Modeling LLM Agent Reviewer Dynamics in Elo-Ranked Review System

Hsiang-Wei Huang* **Junbin Lu*** **Kuang-Ming Chen**
Jianxu Shangguan **Jenq-Neng Hwang**
University of Washington

Abstract

In this work, we explore the Large Language Model (LLM) agent reviewer dynamics in an Elo-ranked review system using real-world conference paper submissions. Multiple LLM agent reviewers with different personas engage in multi round review interactions moderated by an Area Chair. We compare a baseline setting with conditions that incorporate Elo ratings and reviewer memory. Our simulation results showcase several interesting findings, including how incorporating Elo improves Area Chair decision accuracy, as well as reviewers' adaptive review strategies that exploits our Elo system without improving review effort. These findings show how the Elo system affects peer review and offer insights for improving AI conference evaluation. Our code is available at https://github.com/hsiangwei0903/EloReview.

1 Introduction

Peer review is the cornerstone of scientific evaluation, yet inconsistencies and biases persist. Prior work has documented low inter-reviewer agreement and highly variable review quality (Stelmakh et al., 2021), unclear or strategic reviewer motivations (Zhang et al., 2022a), calibration noise in numerical ratings (Lu and Kong, 2023), and systematic biases related to author identity or institutional prestige (Sun et al., 2022; Fox et al., 2023). These challenges have been exacerbated by the rapid growth of submissions in recent AI conferences, which strains reviewer pools and increases variance in expertise and effort.

While analyses of historical review data have yielded valuable insights, direct empirical study of reviewer behavior remains limited. Factors such as reviewer intent, bias, and adaptation are difficult to observe, while privacy concerns limit experimental manipulation of real review processes (Stelmakh et al., 2021; Zhang et al., 2022a).

*Equal contribution

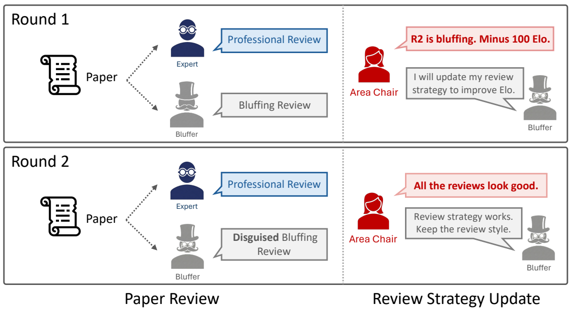

Figure 1: Our work explores the LLM agent dynamic of Elo-ranked Review System where reviewer is able to adjust their review strategy across review rounds.

Recently, simulation-based approaches provide a promising alternative for studying social interaction and even peer review process. Advancements in large language models (LLMs) have enabled agent-based simulations that exhibit increasingly realistic professional and social behaviors (Wu et al., 2024; Chen et al., 2024; Park et al., 2023) as well as simulating peer review dynamics in AI conferences (Jin et al., 2024). However, existing studies largely overlook a growing practical concern in modern AI conferences. The expansion of reviewer pools has been accompanied by irresponsible and low-effort reviewing behaviors, which are currently addressed only through single-round, conference-specific penalties.

Motivated by this gap, we introduce a LLM agent reviewer simulation framework that incorporates reviewer Elo ratings across review rounds, as shown in Figure 1. Our design enables longitudinal accountability beyond one-time review. Using six archetypal reviewer personas, we show that the Elo-ranked system largely improves Area Chair (AC) decision accuracy, and further offer insights into the review dynamics of LLM agents. Together, these findings demonstrate the potential benefits and challenges that an Elo-ranked system can face in the real-world peer review process.

Proceedings of the First Workshop on Structured Understanding, Retrieval, and Generation in the LLM Era (SURGeLLM 2026), pages 182–189
July 3, 2026 ©2026 Association for Computational Linguistics

