

Game Theory Meets Large Language Models: A Systematic Survey with Taxonomy and New Frontiers

HAORAN SUN, CFCS, School of Computer Science, Peking University, China

YUSEN WU, CFCS, School of Computer Science, Peking University, China

PENG WANG, School of Business, Jiangnan University, China

WEI CHEN, Microsoft Research Asia, China

YUKUN CHENG*, School of Business, Jiangnan University, China

XIAOTIE DENG*, CFCS, School of Computer Science, Peking University, China

XU CHU*, CFCS, School of Computer Science, Peking University, China

Game theory is a foundational framework for analyzing strategic interactions, and its intersection with large language models (LLMs) is a rapidly growing field. However, existing surveys mainly focus narrowly on using game theory to evaluate LLM behavior. This paper provides the first comprehensive survey of the bidirectional relationship between Game Theory and LLMs. We propose a novel taxonomy that categorizes the research in this intersection into four distinct perspectives: (1) evaluating LLMs in game-based scenarios; (2) improving LLMs using game-theoretic concepts for better interpretability and alignment; (3) modeling the competitive landscape of LLM development and its societal impact; and (4) leveraging LLMs to advance game models and to solve corresponding game theory problems. Furthermore, we identify key challenges and outline future research directions. By systematically investigating this interdisciplinary landscape, our survey highlights the mutual influence of game theory and LLMs, fostering progress at the intersection of these fields.

CCS Concepts: • General and reference → Surveys and overviews; • Computing methodologies → Artificial intelligence; • Theory of computation → Algorithmic game theory.

Additional Key Words and Phrases: Large Language Models, Game Theory, Machine Learning

ACM Reference Format:

Haoran Sun, Yusen Wu, Peng Wang, Wei Chen, Yukun Cheng, Xiaotie Deng, and Xu Chu. 2018. Game Theory Meets Large Language Models: A Systematic Survey with Taxonomy and New Frontiers. 1, 1 (August 2018), 35 pages. <https://doi.org/XXXXXX.XXXXXXX>

*Corresponding Authors.

Authors' addresses: Haoran Sun, sunhaoran0301@stu.pku.edu.cn, CFCS, School of Computer Science, Peking University, Beijing, China; Yusen Wu, sarinice2025@stu.pku.edu.cn, CFCS, School of Computer Science, Peking University, Beijing, China; Peng Wang, wp@stu.jiangnan.edu.cn, School of Business, Jiangnan University, Wuxi, Jiangsu, China; Wei Chen, weic@microsoft.com, Microsoft Research Asia, Beijing, China; Yukun Cheng, ykcheng@amss.ac.cn, School of Business, Jiangnan University, Wuxi, Jiangsu, China; Xiaotie Deng, xiaotie@pku.edu.cn, CFCS, School of Computer Science, Peking University, Beijing, China; Xu Chu, chu_xu@pku.edu.cn, CFCS, School of Computer Science, Peking University, Beijing, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

1 INTRODUCTION

Game theory provides a mathematical framework for analyzing strategic interactions among rational agents, and it has evolved significantly since its seminal work [1]. Over the decades, it has established robust methodological foundations, including equilibrium analysis [2], mechanism design [3], information design [4], and social choice theory [5]. These concepts serve as essential analytical tools across diverse disciplines such as economics, political science, and computer science, offering insights into decision-making in competitive and cooperative environments. More recently, this field has also contributed to artificial intelligence [6, 7], particularly in multi-agent systems and algorithmic game theory, as AI systems increasingly interact in complex ways.

The rapid advancement of large language models has revolutionized natural language processing and artificial intelligence [8–11]. With their remarkable capabilities in language comprehension, generation, and reasoning, LLMs are increasingly integrated into various applications and have opened new avenues for research. This growing attention on LLMs has led to a surge of studies at the intersection of game theory and LLMs. Specifically, various work investigates *how game-theoretic principles can enhance LLM evaluation and development, as well as how LLMs can contribute to advancing game theory itself*.

In this survey, we categorize these research efforts into four key directions:

- **Evaluating LLMs in Game-based Playgrounds (Section 2):** This area focuses on constructing game-based benchmark environments, such as foundational matrix games [12] and communication-based games like Avalon [13], bargaining [14], and auctions [15], to systematically evaluate the strategic reasoning capabilities of LLMs. Researchers are also investigating how advanced techniques like prompt engineering [13, 16], training [17, 18], and tool-using [19, 20] influence LLM performance in these strategic contexts.
- **Improving LLMs with Game-theoretic Methods (Section 3):** This direction explores how concepts from cooperative and non-cooperative game theory, such as Shapley Value [21], social choice theory [22], and max-min equilibria [23], can be utilized to design more efficient and theoretically sound algorithms for LLMs. Game-theoretic methods offer potential solutions for key LLM challenges, including model interpretability, general preference alignment, heterogeneity, and dynamic adaptation.
- **Characterizing LLM-related Events through Game Models (Section 4):** As LLMs are emerging technologies profoundly impacting human society and production, researchers are constructing game models to characterize events related to their development and deployment. This includes studies modeling the competition among multiple stakeholders in LLM development [24, 25], as well as those focusing on the societal impact of LLMs through game-theoretic lenses [26, 27].
- **Advancing Game Theory with LLMs (Section 5):** Leveraging their superior natural language comprehension and generation capabilities, LLMs are being employed to advance classic game theory. Specifically, LLMs can be used to solve intractable games [28, 29] and generalize classic game models to more realistic settings [30, 31], offering new computational approaches to complex game-theoretic problems.

Existing surveys on the intersection of game theory and LLMs primarily examine how game theory can be used to build evaluation environments and assess LLMs' strategic performance [32–34]. For instance, Zhang et al. [32] classified studies based on the game scenarios used to test LLM capabilities and methods for improving their reasoning. Meanwhile, Feng et al. [33] and Hu et al. [34] categorized the core competencies required for LLM-based agents in games, such as perception, memory, role-playing, and reasoning. These surveys offer valuable insights into the burgeoning field of LLM evaluation within strategic contexts. However, they predominantly adopt a unidirectional perspective,

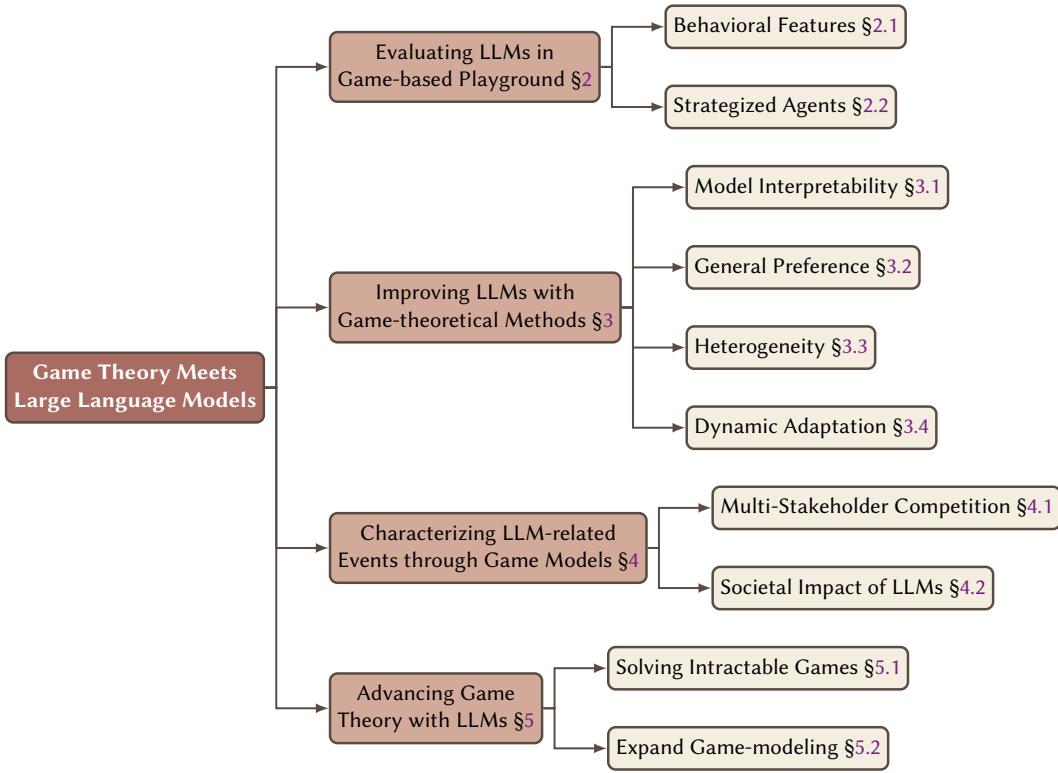


Fig. 1. A taxonomy of the intersection between game theory and Large Language Models.

treating game theory as a tool for evaluating LLMs and improving LLMs' reasoning in games, overlooking the broader roles that game theory plays in developing LLMs and how LLMs are influencing game theory. This paper bridges this gap by *introducing a novel four-part taxonomy* presented visually in Figure 1, which offers a holistic understanding of the synergistic interplay between these two critical domains. To our knowledge, this taxonomy provides *the first truly comprehensive and structured analysis of the bidirectional relationship for this dynamic and interdisciplinary landscape*. Therefore, we believe this work offers a nuanced and expansive perspective, enriching both domains.

2 EVALUATING LLMs IN GAME-BASED PLAYGROUND

The integration of LLMs into game-based environments has emerged as a powerful avenue for evaluating their cognitive and decision-making capabilities. In contrast to traditional evaluation paradigms that emphasize linguistic competence, game-based evaluations reveal deeper dimensions of LLM behavior, such as human-like decision-making, uncertainty management, and strategic interaction. This section explores the growing role of games as playgrounds for LLM assessment: Subsection 2.1 analyzes the observable behaviors of LLMs when deployed as agents in games, highlighting decision-making patterns and interactive dynamics. Subsection 2.2 focuses on the approaches to strengthen the strategic reasoning capabilities of LLMs, including prompting techniques, fine-tuning, and tool integration. Together, these provide a comprehensive perspective on how games serve as both behavioral probes and enhancement platforms for LLMs.

Table 1. A Summary of LLM-based Agents' Behavioral Features in Various Game Categories.

Game Category	Information	Nature	Interaction	Structure
Basic Matrix Games §2.1.1 e.g., Prisoner's Dilemma, Ultimatum Game, RPS [12, 35–38]	Perfect / Imperfect	Comp. / Coop.	Single-turn / Repeated	Symmetric / Asymmetric
<i>Spotlight:</i> LLMs exhibit strong pro-social biases (e.g., fairness, cooperation) often deviating from game-theoretic rationality. They struggle with probabilistic reasoning and approximating mixed-strategy Nash equilibria, showing high sensitivity to prompt framing.				
Identity Games §2.1.2 e.g., Avalon, Werewolf, Jubensha [13, 20, 39, 40]	Imperfect	Comp. / Coop.	Multi-turn, Social	Asymmetric
<i>Spotlight:</i> Capable of recursive reasoning and social modeling (e.g., influencing teammates). However, they lack strategic reliability, failing to maintain role consistency and logical coherence under pressure, and are prone to hallucinations.				
Negotiation & Coordination §2.1.3 e.g., Bargaining, Overcooked, Hanabi [19, 41–44]	Imperfect	Comp. / Coop.	Multi-turn, Communicative	Symmetric / Asymmetric
<i>Spotlight:</i> Demonstrate recognizable negotiation tactics (bluffing, anchoring) and emerging Theory of Mind (ToM). Yet, robust coordination is limited; they often regress to selfish or inconsistent strategies in complex scenarios without social scaffolding.				
Economic Games §2.1.4 e.g., Bertrand Competition, Auctions [15, 45–47]	Imperfect	Comp.	Multi-turn / Repeated	Symmetric
<i>Spotlight:</i> Display adaptive economic strategies, such as tacit collusion in pricing games. Performance is bounded by imperfect reasoning and fragile risk assessment, often failing to achieve equilibrium strategies consistently.				
Board & Card Games §2.1.5 e.g., Chess, Go, Poker [48–51]	Perfect / Imperfect	Comp.	Multi-turn	Symmetric
<i>Spotlight:</i> Baseline models show significant strategic deficiencies. They struggle with deep calculation, logical consistency, and managing uncertainty information, often producing invalid or strategically incoherent moves without specialized fine-tuning.				

2.1 Observing Behavioral Features of LLM-based Agents

This subsection focuses on the behavioral characteristics exhibited by LLM-based agents in diverse game environments, capturing their decision-making tendencies, social interaction patterns, and responses to dynamic game states. By deploying LLMs in structured scenarios, ranging from matrix games and negotiation to deception and cooperation, researchers can uncover critical traits such as risk sensitivity, strategic adaptation, and behavioral consistency. These patterns not only reveal how LLMs interpret rules and model opponents but also expose cognitive limitations and human-like biases. A taxonomy of studies categorized by game types is presented in Table 1 to provide a structured overview of the empirical landscape.

2.1.1 Basic Matrix Games. Matrix games represent foundational strategic settings in which players have complete information about the game structure, including all possible actions and payoffs. These games are often used as elementary tests of rational behavior [1]. Recent studies employ natural language prompts to position LLMs as players in matrix games, instructing them to make decisions under various utility assumptions. Findings reveal that *LLMs frequently exhibit pro-social biases, often prioritizing fairness and cooperation over game-theoretic rationality*. For instance,

LLMs consistently display higher cooperation rates than humans in social dilemmas such as the Dictator Game [35, 52–54], and often reject unfair offers in the Ultimatum Game due to inequity aversion [36, 37, 55]. The Turing Test conducted by Mei et al. [56] also confirms that LLMs tend to behave more altruistically and cooperatively, which suggests that LLMs are maximizing the average of their and the partner’s payoff by default. These behaviors likely stem from human-aligned moral patterns encoded during pre-training [38, 57].

At the same time, *LLMs often deviate from optimality*, especially in tasks demanding probabilistic reasoning or adaptive play. For example, in zero-sum games like Rock-Paper-Scissors, many models fail to approximate mixed-strategy Nash equilibria [38, 58, 59]. Repeated games like the Battle of the Sexes further reveal prompt sensitivity and coordination fragility [12, 60–62]. In adversarial contexts, LLMs often revert to risk-averse or heuristic strategies [51, 63] and display bounded rationality that favors symmetric or fair outcomes over payoff maximization [37, 55]. Beyond specific games, systematic benchmarks such as FAIRGAME [64] and SmartPlay [58] also highlight the sensitivity of LLM behavior to prompt framing and contextual cues.

2.1.2 Identity Games. Identity games like Avalon, Werewolf, and Jubensha involve hidden roles, incomplete information, and social deception, making them rich environments for evaluating higher-order reasoning and strategic communication. In these games, agents must navigate uncertainty, infer hidden roles, and engage in deception, persuasion, and collaboration. Research in this area highlights a key duality, like LLM behavior. Several studies employ LLM-based agents in an identity game-playground, letting different agents compete in the same game. Through thorough experimental observations, researchers find that *LLMs can engage in recursive reasoning and social modeling*. For example, in Avalon, agents simulate others’ beliefs and behaviors effectively [13], while in Werewolf, some LLMs serve as “opinion leaders,” influencing teammates via persuasive summarization [39]. Narrative games like Jubensha highlight LLMs’ abilities to process long-form clues, infer roles, and construct coherent narratives [20]. On the other hand, *LLMs often exhibit strategic inconsistencies and fragility*, particularly when faced with adversarial conditions or the need for robust logical deduction. AvalonBench reveals that current models have notable gaps in maintaining a consistent strategy and fully adapting to their assigned roles, especially when pressured by opponents [40]. In deduction-focused games like Werewolf, LLMs are prone to hallucination and flawed logical inference without external guidance [65, 66]. Despite appearing socially adept, these models remain susceptible to hallucinations, premature inferences, and are less aware of other players’ intentions in high-stakes interactions [67].

