Shiny App GitHub

Enhancing Athlete Selection for Artistic Gymnastics at the Paris 2024 Olympics: A Data-Driven Approach

Sahil Singh[1] Raymond Lee[1]

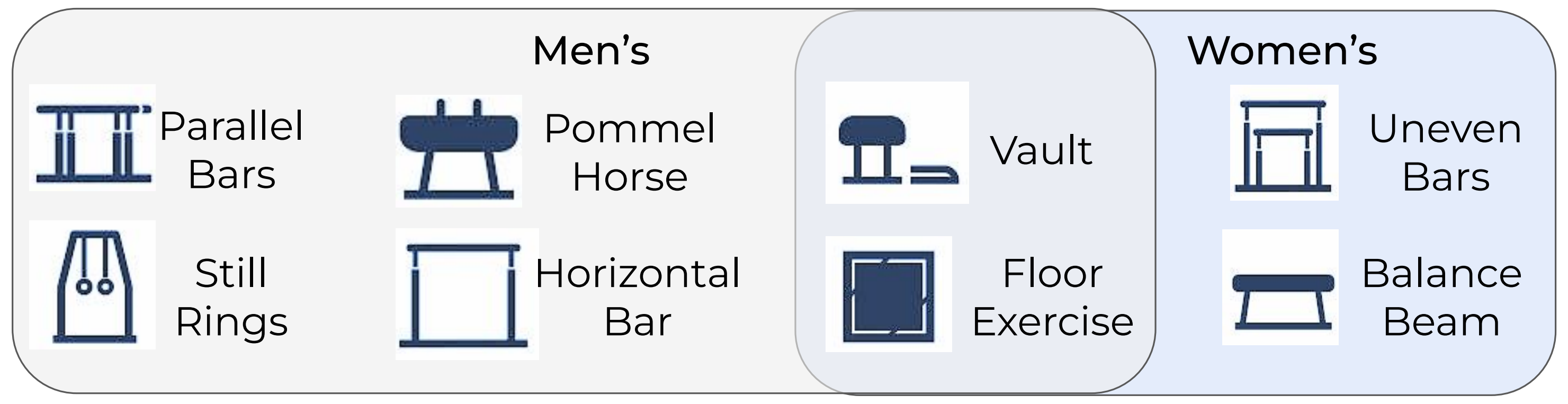
[1] Department of Statistics and Data Science, Yale University

Introduction

Athlete selection for gymnastics events has inherent complexities, including evolving rules, subjective scoring, and the need to balance various factors such as performance consistency, injury history and team combinations. In this work, we focus on addressing the critical task of selecting artistic gymnasts for Team USA at the Paris 2024 Olympics. This is helpful for providing a data-driven approach for team selection, offering enhanced decision-making capabilities. We make the following contribution: Development of an interactive tool using historical performance data and statistical analysis techniques to assist in making informed selections. Its significance lies in optimizing team composition and potentially impacting success at the sporting event. By delving into the challenges and intricacies of gymnastics selection, our work underscores the transformative potential of data-driven approaches in enhancing athlete selection strategies for Olympic sports.

Rules, Scoring and Competition Format

At the Paris 2024 Olympics, Artistic Gymnastics teams consist of five gymnasts, who collectively compete on the following apparatuses:



Men's: Parallel Bars, Pommel Horse, Still Rings, Horizontal Bar

Women's: Vault, Uneven Bars, Floor Exercise, Balance Beam

Additionally, both men and women have individual all-around events and team events.

Scoring is based on two main components, alongside deductions for errors:

- **Execution Score** (E-score): evaluated on a traditional 10.0 system, deducts points for errors
- **Difficulty Score** (D-score): starts at zero and increases based on the difficulty of elements performed in the routine

Hence, the final score is calculated by:

***Final Score** = [Difficulty + Execution] - Any neutral deductions*

The competition starts with the qualification round. Here, all gymnasts compete in their events with no medals awarded. This round determines advancement to the finals in team, individual all-around, and individual apparatus competitions, with medals awarded for each.