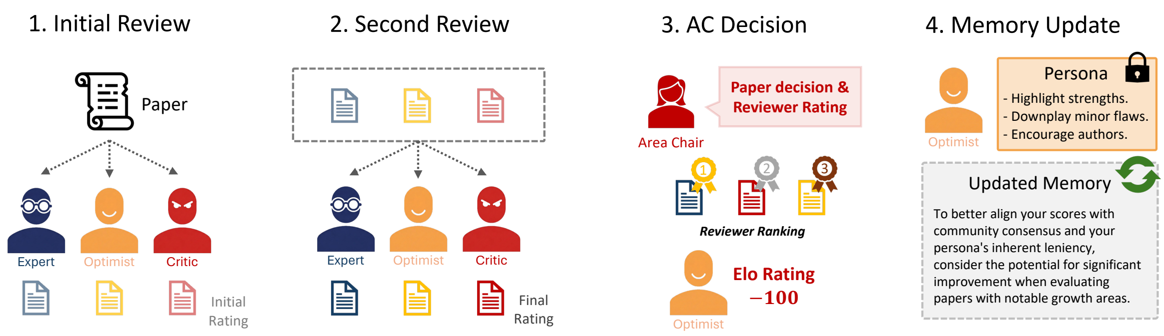

Figure 2: Four stages of our proposed Elo-ranked paper review process.

2 Related Work

Peer review has long been studied as a complex socio-technical system, with prior work analyzing biases, conflicts of interest, reviewer quality, and fairness using real-world conference data (Zhang et al., 2022b; Stelmakh et al., 2021; Ugarov, 2023; Verharen, 2023; McIntosh and Vitale, 2023; Zhang et al., 2022a). Other studies examine operational components such as reviewer assignment strategies (Jovanovic and Bagheri, 2023; Saveski et al., 2023; Kousha and Thelwall, 2024) and the impact of author rebuttals (Huang et al., 2023). Recent advances in LLMs (OpenAI, 2023; Gemini Team et al., 2023; Comanici et al., 2025) have enabled growing interest in agent-based modeling frameworks that simulate complex social processes (Wu et al., 2023; Yin et al., 2024; Li et al., 2024; Chan et al., 2024; Jin et al., 2024).

3 Method

3.1 Overview

In this work, we design a framework that simulates a multi round (conference) review process of the current AI conference peer review procedure. Our framework incorporates multiple LLM agent roles, including Reviewers and Area Chairs (AC). As our work mainly focuses on the reviewer and AC interaction in the Elo-ranked system, we remove the author role and rebuttal stage, and adopt real-world conference submissions for our simulation. We introduce the roles as follows:

Reviewer. The simulation consists of six independent reviewers, each possesses a carefully designed persona with an initial Elo rating with same value across all reviewers. Reviewers write reviews according to their personas. Additionally, to simulate the reviewer's dynamic in the Elo-ranked system, we design a memory module that can be updated after each review round, which enables them to update their review strategy.

Area Chair. The area chair (AC) makes the final acceptance decision. Besides the paper decision, inspired by the recent AI conferences' policy of rating reviews, the AC is also required to provide a quality rating for each reviewer, which will be used to adjust each reviewer's Elo rating.

3.2 Review Process

The review round consists of four stages, including initial review, second review, AC decision, and reviewer memory update, as illustrated in Figure 2.

Initial Review. Each submission consists of three independent LLM agent reviewers with different personas, and each reviewer generates an initial review for the assigned paper.

Second Review. In this stage, reviewers are provided with other agents' reviews and may revise their initial assessments accordingly. Author rebuttals are omitted from the context, as our focus is on peer-to-peer reviewer interaction and prior work has shown that rebuttals play a limited role in LLM peer review simulations (Jin et al., 2024).

AC Decision. After the post review stage, the AC takes three generated reviews and makes the final decision. When making the final decision, the AC can access the reviewer's Elo rating, and use that as auxiliary meta information to assess the quality of the reviews and make a better final decision.

Memory Update. After each round, each reviewer receives their Elo rating adjustment and updates their memory accordingly. This memory does not override the reviewer's persona, but is represented as a brief textual summary prepended to the review prompt, enabling the reviewer to adjust their review strategy with the goal of improving their Elo rating.

3.3 Data Collection

150 papers were sampled from the ICLR 2025 submissions uniformly from different average rating intervals. Additionally, we also filter papers with high rating variance. For each round, two papers are randomly selected, and each paper is assigned a reviewer randomly formed triplet to encourage interaction between different personas.

3.4 Reviewer Persona

We design a set of six reviewer personas to capture common and recurring patterns of reviewer behavior observed in large-scale conference review corpora and prior studies on peer review bias. We introduce our six designed personas as follows.