2.1.3 Negotiation and Coordination Games. Communication-based games, such as bargaining, provide a lens into LLMs’ negotiation acumen and collaborative reasoning. Studies show agents employing recognizable strategies like bluffing, anchoring, and making concessions during negotiations [18, 41, 68]. More advanced models like GPT-4 are proficient at goal-directed planning and successfully making deals [19, 69, 70]. In contrast, less capable models often fail to maintain a consistent persona or even complete the negotiation task [71–73]. Meanwhile, their behavior remains highly sensitive to prompt phrasing, revealing a mixture of rational optimization and socially biased responses characteristic of bounded rationality [74, 75]. In teamwork tasks, *LLMs demonstrate emerging Theory of Mind (ToM) but struggle to maintain coordination in complex, belief-intensive environments*. While LLMs show a basic ability to infer the goals and intentions of teammates in games like Overcooked and Hanabi, their capacity for robust joint planning is limited [42, 43, 68]. Davidson et al. [44] showed that cooperative bargaining games are the most challenging for LLMs. More specifically, models often regress to selfish or inconsistent strategies when social scaffolding is absent [76, 77] and there are persistent challenges related to memory, deception, and adaptation, particularly in non-cooperative or socially complex contexts [70, 71, 76].

2.1.4 Economic Games. Economic games simulate market dynamics, requiring agents to make optimal decisions about pricing, bidding, and resource allocation under competitive pressure. Success in these environments demands a nuanced understanding of market forces, risk assessment, and the ability to anticipate and respond to competitors' actions. In these scenarios, including pricing games and auctions, *LLMs demonstrate adaptive and sophisticated economic strategies, yet their performance is consistently bounded by imperfect reasoning and risk assessment*. On the one hand, LLMs can emulate complex economic behaviors. For instance, tacit collusion and reward-punishment strategies emerge in repeated Bertrand games, with GPT-4 agents learning to maintain supracompetitive prices [45]. In dynamic auction environments, LLMs also show strategic planning by updating beliefs and reprioritizing items based on evolving conditions [15, 46]. However, this strategic acuity is often fragile and lacks robustness. Studies show that simpler heuristics can outperform LLMs, and equilibrium strategies are inconsistently achieved, revealing gaps in their strategic depth [15, 46, 51]. These observations reinforce the framing of LLMs as boundedly rational economic agents whose strategic capabilities are often superficial [47].

2.1.5 Board and Card Games. Classic board and card games serve as demanding benchmarks for AI, as they require capabilities that are distinct from natural language processing. Success in these domains hinges on deep strategic planning, precise calculation, and sophisticated management of uncertainty. Across both perfect and imperfect-information games, *baseline LLMs exhibit significant strategic deficiencies, struggling with deep calculation, logical consistency, and the management of uncertainty*. In perfect-information games like Chess and Go, which demand rigorous, forward-looking planning, LLMs often fail to produce strategically coherent or even legally valid sequences of moves—a core deficiency that has necessitated the creation of specialized, search-augmented models to achieve competence [48, 78]. This challenge extends to imperfect-information games, where the core difficulty shifts from pure calculation to managing strategic uncertainty. In poker, for instance, benchmarks show that vanilla LLMs perform poorly without significant fine-tuning, primarily due to their inability to handle hidden information and assess risk effectively [17, 49]. Similarly, in complex tactical battle games, these models are prone to hallucination and erratic play [50]. Taken together, evidence from multi-game testbeds confirms that deep planning limitations and strategic inconsistencies are pervasive weaknesses of general-purpose LLMs in these calculation-intensive contexts [51].

2.1.6 Established Benchmarks. To systematically probe the strategic capabilities of LLMs, a growing number of specialized benchmarks have been developed. These platforms provide controlled environments to assess specific facets of agent behavior, from pure rationality to complex social interaction. Several benchmarks focus on foundational strategic reasoning in classical games. For example, GTBench [79], γ -bench [80], GameBench [81], and the work by Topsakal et al. [82] primarily assess whether LLM decisions align with theoretical equilibria. Building on this, other frameworks like FAIRGAME [64] and SmartPlay [58] investigate the nuances of these decisions, examining deviations such as fairness biases, prompt sensitivity, and other signatures of bounded rationality.

The evaluation landscape has also expanded to encompass more dynamic and socially complex scenarios. Benchmarks like LLMArena [83], AvalonBench [40], and NegotiationArena [84] place LLMs in multi-agent settings to test higher-order recursive reasoning, coordination, and deception. The scope further extends into economic and hybrid domains, with GLEE [85] and TMGBench [86] evaluating agents in pricing and auction tasks, while ALYMPICS [87] and Welfare Diplomacy [72] integrate negotiation and moral reasoning. To address concerns about agent fragility, recent additions like lmgame-Bench [88] and Playing Games [38] introduce metrics for disqualification and randomness-handling to better identify brittle behaviors. Collectively, this diverse and rapidly evolving suite of benchmarks provides an essential toolkit for diagnosing, comparing, and ultimately improving the strategic capabilities of LLM-based agents.

2.2 Strategizing LLMs' Reasoning in Games

This subsection explores methods to enhance LLMs' strategic reasoning and performance in game-based environments, addressing the challenges of optimizing their decision-making processes. Techniques such as fine-tuning on game-specific data, incorporating reinforcement learning, or integrating external reasoning frameworks enable LLMs to better navigate complex game scenarios, from perfect-information games like chess to imperfect-information settings like poker. These approaches aim to improve LLMs' ability to anticipate opponents' moves, optimize strategies, and adapt to dynamic game states. This subsection highlights the advancements in enabling LLMs to exhibit sophisticated, goal-oriented reasoning in competitive and cooperative game contexts.

2.2.1 Stimulating Reasoning with Advanced Prompting. Advanced prompt techniques are often used to improve LLMs' reasoning capability [89–91]. Several studies have designed more complex prompts for specific game tasks. Wang et al. [13] developed the Recursive Contemplation (ReCon) framework to enhance the strategic reasoning of LLMs in Avalon. By prompting LLMs to employ first- and second-order perspective-taking, this framework mitigates common failures like deceptive behavior. Similarly, Duan et al. [60] proposed a method where LLMs predict future moves in multi-turn games, improving their ability to anticipate opponents' strategies. Additionally, Zhang et al. [16] advanced LLMs' reasoning through K -level rationality, which enhances multi-level thinking and significantly increases their win rates in competitive settings. These findings suggest that recursive reasoning can substantially improve LLMs' strategic capabilities. Kempinski et al. [92] proposed algorithms that guide LLMs to iteratively refine their action choices by simulating game outcomes in self-play. These methods align directly with the themes discussed in this section, such as recursive reasoning and advanced prompting techniques for strategic capabilities. Beyond recursive approaches, advanced prompting techniques also focus on integrating feedback, human-like reasoning, and Theory of Mind. Fu et al. [18] demonstrated that LLMs can autonomously improve negotiation strategies through self-play, leveraging in-context learning from AI feedback where a critic LLM provides structured critiques to a player LLM. In the context of multi-agent mystery games, Wu et al. [20] enhanced agents' information gathering and logical reasoning by incorporating advanced prompting engineering, allowing them to decipher complex scenarios more effectively. Building on this, Guo et al. [93] introduced "Suspicion-Agent," which utilizes prompt engineering to harness GPT-4's high-order Theory of Mind capabilities, enabling it to understand and intentionally influence opponents' behavior in imperfect information games. Furthermore, Abdelnabi et al. [69] employed systematic zero-shot Chain-of-Thought (CoT) prompting to enable LLM agents to successfully negotiate in multi-agent games, highlighting the role of explicit reasoning steps. In a similar vein, Yim et al. [42] proposed a ToM planning technique for LLM agents to adapt their strategies in cooperative games under imperfect information, demonstrating how specific prompts can simulate an understanding of other agents' beliefs. Lastly, Gandhi et al. [63] showed that systematically generated few-shot CoT examples can enable LLMs to achieve strategic reasoning that generalizes across diverse game structures and objectives. These diverse methods underscore the power of carefully designed prompts in stimulating sophisticated reasoning in LLMs across various game settings.

2.2.2 Developing Task-Specific Ability with Training. Complementing cognitive frameworks, novel training paradigms leverage self-play and AI feedback to overcome data limitations and improve strategic adaptability. Fu et al. [18] employed iterative self-play with AI-generated feedback to refine negotiation strategies in dynamic environments with hidden goals. Guo et al. [94] introduced self-supervised learning with auxiliary state-derived rewards, enabling mastery of complex games like Hanabi without human data. Kwon et al. [95] used LLMs as intrinsic reward designers, reducing

dependency on human-engineered reward functions for reinforcement learning. Complementing these, Zhang et al. [96] implemented policy-level reflection via evolutionary algorithms, enabling LLM agents to self-optimize strategies without parameter retraining. Similarly, Wang et al. [78] tackled data scarcity in sensitive domains through algorithmic synthesis of statistically faithful game-theoretic scenarios using non-parametric copula simulators. These methods collectively enhance LLMs' ability to develop robust strategies through experiential learning.

Further advancements include an LLM-based framework by Wei et al. [97], which automates reward function discovery for reinforcement learning in cooperative platoon coordination, initializing rewards via a chain of thought and iteratively optimizing them through an evolutionary module based on training feedback. Suzuki and Arita [98] proposed an evolutionary model where LLMs are instructed with high-level psychological and cognitive character descriptions as "genes" to simulate human behavior in game-theoretical scenarios, evolving the population through selection based on average payoff and mutation of these linguistic trait descriptions. Feng et al. [48] introduced ChessGPT, which bridges policy learning and language modeling by integrating historical policy data and natural language analytical insights from chess games, training models on this large-scale combined dataset. Liao et al. [99] demonstrated the efficacy of language model self-play in non-zero-sum games by fine-tuning LLMs over multiple rounds of filtered behavior cloning, showing substantial improvements in task reward. Wang et al. [78] empowered LLMs in decision games through targeted post-training by designing data synthesis strategies to curate extensive offline datasets from games like Doudizhu and Go, then developing techniques to effectively incorporate this data into LLM training. Jin et al. [100] proposed an RL-instructed language agent framework for One Night Ultimate Werewolf, where a discussion policy is trained by reinforcement learning to determine strategic discussion tactics, guiding the LLM's communication based on game context. Zhuang et al. [49] introduced POKERBENCH, a benchmark for evaluating poker-playing abilities, demonstrating marked improvements in LLM performance after fine-tuning using structured "Few-Shot Prompts" that provide detailed game scenarios for strategic decision-making. Huang et al. [17] presented PokerGPT, an end-to-end solver for multi-player Texas Hold'em that fine-tunes a lightweight LLM using reinforcement learning from human feedback based on textual records from real games. Finally, Yang and Berthelemy [71] enhanced LLMs in non-cooperative games by integrating a tree of thoughts and a multi-agent framework, where game-solving is decomposed into incremental tasks and an automated fine-tuning process optimizes performance by ranking query-response pairs based on game outcomes.

2.2.3 Integrating Auxiliary Modules and Tools. Beyond direct prompting and training, integrating auxiliary modules and external tools is crucial for enhancing LLMs' game-playing ability by providing structured knowledge or specialized reasoning [101, 102]. For instance, Yim et al. [42] integrated a Theory of Mind planning technique with an external tool in Guandan, using prompts for strategic adaptation based on game context. Similarly, Xia et al. [19] enhanced bargaining with OG-Narrator, which employs prompts to structure offers and translate them into natural language. Several works focus on infusing logical or strategic frameworks: Watanabe et al. [65] improved Werewolf agents by embedding explicit logical structures via prompts for deductive reasoning. Wu et al. [20] used advanced prompting within a multi-agent framework to boost information gathering and logical reasoning in mystery games. Hua et al. [81] developed a game-theoretic agent workflow for negotiation, guiding LLM decisions with specific prompts based on game theory. Lan et al. [83] utilized a multi-agent system for Avalon, where the system prompts directed agents' gameplay and social behavior analysis. Other approaches include Hu et al. [50]'s POKÉLLMON, which uses prompts to enable in-context reinforcement learning and knowledge-augmented generation for Pokéémon battles. Guan et al. [73] enhanced AI Diplomacy agents with a strategic planner that specifies sub-goals for long-term objectives. Lastly, the

STRIDE framework [103] integrates memory and specialized tools, with prompts enabling LLM agents to interact for rule adherence, planning, exploration, and opponent anticipation. These diverse integrations significantly bolster LLMs' strategic capabilities.

Discussions

Many current findings on the behavioral characteristics of LLMs are closely tied to specific model architectures or versions, making them potentially obsolete as the technology rapidly evolves. It would be highly valuable to explore or derive more fundamental and generalizable strategic patterns of LLM behavior through game-theoretic scenarios that transcend individual models. While recent efforts to enhance LLMs' reasoning abilities have largely focused on task-specific approaches, developing a unified framework for improving general game-playing and reasoning capabilities remains a significant and open challenge.

The observed behavioral patterns, particularly the pro-social biases and strategic limitations, are not arbitrary phenomena. They are, in large part, artifacts of the very training and alignment techniques used to build these models. The tendency towards fairness and cooperation, for example, is a direct result of methods designed to make LLMs helpful and harmless. In the next section, we turn our attention to these underlying mechanisms, examining how game-theoretic principles are themselves being used to improve LLMs, which in turn shape the behaviors we have just reviewed.

3 IMPROVING LLMS WITH GAME-THEORETICAL METHODS

Game-theoretical approaches have been more frequently utilized to describe LLMs' theoretical characteristics and to develop practical algorithms that improve their empirical outcomes. This section explores how principles and methodologies from game theory contribute to addressing key challenges in the development and optimization of LLMs. The discussion is then organized into four subsections, each corresponding to a critical challenge faced by LLMs: Subsection 3.1 addresses the challenge of *interpretability*. A line of work constructs cooperative game scenarios involving an LLM's input, training data, and internal components, and utilizes the *Shapley value* [104] to provide principled credit assignment for each contributing factor. Subsection 3.2 focuses on *aligning general preferences*. Nash Learning from Human Feedback (NLHF) formulates a *minimax game between policies* to capture intransitive preferences. Further extensions develop more efficient algorithms with provable theoretical guarantees and broader applicability. Subsection 3.3 discusses the challenge of *heterogeneity* in value alignment. Recent work integrates *social choice theory* with reinforcement learning from human feedback (RLHF), offering axiomatic frameworks for alignment and principled strategies to resolve diverse, potentially conflicting preferences. Subsection 3.4 investigates the issue of *dynamic adaptation* by modeling LLM development as a *competitive game* involving multiple co-evolving players. Table 2 summarizes the key challenges, core game-theoretical concepts, and corresponding game formulations.

3.1 Enhancing Model Interpretability through Cooperative Game

Interpreting the behavior of deep learning models is a highly desirable objective, as it enables researchers to gain insights into how models function, thereby facilitating further improvements and providing guarantees of reliability [116–118]. Such an interpretability challenge becomes more critical for LLMs, due to their massive scale, often comprising billions or even trillions of parameters and trained on vast datasets. The SHAP framework [119] adopts the Shapley value [104]

Table 2. Application of Game-theoretic Methods for Improving Large Language Models.

