Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Sister Conferences Best Papers Track

Soft Condorcet Optimization for Ranking of General Agents (Extended Abstract)*

Marc Lanctot[1], Kate Larson[1,2], Michael Kaisers[1], Quentin Berthet[1], Ian Gemp[1], Manfred Diaz[3,†], Roberto-Rafael Maura-Rivero[1,4], Yoram Bachrach[5,‡], Anna Koop[1], Doina Precup[1]
[1]Google DeepMind
[2]University of Waterloo
[3]MATS Research Fellow
[4]LSE
[5]Meta

Abstract

Driving progress of AI models and agents requires comparing their performance on standardized benchmarks; for *general* agents, individual performances must be aggregated across a potentially wide variety of different tasks. In this extended abstract, we describe a ranking scheme inspired by social choice frameworks, called Soft Condorcet Optimization (SCO), to compute the optimal ranking of agents: the one that makes the fewest mistakes in predicting the agent comparisons in the evaluation data. This optimal ranking is the maximum likelihood estimate when evaluation data (which we view as votes) are interpreted as noisy samples from a ground truth ranking, a solution to Condorcet's original voting system criteria. SCO ratings are maximal for Condorcet winners when they exist, which we show is not necessarily true for the classical rating system Elo. In practice, SCO serves as an accurate approximation to the Kemeny-Young voting method, excels in the sparse data regime, and provides the best approximation to the optimal ranking compared to every baseline on a Diplomacy player ranking problem with 31,094 games and 52,958 agents.

1 Introduction

Progress in the field of artificial intelligence has been driven by measuring the performance of agents on common benchmarks and challenge problems [Samuel, 1959; Campbell *et al.*, 2002; Tesauro, 1995; Syed, ; Silver *et al.*, 2016; Vinyals *et al.*, 2019; (FAIR) *et al.*, 2022]. In machine learning, common benchmarks like the UCI data set repository allowed direct comparisons of supervised learning algorithms [Kelly *et al.*,]. Competitions, such as ImageNet, led to breakthroughs in deep learning [Krizhevsky *et al.*, 2012]. All of these examples require comparing agents (or models). Original success stories such as DeepBlue, TD-Gammon, and AlphaGo focused on a single domain. In the past ten years, agents have become increasingly more generally capable. AlphaZero extended application of AlphaGo to chess and Shogi [Silver *et al.*, 2018]. The Arcade Learning Environment [Bellemare *et al.*, 2013], which steered much of the agent development in deep reinforcement learning, evaluated agents across 57 different Atari games. Recently, language models have been evaluated across suites of tasks such as in HELM [Liang *et al.*, 2022], BIG-bench [Srivastava *et al.*, 2023] AgentBench [Liu *et al.*, 2023], and via a public leaderboard such as Chatbot Arena driven by human voting [Chiang *et al.*, 2024]. Answering simple questions for these generally capable agents, such as "Which is the best agent?" or "Is agent X better than agent Y?" or "What is the relative ranking of agents X, Y, and Z?" become increasingly more difficult when aggregating over many different contexts: how agents are scored may vary wildly across tasks, data collected for evaluation may not be balanced evenly across tasks (or agents), and classical rating systems were simply not designed for this use case.

To address these problems, recent ranking methods such as Vote'N'Rank [Rofin *et al.*, 2023] and Voting-as-Evaluation (VasE) [Lanctot *et al.*, 2023] use voting methods to aggregate results across tasks. Using computational social choice as a basis for ranking agents has several benefits: well-studied consistency properties of the voting methods are inherited, they do no require score normalization, and are less sensitive to score values and agent population than game-theoretic evaluation schemes [Balduzzi *et al.*, 2018]. However, classic voting schemes, and related tournament solutions, typically assume that the data (e.g. agent comparisons) is complete. This assumption is not necessarily valid in the agent evaluation setting. While there has been research on identifying "necessary and possible winners" when there is incomplete voting or comparison data, the results are mixed [Pini *et al.*, 2011; Xia and Conitzer, 2011; Aziz *et al.*, 2015] and most of the findings are focused on identifying top-ranked agents, not ranking all agents as is our focus.