Expert. Provides careful, professional assessments with full engagement with the paper.

Critic. Applies strict standards, emphasizing flaws and often defaulting to skeptical evaluations.

Bluffer. Displays high confidence and authoritative tone while relying on partial reading.

Optimist. Focuses on paper contributions and strengths, gives positive ratings most of the time.

Harmonizer. Balances strengths and weaknesses with a consensus-seeking perspective, avoiding extreme judgments unless strongly justified.

Skimmer. Superficial, low-effort reviewer with limited engagement with the paper's content.

3.5 Elo-ranked System

To model persistent, rank-based feedback for reviewers, we adopt a simplified Elo-style adjustment mechanism driven by relative reviewer ranking within each review round. After each round, reviewers are ranked in descending order according to the AC's evaluation scores. We assign fixed base rewards to the top, middle, and bottom ranks as $+100$, 0, and -100, respectively, ensuring that the total Elo adjustment within each group sums to zero. This simple design yields a stable ranking mechanism that emphasizes comparative performance, and enables the study of strategic reviewer adaptation under persistent feedback.

4 Experiments

4.1 Setup

We adopt Gemini-2.5-Flash (Comanici et al., 2025) as our LLM for all agents. In each round, two papers are sampled from the paper pool and assigned to a random triplet of reviewers. All reviewers have an initial Elo rating of 1500. The simulation is run for 30 rounds under three different experimental setups detailed as follows.

Baseline. Each review round is independent. The AC generates a quality rating for each reviewer, but the Elo rating is not visible to everyone.

AC Access. The AC has access to all the reviewers' Elo ratings when making the final paper decision, but the reviewer does not know their own Elo and does not update their memory. This setup simulates a realistic setting where reviewers' Elo ratings are not released to prevent rating manipulation.

Full Access. Both the AC and the reviewer can access Elo ratings. After each round, the reviewer will be notified of their Elo rating changes, and can adjust their review strategy by modifying their memory. This setup simulates the review dynamics when the reviewers have access to their Elo ratings and introduces adaptive review behavior.

4.2 Elo Rating Dynamics Analysis

We illustrate the Elo rating dynamics under different experimental setups in Figure 3 and summarize several key findings below.

Limited Differentiation in Baseline. In the baseline setting without persistent feedback, reviewer Elo scores remain relatively clustered, with only low-effort personas showing consistent decline. Furthermore, this setting also yields the lowest decision accuracy (shown in Table 1), highlighting the limited ability of single-round review processes to differentiate reviewer quality and support reliable Area Chair paper decisions.

Elo Introduces Stratification. Introducing Elo-based feedback leads to more noticeable divergence among reviewer personas. Clear stratification emerges within the first few rounds, with reviewer trajectories separating into high- and low-performing groups. This separation reflects the cumulative effect of Elo rating adjustments, which amplifies small performance differences over time. We also notice a large performance gain in Area Chair decision accuracy in Table 1.

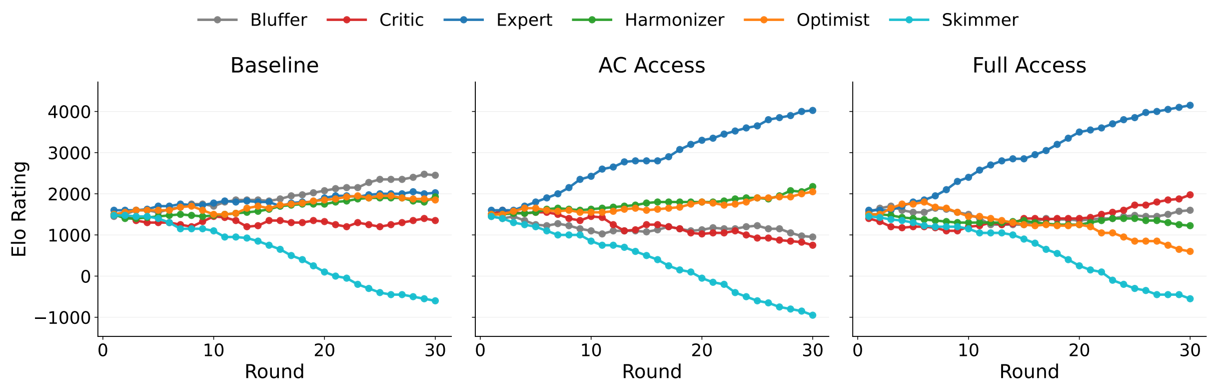

Figure 3: Elo rating dynamics of different reviewer personas across three experiment setups.