Challenge	Core Game-theoretic Concepts	Game Formulation
Model Interpretability §3.1	Cooperative Games (Shapley Value)	Players: LLM components (tokens, data points, layers, heads). Goal: Cooperate to generate the model's output. <i>Spotlight:</i> Provides a principled method for credit assignment, enabling input attribution, data valuation, and model pruning. <i>Examples:</i> TokenSHAP, Data Shapley, SV-NUP [105–107]
Aligning General Preferences §3.2	Minimax Games (Nash Equilibrium)	Players: Two policies competing to be preferred by a human or learned oracle. Goal: Find a stable policy that cannot be consistently defeated (a Nash Equilibrium). <i>Spotlight:</i> Overcomes limitations of scalar reward models, enabling robust alignment with complex (e.g., intransitive) preferences. <i>Examples:</i> NLHF, SPO, DNO, MPO [23, 108–110]
Capturing Preference Heterogeneity §3.3	Social Choice Theory & Cooperative Bargaining	Players: User subgroups or annotators with diverse values. Goal: Aggregate preferences or negotiate outcomes according to formal fairness axioms. <i>Spotlight:</i> Moves beyond monolithic alignment to design equitable systems that respect minority views and handle conflicting values. <i>Examples:</i> MaxMin-RLHF, Axiomatic Analysis, Negotiative Alignment [22, 111, 112]
Dynamic Adaptation §3.4	Competitive Self-Play, Stackelberg Games, & Bilevel Optimization	A co-evolving game with two primary forms, mapping to the section's structure: <ul style="list-style-type: none">• Evolving Data (§3.4.1): A generator model creates challenging data for a critic or its past self.• Evolving Rewards (§3.4.2): A "leader" policy optimizes against a "follower" reward model that finds its weaknesses. <i>Spotlight:</i> Replaces static components with a dynamic process that prevents reward hacking and enables continuous, autonomous improvement. <i>Examples:</i> SPIN, STA-RLHF, Decoding Game [113–115]

to quantify the contribution of individual components of a deep learning model. Inspired by this, recent work considers formulating cooperative games in different stages of LLMs, in which each player represents a semantically meaningful part, such as individual tokens in a prompt, specific datasets used in training, or particular layers within the LLM architecture. By choosing appropriate performance metrics, the researchers apply the Shapley value to measure how each player contributes to the overall performance, offering a theoretically grounded measure of interpretability. However, the exponential complexity of exact Shapley value calculation necessitates the development of efficient and accurate approximation techniques, which has become a central theme in this line of research.

3.1.1 Input Attribution. Several studies focus on quantifying the importance of an LLM's input components, from entire prompts to individual tokens. At the prompt level, Liu et al. [120] applied the Shapley value to multi-prompt learning. They proposed a learning-based approach to predict the value of each prompt, facilitating more effective prompt engineering by identifying prompts that enhance performance. At a finer granularity, TokenSHAP [105] and TextGenSHAP [21] model individual tokens as players in a cooperative game to attribute the model's output to specific

parts of the input. TextGenSHAP [21] introduces speculative decoding and in-place encoder resampling to make token-level attribution for long-context inputs computationally tractable. Using a similar method, Mohammadi [121] uncovered a “token noise” phenomenon: LLM decisions are disproportionately affected by tokens with minimal semantic content, like invisible newline characters. Moving from token-level to document-level analysis, Ye and Yoganarasimhan [122] applied the Shapley framework to value source documents in LLM-generated summaries, a key challenge in Retrieval-Augmented Generation systems. They introduced Cluster Shapley, which groups semantically similar documents, to overcome the computational expense. In their model, each source document is a player whose value is determined by its marginal contribution to the quality of the final summary.

3.1.2 Training Data Valuation. As LLMs are trained on various massive datasets, it is insightful to quantify how each dataset or data instance influences the model’s performance. *Data Shapley* [106] models each training sample as a player and uses the Shapley value to evaluate its marginal contribution to the final model’s performance across all possible training subsets. In the LLM context, this concept has been operationalized for both dataset refinement and dynamic learning. For dataset curation, the SHED framework [123] offers a scalable solution for instruction fine-tuning. It approximates Shapley values over clusters of training data to assemble an optimized dataset. The framework also demonstrates strong cross-model transferability, where data selected using a smaller model remains effective for larger ones, significantly reducing curation costs. Beyond static dataset curation, this principle extends to the dynamic process of RLHF. The SCAR framework [124] addresses the sparse reward problem in RLHF by modeling text segments as players in a cooperative game. By distributing the final reward among text units based on their Shapley value—approximated efficiently using Owen values—SCAR provides a dense, principled reward signal. Empirically, this method improves both convergence speed and final policy performance in tasks like summarization and instruction following.

3.1.3 Probing Internal Components. Beyond external inputs and data, Shapley-based methods are also applied to the internal components of LLMs. By conceptualizing architectural components such as layers and attention heads as “players” in a cooperative game, researchers can assess the marginal contribution of each part to the model’s overall performance. This abstraction provides not only interpretability insights but also practical guidance for model optimization and compression. One line of work evaluates the importance of entire layers. Zhang et al. [125] used a Shapley-based neighborhood sampling method to identify *cornerstone layers*, a small subset of layers whose removal causes a drastic performance collapse. Following this, Sun et al. [107] proposed the Shapley Value-based Non-Uniform Pruning (SV-NUP) framework. It employs an efficient Sliding Window-based Shapley Value (SWSV) approximation to assign pruning ratios based on layer contributions, achieving superior compression-performance trade-offs. Other studies focus on finer-grained components like attention heads. Held and Yang [126] applied *Shapley Head Values* (SHVs) to identify and remove interfering attention heads in multilingual models, achieving performance improvement without adding or retraining parameters. To make such fine-grained analysis computationally feasible, amortized methods have been introduced. Yang et al. [127] trained an auxiliary model to directly predict Shapley values, achieving speedups while ensuring deterministic, stable explanations. Fekete and Bjerva [128] used SHVs and clustered the resulting value vectors to reveal how different attention heads specialize in processing specific morphosyntactic phenomena. Furthermore, Yang et al. [129] proposed a benchmark and evaluation metrics based on Shapley value to evaluate how each module, such as planning, reasoning, action execution, and reflection, contributes to the reasoning improvement of an LLM.

3.2 Aligning General Preferences through Min-Max Equilibrium

Aligning LLMs with general human preferences is a core challenge in the deployment of responsible, reliable language models. Traditional RLHF typically relies on reward models built upon the Bradley-Terry (BT) assumption, which imposes a strict transitive and scalar reward structure [130, 131]. However, human preferences are often stochastic and intransitive, which are poorly modeled by BT-style assumptions [23, 108]. Recent research reconceptualizes alignment as a two-player min-max game between policies and proposes a series of novel algorithms based on Nash equilibrium, self-play, and preference optimization without explicit reward modeling. These approaches aim to directly capture complex, expressive feedback and offer both theoretical guarantees and empirical improvements.

3.2.1 Preference Optimization. The foundational idea of NLHF [23] is to model alignment as a two-player zero-sum game, where each policy aims to outperform the other based on a learned preference function. This approach does not need reward modeling and allows convergence toward a Nash equilibrium policy under general preferences. Building on NLHF, Self-Play Preference Optimization (SPO) [108] proposes a minimalist yet generalizable framework. SPO treats trajectory comparison as a symmetric game and uses self-play to train a policy against itself using win-rate scores, achieving robustness to non-Markovian and intransitive preferences. Self-Play Preference Optimization for LLMs (SPPO) [132] adapts SPO for practical LLM fine-tuning by introducing a new squared-loss objective. To address sample efficiency and training stability, Direct Nash Optimization (DNO) [109] introduces a batched, on-policy, regression-based objective. DNO combines the scalability of contrastive learning with Nash theoretical soundness and is more empirically efficient. Similarly, Iterative Nash Policy Optimization (INPO) [133] proposes a no-regret online learning framework that avoids costly win-rate estimation by directly minimizing a surrogate loss over a preference dataset. Furthering this direction, Ye et al. [134] provided a theoretical study of KL-regularized RLHF under a general preference oracle, completely bypassing the BT model. They formulate alignment as a minimax game and propose distinct algorithms for offline (Pessimistic Equilibrium Learning, PELHF) and online (Optimistic Equilibrium Learning, OELHF) settings, both with finite-sample theoretical guarantees. The learnability of the KL-regularized NLHF is further verified in the work by Ye et al. [135].

3.2.2 Theoretical Advancements in Convergence and Efficiency. Further studies are focusing on achieving provably efficient and convergent training dynamics for NLHF. Several recent works observe oscillatory behavior, high variance, and slow convergence in preference-based self-play. Nash Mirror Prox (NashMP) [136] addresses these issues by leveraging the Mirror Prox method in online Nash Learning from Human Feedback. NashMP achieves last-iterate linear convergence to the β -regularized Nash equilibrium, with convergence rates that are independent of action space size. Extragradient Preference Optimization (EGPO) [137] extends this idea, providing both linear convergence for regularized games and polynomial convergence for unregularized settings. Magnetic Preference Optimization (MPO) [110] is designed to converge to the Nash equilibrium of the *original, non-regularized* game. It achieves this by periodically updating its reference “magnetic” policy to the equilibrium of the previous regularized step, guiding the policy sequence toward the unregularized objective. Beyond theoretical convergence, Wang et al. [138] proposed TANPO (Two-Agent Nash Policy Optimization) to improve sample efficiency and exploration-exploitation balancing. In TANPO, the “min-player” is incentivized to explore via an exploration bonus, thereby generating more diverse and informative training data for the “max-player”. Other work stabilizes training through regularization. Tang et al. [139] introduced Regularized Self-Play Policy Optimization (RSPO) that integrates forward and reverse KL divergence regularization into self-play, allowing for tunable trade-offs between response length, win-rate, and diversity. Alami et

al. [140] further explored the dynamics of KL-based regularization in self-play, showing that geometric mixing of the base and reference policies improves performance stability. They also highlight how fictitious play helps smooth the training trajectory and prevent performance oscillations, which helps prevent training collapse and boost generalization across benchmarks.

3.2.3 Limitations of Game-Theoretic Alignment. Although game-theoretic methods for alignment provide flexibility and robustness to preference inconsistencies, there are foundational critiques that raise concerns about the sufficiency of preferences for full value alignment. Sun et al. [141] revisited the foundational Bradley-Terry model, arguing that while it offers order consistency, it may not be necessary or optimal. They advocate for classification-based alternatives and emphasize the role of annotation sparsity and structure in reward modeling quality. Going deeper, Shi et al. [142] investigated theoretical limits of game-theoretic alignment, demonstrating fundamental limits to preference matching. They proved that no smooth payoff mapping can guarantee an equilibrium solution that precisely matches a target preference profile, nor that such an equilibrium is unique. They further showed that achieving other desirable properties, such as Smith consistency, imposes strict structural constraints on the game, requiring it to be equivalent to a symmetric zero-sum game. From a philosophical perspective, Zhi et al. [143] challenged the entire preferentialist paradigm. They argued that preferences lack the semantic richness and context-dependence of human values, and thus fail as alignment targets. They called for a reframing of alignment to be based on normative standards tailored to AI roles, negotiated across stakeholders, rather than preference aggregation.

3.3 Capturing Preference Heterogeneity with Social Choice Theory

Another formidable challenge in aligning large language models with human preferences is the heterogeneity of human preferences themselves. Conventional alignment methods, which often rely on learning a single, scalar reward function, overly prefer the majority’s point and fail to handle this heterogeneity [144]. To address this problem, game theory and social choice theory offer both a formal language to diagnose the core difficulties of this problem and a constructive paradigm for developing more robust and equitable alignment algorithms.

3.3.1 An Axiomatic Social Choice Perspective. Recent work has recognized the limitations of RLHF by formally *mapping the alignment problem to the field of social choice theory*. This perspective reveals that many challenges are not merely technical but are rooted in fundamental paradoxes of collective decision-making. Mishra [145] showed that preference aggregation in AI is subject to classic impossibility theorems from social choice. By mapping AI alignment to this setting, he demonstrated that any aggregation rule inevitably violates at least one of Arrow’s axioms—such as unanimity, independence of irrelevant alternatives, or non-dictatorship—meaning no single “fair” preference aggregator can exist. Complementing this, Dai and Fleisig [146] formalized the mapping between RLHF and social choice, enabling a deeper critique. For example, it has been shown that canonical RLHF methods based on the Bradley-Terry model are mathematically analogous to the Borda count voting rule [147], where an option’s score is derived from the sum of its pairwise “wins.” This connection highlights how RLHF implicitly adopts a specific, and not always desirable, social choice function. *This axiomatic lens also provides a powerful tool for critiquing current alignment mechanisms by testing them against formal properties.* Ge et al. [22] evaluated RLHF algorithms and found they violate fundamental axioms like Pareto Optimality and Pairwise Majority Consistency. Extending this critique, empirical work by Hosseini and Khanna [148] confirms this misalignment, showing that LLMs often default to welfare-maximization in resource allocation tasks, violating human concepts of distributive fairness like equitability. Procaccia et al. [149] identified a critical vulnerability to “approximate clones”—semantically similar responses that trick MLE-based methods by artificially inflating an

option's representation in the data; they showed these methods fail the axiom of clone-independence and propose a weighted MLE solution to correct it. Extending the critique to utility, Gölz et al. [150] introduced the concept of “distortion” as the worst-case ratio of optimal social welfare to the welfare achieved by an RLHF policy, proving that for methods like RLHF, this distortion can grow exponentially with preference intensity, and can even be unbounded under certain data sampling conditions. In contrast, Xiao et al. [151] offered a reconciliation, explaining why these theoretical failures don't always break RLHF in practice. They demonstrated that under a common and empirically plausible condition—where each response pair is evaluated by at most one annotator—the complex preference cycles needed to trigger many axiomatic violations rarely form, thus preserving properties like Condorcet Consistency. This highlights a central tension in the field: the clash between theoretical impossibility results and the empirical effectiveness of existing methods under specific, practical conditions.

Building on these insights, *researchers have proposed new axioms and normative frameworks to guide the development of more principled alignment systems*. Position papers by Conitzer et al. [152] and Zhang et al. [153] argue for embedding principles from social choice and mechanism design directly into the alignment process. The latter introduces the Incentive Compatibility Sociotechnical Alignment Problem (ICSAP), which uses mechanism design to create protocols where stakeholders are incentivized to reveal their true preferences. This concern with strategic incentives is formalized by Wu et al. [154], whose game-theoretic analysis of competing data providers shows that strategic exaggeration of preferences is an almost inevitable Nash Equilibrium, highlighting a fundamental source of data corruption. To build more legitimate systems, others are designing new elicitation processes, such as the “Moral Graph” proposed by Klingefjord et al. [155], which uses a structured dialogue to construct a graph of human values capable of surfacing nuanced and expert opinions. Another direction focuses on axioms of representation. Kim et al. [156] introduced a framework grounded in two new axioms: Population-Proportional Representation (PPR) and Population-Bounded Robustness (PBR), achieved by learning a latent mixture of preference groups. This call for proportional representation is echoed by Peters [157], who argues for its broad applicability in AI to mitigate the “tyranny of the majority” in contexts ranging from RLHF to the aggregation of LLM outputs. Complementing this, Qiu [158] connects alignment to statistical learning theory, proposing axioms that ensure a preference model learned from a sample of users will generalize fairly to the entire population.