In this extended abstract, we summarize a recently-presented paper at AAMAS 2025 [Lanctot *et al.*, 2025] by describing a scheme for general agents inspired by the interpretation of voting rules as maximum likelihood estimators [Conitzer and Sandholm, 2005]. Starting with Condorcet's original model of voting [Brandt *et al.*, 2016, Chapter 8], Young showed that the maximum likelihood estimate

*Extended abstract of this full paper [Lanctot *et al.*, 2025].
†Work performed while at Mila and University of Montreal.
‡Work performed while at Google DeepMind.

(MLE) of the true ranking is the one that minimizes the sum of Kendall-tau distances to all the votes. Soft Condorcet Optimization (SCO) solves an optimization problem that assigns a numerical *rating* (score) θ_a to each agent (alternative) a. SCO then treats these ratings as a parameter vector, the votes as a data set, and defines a differentiable loss function as the objective, which can be optimized in several different ways. The final ranking of agents is obtained by sorting the ratings.

2 Background

In this section, we describe the building blocks and introduce some basic terminology required to understand our method.

2.1 Evaluation of General Agents

The problem of evaluating agents is that of ranking agents according to their skill. Skill can be determined in several ways. In the **Agent-versus-Task (AvT)** setting, agents compete individually in different tasks and compare outcomes (scores) to each other in each task (*e.g.*, language models and the various metrics for assessing their abilities). In the **Agent-versus-Agent (AvA)** setting, agents directly compete against each other (*e.g.*, online games such as chess or Diplomacy).

Classical Evaluation

Elo is a classic rating system that uses a simple logistic model learned from win/loss/draw outcomes [Elo, 1978]. A rating, r_i, is assigned to each player i such that the probability of player i beating player j is predicted as $\hat{p}_{ij} = \frac{1}{1+10^{(r_j - r_i)/400}}$. While Elo was designed specifically to rate players in the two-player zero-sum, perfect information game of Chess, it has been widely applied to other domains, including in evaluation of large language models [Zheng *et al.*, 2023]. TrueSkill [Herbrich *et al.*, 2006] and bayeselo [Coulom, 2005] are rating systems based on similar foundations (Bradley-Terry models of skill) that also model uncertainty over ratings using Bayesian methods.

Elo has a number of positive qualities, but also a number of well-know drawbacks. Of particular interest is the incompatibility of Elo with the concept of a Condorcet winner from social choice theory.

Voting as Evaluation of General Agents

Another way to evaluate agents is to use social choice theory recently proposed as Voting-as-Evaluation (VasE) [Lanctot *et al.*, 2023]. In VasE, the alternatives correspond to general agents and votes to assessments of their performance. In the AvT setting, agents are ordered based on their performance on different tasks (such as each game in the Atari Learning Environment [Bellemare *et al.*, 2013] or on different benchmarks for large language models [Liu *et al.*, 2023]). In the AvA setting, agents compete directly in a multiagent environment, such as multiplayer games (like chess or poker), and each game outcome corresponds to a ranking over a subset of agents. Chatbot Arena [Zheng *et al.*, 2023], where language models compete head-to-head to provide the best answer to the same question, is another instance of the AvA setting. Casting agent evaluation as an application of computational social choice provides the benefit of robustness in the form of Condorcet consistency, clone/composition consistency, agenda consistency, and/or population consistency depending on the choice of voting rule used to evaluate agents. However, it is unclear how well these methods perform when the data is missing or unevenly distributed, which often occurs in the agent evaluation setting. SCO is motivated similarly to VasE and, as such, will be empirically assessed mainly for agent evaluation. To make the connection to ideas from the social choice literature, we will often use voting language when discussing SCO and its use in evaluation. Relevant terminology is introduced later in the paper.

2.2 Permutation and Ranking Distances

We now define distance metrics over rankings that will play a key role when describing our method and loss function. Informally, the Kendall-tau distance counts the number of pairwise disagreements between permutations.