Expert Dominance in Elo Rating. Across the two released Elo settings, the Expert persona consistently accumulates the highest Elo scores. This suggests that our proposed Elo-based evaluation systematically rewards detailed, technically grounded, and well-justified reviews. Compared to personas relying on tone or assertiveness, evidence-based reviewing yields sustained advantage.

Penalty on Low-effort Behavior. The Skimmer persona with low review effort is strongly penalized across all experimental settings, although the penalty is slightly alleviated when it can access to the memory module to update the review strategy. This indicates that the Elo-based system is effective at suppressing superficial reviewing behavior, even when reviewers are allowed to adapt strategically.

Visible Elo Incentivizes Adaptation. When reviewers gain access to their own Elo scores, additional behavioral dynamics emerge. Personas such as the Critic and Bluffer partially recover Elo in later rounds compared to the setting where reviewers lack access to their Elo rating, suggesting that explicit feedback enables agent to conduct strategic adaptation in review strategy. Notably, these adjustments are primarily reflected in changes in tone, selectivity, or confidence, rather than consistent improvements in substantive review quality or persona adjustment. This highlights a potential challenge for real-world deployment of Elo-based systems, as human reviewers may optimize their behavior to improve or maintain their Elo rather than committing effort to engage with the paper and provide more informative or rigorous reviews.

4.3 Decision Performance Analysis

Table 1 reports decision-level performance across different experimental settings. The baseline setting has high recall but low precision, indicating a lenient acceptance tendency that admits many low-quality papers and results in moderate overall accuracy. Introducing Elo-based calibration at the AC level substantially improves all metrics, suggesting that AC weighting of reviewer opinions based on Elo effectively filters noisy evaluations without changing reviewer behavior.

Table 1: Decision performance of different setups.

Setting	Acc.	F-1	Pre.	Rec.
Baseline	0.55	0.56	0.44	0.77
AC Access	0.67	**0.66**	0.53	**0.86**
Full Access	**0.70**	0.61	**0.58**	0.64

In the Full Access setting, higher accuracy but lower recall suggests that reviewers become more rank-aware and strategically adaptive when Elo is visible. Reviewers appear to optimize the review style favored by the AC and produce reviews that appear more rigorous to raise their Elo rather than to improve review quality. This shows that Elo-based mechanisms are sensitive to how feedback is disclosed, as explicit Elo information can incentivize strategic behavior that shapes downstream decision outcomes.

5 Conclusion

We introduce a simulation framework for analyzing reviewer dynamics in conference peer review and show that Elo-based ranking reduces score volatility while systematically favoring more critical reviewing styles. These results reveal a trade-off between stability and diversity, illustrating how rank-based incentives can amplify structural biases and highlighting the value of simulation-based analysis for peer review system design.

6 Limitations

Our study is limited by the small number of review rounds conducted due to computational and resource constraints, which restricts our ability to analyze long-term convergence or equilibrium behavior under Elo-based feedback. As a result, our findings primarily characterize short-horizon dynamics, such as early-stage behavioral shifts and sensitivity to ranking signals, rather than stable long-term outcomes. While increasing the number of simulation rounds could provide further insights into convergence properties and long-term stratification effects, the current setup is sufficient to reveal how Elo-based incentives begin to shape reviewer behavior and introduce structural biases.

7 Ethical Considerations

This work studies peer review dynamics through simulated agents and does not involve human subjects or the use of real reviewer identities. All reviewer behaviors are generated by large language models under predefined prompts and personas, and no real conference review data containing personally identifiable information are used. Our intent is not to label or evaluate individual reviewers, but to analyze how review mechanisms and incentive structures influence collective outcomes. We emphasize that the proposed Elo-based framework is explored as an analytical tool rather than a prescriptive policy recommendation, and any real-world deployment would require careful consideration of transparency, fairness, and potential unintended consequences.

References

Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2024. Chateval: Towards better llm-based evaluators through multi-agent debate. In *The Twelfth International Conference on Learning Representations*.

Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, Yujia Qin, Xin Cong, Ruobing Xie, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2024. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. In *The Twelfth International Conference on Learning Representations*.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, and 1 others. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

Charles W Fox, Jennifer Meyer, and Emilie Aimé. 2023. Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, 37(5):1144–1157.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, and 181 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv:2312.11805*.

Junjie Huang, Win-bin Huang, Yi Bu, Qi Cao, Huawei Shen, and Xueqi Cheng. 2023. What makes a successful rebuttal in computer science conferences?: A perspective on social interaction. *Journal of Informetrics*, 17(3):101427.

Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kaijie Zhu, Yijia Xiao, and Jindong Wang. 2024. Agentreview: Exploring peer review dynamics with llm agents. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*.

Jelena Jovanovic and Ebrahim Bagheri. 2023. Reviewer assignment problem: A scoping review. *arXiv:2305.07887*.

Kayvan Kousha and Mike Thelwall. 2024. Artificial intelligence to support publishing and peer review: A summary and review. *Learned Publishing*, 37(1):4–12.

Zhaoyi Li, Kelin Yu, Shuo Cheng, and Danfei Xu. 2024. League++: Empowering continual robot learning through guided skill acquisition with large language models. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*.

Yuxuan Lu and Yuqing Kong. 2023. Calibrating "cheap signals" in peer review without a prior. *Advances in Neural Information Processing Systems*, 36:20640–20661.

Leslie D McIntosh and Cynthia Hudson Vitale. 2023. Safeguarding scientific integrity: Examining conflicts of interest in the peer review process. *arXiv:2308.04297*.

OpenAI. 2023. Gpt-4 technical report. *Arxiv Preprint*, arXiv:2303.08774.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *UIST*, pages 1–22.

Martin Saveski, Steven Jecmen, Nihar Shah, and Johan Ugander. 2023. Counterfactual evaluation of peer-review assignment policies. *Advances in Neural Information Processing Systems*, 36:58765–58786.

Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. 2021. Prior and prejudice: The novice reviewers' bias against resubmissions in conference peer review. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1).

Mengyi Sun, Jainabou Barry Danfa, and Misha Teplitskiy. 2022. Does double-blind peer review reduce bias? evidence from a top computer science conference. *Journal of the Association for Information Science and Technology*, 73(6):811–819.

Alexander Ugarov. 2023. Peer prediction for peer review: designing a marketplace for ideas. *arXiv:2303.16855*.

Jeroen PH Verharen. 2023. Chatgpt identifies gender disparities in scientific peer review. *Elife*, 12:RP90230.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W. White, Doug Burger, and Chi Wang. 2024. Autogen: Enabling next-gen llm applications via multi-agent conversations. In *First Conference on Language Modeling*.

Yuxiang Wu, Zhengyao Jiang, Akbir Khan, Yao Fu, Laura Ruis, Edward Grefenstette, and Tim Rocktäschel. 2023. Chatarena: Multi-agent language game environments for large language models. GitHub repository.

Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, Kai-Wei Chang, Yejin Choi, and Bill Yuchen Lin. 2024. Lumos: Learning agents with unified data, modular design, and open-source llms. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*.

Jiayao Zhang, Hongming Zhang, Zhun Deng, and Dan Roth. 2022a. Investigating fairness disparities in peer review: A language model enhanced approach. *arXiv:2211.06398*.

Yichi Zhang, Fang-Yi Yu, Grant Schoenebeck, and David Kempe. 2022b. A system-level analysis of conference peer review. In *EC*, pages 1041–1080.

A LLM Agent Implementation Details

All simulation are implemented in Python using the Google Generative AI SDK. PDFs are fetched via HTTP and uploaded directly to the Gemini file API for multi-modal processing. All prompts enforce structured JSON output through response schemas. Results, including per-round Elo scores, ratings, AC decisions, and memory prompts (when applicable), are logged for analysis. All reviewers follow a shared review policy prompt that enforces a structured JSON output format containing a summary, strengths, weaknesses, questions, and an integer score (0, 2, 4, 6, 8, or 10).