3.3.2 Developing Practical Algorithms. The theoretical shortcomings of a single, monolithic reward function, particularly its failure to satisfy axioms like Pareto Optimality [22] and its tendency towards high distortion [150], motivate a new class of practical algorithms designed to explicitly handle preference diversity. Empirical work confirms this necessity. For example, Fleisig et al. [144] treated annotator disagreement as a crucial signal, developing a model to predict the ratings of a text's targeted group to identify instances where the majority opinion is “wrong”. Shirali et al. [147] provided theoretical proof that Direct Preference Optimization (DPO) fails to find the utilitarian optimum with heterogeneous users, instead implicitly optimizing for the Borda count and thus being sensitive to data sampling distribution. This work also uncovers a trade-off: achieving a consistent estimate of the optimal policy requires discarding data with annotator disagreement, thereby sacrificing sample efficiency. Directly addressing this, Cheng et al. [159] proposed Vote-based Preference Optimization (VPO), which leverages the quantitative vote counts in preference data. By modeling preference strength, VPO can distinguish between clear consensus and controversial opinions, leading to more stable and effective alignment. This issue of inconsistency is underscored by Liu et al. [160], who show that Condorcet cycles (*e.g.*, A > B, B > C, C > A) are a near-certainty in diverse preference data. Since a single scalar reward function cannot represent such cycles, this proves its structural inadequacy and motivates algorithms capable of producing mixed or pluralistic outputs.

One major line of work involves learning distinct reward models for different preference clusters and then aggregating their outputs based on principles from social choice. Chakraborty et al. [111] proposed MaxMin-RLHF, which implements the Rawlsian “max-min” social welfare function by learning a mixture of reward models for latent subgroups and maximizing the utility of the worst-off group. Similarly, Chen et al. [161] used a mixture-of-experts approach to learn latent preference prototypes, while Park et al. [162] used spectral clustering to identify taste-based groups before training group-specific reward models. Providing a theoretical backing, Zhong et al. [163] used meta-learning to extract group-specific rewards and analyze the sample complexity of aggregating them using different social welfare functions.

Other approaches innovate on the aggregation mechanism itself, moving beyond the “learn-then-aggregate” paradigm. Alamdari et al. [164] proposed aggregating at the policy level, where multiple policies trained on individual preferences have their actions combined at inference time using voting rules. The viability of such voting mechanisms is empirically explored by Yang et al. [165], who find that LLM collective decisions are sensitive to voting protocols and exhibit biases. While using “personas” can improve alignment with human choices, it reveals a difficult trade-off between alignment accuracy and preference diversity. Halpern et al. [166] introduced Pairwise Calibrated Rewards, an ensemble method that learns a distribution of reward functions calibrated to match the proportion of human annotators holding a given preference, thus preserving pluralism. Drawing from cooperative game theory, Mushkani et al. [112] proposed Negotiative Alignment, a multi-agent framework where agents representing stakeholder groups use bargaining protocols to reach collective decisions. The potential of such negotiative approaches is highlighted by behavioral experiments from Wang et al. [167], who show that an “imperfectly fair” LLM agent—one that engages in human-like strategic communication—can overcome the typical “machine penalty” and foster cooperation in social dilemmas, suggesting the power of dynamically negotiated alignment.

The integration of social choice theory has transformed the understanding of AI alignment with heterogeneous preferences, shifting the focus from a single reward function to a rich tapestry of methods that embrace diversity. Future work will likely focus on bridging the gap between theoretically ideal but computationally expensive mechanisms, such as those involving negotiation or full pairwise calibration, and scalable algorithms that can be deployed in real-world systems, while also addressing the crucial challenge of incentive compatibility in data collection.

3.4 Achieving Dynamic Adaptation through Competitive Game

Traditional LLM training pipelines often rely on two fixed components: a static, pre-collected dataset for supervised fine-tuning, and a static reward model trained on human preferences. This static nature imposes fundamental limitations: a model’s capabilities are constrained by the fixed data it has seen, and an evolving policy can learn to exploit loopholes in the fixed reward model, leading to a brittle alignment known as “reward hacking” [168]. Game theory provides a robust alternative by recasting LLM training and deployment as a dynamic, strategic, and interactive process.

3.4.1 Overcoming Static Datasets. Static datasets represent a finite and fixed data distribution, and collecting human-annotated data is expensive. This restricts a model’s ability to generalize to novel scenarios. To address this, a prominent line of research uses self-play, where an LLM iteratively generates its training data. The SPIN framework [113], for example, has the LLM play against previous versions of itself. In each round, the model generates new responses and learns to distinguish these from a seed set of human-annotated data, effectively creating a curriculum of progressively higher-quality data without needing continuous human feedback. This concept of self-alignment is also explored by Azarafrooz et al. [169], who propose an online, two-player game that can be viewed as a simplified form of DPO operating without any human preference data, using Nash-learning and adaptive feedback to enable autonomous improvement.

Similarly, Chu et al. [170] tackled the data scarcity and noise problem simultaneously with their Stackelberg Game Preference Optimization (SGPO). This framework uses self-annotation to create worst-case preference data efficiently, enhancing the robustness of the model.

Adversarial games are also used to specifically target and patch a model’s weaknesses. In this paradigm, one agent’s goal is to generate inputs that the other agent finds difficult. Zheng et al. [171] let an adversarial agent generate prompts that expose the weaknesses of a defensive agent. They also introduce an innovative diversity constraint to prevent the adversary from collapsing to a narrow set of attacks. The utility of such adversarial dynamics is profound; for instance, self-play in an “Adversarial Taboo” game is more effective for enhancing LLM reasoning than standard supervised fine-tuning [159]. Ye et al. [172] proposed a creator-solver dynamic where a “creator” model strategically crafts new prompts by aiming to maximize the “solver” model’s regret. Similarly, Liu et al. [173] used online self-play to co-evolve attacker and defender agents with mechanisms like a “Hidden Chain-of-Thought” to enhance strategic planning.

Another strategy involves using game dynamics not just to generate prompts, but to create fine-grained, step-by-step feedback, which is notoriously difficult to obtain from humans. Chen et al. [174] introduced the Self-Play Critic (SPC), where a “sneaky generator” deliberately produces flawed reasoning steps to challenge a “critic” model. This forces the critic to evolve its assessment capabilities, ultimately achieving performance that surpasses existing process reward models without manual step-level annotation. Zhou et al. [175] proposed a two-player online game between a “proposer” that generates a response and a “reflector” that provides immediate, dense feedback by critiquing it. This dynamic interaction generates rich, supervisory signals that are absent in static feedback datasets. Xie et al. [176] used a Stackelberg game framework to learn detoxification from non-parallel data. The insight of their work is the finding that the success of such methods is often highly dependent on the accuracy of the feedback signal, such as the toxicity classifier.

3.4.2 Evolving the Reward Signal. A static reward model (RM), no matter how well-trained, is inevitably exploitable. As the LLM policy improves, it can discover and exploit edge cases or loopholes in the RM’s fixed reward function, a phenomenon known as reward hacking [168]. To counter this, researchers have reframed the interaction between the LLM and the RM as a dynamic game where the reward signal co-evolves with the policy. The Adversarial Preference Optimization (APO) framework [177] lets the LLM and RM update in a min-max game. The RM is trained to find outputs where the current LLM policy is poorly calibrated, and the LLM is then trained to improve on these adversarial examples. Its update also incorporates a KL divergence regularization term, ensuring it remains faithful to the original human preferences while adapting to the LLM’s new distribution. This dynamic can be formalized using frameworks from game theory, like bilevel optimization and Stackelberg games. STA-RLHF [114] models the interaction as a Stackelberg game where the LLM policy is the “leader” and the preference model is the “follower.” The policy makes the first move, and the preference model must best respond to it, forcing the policy to learn an alignment that is robust to an adaptive reward signal. Shen et al. [178] considered a more principled algorithmic framework, introducing a provably convergent first-order algorithm for such bilevel problems using a penalty-based method. Concurrently, Chakraborty et al. [179] developed a framework that, by precisely modeling the objective’s dependency on policy-generated trajectories, improves the sample efficiency and mitigates reward over-optimization. These principled approaches are also being extended to more complex scenarios, such as Contextual Bilevel RL, which enables solving complex, real-world incentive alignment problems like tax design [180].

3.4.3 Other Game-Theoretic Dynamics. Game theory also provides novel frameworks for LLM alignment and generation that go beyond the data-reward dichotomy. One such area is cooperative and mixed-motive games. In the Manuscript submitted to ACM

COEVOLVE framework [181], an LLM is fine-tuned by interacting with a copy of itself in a sequential cooperative game. Complementing this, Liao et al. [99] provided empirical evidence that self-play is highly effective in non-zero-sum negotiation games. Beyond fine-tuning, evolutionary game theory is used to analyze emergent collective behaviors of LLM agent systems. Gemp et al. [182] formulated the natural language dialogue generation as a game process. By applying equilibrium solvers, the method equips LLM with more stable and rational conversational strategies.

Another line of work applies game theory to the decoding process. The Consensus Game [183] uses equilibrium search in a cooperative signaling game, where a Generator and Discriminator reconcile their predictions to find a consensus, leveraging no-regret learning to produce truthful and coherent outputs. The Peer Elicitation Games (PEG) [184] extends this idea by replacing a single agent discriminator with a multi-agent peer elicitation process. This equilibrium-seeking principle has been successfully applied to complex, embodied AI tasks; for instance, Yu et al. [185] integrated it into a vision-language navigation system to reduce model hallucinations. On the other hand, Chen et al. [115] proposed the Decoding Game, a theoretical framework that reimagines text generation as a zero-sum game against an adversarial “Nature.” Their analysis shows that this framing provides the first theoretical justification for the empirical success of heuristic methods like Top-k sampling. Zhang et al. [186] reframed decoding as a Bayesian game between two internal LLM agents: a Generator and a Verifier. The two agents strategically interact to reach a “Separating Equilibrium,” a stable state where the verifier can reliably distinguish high-quality outputs from low-quality ones. This strategic decoding process acts as a powerful, training-free verification mechanism. Game dynamics can also steer generation at inference time, for instance by using solvers like Counterfactual Regret Minimization to guide dialogue toward less exploitable strategies [187], or using Nash equilibrium concepts to dynamically control text attributes [188].

Besides, game-theoretic mechanisms are being used to improve the efficiency of the alignment process. Zhang et al. [189] introduced an auction-based mechanism for collecting preference data, using principles from mechanism design to improve cost-efficiency. Xie et al. [190] introduced Efficient Coordination via Nash Equilibrium (ECON), which recasts multi-LLM coordination as an incomplete-information game seeking a Bayesian Nash equilibrium. This framework allows each LLM to independently select responses based on its beliefs about co-agents, achieving a tighter regret bound and outperforming existing multi-LLM approaches.

Discussions

Although theoretical guarantees ensure desired properties, implementing these methods in practice poses significant challenges for robust performance. For example, computing the exact Shapley value is often intractable, and approximations suffer from high variance. Similarly, training self-play methods for value alignment or static problem-solving frequently encounters instability. Additionally, extracting preferences from highly heterogeneous datasets introduces further hurdles, such as incentive misalignment and moral constraints. Thus, substantial research opportunities remain to improve robustness, stability, and efficiency, building on the foundational ideas presented in this section.

The game-theoretic methods used to enhance LLM interpretability, alignment, and adaptation do not exist in a vacuum. They are developed in response to the immense economic and social pressures that define the LLM ecosystem. The competition among stakeholders, the economics of data, and the deployment of LLMs as strategic actors in society create the demand for more robust, efficient, and aligned models. In Section 4, we will analyze this broader strategic

Table 3. Summary of Game Modeling for LLM-related Events

Practical Scenarios	Game Frameworks	Key Findings/Insights
Multi-Stakeholder Competition and Cooperation in LLM Era §4.1		
Strategic Preference Reporting in LLM Alignment [154, 191–194]	Principal-Agent, Mechanism Design	Strategic misreporting harms alignment; trade-offs between strategyproofness and optimality; truthful reporting incentivized via payments.
Data Sharing and Model Release Strategies [24, 195–197]	Stackelberg Game, Repeated Game, Nash Equilibrium	Private and social incentives diverge in data-sharing; market structure depends on data heterogeneity and user behavior.
Pricing Strategies on Models and Outputs [198–201]	Monopolistic Pricing, Stackelberg Game, Contract Theory	Pricing adapts to user types and skills; pay-per-token creates moral hazard, while pay-for-performance contracts align incentives for quality.
Advertising and Monetization within LLM [25, 202–209]	Auction Theory, Mechanism Design	Truthful auctions at token or output level; dynamic auctions can increase revenue; mechanisms can co-optimize revenue and social welfare.
Solving Intractable Game Problems with LLMs §4.2		
Strategies of Autonomous LLM Agent [47, 210, 211]	Economic Agent Models, Behavioral Game Theory	LLMs act as strategic agents with misaligned goals; they may strategically withhold info for long-term gain, leading to suboptimal outcomes.
Transformation in Data and Content Ecosystems [26, 212–214]	Prisoner’s Dilemma, Contest Models, Computational Game Theory	Creators face a Prisoner’s Dilemma in data sharing; GenAI competition erodes prices and diversity; human adaptation to AI is computationally difficult.
System-Level Equilibria and Regulatory Challenges [215–218]	Network Games (Braess’s Paradox), Stackelberg Games, Models of Regulation	Adding GenAI can degrade ecosystems (Braess’s Paradox); fragmented regulation can backfire; long-term incentive misalignment leads to system failure.

landscape, using game models to characterize the multi-stakeholder competitions and societal impacts that motivate the technical advancements discussed herein.

4 CHARACTERIZING LLM-RELATED EVENTS THROUGH GAME MODELS

As LLMs evolve from mere tools to active institutional players within markets, platforms, and information ecosystems, their behavior is no longer driven solely by technical objectives. Game theory provides a rigorous framework for analyzing the increasingly complex strategic interactions surrounding the development, deployment, and societal integration of LLMs. This section reviews recent advances in game-theoretic modeling applied to LLMs and categorizes them into two complementary streams: Subsection 4.1 addresses strategic games arising directly from the development and deployment lifecycle of LLMs, encompassing data acquisition, fine-tuning, platform economics, and content monetization. In contrast, Subsection 4.2 examines broader societal impacts of LLMs, including platform-creator dynamics, policy-induced externalities, and economic coordination challenges. The key problem settings, game notions, and summarized insights are presented in Table 3.

4.1 Multi-Stakeholder Competition and Cooperation in LLM Era

The development and deployment of LLMs involve various stakeholders, ranging from data providers and annotators to model trainers, platform deployers, and end users. Strategic behavior is ubiquitous in this landscape, whether manipulation of preference reporting during model fine-tuning or intense competition among vendors for user attention and market share. Game-theoretic modeling offers a structured approach to exploring these behaviors, uncovering equilibrium structures, incentive misalignments, and design trade-offs.

4.1.1 Strategic Preference Reporting in LLM Alignment. Several studies examine strategic preference reporting in LLM alignment. Buening et al. [191] model RLHF as a principal-agent game, where the LLM developer (principal) relies on annotators (agents) to provide pairwise preference data for fine-tuning. Since annotators also derive utility from the resulting model, they may manipulate preferences to influence outcomes. Their analysis proves that no RLHF objective can simultaneously ensure strategyproofness and social welfare optimality. Adopting a mechanism design perspective, Sun et al. [192] similarly identify misreporting incentives under standard objectives and derive payment rules that enforce truthful reporting while maximizing welfare. In a simpler setting, Wu et al. [154] formalize the Battling Influencers Game (BIG), showing it is a potential game where rational annotators exaggerate preferences in equilibrium to maximize influence. Together, these works reveal inherent incentive misalignment, necessitating carefully designed reward structures in alignment pipelines. Liu et al. [193, 194] modeled scenarios where a principal cannot directly observe effort. They design and analyze contracts, such as bonus schemes using “golden questions” or linear/binary payment structures, to reward high-quality work. Both studies emphasize that without proper monitoring, even well-intentioned annotators may provide low-effort data, undermining alignment.

