## Summary

We are expanding the window used to calculate reputations and ranks from the latest 20 weekend rounds to all resolved rounds in the last year. This will take effect after scoring on March 3rd 2023.

## What are model scores?

Model scores (e.g. "reputation" / "rank") are leaderboard metrics that can show how good a model is. The goals of any model score are:
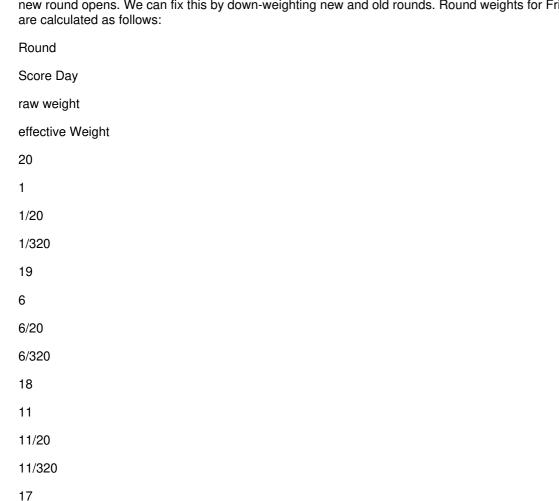
1. Measure long-term performance of a model

2. Provide a way to rank the leaderboard

3. Encourage submitting every round

4. Enliven day-to-day participation

## Current characteristics of model scores

1. 20-week weighted average of each metric; this achieves goals 1 and 2.

2. Missing corr scores are filled with -0.1; this achieves goal 3, but can result in many models w/ the same negative score, which appears as a bug in the leaderboard.

3. Uses resolved and unresolved rounds weighted by their % scoring progress; this achieves goal 4, but adds complexity just to decrease volatility of the score.

## How round weights lowered volatility

Round weights not only enable daily updates, they reduce volatility that comes with changing a score every day. The 4 unresolved rounds have higher volatility than resolved rounds and will push out a non-volatile resolved round every time a new round opens. We can fix this by down-weighting new and old rounds. Round weights for Friday (under weekly rounds) are calculated as follows:

| Round | Score Day | raw weight | effective Weight |
|---|---|---|---|
| 20 | 1 | 1/20 | 1/320 |
| 19 | 6 | 6/20 | 6/320 |
| 18 | 11 | 11/20 | 11/320 |
| 17 | 16 | 16/20 | |

16/320

16

20

1

20/320

…

4

20

19/20

19/320

3

20

14/20

14/320

2

20

9/20

9/320

1

20

4/20

4/320

As you can see, a newly scoring round only has 1/320th of the total weight in the model score; the same is true of the oldest resolved round on the day before a round resolution day. This means that new rounds entering the window and old rounds leaving the window don't cause wild swings in the score; the same would be true if we simply included more rounds. This complexity is unwarranted and is why we are moving away from this method of calculating reputation.

## Calculating model scores without round weights

With daily tournaments, rounds began resolving starting 2022-11-25. This means we can now update the model scores every day by:

- using unweighted average of only resolved rounds in the last year; this will:
- avoid complexity of round weights
- improve estimation of long-term value (~5 months vs 1 year)
- naturally reduce volatility in the score as daily rounds resolve
- each round had 1/52 weight before 2022-11-25
- each round will have 1/260 weight starting 2023-11-25
- each round had 1/52 weight before 2022-11-25
- each round will have 1/260 weight starting 2023-11-25
- avoid complexity of round weights
- improve estimation of long-term value (~5 months vs 1 year)

- naturally reduce volatility in the score as daily rounds resolve

- each round had 1/52 weight before 2022-11-25

- each round will have 1/260 weight starting 2023-11-25

- each round had 1/52 weight before 2022-11-25

- each round will have 1/260 weight starting 2023-11-25

- fill missing rounds/scores with 0; this will:

- avoid users having the same non-zero score

- be a non-monetary way to encourage submitting all days

- weekly submitters will only be able to achieve 1/5th of the score that a daily submitter could

- weekly submitters will only be able to achieve 1/5th of the score that a daily submitter could

- avoid users having the same non-zero score

- be a non-monetary way to encourage submitting all days

- weekly submitters will only be able to achieve 1/5th of the score that a daily submitter could

- weekly submitters will only be able to achieve 1/5th of the score that a daily submitter could

Using the above rules, we still achieve the 4 goals of model scores, but simpler, cleaner and better than using round weights. Eventually we will more easily be able to offer multiple time horizons like 1-month, 3-month, and/or 6-month.

This change will be fully backfilled and deployed after the scoring pipeline finishes on March 3rd 2023. The first live 1-year reputation will be calculated Saturday March 4th, 2023.