Definition 1. *Let S_1, S_2 be finite sets of elements such that $S_1 \subseteq S_2$. Let π_1, π_2 be permutations over elements in S_1 and S_2, respectively. The **Kendall-tau distance** between two permutations is defined as*

$$K_d(\pi_1, \pi_2) = \sum_{\{i,j\} \in \mathcal{C}_2(S_1)} \bar{K}_{i,j}(\pi_1, \pi_2), \tag{1}$$

where $\mathcal{C}_2(S)$ is the set of unordered pairs of S (combinations of size 2), $\bar{K}_{i,j}(\pi_1, \pi_2) = 0$ if i and j are in the same order in π_1 and π_2, and $\bar{K}_{i,j}(\pi_1, \pi_2) = 1$ otherwise.

This definition allows one set to be a subset of the other, which is more general than the standard definition; this distinction is necessary for the evaluation metric used in Section 4.1, and corresponds to the standard definition when $S_1 = S_2$.

Since the maximum distance is $\binom{|S_1|}{2} = \frac{|S_1|(|S_1|-1)}{2}$, this can be easily normalized to be in $[0, 1]$:

Definition 2. *Let S_1, S_2 be finite sets of elements such that $S_1 \subseteq S_2$. Let π_1, π_2 be permutations over elements in S_1 and S_2, respectively. The **normalized Kendall-tau distance** π_1 and π_2 is defined as*

$$K_n(\pi_1, \pi_2) = \frac{2K_d(\pi_1, \pi_2)}{|S_1|(|S_1| - 1)}. \tag{2}$$

2.3 Social Choice Theory

A **voting scheme** is defined as $\langle A, V, f \rangle$ where $A = \{a_1, \ldots, a_m\}$ is the set of **alternatives** (agents), $V = \{v_1, \ldots, v_n\}$ is the set of **voters**, and f is the **voting rule** that determines how votes are aggregated. Voters have **preferences** over alternatives: $a_1 \succ_{v_i} a_2$ indicates voter v_i strictly prefers alternative a_1 over alternative a_2. In this paper, we assume strict preferences only, however the ideas can easily be extended to the case that allow weak preference (including ties/indifference between alternatives, *e.g.*, $a_1 \succeq s_2$). These preferences induce total orders over alternatives, which we denote by $\mathcal{L}$. A **preference profile**, $[\succ] \in \mathcal{L}^n$, is a vector specifying the preferences of each voter in V.

The central problem of social choice theory is how to aggregate preferences of a population so as to reach some collective decision. A voting rule that determines the "winner"

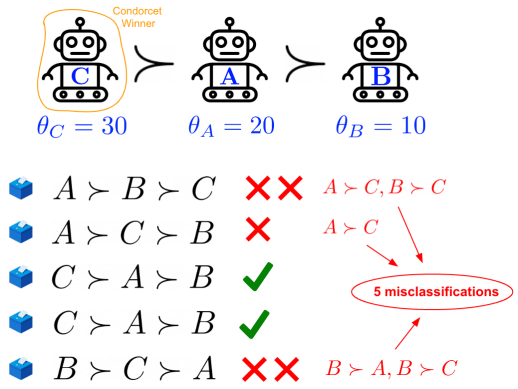

Figure 1: An example calculation of the Kendall-tau distances of each vote to the optimal ranking $R = C \succ A \succ B$ in an example preference profile with five votes. The sum of the distances is 5. This sum for any other ranking $R' \neq R$ is greater than 5.

(a non-empty subset, possibly with ties), is a **social choice function** (SCF). A voting rule that returns an aggregate ranking (total order) is a **social welfare function** (SWF). Much of the social choice literature focuses on understanding what properties different voting rules support.

The preference count $N(x, y)$, $x, y \in A$ is the number of voters in $[\succ]$ that strictly prefer x to y. The vote margin is the difference in preference count: $\delta(x, y) = N(x, y) - N(y, x)$. A weak **Condorcet winner** wins or ties in every head-to-head pairing. Formally, given a preference profile, $[\succ]$, a weak Condorcet winner is an alternative $a^* \in A$ such that $\forall a' \in A, \delta(a^*, a') \geq 0$. If the inequality is strict for all pairs except (a^*, a^*) then we call it a strong Condorcet winner.