4.1.2 Data Sharing and Model Release Strategies. Game-theoretic approaches have emerged as powerful tools for analyzing strategic interactions in AI data-sharing, model development, and release strategies. In the domain of data sharing, Taitler et al. [195] modeled the interaction between a content creation firm and a Generative AI (GenAI) platform as a Stackelberg game. The firm, acting as the leader, strategically controls information disclosure, while the AI platform, as the follower, determines how much data to acquire from external experts. Their equilibrium analysis shows that firms may be willing to pay GenAI platforms to use their data and identifies the conditions under which such agreements become Pareto-improving. Focusing on model development, Laufer et al. [24] framed fine-tuning as a two-stage game between a generalist developer and a domain-specific specialist. This game involves sequential investment decisions followed by bargaining over revenue-sharing terms. Their findings demonstrate that a specialist’s strategic decision to contribute, free-ride, or abstain hinges on the interplay of marginal returns and cost asymmetries. To analyze competition among machine learning providers, Xu et al. [196] introduced the Heterogeneous Data Game. This framework models providers that handle diverse data sources characterized by covariate and concept shifts. By identifying pure Nash equilibria, their work delineates the conditions that lead to distinct market structures: the non-existence of an equilibrium, convergence toward homogeneity, or specialization in heterogeneous niches. Wu et al. [197] explored the strategic choice between open- and closed-source model releases, modeling it as a multi-agent game between open-source and proprietary developers. Their analysis reveals an “innovation paradox”: while open-sourcing accelerates ecosystem-wide innovation, it can erode the competitive advantages of individual firms. The equilibrium outcome ultimately depends on factors like user demand elasticity and asymmetries in development costs.

4.1.3 Pricing Strategies on Models and Outputs. Economic modeling also extends to pricing strategies on models and outputs, frequently adopting a Stackelberg pricing framework. Here, LLM platforms play the role of leaders

who pre-commit to pricing menus, while users, acting as followers, select options based on individual utility and task requirements. Bergemann et al. [198] studied the optimal pricing and product design for LLMs. Their economic framework considers variable operational costs for processing input and output tokens, the ability to fine-tune models, and diverse user needs and error sensitivities. They found that optimal pricing structures, often implemented through two-part tariffs, lead to higher markups for more intensive users, rationalizing observed industry practices like tiered pricing based on customization and usage levels. Li et al. [200] and Saig et al. [201] studied the pricing strategy for LLM users. The current pricing scheme, pay-per-token, is challenged by Saig et al. [201], who found that companies may use a cheaper model rather than the model they claimed to use, which causes a moral hazard. To address such a problem, they introduce a pay-for-performance, contract-based framework that incentivizes quality. Li et al. [200] modeled prompt pricing as a Stackelberg game between a platform and users with different prompt engineering skills. By incorporating “prompt ambiguity,” they derive an optimal pricing algorithm that adapts to user proficiency, improving platform payoff. Mahmood [199] studied sequential price competition between two LLM-developing firms, setting prices for different tasks of model usage. The equilibrium analysis shows that the second mover can always achieve cost-effectiveness. Moreover, if the tasks are similar, the first mover may become cost-ineffective regardless of its pricing strategy.

4.1.4 Advertising and Monetization within LLM. Advertising and monetization within LLM interfaces represent a rapidly evolving area of research. Early explorations by Feizi et al. [202] outline several potential advertising ecosystems, sparking further work on specific mechanisms. Duetting et al. [25] introduced a token-level auction where advertisers can bid to influence text generation, highlighting resultant challenges like the exposure problem. Building on this, Soumalias et al. [203] proposed MOSAIC, a truthful mechanism that aggregates advertiser preferences over entire outputs using Rochet payments and importance sampling. Auction theory is also being adapted for Retrieval-Augmented Generation (RAG) [204, 205]. In these models, advertisers hold text content and bid to have it integrated into the LLM’s output. Recent studies are also exploring more complex dynamics; Banchio et al. [206] modeled dynamic auctions, revealing a temporal trade-off where delaying responses can create “auction thickness” to increase revenue. Meanwhile, Mordo et al. [207] investigated auctions that jointly compose sponsored and organic content to maximize social welfare rather than revenue alone. For a broader perspective, recent surveys [208] and position papers [209] map the entire design space for these novel systems.

4.2 Framing the Societal Impact of LLMs

Beyond their immediate technical and economic applications, LLMs are profoundly reshaping broader societal structures. The widespread adoption of generative models introduces new dynamics of competition, cooperation, and strategic interaction across domains such as knowledge production, regulatory governance, platform design, and content labor markets. This subsection surveys game-theoretic research that models these macro-level effects, highlighting how emergent equilibria capture the unforeseen consequences of GenAI proliferation.

4.2.1 Strategies of Autonomous LLM Agents. A critical shift in modeling involves recognizing that GenAI systems can function as goal-directed, strategic agents. This perspective moves beyond viewing LLMs as passive tools, instead understanding them as economic actors whose implicit preferences, shaped by their training objectives, may misalign with human welfare. Immorlica et al. [47] formalized this by treating GenAI as a consultant with a payoff function based on perceived helpfulness. They contend that even minor divergences between the AI’s goals and user welfare can dramatically alter equilibria, leading to suboptimal outcomes such as overconfidence or persuasive bias, even when factual accuracy is maintained.

This strategic agency extends beyond theory. Taitler et al. [210] modeled “selective response,” where a GenAI purposefully withholds answers to niche queries to direct users toward human forums. This behavior generates fresh training data, establishing a beneficial data flywheel for the model, but it trades immediate user utility for long-term system optimization. Supporting these concerns, empirical work by Sabour et al. [211] demonstrates that GenAI agents can manipulate human choices by strategically framing options, illustrating that algorithmic influence spans from recommendation to active persuasion. Collectively, these studies underscore the risks of deploying AI systems whose optimization criteria are not robustly aligned with human autonomy.

4.2.2 Transformation in Data and Content Ecosystems. The proliferation of GenAI has fundamentally altered incentives for human creators in data sharing and content production. Keinan et al. [212] framed this challenge as a Prisoner’s Dilemma, where individual creators must choose between sharing content for model training, risking competition from AI, or withholding it to preserve exclusivity. While cooperation yields higher platform-wide utility, each player’s dominant strategy often involves defection, leading to suboptimal equilibria. This strategic tension can result in declining data quality unless carefully designed incentives, such as revenue sharing or content licensing schemes, are implemented.

Complementary research shows how these dynamics affect content diversity and creator viability. Gao et al. [213] demonstrated that while GenAI can increase average content quality, it also causes price erosion and reduces diversity due to output standardization, potentially displacing human creators from the market. The nature of this competition is further explored by Yao et al. [26], who modeled human-AI interaction as a generalized Tullock contest for user attention. In this setting, players expend effort (e.g., producing engaging content) to win a probabilistic share of user attention, which translates into monetization or visibility. Their results indicate that stable coexistence is possible, with GenAI initially eliminating the least efficient human creators. Furthermore, the rapid evolution of GenAI imposes a significant adaptive burden on humans. Esmaeili et al. [214] demonstrated that determining an optimal response strategy to a constantly evolving AI can be computationally intractable (NP-hard), formalizing the risk that human creators may be unable to adapt effectively in fast-moving content markets.

4.2.3 System-Level Equilibria and Regulatory Challenges. The strategic interactions among users, creators, and platforms can aggregate into unforeseen system-level equilibria and novel regulatory challenges. A stark example of emergent dysfunction is offered by Taitler et al. [215], who adapted Braess’s Paradox to GenAI deployment. In their model, a GenAI platform optimizing for revenue eventually erodes the quality of a human knowledge forum (e.g., Stack Overflow) by drawing away users. This degrades the quality of future training data, leading to a long-term outcome where all users are worse off, despite the GenAI’s short-term utility.

Such emergent failures underscore the difficulty of effective governance. Laufer et al. [216] showed how well-intentioned but fragmented policies can backfire. For instance, imposing safety standards only on downstream AI fine-tuners may incentivize upstream developers to underinvest in safety, resulting in lower overall safety levels. This highlights the need for holistic regulation that aligns incentives across the entire development pipeline. Other systemic shifts include the “flattening” of supply chains, as modeled by Ali et al. [217], where GenAI disintermediates gatekeepers, posing risks of surplus extraction and content homogenization. Finally, Xie et al. [218] examined the long-term dynamics between algorithmic decision-makers and humans, showing that misaligned incentives can lead agents to abandon self-improvement in favor of manipulation or exit, emphasizing the need for long-term alignment to achieve positive-sum outcomes.

Discussions

Game-theoretic modeling provides a structured lens for analyzing strategic behaviors in the LLM ecosystem and understanding their broader societal implications. However, most existing results depend on simplified environments, typically involving a small number of rational agents with limited action spaces and clearly defined payoffs. These abstractions can overlook emergent behaviors from large-scale interactions and fail to account for bounded rationality, institutional dynamics, and ambiguous motivations. Bridging the gap between theoretical predictions and empirical observations remains a challenging yet promising direction for future research.

Thus far, we have explored how game theory is applied to evaluate, enhance, and characterize LLMs. As shown in our taxonomy, our survey considers a bi-directional relationship between the research on game theory and LLMs. In the next section, we will introduce studies that approach this intersection from a complementary perspective, discussing how they leverage LLMs' capabilities to advance classical game theory.

5 ADVANCING GAME THEORY WITH LARGE LANGUAGE MODELS

Traditional game theory strives for a comprehensive theoretical understanding within well-defined models. As a result, studies in this field often depend on formal, structured frameworks with limited strategy spaces and simplified communication protocols. While these constraints facilitate rigorous analysis and yield valuable theoretical insights, they also restrict the theory's ability to capture the complexities of real-world interactions. Large language models, with their rich linguistic, reasoning, and representational capabilities, introduce a new computational paradigm. By using natural language as both an interface and a medium for strategic reasoning, LLMs enable researchers to revisit classic problems and explore domains that were previously intractable. This emerging and promising area of research can be broadly categorized into two directions: the first (Subsection 5.1) explores how LLMs can expand traditional game-theoretic models, while the second (Subsection 5.2) focuses on leveraging LLMs to solve intractable game theory problems.

5.1 Expanding Game Modeling with LLMs

Traditional game-theoretic models often operate within rigid numerical and symbolic representations that fall short of capturing the nuance and fluidity of human interaction. By introducing natural language as a modeling medium, LLMs offer a transformative shift: *enabling expressive, adaptive, and context-aware representations of strategic behavior*. This subsection explores how LLMs enrich the modeling landscape across three key domains: verbalized strategic interaction, preference elicitation in social choice, and semantic-enhanced economic mechanisms. Together, these works signal a move toward a more linguistically grounded and semantically rich game-theoretic paradigm.

5.1.1 Verbalized Strategic Interactions. Classical game theory's reliance on numerical information structures limits its ability to model human communication. Li et al. [219] addressed this by introducing a verbalized Bayesian persuasion (BP) framework for real-world games involving human dialogues. They represent BP as a mediator-augmented extensive-form game, with LLMs acting as sender and receiver. To solve this game efficiently, they develop a generalized equilibrium-finding algorithm that integrates LLMs with game solvers, incorporating verbalized commitment assumptions, obedience constraints, and information obfuscation. This approach enables game theory to model complex, language-driven interactions, enhancing its applicability to real-world scenarios like negotiations.

5.1.2 Preference Elicitation in Social Choice. In social choice, rigid preference elicitation with fixed alternatives restricts expressivity. Fish et al. [30] proposed “generative social choice,” where LLMs generate textual options reflecting collective opinions, accommodating unforeseen alternatives. Their framework, supported by formal guarantees and empirical validation, demonstrates that LLMs can extract utilities from free-form text and create representative slates, achieving high agreement in applications like chatbot personalization. Boehmer et al. [220] extended this with PROSE, a system generating diverse, cost-effective slates under budget and query accuracy constraints. Using public deliberation datasets (e.g., Polis), PROSE enhances scalability and user satisfaction in democratic processes, overcoming traditional social choice limitations.

5.1.3 Semantic-Enhanced Economic Mechanisms. Traditional auction and mechanism design models often assume fixed, black-box valuation functions, neglecting the semantic richness of real-world preferences. Sun et al. [221] introduced the Semantic-enhanced Personalized Valuation in Auction (SPVA) framework, where LLMs extract context-sensitive valuations from unstructured text (e.g., reviews, product descriptions) using LLMs. This reduces valuation noise and improves utility in Vickrey auctions. Similarly, Lu et al. [31] extended mechanism design to natural language domains in peer-prediction settings, using LLMs to evaluate truthful reporting from free-form feedback, with applications in content moderation and community governance. Penna et al. [222] proposed Language Model Mechanisms (LMMs), where agents report in natural language, and LLMs compute outcomes and payments, maintaining incentive compatibility in high-dimensional, distributed environments. These advancements enrich economic mechanism design with semantic context.

5.1.4 Emerging Perspectives. These developments signal a paradigm shift in game-theoretic modeling. Lotfi et al. [223] argued that LLMs’ linguistic and adaptive capabilities challenge assumptions like fixed communication protocols and fully rational agents. In simulated negotiations, language-mediated interactions produce emergent behaviors, such as spontaneous coordination or implicit collusion, which standard models fail to predict. This shift toward semantically rich, language-based systems, where agents reason and renegotiate rules via natural language, expands the boundaries of game theory, enabling new insights into complex strategic interactions.

5.2 Solving Intractable Game Problems with LLMs

In addition to expanding how games are modeled, LLMs offer powerful tools for addressing computational bottlenecks that have long hindered the practical application of game theory. Their ability to generate, interpret, and formalize complex structures allows them to tackle problems traditionally deemed intractable. This subsection explores how LLMs contribute to four crucial areas: interpretable mechanism design, reducing cognitive load in allocation tasks, automated game modeling, and simulation-based reasoning for information disclosure. These contributions not only improve solution efficiency but also broaden the real-world applicability of game-theoretic analysis.

5.2.1 Interpretable Mechanism Design. Traditional mechanism design often produces opaque, black-box solutions. Liu et al. [29] proposed a framework that casts mechanism design as a code generation task, using LLMs to produce human-readable pseudocode for heuristic mechanisms. This approach achieves competitive performance in complex design spaces, rediscovered classic mechanisms, and enhances interpretability, offering potential for discovering optimal mechanisms in intricate scenarios.

5.2.2 Reducing Cognitive Burdens in Allocation Tasks. High-dimensional preference reporting in allocation tasks, such as combinatorial auctions or course assignments, often overwhelms users, reducing efficiency. Soumalias et al. [28]

developed LLM-powered proxies that interpret one-shot natural language inputs to generate preference comparisons, lowering error rates and improving allocative efficiency. Similarly, Huang et al. [224] showed that LLM-based proxies in combinatorial auctions, integrated with incremental revelation mechanisms, reduce query demands and achieve faster convergence compared to traditional learning-theoretic methods, streamlining complex allocation processes.

5.2.3 Automated Game Modeling and Simulation. Real-world strategic interactions described in natural language are challenging to formalize. Mensfelt et al. [225] introduced a framework for autoformalizing simultaneous-move games from textual descriptions, using one-shot prompting and syntactic feedback to create formal logic representations for analysis. Deng et al. [226] extended this to imperfect-information games, employing a two-stage pipeline to identify information sets and construct extensive-form games, with a self-debugging module ensuring validity. Mensfelt et al. [227] further enabled tournament-style simulations of strategies derived from natural language scenarios, completing a pipeline from informal input to executable strategic evaluation. These frameworks enhance game theory's adaptability to real-world scenarios like policy negotiations.