Athlete Model

The athlete model determines how to predict a gymnast's possible future scores based on previous competition scores. We choose a simple yet robust method for modeling an athlete using a Normal distribution, where for athlete *i* on apparatus *k*, we predict their score $S_{i,k}$ to be distributed as:

$$S_{i,k} \sim Normal(\mu = \overline{x_{i,k}}, \sigma^2 = var(x_{i,k}))$$

where $x_{i,k}$ is all of the previous data points of the scores for athlete *i* and apparatus *k* in previous competitions. We also assume:

- If there is no data for an athlete on some apparatus, we assume that the athlete is unable to compete in said apparatus
- If only one data point available (variance=0), we set the variance to the global variance average = 1.3

Simulation Procedure

We use simulation to optimize the performance of all teams rather than just USA to present a more realistic competitive environment. Our procedure is as follows:

```
1. set reasonable_set = list of good athletes (top-5 in any apparatus)
2. set all_simulation_results = dictionary of simulation results
3. initialize best_teams = assignments of best teams for country (initialized randomly)
4. for country in countries (12):
   a. for team in combinations(reasonable_set[country], n=5):
      i.   assign gymnasts in team to apparatuses
      ii.  simulate quals and finals
      iii. set curr_results = results of simulation
      iv.  set all_simulation_results[team] = curr_results
   b. set best_teams[country] = team with best result (custom-defined medal weighting)
5. repeat Step 4 to optimize each team a 2nd time
```