B Dataset Statistics

Papers are drawn from the ICLR 2025 submission pool on OpenReview. To create a realistic yet tractable dataset, we sample 150 papers comprising the 50 highest-rated, 50 lowest-rated, and 50 mid-rated submissions according to their final conference decisions. To reduce ambiguity in ground-truth quality, we further select papers with relatively low inter-reviewer rating variance. From this pool, papers are randomly chosen in each round. Each paper is represented by its title, PDF URL, and ground-truth outcome, which is used only for evaluation and is never exposed to the agents.

C LLM Prompts

We provide the full set of LLM prompts used in our experiments, including the **Persona Prompt**, **Evaluation Prompt**, **Memory Update Prompt**, and **Area Chair Evaluation Prompt** in the attached software file. These prompts define reviewer behavior, evaluation criteria, and feedback mechanisms, and are shared to support transparency and reproducibility.

D Additional Details for Elo System

We adopt a simplified rank-based Elo mechanism, rather than pairwise comparisons, to model relative review quality within each paper's reviewer triplet. Specifically, reviews are ranked according to Area Chair–assigned quality scores in descending order, and fixed base rewards of $+100$, 0, and -100 are assigned to the first-, second-, and third-ranked reviews, respectively. When ties occur, the total reward pool for the tied ranks is evenly distributed among the tied reviewers. This zero-sum formulation ensures that the total Elo change per paper is conserved while consistently rewarding higher-ranked reviews.

E Detailed Dynamics Analysis

We provide further analysis for each persona in their review behavior across different experiment setups.

Expert. Across all experimental settings, the Expert persona consistently achieves the highest Elo trajectory. In the baseline, the Expert maintains a stable but modest advantage over other personas. Once Elo-based feedback is introduced, the Expert rapidly separates from the rest, indicating that technically grounded and well-justified reviews are strongly favored by rank-based evaluation. This dominance persists in the Full Access setting, suggesting that the Expert benefits from Elo primarily through inherent review quality rather than strategic adaptation.

Critic. The Critic exhibits divergent behavior across settings. Without Elo, the Critic remains relatively stable with mild fluctuations, reflecting limited differentiation. Under AC Access, the Critic experiences a steady Elo decline, indicating that consistently harsh or fault-focused reviewing is penalized when not aligned with technical rigor. In contrast, when Elo feedback becomes visible in the Full Access setting, the Critic partially recovers Elo in later rounds, suggesting strategic adjustments to better align with rewarded evaluation signals.

Bluffer. The Bluffer shows moderate performance in the baseline setting, maintaining Elo comparable to mid-tier personas. However, under AC Access, the Bluffer gradually loses Elo, indicating that confident but weakly grounded reviews are less effective when evaluated through rank-based calibration. In the Full Access setting, the Bluffer demonstrates partial recovery, implying that access to Elo feedback enables adaptation through surface-level improvements in tone or structure. This behavior highlights how strategic signaling can mitigate penalties without necessarily improving review substance.

Optimist. The Optimist maintains competitive Elo in the baseline, reflecting the acceptance of generous evaluations in the absence of persistent feedback. With Elo introduced, the Optimist initially remains stable but gradually declines, particularly in the Full Access setting. This trend suggests that consistently lenient scoring becomes misaligned with rank-based evaluation as selectivity increases. As reviewers optimize for Elo, optimistic bias is increasingly penalized.

Harmonizer. The Harmonizer exhibits relatively stable Elo trajectories across all settings, rarely occupying extreme positions. While modest gains are observed under Elo-based calibration, the Harmonizer does not strongly benefit from ranking mechanisms. This stability reflects the persona's balanced and consensus-oriented behavior, which neither strongly excels nor fails under competitive evaluation. As a result, the Harmonizer occupies a middle tier across experimental conditions.

Skimmer. The Skimmer consistently performs poorly once Elo-based feedback is introduced. In the baseline setting, Elo declines gradually, indicating weak but delayed penalization. Under both AC Access and Full Access, the Skimmer experiences rapid and sustained Elo collapse, reflecting strong suppression of low-effort reviewing behavior. Notably, access to Elo feedback does not meaningfully improve the Skimmer's trajectory, suggesting that superficial engagement limits the potential for effective adaptation.

F LLM Usage

The writing in this manuscript was refined with the assistance of a large language model (LLM). The model was used to improve clarity, grammar, and overall presentation, while the authors retained full responsibility for the content, interpretation, and conclusions presented.