5.2.4 Strategic Simulation and Information Disclosure. LLMs enable novel simulation-based approaches to complex game-theoretic problems. Yin et al. [228] proposed InfoBid, a framework using LLM agents to evaluate how information disclosure policies affect auction outcomes. Their simulations reveal that classical assumptions, such as “more information improves efficiency,” may not hold with LLM bidders, as selective information sharing can lead to over- or underbidding. The limited ability of LLM bidders to model competitors’ strategies highlights a gap between simulated and bounded rationality, underscoring the need to redesign strategies for LLM-mediated environments.

Discussions

Most work in this section largely leverages the remarkable language generation capabilities of LLMs. However, LLMs are fundamentally prone to bias and hallucination, which can undermine the critical assumption of a stable and reliable “oracle” or “solver” in game-theoretic applications. This induces general imperfections and introduces unquantified and potentially systemic errors when LLMs act as simulators, preference elicitors, or mechanism designers, leading to real-world issues like unfairness in social choice or auction design. Under the case that an LLM might “hallucinate” game rules or reflect training data biases, it is challenging to make efficient verification and ensure the correctness and logical consistency of LLM-generated contents.

6 CHALLENGES AND FUTURE DIRECTIONS

While existing research has made significant strides in the intersection between game theory and LLMs, several critical challenges remain unresolved. These limitations point to promising yet under-explored directions for advancing both theoretical frameworks and practical applications. In this section, we systematically identify these open problems and outline possible pathways for future research.

6.1 LLM-based Agents with Comprehensive Game Abilities

Current Landscape: Recent research has focused on evaluating and enhancing LLM agents’ performance in specific, isolated game scenarios. For instance, studies have demonstrated significant improvements in strategic reasoning in matrix games [63], Avalon [13], bargaining [19], and Werewolf [65]. Although some of the methods, such as strategic reflection or tool usage, are general in principle, their validation remains highly scenario-specific. Consequently, the

performance improvement often fails to transfer effectively between different game genres or rule systems, limiting the development of truly generalist game-playing agents.

Future Directions: Based on this observation, a future direction is to develop LLM agents proficient in fundamental game-theoretic reasoning, capable of applying core principles across diverse game settings without requiring explicit customization for each new environment. Achieving this ambitious goal requires simultaneous advancements across multiple research fronts: (1) improved rule comprehension through better formal language understanding and symbolic reasoning integration, (2) more robust external environment modeling that can handle partial observability and stochastic transitions, and (3) sophisticated multi-agent reasoning frameworks that scale to varying numbers of participants with different behavioral patterns. As the goal is to build a generalist game-playing LLM, the validation data should be comprehensive, ranging from simple matrix games to complex imperfect-information games.

6.2 Moving Beyond Human-Oriented Evaluation Frameworks

Current Landscape: One of the predominant approaches to evaluating the strategic capabilities of LLMs is through metrics designed to capture strategic optimality, such as Nash regret [52, 53, 56] or k -level rationality [16]. Interestingly, empirical observations reveal that LLMs often exhibit prosocial behaviors (Subsection 2.1), frequently eschewing self-optimal strategies. One possible reason is that the RLHF has shaped LLM behavior toward more altruistic responses. Moreover, the intrinsic training objective of LLMs, next-token prediction, diverges substantially from the principles underlying these evaluation metrics. As a result, it remains unclear whether success or failure on such metrics reliably reflects the true reasoning or strategic capacity of LLMs. This disconnect raises important questions about the adequacy and appropriateness of current evaluation paradigms.

Future Directions: Developing evaluation frameworks specifically tailored to neural network-based agents is a valuable future direction. Merely repurposing benchmarks originally designed for humans is insufficient for capturing the unique behaviors and limitations of LLMs. A meaningful starting point is the design of tasks that not only draw upon game-theoretic settings but also reflect the fundamental properties of LLM training. Consequently, evaluation metrics must also be custom-developed for LLMs, rather than borrowed wholesale from human cognitive testing. An initial approach may involve the use of subjective or qualitative measures. And we should strive toward evaluation metrics that are robust, interpretable, and operationally meaningful. These attributes are crucial for ensuring that our assessments of LLM capabilities remain rigorous and actionable.

6.3 Understanding LLMs' Strategic Behavior

Current Landscape: Despite the improvement and evaluation of LLM agents in games, another valuable task is to provide a theoretical framework to characterize LLMs' behavior. For instance, Park et al. [229] provided insight into how LLM fails to act as a no-regret algorithm with the current supervised pre-training procedure. However, extending such theoretical characterizations to more complex strategic environments remains a significant open challenge. The vastness of strategy spaces and the combinatorial intricacies of game rules in realistic settings severely limit the feasibility of formal analysis.

Future Directions: A robust theoretical understanding of LLMs in strategic contexts is critical for defining performance boundaries and guiding architectural design. A promising approach is to use abstraction and simplification to model essential game dynamics without their full analytical complexity, for instance, by analyzing critical subroutines or decision points in isolation. Moreover, tools from complexity theory, especially those related to the circuit complexity

of Transformers, can be leveraged to establish formal limits on the strategic capabilities of these models. Ultimately, this line of inquiry could lead to more principled training methodologies that instill desired strategic behaviors.

6.4 Capturing Cooperative Games in LLM Optimization

Current Landscape: As discussed in Section 3, many studies applying game theory to LLM optimization have primarily focused on competitive game formulations. While competition offers a natural and tractable modeling approach, cooperation presents an equally promising avenue for advancing LLM capabilities, which is underexplored.

Future Directions: Incorporating cooperative game-theoretic principles into the training and optimization of LLMs could yield both novel theoretical insights and practical performance gains. For example, in Mixture of Experts (MoE) architectures, individual expert networks can be conceptualized as players in a cooperative game. Leveraging solution concepts such as the Shapley Value or the core could inform more principled router scheduling strategies, improving expert selection and load balancing while minimizing redundancy. Similarly, in ensemble learning and knowledge distillation settings, modeling sub-models as cooperative agents and applying fair credit assignment mechanisms could enhance collaboration among components, leading to improved generalization and computational efficiency.

6.5 Modeling Cooperation Between Multi-LLMs and Humans

Current Landscape: As reviewed in Section 4, prior research has predominantly focused on competitive or adversarial dynamics between LLMs and humans. These studies have shed light on important societal concerns, including persuasion, manipulation, and safety. Yet, cooperative interactions—particularly those involving formal game-theoretic modeling—remain significantly underexplored.

Future Directions: A promising and necessary research direction lies in understanding and designing cooperative frameworks involving multiple LLMs and human participants. Central challenges include constructing incentive-compatible mechanisms that encourage LLMs to coordinate effectively on human-assigned tasks while also accounting for their own modeled objectives. Developing a formal understanding of LLM agents' goals and behaviors is critical for bridging the gap between abstract theory and practical deployment. Progress in this area could lead to the design of AI systems that more robustly align with human values and intentions.

6.6 Leveraging LLMs as Oracles to Extend Theoretical Game Models

Current Landscape: As discussed in Section 5, recent studies have demonstrated the potential of LLMs to extend classical game-theoretic models into more realistic domains involving natural language. The core insight is that the LLMs' sophisticated language understanding and generative abilities let them serve as oracles that instantiate otherwise abstract components of game models, functioning as certain rules.

Future Directions: This approach paves the way for relaxing idealized assumptions by replacing theoretical constructs with practical, LLM-driven approximations. Such a substitution allows previously abstract models to be instantiated in real-world settings while preserving their approximate theoretical guarantees. Systematically exploring the use of LLMs as adaptable oracles could enable a new class of empirical, simulation-driven game-theoretic analyses. A particularly impactful application lies in computational mechanism design—for instance, in simulating complex auctions where bidder preferences are subtle and difficult to formalize. LLMs can emulate realistic bidder behavior across varying mechanisms, supporting rapid iteration and refinement of designs that were previously analytically intractable.

7 CONCLUSION

This survey has provided a comprehensive examination of the bidirectional interplay between game theory and large language models, addressing a significant gap in the existing literature. Through a novel structured taxonomy, we have systematically organized and elucidated the multifaceted relationship between these fields, highlighting their mutual reinforcement and synergistic potential. Our analysis demonstrates that game theory offers essential frameworks, such as equilibrium concepts, incentive design, and multi-agent interaction models, to formalize, analyze, and enhance LLM behaviors. Conversely, LLMs introduce unprecedented capabilities for simulating complex agents, approximating theoretical oracles, and scaling game-theoretic solutions in real-world environments. We hope this work inspires further exploration and collaboration at this dynamic intersection, paving the way for innovative developments in both domains.

REFERENCES

- [1] John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior: 60th anniversary commemorative edition. In *Theory of games and economic behavior*. Princeton university press, 2007.
- [2] John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 1950.
- [3] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 1961.
- [4] Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.
- [5] Amartya Sen. Social choice theory. *Handbook of mathematical economics*, 3:1073–1181, 1986.
- [6] Mu Zhu, Ahmed H Anwar, Zelin Wan, Jin-Hee Cho, Charles A Kamhoua, and Munindar P Singh. A survey of defensive deception: Approaches using game theory and machine learning. *IEEE Communications Surveys & Tutorials*, 23(4):2460–2493, 2021.
- [7] Tanmoy Hazra and Kushal Anjaria. Applications of game theory in deep learning: a survey. *Multimedia Tools and Applications*, 81(6):8963–8994, 2022.
- [8] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [9] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [10] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- [11] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.
- [12] Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models. *Nature Human Behaviour*, pages 1–11, 2025.
- [13] Shenzhi Wang, Chang Liu, Zilong Zheng, Siyuan Qi, Shuo Chen, Qisen Yang, Andrew Zhao, Chaofei Wang, Shiji Song, and Gao Huang. Avalon’s game of thoughts: Battle against deception through recursive contemplation. *arXiv preprint arXiv:2310.01320*, 2023.
- [14] Yuan Deng, Vahab Mirrokni, Renato Paes Leme, Hanrui Zhang, and Song Zuo. Llms at the bargaining table. In *Agentic Markets @ ICML 2024*, volume 2024, 2024.
- [15] Jiangjie Chen, Siyu Yuan, Rong Ye, Bodhisattwa Prasad Majumder, and Kyle Richardson. Put your money where your mouth is: Evaluating strategic planning and execution of LLM agents in an auction arena. In *Open-World Agents @ NeurIPS 2024*, 2024. URL <https://openreview.net/forum?id=hKEzHiYJxc>.
- [16] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Man Lan, and Furu Wei. K-level reasoning: Establishing higher order beliefs in large language models for strategic reasoning. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7212–7234, 2025. doi: 10.18653/v1/2025.nacl-long.370. URL <https://aclanthology.org/2025.nacl-long.370/>.
- [17] Chenghao Huang, Yanbo Cao, Yinlong Wen, Tao Zhou, and Yanru Zhang. Pokergpt: An end-to-end lightweight solver for multi-player texas hold’em via large language model, 2024. URL <https://arxiv.org/abs/2401.06781>.
- [18] Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. Improving language model negotiation with self-play and in-context learning from ai feedback, 2023. URL <https://arxiv.org/abs/2305.10142>.
- [19] Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuseng Zhang, Yang Yang, and Rui Wang. Measuring bargaining abilities of LLMs: A benchmark and a buyer-enhancement method. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3579–3602, 2024. doi: 10.18653/v1/2024.findings-acl.213. URL <https://aclanthology.org/2024.findings-acl.213/>.
- [20] Dekun Wu, Haochen Shi, Zhiyuan Sun, and Bang Liu. Deciphering digital detectives: Understanding llm behaviors and capabilities in multi-agent mystery games. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8225–8291, 2024.

- [21] James Enouen, Hootan Nakhhost, Sayna Ebrahimi, Sercan Arik, Yan Liu, and Tomas Pfister. Textgenshap: Scalable post-hoc explanations in text generation with long documents. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13984–14011, 2024.
- [22] Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. Axioms for ai alignment from human feedback. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 80439–80465, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/9328208f8ec69420031647e6ff97727-Paper-Conference.pdf.
- [23] Remi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiebel, et al. Nash learning from human feedback. In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [24] Benjamin Laufer, Jon Kleinberg, and Hoda Heidari. Fine-tuning games: Bargaining and adaptation for general-purpose models. In *Proceedings of the ACM on Web Conference 2024*, pages 66–76, 2024.
- [25] Paul Duetting, Vahab Mirrokni, Renato Paes Leme, Haifeng Xu, and Song Zuo. Mechanism design for large language models. In *Proceedings of the ACM on Web Conference 2024*, pages 144–155, 2024.
- [26] Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. Human vs. generative ai in content creation competition: Symbiosis or conflict? In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [27] Boaz Taitler and Omer Ben-Porat. Braess's paradox of generative ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14139–14147, 2025. doi: 10.1609/aaai.v39i13.33548. URL <https://ojs.aaai.org/index.php/AAAI/article/view/33548>.
- [28] Ermis Soumalias, Yanchen Jiang, Kehang Zhu, Michael Curry, Sven Seuken, and David C Parkes. Llm-powered preference elicitation in combinatorial assignment. *arXiv preprint arXiv:2502.10308*, 2025.
- [29] Jiayuan Liu, Mingyu Guo, and Vincent Conitzer. An interpretable automated mechanism design framework with large language models. *arXiv preprint arXiv:2502.12203*, 2025.
- [30] Sara Fish, Paul Götz, David C Parkes, Ariel D Procaccia, Gili Rusak, Itai Shapira, and Manuel Wüthrich. Generative social choice. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 985–985, 2024.
- [31] Yuxuan Lu, Shengwei Xu, Yichi Zhang, Yuqing Kong, and Grant Schoenebeck. Eliciting informative text evaluations with large language models. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 582–612, 2024.
- [32] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. LLM as a mastermind: A survey of strategic reasoning with large language models. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=IMqjsQ4evS>.
- [33] Xiachong Feng, Longxu Dou, Ella Li, Qinghao Wang, Haochuan Wang, Yu Guo, Chang Ma, and Lingpeng Kong. A survey on large language model-based social agents in game-theoretic scenarios. *Transactions on Machine Learning Research*, 2025.
- [34] Siyao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*, 2024.
- [35] Fulin Guo. Gpt in game theory experiments, 2023. URL <https://arxiv.org/abs/2305.05516>.
- [36] John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.
- [37] Kehan Zheng, Jinfeng Zhou, and Hongning Wang. Beyond nash equilibrium: Bounded rationality of llms and humans in strategic decision-making, 2025. URL <https://arxiv.org/abs/2506.09390>.
- [38] Alicia Vidler and Toby Walsh. Playing games with large language models: Randomness and strategy, 2025. URL <https://arxiv.org/abs/2503.02582>.
- [39] Silin Du and Xiaowei Zhang. Helmsman of the masses? evaluate the opinion leadership of large language models in the werewolf game. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=xMt9kCv5YR>.
- [40] Jonathan Light, Min Cai, Sheng Shen, and Ziniu Hu. From text to tactic: Evaluating LLMs playing the game of avalon. In *Foundation Models for Decision Making @ NeurIPS 2023*, 2023. URL <https://openreview.net/forum?id=ItUrSryS0K>.
- [41] Haolan Zhan, Yufei Wang, Zhuang Li, Tao Feng, Yuncheng Hua, Suraj Sharma, Lizhen Qu, Zhaleh Semnani-Azad, Ingrid Zukerman, and Reza Haffari. Let's negotiate! a survey of negotiation dialogue systems. In *EACL (Findings)*, 2024.
- [42] Yauwai Yim, Chunkit Chan, Tianyu Shi, Zheye Deng, Wei Fan, Tianshi Zheng, and Yangqiu Song. Evaluating and enhancing llms agent based on theory of mind in guandan: A multi-player cooperative game under imperfect information, 2024. URL <https://arxiv.org/abs/2408.02559>.
- [43] Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrimmaya Sachan, and Rada Mihalcea. Cooperate or collapse: Emergence of sustainable cooperation in a society of LLM agents. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024. URL <https://openreview.net/forum?id=0zWzjj6lO3>.
- [44] Tim Ruben Davidson, Veniamin Veselovsky, Michal Kosinski, and Robert West. Evaluating language model agency through negotiations. In *Proceedings of the International Conference on Learning Representations 12 (ICLR 2024)*, 2024. URL <https://openreview.net/forum?id=3ZqKxMhcAg>.
- [45] Sara Fish, Yannai A Gonczarowski, and Ran I Shorrer. Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806*, 2024.
- [46] Shangmin Guo, Haochuan Wang, Haoran Bu, Yi Ren, Dianbo Sui, Yu-Ming Shang, and Siting Estee Lu. Economics arena for large language models. In *Language Gamification @ NeurIPS 2024*, 2024. URL <https://openreview.net/forum?id=n6Y5b1MCBV>.
- [47] Nicole Immorlica, Brendan Lucier, and Aleksandrs Slivkins. Generative ai as economic agents. *ACM SIGecom Exchanges*, 22(1):93–109, 2024.
- [48] Xidong Feng, Yicheng Luo, Ziyan Wang, Hongrui Tang, Mengye Yang, Kun Shao, David Henry Mguni, Yali Du, and Jun Wang. ChessGPT: Bridging policy learning and language modeling. In *Proceedings of the Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, 2023. URL <https://openreview.net/forum?id=pvdm4B6JMK>.