3 Soft Condorcet Optimization

Soft Condorcet Optimization (SCO) is a ranking scheme for evaluation of general agents inspired by social choice theory. An SCO ranking is derived from and represented by **SCO ratings**, θ_a, for each alternative $a \in A$. The rating $\theta_a \in [\theta_{\min}, \theta_{\max}]$ serves only to determine a's relative order compared to other alternatives in the ranking, such that $a \succ b$ if and only if $\theta_a > \theta_b$. The ranking is induced by numerically sorting these ratings (*e.g.*, see Figure 1). We then formulate an optimization problem by carefully constructing a loss function that that penalizes discrepancies or misclassifications in the ordinal relationships between alternatives.

3.1 SCO Ratings and the Sigmoid Loss

In this section, we derive how SCO ratings can be computed by minimizing a **sigmoid loss** with gradient descent. In the full paper [Lanctot *et al.*, 2025] we also show that it can be optimized by sigmoidal programming, and describe an alternative convex loss function (*i.e.,* Fenchel-Young loss via perturbed optimizers [Berthet *et al.*, 2020]).

For some set of ratings $\boldsymbol{\theta}$, this loss function quantifies the amount of error: the level of disagreement between the specific values of each rating and the data (preference profiles obtained through agent evaluation). The goal is then to find an assignment of ratings that minimizes this loss, *i.e.*, the set of ratings that best explain the preference data.

Let $\Pi(A)$ denote the set of permutations over alternatives A. We will call the data we work with "votes" to emphasize that we are working with (partial) rankings over alternatives, and borrow terminology from preferences and social choice. In particular, we view the entire dataset to be a collection of votes and so will refer to it as *preference profile*, $[\succ]$. Let $v \in [\succ]$ refer to each vote in the profile, where v is a permutation over subsets of A, with length $|v|$. For a vote v, denote $v[i]$ as the alternative in position i such that vote v is represented as: $v[0] \succ v[1] \succ \cdots \succ v[|v| - 1]$.

The ultimate goal is to find an **optimal ranking** R:

Definition 3. *Given some a profile $[\succ]$ (i.e., set of votes), an* **optimal ranking** *minimizes the sum of the Kendall-tau distances*

$$R = \operatorname*{argmin}_{R \in \Pi(A)} \sum_{v \in [\succ]} K_d(v, R).$$

Let index pairs $I_2(v) = \{(i, j) \mid i, j \in \{0, 1, \cdots, |v| - 1\}$ and $i < j\}$. Given a preference profile and ratings $\boldsymbol{\theta}$, we define a **discrete loss**:

$$L([\succ], A, V, \boldsymbol{\theta}) = \sum_{v \in [\succ]} \sum_{(i,j) \in I_2(v)} D_v(\theta_{v[i]}, \theta_{v[j]}), \qquad (3)$$

where $D_v(\theta_a, \theta_b)$ is a function that measures the *discrepancy* between the positions of alternatives a and b:

$$D_v(\theta_a, \theta_b) = \begin{cases} 1 & \text{if } \theta_b - \theta_a > 0 \text{ in } v; \\ 0 & \text{otherwise,} \end{cases}, \qquad (4)$$

then the minimum of the discrete loss function corresponds to a ratings assignment $\boldsymbol{\theta}$ such that ranking induced by $\boldsymbol{\theta}$ minimizes the sum of the Kendall-tau distances to all the votes.

Since D is a step function discontinuous at $\theta_a = \theta_b$, it is not differentiable in $\boldsymbol{\theta}$. To solve this, we replace D with a smooth approximation, *i.e.*, the logistic function

$$\tilde{D}_v(\theta_a, \theta_b) = \sigma(\theta_b - \theta_a) = \frac{1}{1 + e^{(\theta_a - \theta_b)/\tau}}, \qquad (5)$$

leading to a soft Kendall-tau (differentiable) **sigmoid loss**:

$$\tilde{L}([\succ], A, V, \boldsymbol{\theta}) = \sum_{v \in [\succ]} \sum_{(i,j) \in I_2(v)} \tilde{D}_v(\theta_{v[i]}, \theta_{v[j]}). \qquad (6)$$

The sigmoid loss is a smooth, differentiable approximation to the discrete Kendall-tau distance.