Optimization Definition

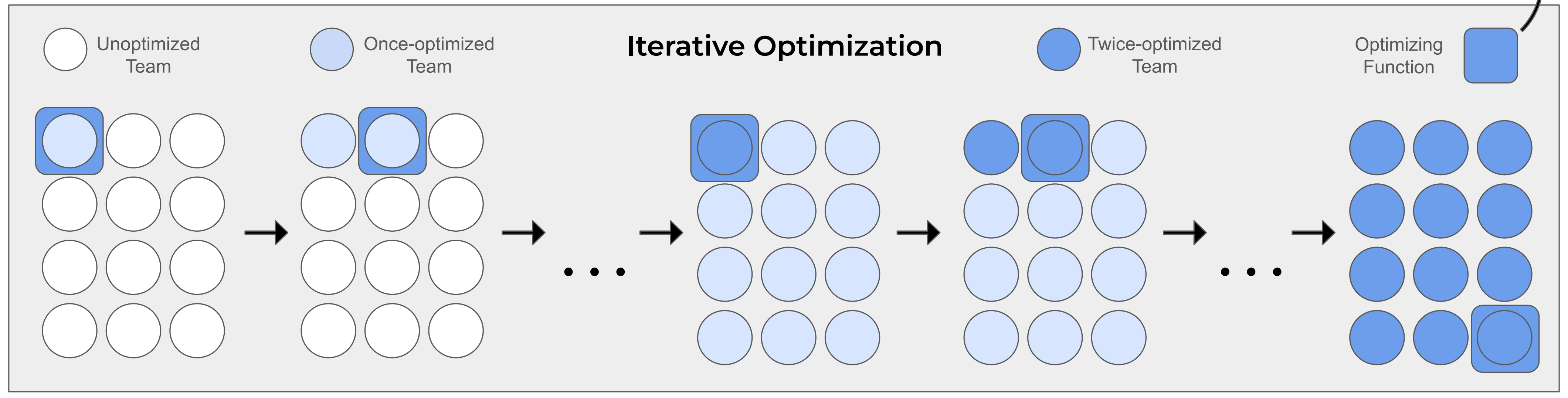


Figure 1. Iterative Optimization

Results and Discussion

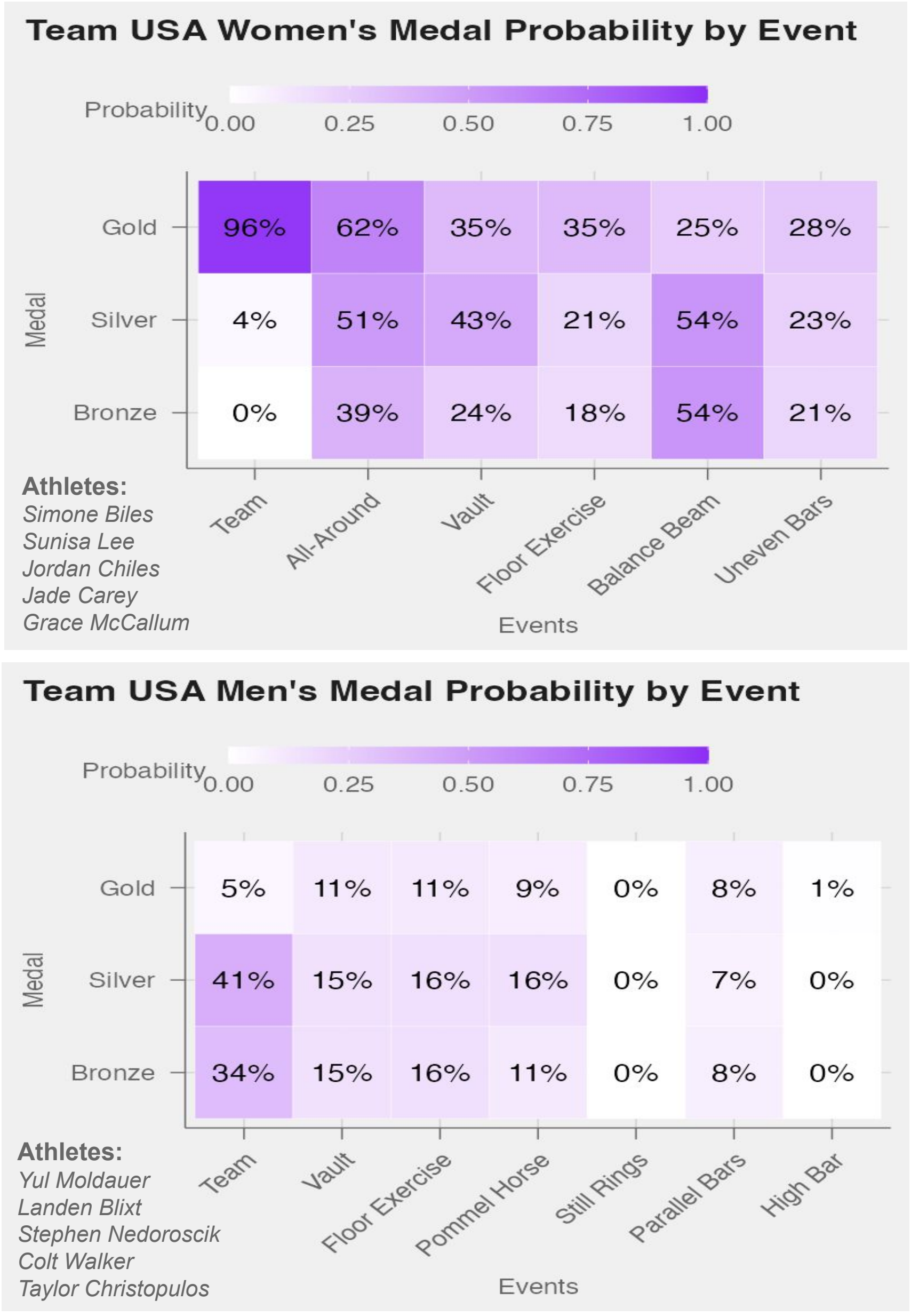


Figure 2. Team USA Medalling Probabilities

In total, we ran and recorded results for 43,000+ simulations for men and 7,500+ simulations for women athletes. We identified specific athletes who consistently excelled in particular events (Simone Biles on Vault and Floor Exercise, Sunisa Lee on Uneven Bars). These insights help optimize team composition by identifying the strengths of individual athletes.