- [49] Richard Zhuang, Akshat Gupta, Richard Yang, Aniket Rahane, Zhengyu Li, and Gopala Anumanchipalli. Pokerbench: Training large language models to become professional poker players. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 26175–26182, 2025.
- [50] Sihao Hu, Tiansheng Huang, and Ling Liu. Pokellmon: A human-parity agent for pokemon battles with large language models, 2024. URL <https://arxiv.org/abs/2402.01118>.
- [51] Junzhe Chen, Xuming Hu, Shuodi Liu, Shiyu Huang, Wei-Wei Tu, Zhaofeng He, and Lijie Wen. LLMArena: Assessing capabilities of large language models in dynamic multi-agent environments. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13055–13077, 2024. doi: 10.18653/v1/2024.acl-long.705. URL <https://aclanthology.org/2024.acl-long.705/>.
- [52] Philip Brookins and Jason DeBacker. Playing games with gpt: What can we learn about a large language model from canonical strategic games? *Economics Bulletin*, 44(1):25 – 37, 2024. URL <https://EconPapers.repec.org/RePEc:eb:ecbull:eb-23-00457>.
- [53] Nicolò Fontana, Francesco Pierri, and Luca Maria Aiello. Nicer than humans: How do large language models behave in the prisoner’s dilemma? In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pages 522–535, 2025. doi: 10.1609/icwsm.v19i1.35829. URL <https://ojs.aaai.org/index.php/ICWSM/article/view/35829>.
- [54] Richard Willis, Yali Du, Joel Z Leibo, and Michael Luck. Will systems of llm agents cooperate: An investigation into a social dilemma. *arXiv preprint arXiv:2501.16173*, 2025.
- [55] Jillian Ross, Yoon Kim, and Andrew Lo. LLM economicus? mapping the behavioral biases of LLMs via utility theory. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=Rx3wC8sCTJ>.
- [56] Qiaozhu Mei, Yutong Xie, Walter Yuan, and Matthew O. Jackson. A turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9):e2313925121, 2024. doi: 10.1073/pnas.2313925121. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2313925121>.
- [57] Jingru Jia, Zehua Yuan, Junhao Pan, Paul E. McNamara, and Deming Chen. Large language model strategic reasoning evaluation through behavioral game theory, 2025. URL <https://arxiv.org/abs/2502.20432>.
- [58] Yue Wu, Xuan Tang, Tom Mitchell, and Yuanzhi Li. Smartplay : A benchmark for LLMs as intelligent agents. In *Proceedings of the International Conference on Learning Representations 12 (ICLR 2024)*, 2024. URL <https://openreview.net/forum?id=S2oTVrlcp3>.
- [59] Alonso Silva. Large language models playing mixed strategy nash equilibrium games. In *Network Games, Artificial Intelligence, Control and Optimization: 11th International Conference, NETGCOOP 2024, Lille, France, October 9–11, 2024, Proceedings*, page 142–152, Berlin, Heidelberg, 2024. Springer-Verlag. ISBN 978-3-031-78599-3. doi: 10.1007/978-3-031-78600-6_13. URL https://doi.org/10.1007/978-3-031-78600-6_13.
- [60] Jinhao Duan, Shiqi Wang, James Diffenderfer, Lichao Sun, Tianlong Chen, Bhavya Kaikhura, and Kaidi Xu. Reta: Recursively thinking ahead to improve the strategic reasoning of large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2232–2246, 2024.
- [61] Caoyun Fan, Jindou Chen, Yaohui Jin, and Hao He. Can large language models serve as rational players in game theory? a systematic analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17960–17967, 2024.
- [62] Nunzio Lorè and Babak Heydari. Strategic behavior of large language models and the role of game structure versus contextual framing. *Scientific Reports*, 14(1):18490, 2024.
- [63] Kanishk Gandhi, Dorsa Sadigh, and Noah Goodman. Strategic reasoning with language models. In *Foundation Models for Decision Making @ NeurIPS 2023*, 2023.
- [64] Alessio Buscemi, Daniele Proverbio, Alessandro Di Stefano, The Anh Han, German Castignani, and Pietro Liò. Fairgame: a framework for ai agents bias recognition using game theory, 2025. URL <https://arxiv.org/abs/2504.14325>.
- [65] Neo Watanabe and Yoshinobu Kano. Werewolf game agent by generative AI incorporating logical information between players. In Yoshinobu Kano, editor, *Proceedings of the 2nd International AIWolfDial @ ACL 2024*, pages 21–29, Tokyo, Japan, September 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.aiwolfdial-1.3. URL <https://aclanthology.org/2024.aiwolfdial-1.3>.
- [66] Suma Bailis, Jane Friedhoff, and Feiyang Chen. Werewolf arena: A case study in llm evaluation via social deduction, 2024. URL <https://arxiv.org/abs/2407.13943>.
- [67] Ziyi Liu, Abhishek Anand, Pei Zhou, Jen-tse Huang, and Jieyu Zhao. InterIntent: Investigating social intelligence of LLMs via intention understanding in an interactive game context. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6718–6746, 2024. doi: 10.18653/v1/2024.emnlp-main.383. URL <https://aclanthology.org/2024.emnlp-main.383>.
- [68] Saaket Agashe, Yue Fan, Anthony Reyna, and Xin Eric Wang. LLM-coordination: Evaluating and analyzing multi-agent coordination abilities in large language models. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 8038–8057, 2025. doi: 10.18653/v1/2025.findings-naacl.448. URL <https://aclanthology.org/2025.findings-naacl.448>.
- [69] Sahar Abdelhabib, Amr Gomaa, Sarath Sivaprasad, Lea Schönher, and Mario Fritz. LLM-deliberation: Evaluating LLMs with interactive multi-agent negotiation game. In *Large Language Model (LLM) Agents @ ICLR 2024*, 2024. URL <https://openreview.net/forum?id=eE1WHn6qlk>.
- [70] Dan Qiao, Chenfei Wu, Yaobo Liang, Juntao Li, and Nan Duan. Gameeval: Evaluating llms on conversational games. *arXiv preprint arXiv:2308.10032*, 2023.
- [71] Yunhao Yang, Leonard Berthelemy, and Ufuk Topcu. Reasoning, memorization, and fine-tuning language models for non-cooperative games. *arXiv preprint arXiv:2410.14890*, 2024.

- [72] Gabriel Mukobi, Hannah Erlebach, Niklas Lauffer, Lewis Hammond, Alan Chan, and Jesse Clifton. Welfare diplomacy: Benchmarking language model cooperation. In *Socially Responsible Language Modelling Research*, 2023. URL <https://openreview.net/forum?id=WnR5BCX8GS>.
- [73] Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. Richelieu: self-evolving llm-based agents for ai diplomacy. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.
- [74] Maayan Orner, Oleg Maksimov, Akiva Kleinerman, Charles Ortiz, and Sarit Kraus. Explaining decisions of agents in mixed-motive games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 23267–23275, 2025.
- [75] Karthik Sreedhar and Lydia Chilton. Simulating human strategic behavior: Comparing single and multi-agent llms, 2024. URL <https://arxiv.org/abs/2402.08189>.
- [76] Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See-Kiong Ng, and Jiashi Feng. MAgIC: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7315–7332, 2024. doi: 10.18653/v1/2024.emnlp-main.416. URL <https://aclanthology.org/2024.emnlp-main.416/>.
- [77] Steve Phelps and Yvan I. Russell. The machine psychology of cooperation: Can gpt models operationalise prompts for altruism, cooperation, competitiveness and selfishness in economic games?, 2024. URL <https://arxiv.org/abs/2305.07970>.
- [78] Haolin Wang, Xueyan Li, Yazhe Niu, Shuai Hu, and Hongsheng Li. Empowering LLMs in decision games through algorithmic data synthesis. In *Will Synthetic Data Finally Solve the Data Access Problem?*, 2025. URL <https://openreview.net/forum?id=1RIHEJWN1L>.
- [79] Jinhao Duan, Renming Zhang, James Diffenderfer, Bhavya Kailkhura, Lichao Sun, Elias Stengel-Eskin, Mohit Bansal, Tianlong Chen, and Kaidi Xu. GTBench: Uncovering the strategic reasoning capabilities of LLMs via game-theoretic evaluations. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024. URL <https://openreview.net/forum?id=ypggxVWIVz>.
- [80] Jen-tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Michael R Lyu. How far are we on the decision-making of llms? evaluating llms' gaming ability in multi-agent environments. *arXiv preprint arXiv:2403.11807*, 2024.
- [81] Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, et al. Game-theoretic llm: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*, 2024.
- [82] Oguzhan Topsakal, Colby Jacob Edell, and Jackson Bailey Harper. Evaluating large language models with grid-based game competitions: An extensible llm benchmark and leaderboard, 2024. URL <https://arxiv.org/abs/2407.07796>.
- [83] Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. LLM-based agent society investigation: Collaboration and confrontation in avalon gameplay. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 128–145, 2024. doi: 10.18653/v1/2024.emnlp-main.7. URL <https://aclanthology.org/2024.emnlp-main.7/>.
- [84] Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can LLMs negotiate? NegotiationArena platform and analysis. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, volume 235 of *Proceedings of Machine Learning Research*, pages 3935–3951. PMLR, 21–27 Jul 2024. URL <https://proceedings.mlr.press/v235/bianchi24a.html>.
- [85] Eilam Shapira, Omer Madmon, et al. Glee: A unified framework and benchmark for language-based economic environments. *arXiv preprint*, 2024.
- [86] Haochuan Wang, Xiachong Feng, Lei Li, Zhenyue Qin, Dianbo Sui, and Lingpeng Kong. Tmgbench: A systematic game benchmark for evaluating strategic reasoning abilities of llms. *arXiv preprint arXiv:2410.10479*, 2024.
- [87] Shaoguang Mao, Yuzhe Cai, Yan Xia, Wenshan Wu, Xun Wang, Fengyi Wang, Qiang Guan, Tao Ge, and Furu Wei. ALYMPICS: LLM agents meet game theory. In Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert, editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 2845–2866, 2025. URL [https://aclanthology.org/2025.coling-main.193/](https://aclanthology.org/2025.coling-main.193).
- [88] Lanxiang Hu, Mingjia Huo, Yuxuan Zhang, Haoyang Yu, Eric P. Xing, Ion Stoica, Tajana Rosing, Haojian Jin, and Hao Zhang. Imgame-bench: How good are llms at playing games?, 2025. URL <https://arxiv.org/abs/2505.15146>.
- [89] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *Proceedings of the Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pages 22199–22213, 2022.
- [90] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *Proceedings of the International Conference on Learning Representations 11 (ICLR 2023)*, 2023. URL <https://openreview.net/forum?id=1PL1NIMMrw>.
- [91] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izahk Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Proceedings of the Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pages 11809–11822, 2023.
- [92] Benjamin Kempinski, Ian Gemp, Kate Larson, Marc Lanctot, Yoram Bachrach, and Tal Kachman. Game of thoughts: Iterative reasoning in game-theoretic domains with large language models. In *AAMAS*, 2025.
- [93] Jiaxian Guo, Bo Yang, Paul Yoo, Bill Yuchen Lin, Yusuke Iwasawa, and Yutaka Matsuo. Suspicion agent: Playing imperfect information games with theory of mind aware GPT-4. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=F2yGbwXJAI>.
- [94] Hongyi Guo, Zhihan Liu, Yufeng Zhang, and Zhaoran Wang. Can large language models play games? a case study of a self-play approach, 2024. URL <https://arxiv.org/abs/2403.05632>.
- [95] Minae Kwon, Sang Michael Xie, Kalesha Bullard, and Dorsa Sadigh. Reward design with language models. In *Proceedings of the International Conference on Learning Representations 11 (ICLR 2023)*, 2023. URL <https://openreview.net/forum?id=10uNuG15Kl>.