3.2 Computing SCO Ratings by Gradient Descent

The sigmoid loss can be minimized by stochastic gradient descent [Hazan, 2015]. We compute the gradient for a subset of the votes (*i.e.*, batch $B \subseteq [\succ]$), $\nabla_{\boldsymbol{\theta}} \tilde{L}(B, \boldsymbol{\theta})$, from the batch loss $\tilde{L}(B, \boldsymbol{\theta})$, which resembles equation (6) but summed only over the votes in B using the continuous $\tilde{D}_v$ from equation (5):

$$\tilde{L}(B, \boldsymbol{\theta}) = \sum_{v \in B} \sum_{(i,j) \in I_2(v)} \tilde{D}_v(\theta_{v[i]}, \theta_{v[j]}). \qquad (7)$$

This allows an online form of the algorithm where ratings for players can be updated in a decentralized fashion after receiving the outcome of a single game ($|B| = 1$).

3.3 Theoretical Properties

Given the SCO framework, defined through the loss function introduced in equation 6, the first question to ask is whether its solution does lead to rankings with desired properties.

Theorem 1. *Given the sum of soft Kendall-tau distances:*

$$\tilde{L}([\succ], A, V, \boldsymbol{\theta}) \quad = \sum_{a,b \in A \times A} N(a,b)\sigma(\theta_b - \theta_a)\,, \quad (8)$$

if for preference profile $[\succ]$, voters V, there exists a candidate $c \in A$ that is a Condorcet winner, the loss is monotonically decreasing with θ_c. As a consequence, if $\boldsymbol{\theta}^*$ is a global minimum of $\tilde{L}$ on the ℓ_∞ ball of radius $\theta_{\max}$, then $\theta^*_c = \theta_{\max}$.

For the proof, please see the full paper [Lanctot *et al.*, 2025]. Due to Theorem 1, the global minimum of the proposed loss function corresponds to SCO ratings that top-rank a Condorcet winner, if one exists.

4 Empirical Evaluation

The full paper [Lanctot *et al.*, 2025] shows a counter-example preference profile where SCO top-ranks the single Condorcet winner where both Elo and Fenchel-Young do not. It also shows that SCO rankings are on average 0 to 0.043 away from the optimal ranking in normalized Kendall-tau distance across 865 preference profiles from the PrefLib open ranking archive. It also shows that in a simulated noisy tournament setting, SCO achieves accurate approximations to the ground truth ranking and the best among several baselines when 59% or more of the preference data is missing.

4.1 Diplomacy Game Outcome Prediction

In this subsection we investigate the question of how well SCO rankings predict human Diplomacy game outcomes. This data set consists of all human-played seven-player games taken from the `webDiplomacy.net` spanning 11 years, resulting in a data set of size $m = 52,958$ and $n = 31,049$. As a result, less than 0.01% of the unique player combinations are observed in the data. Finding a single ranking that sufficiently predicts the unseen test data is a difficult due to the extreme sparsity of the preference data.

We reflect the evaluation used in [Lanctot *et al.*, 2023] with a small modification to adopt common practice in the supervised learning setting . First, we create 50 random splits of the data into training sets $\mathcal{D}_R$ and testing sets $\mathcal{D}_T$, with $|\mathcal{D}_R| = 28049$ game outcomes (votes) and $|\mathcal{D}_T| = 3000$ game outcomes. The random splits are such that each alternative in the test set is seen at least once on the training set, but no game outcomes (data points) are shared across train / test split. At each iteration t the method has a ranking denoted R_t learned from data in $\mathcal{D}_R$; we then compute and report the average Kendall-tau distance over the test set: $\text{KTD}_{test}(t) = \frac{1}{3000}\sum_{g \in \mathcal{D}_T} K_d(g, R_t)$.

The resulting $\text{KTD}_{test}(t)$ is shown in Figure 2. We found that batch size had a minor effect on the results, so we show results for batch size 32 only. SCO with sigmoid loss reaches an average Kendall-tau distance of 8.10 after roughly 190000 iterations, and Fenchel-Young loss reaches 8.05 at 600000 iterations. There is an effect of increasing error after reaching this low point, likely due to overfitting. As a point of comparison, Elo and the best VasE method on this data set, VasE(Copeland), achieve a value of 8.34. The next best VasE method was plurality, achieving a value of 8.57. The more complex Condorcet methods cannot be run on this dataset due to their quadratic complexity and $m = 52,958$.