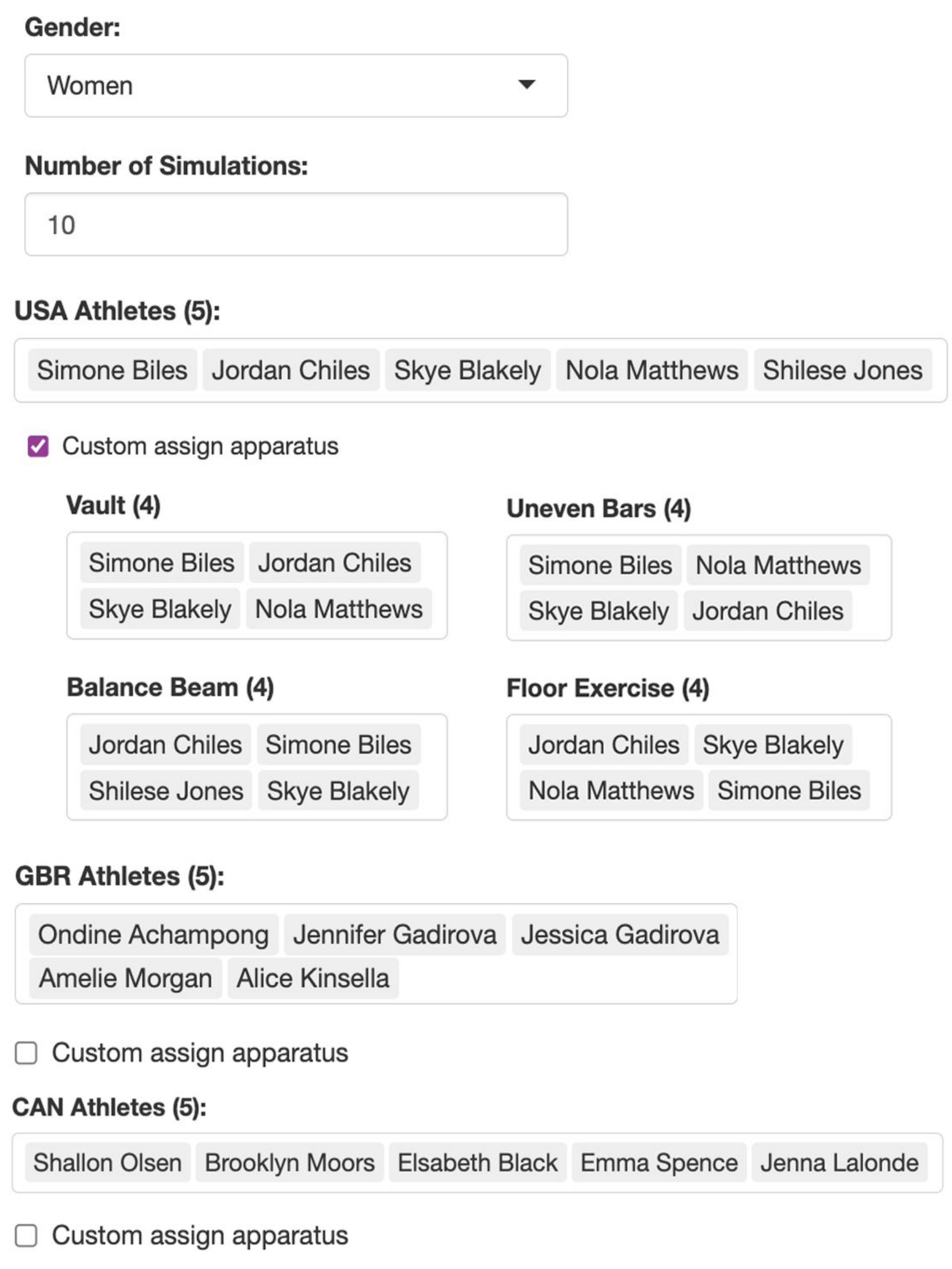
Our simulations consistently showed a strong correlation between the presence of top-performing athletes and the overall success of teams. For instance, our analysis demonstrated that the inclusion of Simone Biles (USA Women's) and Daiki Hashimoto (Japan Men's) significantly boosted their teams' medal prospects.

