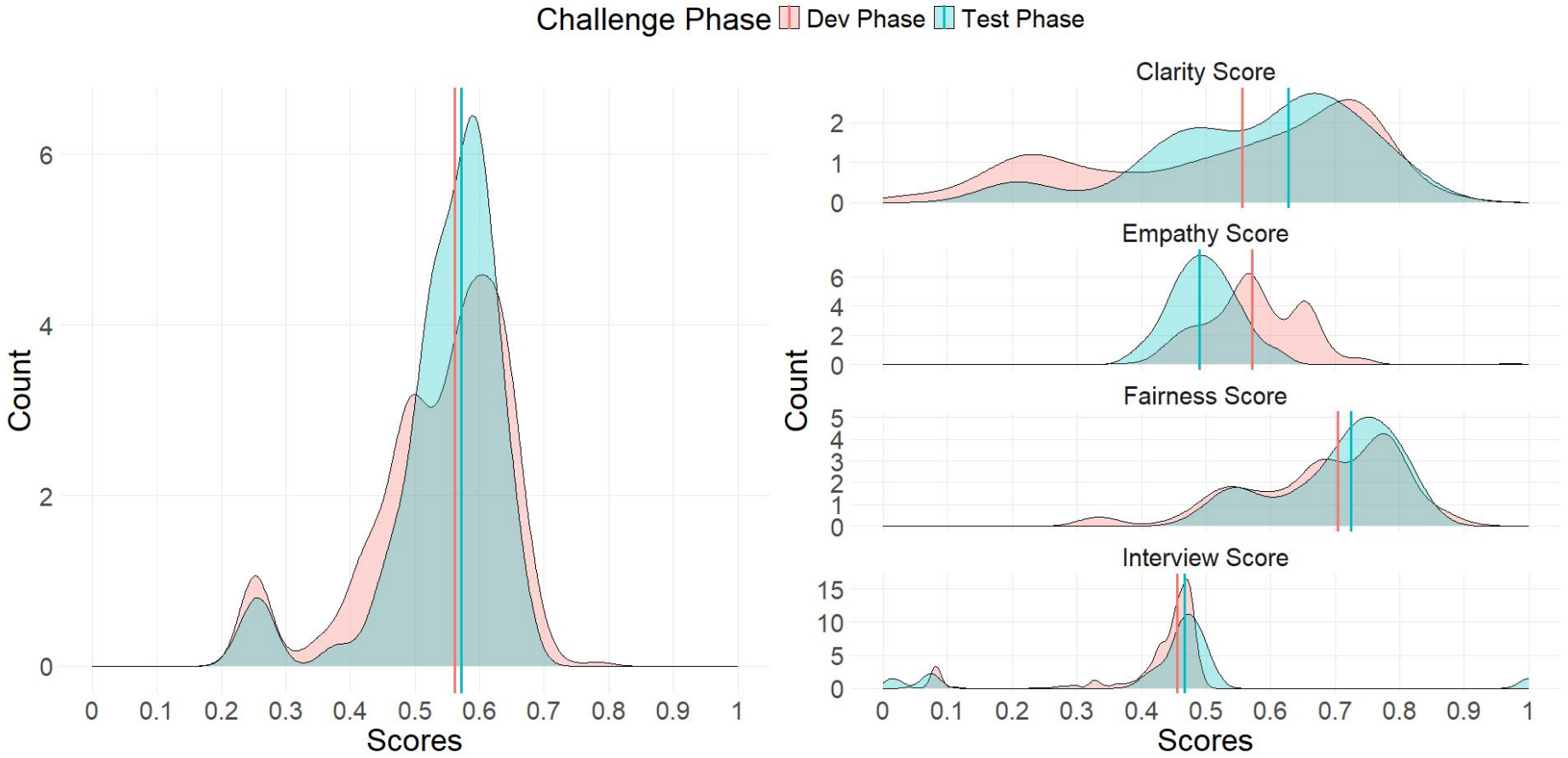
Generating Interview Responses (VA Tech): Generate a likely text response for the 5th question based on 4 previous responses of an interviewee (metric: average cosine)