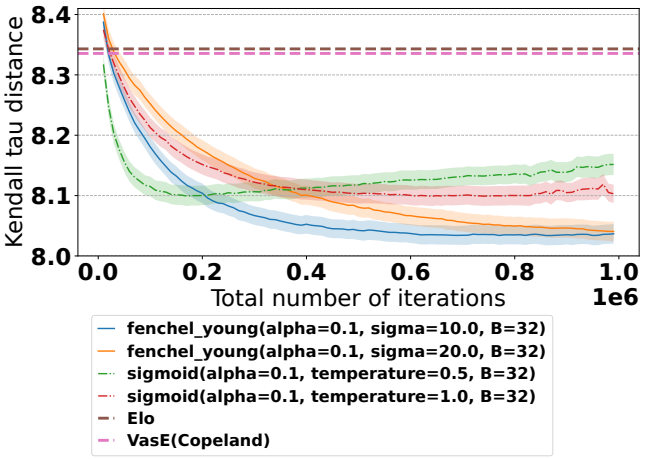

Figure 2: Average $\text{KTD}_{test}(t)$ between rankings and Diplomacy game outcomes across 50 seeds and train/test splits. All runs use batch size 32. Error bars depict 95% conf. intervals.

5 Conclusion and Continued Work

If Condorcet-consistency or distance to the optimal ranking is important, we recommend using the sigmoid loss as it minimizes the distance to it directly; despite being non-convex, in practice it finds the Condorcet winner when it exists $\geq 96\%$ of the time when the number of alternatives $m \leq 500$ and returns almost-optimal rankings on instances where it can be compared [Lanctot *et al.*, 2025, Section 4.2]. Whereas, if assured convergence or weighting the loss functions by win rates (like Elo) is more important, we recommend using the Fenchel-Young loss optimization.

Since publication of the original paper, we have also applied SCO in an online *active evaluation* setting [Lanctot *et al.*, 2026]; SCO is the best-performing method on an Atari learning agent ranking problem (out of 16 baselines), and performs comparably to Elo in a synthetic data setting.

This line of work suggests two avenues of future work: First, the approach of computing rankings by gradient descent on differentiable approximations of ranking distances can be generalized to other (or potentially all) voting methods, e.g. building on the perturbation methods applied in Fenchel-Young losses. Second, deriving a differentiable loss for the Maximal Lotteries voting method, possibly via the zero-sum game that this induces, would unlock novel applications of these methods to language model alignment [Maura-Rivero *et al.*, 2025], robust agent evaluation [Khalaf *et al.*, 2026], and possibly relates to a natural online adaptive process [Brandl and Brandt, 2024].

References

[Aziz *et al.*, 2015] Haris Aziz, Markus Brill, Felix Fischer, Paul Harrenstein, Jérôme Lang, and Hans Georg Seedig. Possible and necessary winners of partial tournaments. *Journal of Artificial Intelligence Research*, 54:493–534, 2015.

[Balduzzi *et al.*, 2018] David Balduzzi, Karl Tuyls, Julien Perolat, and Thore Graepel. Re-evaluating evaluation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[Bellemare *et al.*, 2013] Marc Bellemare, Yavar. Naddaf, Joel Veness, and Micheal Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47:253–279, jun 2013.

[Berthet *et al.*, 2020] Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach. Learning with differentiable pertubed optimizers. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9508–9519. Curran Associates, Inc., 2020.

[Brandl and Brandt, 2024] Florian Brandl and Felix Brandt. A natural adaptive process for collective decision-making. *Theoretical Economics*, 19(2), 2024.

[Brandt *et al.*, 2016] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[Campbell *et al.*, 2002] Murray Campbell, Jr. Hoane, A. Joseph, and Feng-hsiung Hsu. Deep blue. *Artificial Intelligence*, 134(1–2):57–83, 2002.

[Chiang *et al.*, 2024] Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Joseph E. Gonzalez Michael Jordan, and Ion Stoica. Chatbot arena: An open platform for evaluating llms by human preference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PLMR*, 2024.