Consequently, we designed an R Shiny app to serve as a tool to explore a plethora of different strategies, with options to include/exclude certain players or applying custom weights on each medal. As an example, we implemented a strategy to prioritize medals as Gold (60%), Silver (30%) and Bronze (10%).

We utilized our simulation data and interactive tool to find the best Team USA combinations for men and women subject to the strategy selected. Figure 2 shows the results yielded with 96% probability of Gold in women's Team event and 41% probability of Silver in men's Team event, among medal probabilities for other events.

Custom Simulations

In addition to exploring our computed simulations and analyses in the previous sections, we also include in our R Shiny app the capability to run custom simulations (Figure 3).



Gender: Women ▾

Number of Simulations: 10

USA Athletes (5): Simone Biles | Jordan Chiles | Skye Blakely | Nola Matthews | Shilese Jones

☑ Custom assign apparatus

Vault (4): Simone Biles | Jordan Chiles | Skye Blakely | Nola Matthews

Uneven Bars (4): Simone Biles | Nola Matthews | Skye Blakely | Jordan Chiles

Balance Beam (4): Jordan Chiles | Simone Biles | Shilese Jones | Skye Blakely

Floor Exercise (4): Jordan Chiles | Skye Blakely | Nola Matthews | Simone Biles

GBR Athletes (5): Ondine Achampong | Jennifer Gadirova | Jessica Gadirova | Amelie Morgan | Alice Kinsella

☐ Custom assign apparatus

CAN Athletes (5): Shallon Olsen | Brooklyn Moors | Elsabeth Black | Emma Spence | Jenna Lalonde

☐ Custom assign apparatus

Figure 3. Custom Simulation Selection

This capability supports assigning any combination of athletes for any country and putting them in a simulated competition. In addition, it also allows users to have custom apparatus assignments for each country's selected athletes, granted there is data of the athlete on said apparatus.

Multiple simulations with each assignment can be conducted, in which medal counts are recorded and results can be viewed to see in what proportion of runs did athletes win each medal on each apparatus. We offer real-time visualization of the results to compare performances of specific athletes and Team USA.

In our testing, 10 simulations take under 3.5 seconds.

Next Steps

While our simulation model has provided valuable insights, there are opportunities for refinement and enhancement in future research. Possibilities include:

- Differing weights on data points (e.g. more recent data weighs more)
- Negative weights on athlete with higher age (for older competitors)
- Negative weights based on historical health and injury risks

In addition, we also recognize some possible features to incorporate in our interactive custom simulation tool:

- Prioritize medals in specific apparatuses or events
- Faster simulation speeds with multithreading/ multiprocessing
- Easier direct comparison between different custom simulations

References

1. NBC Olympics. (2023). Gymnastics 101: Olympic competition format. Paris 2024 Olympic Games. https://www.nbcolympics.com/news/gymnastics-101-olympic-competition-format
2. Sanfelice Bazanella, A., Campestrini, L., & Eckhard, D. (2012). Iterative optimization. Communications and Control Engineering, 69–88. https://doi.org/10.1007/978-94-007-2300-9_4
3. Singh, S., & Lee, R. (2023). Sahil-681/gymnastics-data-challenge. GitHub. https://github.com/sahil-681/Gymnastics-Data-Challenge

sahil.singh@yale.edu raymond.lee@yale.edu