Rating Item Clarity (VA Tech): Predict the average clarity rating for personality items based on item clarity ratings from human evaluators (metric: correlation)

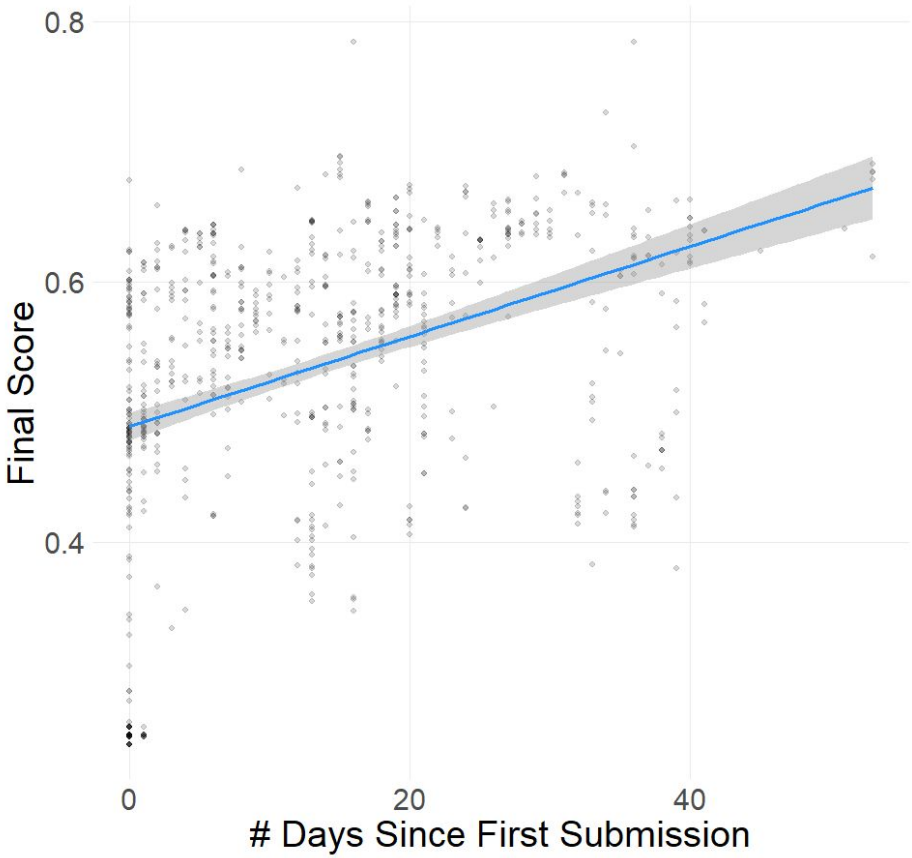
Identifying Fairness Perceptions (VA Tech): Identify which policy out of 2 policies received the majority vote as the fairer option from human evaluators (metric: accuracy)