[Conitzer and Sandholm, 2005] Vincent Conitzer and Tuomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, 2005.

[Coulom, 2005] Rémi Coulom. Bayesian Elo rating, 2005. https://www.remi-coulom.fr/Bayesian-Elo.

[Elo, 1978] Arpad E. Elo. *The Ratings of Chess Players, Past and Present*. Arco Publishing, Inc., 2nd edition, 1978.

[(FAIR) *et al.*, 2022] Meta Fundamental AI Research Diplomacy Team (FAIR), Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

[Hazan, 2015] Elad Hazan. *Introduction to online convex optimization*. Now Publishers Inc., 2015. Foundations and Trends in Optimization, Volume 2, Issue 3-4.

[Herbrich *et al.*, 2006] Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill™: A Bayesian skill rating system. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006.

[Kelly *et al.*,] Markelle Kelly, Rachel Longjohn, and Kolby Nottingham. The UCI machine learning repository. https://archive.ics.uci.edu.

[Khalaf *et al.*, 2026] Hadi Khalaf, Serena L. Wang, Daniel Halpern, Itai Shapira, Flavio du Pin Calmon, and Ariel D. Procaccia. Robust ai evaluation through maximal lotteries, 2026.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105, 2012.

[Lanctot *et al.*, 2023] Marc Lanctot, Kate Larson, Yoram Bachrach, Luke Marris, Zun Li, Avishkar Bhoopchand, Thomas Anthony, Brian Tanner, and Anna Koop. Evaluating agents using social choice theory, 2023.

[Lanctot *et al.*, 2025] Marc Lanctot, Kate Larson, Michael Kaisers, Quentin Berthet, Ian Gemp, Manfred Diaz, Roberto-Rafael Maura-Rivero, Yoram Bachrach, Anna Koop, and Doina Precup. Soft condorcet optimization for ranking of general agents. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2025. Full paper available at https://arxiv.org/abs/2411.00119.

[Lanctot *et al.*, 2026] Marc Lanctot, Kate Larson, Ian Gemp, and Michael Kaisers. Active evaluation of general agents: Problem definition and comparison of baseline algorithms. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2026. Full paper available at https://arxiv.org/abs/2601.07651.

[Liang *et al.*, 2022] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi,

Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. Holistic evaluation of language models. *CoRR*, abs/2211.09110, 2022. See latest results at: https://crfm.stanford.edu/helm/latest/.

[Liu *et al.*, 2023] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. AgentBench: Evaluating llms as agents, 2023.

[Maura-Rivero *et al.*, 2025] Roberto-Rafael Maura-Rivero, Marc Lanctot, Francesco Visin, and Kate Larson. Jackpot! alignment as a maximal lottery, 2025.

[Pini *et al.*, 2011] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Possible and necessary winners in voting trees: majority graphs vs. profiles. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 311–318, 2011.

[Rofin *et al.*, 2023] Mark Rofin, Vladislav Mikhailov, Mikhail Florinsky, Andrey Kravchenko, Tatiana Shavrina, Elena Tutubalina, Daniel Karabekyan, and Ekaterina Artemova. Vote'n'rank: Revision of benchmarking with social choice theory. In Andreas Vlachos and Isabelle Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 670–686, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

[Samuel, 1959] Arthur L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44:206–226, 1959.

[Silver *et al.*, 2016] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

[Silver *et al.*, 2018] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 632(6419):1140–1144, 2018.

[Srivastava *et al.*, 2023] Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, and et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*, 2023.

[Syed,] Aamir Syed, Omar; Syed. Arimaa – a new game designed to be difficult for computers. *International Computer Games Association Journal*, 26:138–139.

[Tesauro, 1995] Gerald Tesauro. Temporal difference learning and TD-gammon. *Communications of the ACM*, 38(3):58–68, 1995.

[Vinyals *et al.*, 2019] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[Xia and Conitzer, 2011] Lirong Xia and Vincent Conitzer. Determining possible and necessary winners given partial orders. *Journal of Artificial Intelligence Research*, 41:25–67, 2011.

[Zheng *et al.*, 2023] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric. P Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena, 2023. See also https://lmsys.org/blog/2023-05-03-arena/ and https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard. Accessed August 13th, 2023.