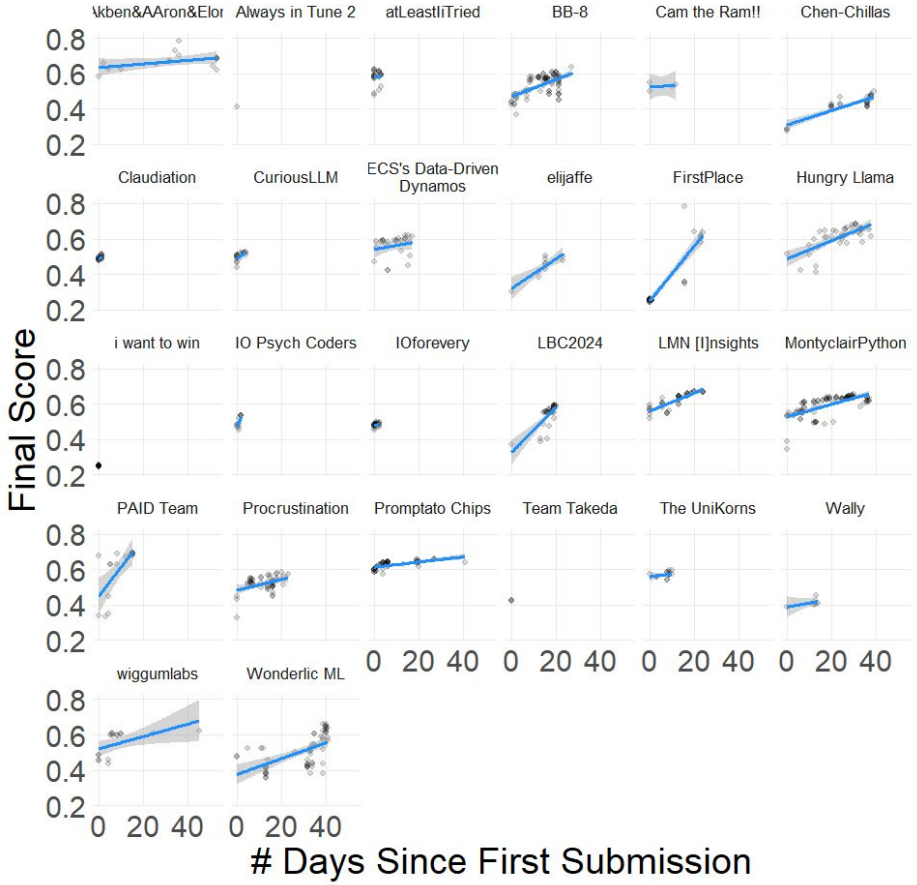
final score = (*empathy* × 0.25) + (*interview* × 0.25) + (*clarity* × 0.25) + (*fairness* × 0.25)

Score Distributions



Progression during Dev Phase





Top 10 Teams on Private Leaderboard

Rank	Team	Final Score
1	????	????
2	????	????
3	????	????
4	????	????
5	Promptato Chips	0.617
6	Team Takeda	0.608
7	LMN [I]nsights	0.600
8	MontyclairPython	0.596
9	ECS's Data-Driven Dynamos	0.584
10	LBC2024	0.571

Something to Learn from Different Approaches

	Empathy	Interview	Clarity	Fairness	Final Score
Team 4	.122	.115	.193	.198	.630
Team 3	.140	**.128**	.185	.190	.643
Team 2	**.152**	.124	.169	**.207**	.652
Team 1	.145	.110	**.204**	**.207**	**.666**

Something to Learn from Different Approaches

	Empathy	Interview	Clarity	Fairness	Final Score
Team 4	.122	.115	.193	.198	.630
Team 3	.140	**.128**	.185	.190	.643
Team 2	**.152**	.124	.169	**.207**	.652
Team 1	.145	.110	**.204**	**.207**	**.666**

Max scores	.152	.128	.204	.207	**.691**

The Finalists

The Top 4 Teams

(From 4th - 1st)

Fourth Place

Wonderlic ML

Final Score = .630

Team Members:

Guglielmo Menchetti
Lea Cleary
Annie Brinza

wonderlic.
the science of selection

Third Place

Hungry Llama

Final Score = .643

Team Members:

Jennifer Gibson
Shane Halder
Blake Hoffman
Hannah Johnson
Joseph Nicolas Luchman
Nick McCann
Selena Tran

ForsMarsh

Second Place

Akben & AAron & Elon

Final Score = .652

Team Members:

Mustafa Akben

Aaron Satko

ELON
UNIVERSITY

First Place

PAID Team

Final Score = .666

Team Members:

Zihao Jia
Mina Son
Philseok Lee

Panelist Q&A

How will LLMs change the science and practice of I-O?

How did you do it, what was your secret sauce?

What would you have done differently?

Where do you see these methods being applied in I-O?

What most impressed you most about the other teams' approaches?

What is your takeaway from participating and winning a ML competition?

What would you like to see in future I-O ML competitions?

Panelist Q&A

Audience, Any Questions for our Panelists?

Closing Remarks