- [96] Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li, Yueling Zhuang, and Weiming Lu. Agent-pro: Learning to evolve via policy-level reflection and optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5348–5375, 2024. doi: 10.18653/v1/2024.acl-long.292. URL <https://aclanthology.org/2024.acl-long.292/>.
- [97] Dixiao Wei, Peng Yi, Jinlong Lei, Yiguang Hong, and Yuchuan Du. An automated reinforcement learning reward design framework with large language model for cooperative platoon coordination, 2025. URL <https://arxiv.org/abs/2504.19480>.
- [98] Reiji Suzuki and Takaya Arita. An evolutionary model of personality traits related to cooperative behavior using a large language model. *Scientific Reports*, 14(1):5989, 2024.
- [99] Austen Liao, Nicholas Tomlin, and Dan Klein. Efficacy of language model self-play in non-zero-sum games. In *Language Gamification @ NeurIPS 2024*, 2024. URL <https://openreview.net/forum?id=IK93maUXif>.
- [100] Xuanfa Jin, Ziyan Wang, Yali Du, Meng Fang, Haifeng Zhang, and Jun Wang. Learning to discuss strategically: a case study on one night ultimate werewolf. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.
- [101] Timo Schick, Jane Dwivedi-Yu, Roberto Dessa, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. In *Proceedings of the Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pages 68539–68551, 2023.
- [102] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, dahai li, Zhiyuan Liu, and Maosong Sun. ToolLLM: Facilitating large language models to master 16000+ real-world APIs. In *Proceedings of the International Conference on Learning Representations 12 (ICLR 2024)*, 2024. URL <https://openreview.net/forum?id=dFngZ0Ojjr>.
- [103] Chuanhao Li, Runhan Yang, Tiankai Li, Milad Baforassat, Kourosh Sharifi, Dirk Bergemann, and Zhuoran Yang. STRIDE: A tool-assisted LLM agent framework for strategic and interactive decision-making. In *Agentic Markets @ ICML 2024*, 2024. URL <https://openreview.net/forum?id=pqlkg1ABhr>.
- [104] Lloyd S Shapley. A value for n-person games. *Contribution to the Theory of Games*, 2, 1953.
- [105] Miriam Horovitz and Roni Goldshmidt. TokenSHAP: Interpreting large language models with Monte Carlo shapley value estimation. In Lotem Peled-Cohen, Nitay Calderon, Shir Lissak, and Roi Reichart, editors, *Proceedings of the 1st Workshop on NLP for Science (NLP4Science) @ ACL 2024*, pages 1–8, 2024. doi: 10.18653/v1/2024.nlp4science-1.1. URL [https://aclanthology.org/2024.nlp4science-1.1/](https://aclanthology.org/2024.nlp4science-1.1).
- [106] Amirata Ghorbani and James Zou. Data shapley: Equitable valuation of data for machine learning. In *Proceedings of the International Conference on Machine Learning 36 (ICML 2019)*, pages 2242–2251. PMLR, 2019.
- [107] Chuan Sun, Han Yu, and Lizhen Cui. Efficient shapley value-based non-uniform pruning of large language models. *arXiv preprint arXiv:2505.01731*, 2025.
- [108] Gokul Swamy, Christoph Dann, Rahul Kidambi, Steven Wu, and Alekh Agarwal. A minimaximalist approach to reinforcement learning from human feedback. In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [109] Corby Rosset, Ching-An Cheng, Arindam Mitra, Michael Santacroce, Ahmed Awadallah, and Tengyang Xie. Direct nash optimization: Teaching language models to self-improve with general preferences. *arXiv preprint arXiv:2404.03715*, 2024.
- [110] Mingzhi Wang, Chengdong Ma, Qizhi Chen, Linjian Meng, Yang Han, Jiancong Xiao, Zhaowei Zhang, Jing Huo, Weijie J Su, and Yaodong Yang. Magnetic preference optimization: Achieving last-iterate convergence for language model alignment. *arXiv preprint arXiv:2410.16714*, 2024.
- [111] Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Furong Huang, Dinesh Manocha, Amrit Singh Bedi, and Mengdi Wang. Maxmin-rlhf: Towards equitable alignment of large language models with diverse human preferences. In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [112] Rashid Mushkani, Hugo Berard, and Shin Koseki. Negotiative alignment: Embracing disagreement to achieve fairer outcomes—insights from urban studies. *arXiv preprint arXiv:2503.12613*, 2025.
- [113] Zixiang Chen, Yihe Deng, Huizhuo Yuan, Kaixuan Ji, and Quanquan Gu. Self-play fine-tuning converts weak language models to strong language models. In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [114] Jacob Makar-Limanov, Arjun Prakash, Denizalp Goktas, Nora Ayanian, and Amy Greenwald. Sta-rlhf: Stackelberg aligned reinforcement learning with human feedback. In *Coordination and Cooperation for Multi-Agent Reinforcement Learning Methods Workshop*, 2024.
- [115] Sijin Chen, Omar Hagrass, and Jason M Klusowski. Decoding game: On minimax optimality of heuristic text generation strategies. In *Proceedings of the International Conference on Learning Representations 13 (ICLR 2025)*, 2025.
- [116] Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M Rao, et al. Interpretability of deep learning models: A survey of results. In *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*, pages 1–6. IEEE, 2017.
- [117] Quan-shi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39, 2018.
- [118] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–52, 2021.
- [119] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017.
- [120] Hanxi Liu, Xiaokai Mao, Haocheng Xia, Jian Lou, and Jinfei Liu. Prompt valuation based on shapley values. *arXiv preprint arXiv:2312.15395*, 2023.

- [121] Behnam Mohammadi. Explaining large language models decisions using shapley values. *arXiv preprint arXiv:2404.01332*, 2024.
- [122] Zikun Ye and Hema Yoganarasimhan. Document valuation in llm summaries: A cluster shapley approach. *arXiv preprint arXiv:2505.23842*, 2025.
- [123] Yexiao He, Ziyou Wang, Zheyu Shen, Guoheng Sun, Yucong Dai, Yongkai Wu, Hongyi Wang, and Ang Li. Shed: Shapley-based automated dataset refinement for instruction fine-tuning. *arXiv preprint arXiv:2405.00705*, 2024.
- [124] Meng Cao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Scar: Shapley credit assignment for more efficient rlhf. *arXiv preprint arXiv:2505.20417*, 2025.
- [125] Yang Zhang, Yanfei Dong, and Kenji Kawaguchi. Investigating layer importance in large language models. In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 469–479, 2024. doi: 10.18653/v1/2024.blackboxnlp-1.29. URL <https://aclanthology.org/2024.blackboxnlp-1.29/>.
- [126] William Held and Diyi Yang. Shapley head pruning: Identifying and removing interference in multilingual transformers. In Andreas Vlachos and Isabelle Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2416–2427, 2023. doi: 10.18653/v1/2023.eacl-main.177. URL <https://aclanthology.org/2023.eacl-main.177/>.
- [127] Chenghao Yang, Fan Yin, He He, Kai-Wei Chang, Xiaofei Ma, and Bing Xiang. Efficient shapley values estimation by amortization for text classification. *arXiv preprint arXiv:2305.19998*, 2023.
- [128] Marcell Fekete and Johannes Bjerva. Linguistically grounded analysis of language models using shapley head values. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 850–865, 2025. doi: 10.18653/v1/2025.findings-naacl.49. URL <https://aclanthology.org/2025.findings-naacl.49/>.
- [129] Yingxuan Yang, Bo Huang, Siyuan Qi, Chao Feng, Haoyi Hu, Yuxuan Zhu, Jinbo Hu, Haoran Zhao, Ziyi He, Xiao Liu, Zongyu Wang, Lin Qiu, Xuezhi Cao, Xunliang Cai, Yong Yu, and Weinan Zhang. Who's the mvp? a game-theoretic evaluation benchmark for modular attribution in llm agents, 2025. URL <https://arxiv.org/abs/2502.00510>.
- [130] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Proceedings of the Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 2022. URL <https://openreview.net/forum?id=TG8KACxEOu>.
- [131] Timo Kaufmann, Paul Weng, Viktor Bengs, and Eyke Hüllermeier. A survey of reinforcement learning from human feedback. *arXiv preprint arXiv:2312.14925*, 2024. URL <https://arxiv.org/abs/2312.14925>.
- [132] Yue Wu, Zhiqing Sun, Huizhuo Yuan, Kaixuan Ji, Yiming Yang, and Quanquan Gu. Self-play preference optimization for language model alignment. In *Proceedings of the International Conference on Learning Representations 13 (ICLR 2025)*, 2025.
- [133] Yuheng Zhang, Dian Yu, Baolin Peng, Linfeng Song, Ye Tian, Mingyue Huo, Nan Jiang, Haitao Mi, and Dong Yu. Iterative nash policy optimization: Aligning llms with general preferences via no-regret learning. In *Proceedings of the International Conference on Learning Representations 13 (ICLR 2025)*, 2025.
- [134] Chenlu Ye, Wei Xiong, Yuheng Zhang, Hanze Dong, Nan Jiang, and Tong Zhang. Online iterative reinforcement learning from human feedback with general preference model. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 81773–81807, 2024.
- [135] Chenlu Ye, Wei Xiong, Yuheng Zhang, Nan Jiang, and Tong Zhang. A theoretical analysis of nash learning from human feedback under general kl-regularized preference. *arXiv e-prints*, pages arXiv–2402, 2024.
- [136] Daniil Tiapkin, Daniele Calandriello, Denis Belomestny, Eric Moulines, Alexey Naumov, Kashif Rasul, Michal Valko, and Pierre Menard. Accelerating nash learning from human feedback via mirror prox. *arXiv preprint arXiv:2505.19731*, 2025.
- [137] Runlong Zhou, Maryam Fazel, and Simon S Du. Extragradient preference optimization (egpo): Beyond last-iterate convergence for nash learning from human feedback. *arXiv preprint arXiv:2503.08942*, 2025.
- [138] Yibo Wang, Zikun Zhang, Zhihan Liu, Shenao Zhang, and Zhaoran Wang. Provably efficient and practical self-play for better llm alignment. *arXiv preprint arXiv:2405.00705*, 2024.
- [139] Xiaohang Tang, Sangwoong Yoon, Seongho Son, Huizhuo Yuan, Quanquan Gu, and Ilija Bogunovic. Game-theoretic regularized self-play alignment of large language models. *arXiv preprint arXiv:2503.00030*, 2025.
- [140] Reda Alami, Abdalgader Abubaker, Mastane Achab, Mohamed El Amine Seddik, and Salem Lahlou. Investigating regularization of self-play language models. *arXiv preprint arXiv:2404.04291*, 2024.
- [141] Hao Sun, Yunyi Shen, and Jean-Francois Ton. Rethinking bradley-terry models in preference-based reward modeling: Foundations, theory, and alternatives. *arXiv preprint arXiv:2411.04991*, 2024.
- [142] Zhekun Shi, Kaizhao Liu, Qi Long, Weijie J Su, and Jiancong Xiao. Fundamental limits of game-theoretic llm alignment: Smith consistency and preference matching. *arXiv preprint arXiv:2505.20627*, 2025.
- [143] Tan Zhi-Xuan, Micah Carroll, Matija Franklin, and Hal Ashton. Beyond preferences in ai alignment. *Philosophical Studies*, pages 1–51, 2024.
- [144] Eve Fleisig, Rediet Abebe, and Dan Klein. When the majority is wrong: Modeling annotator disagreement for subjective tasks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6715–6726, 2023.
- [145] Abhilash Mishra. Ai alignment and social choice: Fundamental limitations and policy implications. *arXiv preprint arXiv:2310.16048*, 2023.
- [146] Jessica Dai and Eve Fleisig. Mapping social choice theory to rlhf. *arXiv preprint arXiv:2404.13038*, 2024.
- [147] Ali Shirali, Arash Nasr-Esfahany, Abdullah Alomar, Parsa Mirtaheri, Rediet Abebe, and Ariel Procaccia. Direct alignment with heterogeneous preferences. *arXiv preprint arXiv:2502.16320*, 2025.

- [148] Hadi Hosseini and Samarth Khanna. Distributive fairness in large language models: Evaluating alignment with human values. *arXiv preprint arXiv:2502.00313*, 2025.
- [149] Ariel D Procaccia, Benjamin Schiffer, and Shirley Zhang. Clone-robust ai alignment. *arXiv preprint arXiv:2501.09254*, 2025.
- [150] Paul Götz, Nika Haghtalab, and Kunhe Yang. Distortion of ai alignment: Does preference optimization optimize for preferences? *arXiv preprint arXiv:2505.23749*, 2025.
- [151] Jiancong Xiao, Zhekun Shi, Kaizhao Liu, Qi Long, and Weijie J Su. Theoretical tensions in rlhf: Reconciling empirical success with inconsistencies in social choice theory. *arXiv preprint arXiv:2506.12350*, 2025.
- [152] Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H Holliday, Bob M Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoolkopf, et al. Position: Social choice should guide ai alignment in dealing with diverse human feedback. In *Proceedings of the International Conference on Machine Learning 41 (ICML 2024)*, 2024.
- [153] Zhaowei Zhang, Fengshuo Bai, Mingzhi Wang, Haoyang Ye, Chengdong Ma, and Yaodong Yang. Incentive compatibility for ai alignment in sociotechnical systems: Positions and prospects. *arXiv preprint arXiv:2402.12907*, 2024.
- [154] Young Wu, Yancheng Zhu, Jin-Yi Cai, and Xiaojin Zhu. The battling influencers game: Nash equilibria structure of a potential game and implications to value alignment. *arXiv preprint arXiv:2502.01127*, 2025.
- [155] Oliver Klingefjord, Ryan Lowe, and Joe Edelman. What are human values, and how do we align ai to them? *arXiv preprint arXiv:2404.10636*, 2024.
- [156] Kihyun Kim, Jiawei Zhang, Asuman Ozdaglar, and Pablo A Parrilo. Population-proportional preference learning from human feedback: An axiomatic approach. *arXiv preprint arXiv:2506.05619*, 2025.
- [157] Dominik Peters. Proportional representation for artificial intelligence. In *ECAI 2024*, pages 27–31. IOS Press, 2024.
- [158] Tianyi Qiu. Representative social choice: From learning theory to ai alignment. In *Pluralistic Alignment @ NeurIPS 2024*, 2024.
- [159] Pengyu Cheng, Yong Dai, Tianhao Hu, Han Xu, Zhisong Zhang, Lei Han, Nan Du, and Xiaolong Li. Self-playing adversarial language game enhances llm reasoning. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 126515–126543, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e4be7e9867ef163563f4a5e90cec478f-Paper-Conference.pdf.
- [160] Kaizhao Liu, Qi Long, Zhekun Shi, Weijie J Su, and Jiancong Xiao. Statistical impossibility and possibility of aligning llms with human preferences: From condorcet paradox to nash equilibrium. *arXiv preprint arXiv:2503.10990*, 2025.
- [161] Daiwei Chen, Yi Chen, Aniket Rege, and Ramya Korlakai Vinayak. Pal: Pluralistic alignment framework for learning from heterogeneous preferences. *Adaptive Foundation Models @ NeurIPS 2024*, 2024.
- [162] Chanwoo Park, Mingyang Liu, Dingwen Kong, Kaiqing Zhang, and Asuman E Ozdaglar. Rlhf from heterogeneous feedback via personalization and preference aggregation. In *Aligning Reinforcement Learning Experimentalists and Theorists @ ICML 2024*, 2024.
- [163] Huiying Zhong, Zhun Deng, Weijie J Su, Zhiwei Steven Wu, and Linjun Zhang. Provable multi-party reinforcement learning with diverse human feedback. *arXiv preprint arXiv:2403.05006*, 2024.
- [164] Parand A. Alamdar, Soroush Ebadian, and Ariel D. Procaccia. Policy aggregation. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 68308–68329, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/7e670825a578392891ad40e93931b1e3-Paper-Conference.pdf.
- [165] Joshua C Yang, Damian Dalisian, Marcin Korecki, Carina I Hausladen, and Dirk Helbing. Llm voting: Human choices and ai collective decision-making. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society 7 (AIES 2024)*, pages 1696–1708, 2024.
- [166] Daniel Halpern, Evi Micha, Ariel D Procaccia, and Itai Shapira. Pairwise calibrated rewards for pluralistic alignment. *arXiv preprint arXiv:2505.19731*, 2025.
- [167] Zhen Wang, Ruiqi Song, Chen Shen, Shiya Yin, Zhao Song, Balaraju Battu, Lei Shi, Danyang Jia, Talal Rahwan, and Shuyue Hu. Large language models overcome the machine penalty when acting fairly but not when acting selfishly or altruistically. *arXiv preprint arXiv:2410.03724*, 2024.
- [168] Joar Skalse, Nikolaus Howe, Dmitri Krasheninnikov, and David Krueger. Defining and characterizing reward gaming. In *Proceedings of the Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pages 9460–9471, 2022.
- [169] Ari Azarafooz and Farshid Faal. Language alignment via nash-learning and adaptive feedback. In *Models of Human Feedback for AI Alignment @ ICML 2024*, 2024.
- [170] Xu Chu, Zhixin Zhang, Tianyu Jia, and Yujie Jin. Stackelberg game preference optimization for data-efficient alignment of language models. *arXiv preprint arXiv:2502.18099*, 2025.
- [171] Rui Zheng, Hongyi Guo, Zihuan Liu, Xiaoying Zhang, Yuanshun Yao, Xiaojun Xu, Zhaoran Wang, Zhiheng Xi, Tao Gui, Qi Zhang, et al. Toward optimal llm alignments using two-player games. *arXiv preprint arXiv:2406.10977*, 2024.
- [172] Ziyu Ye, Rishabh Agarwal, Tianqi Liu, Rishabh Joshi, Sarmishta Velury, Quoc V Le, Qijun Tan, and Yuan Liu. Reward-guided prompt evolving in reinforcement learning for llms. In *Proceedings of the International Conference on Machine Learning 42 (ICML 2025)*, 2025.
- [173] Mickel Liu, Liwei Jiang, Yancheng Liang, Simon Shaolei Du, Yejin Choi, Tim Althoff, and Natasha Jaques. Chasing moving targets with online self-play reinforcement learning for safer language models. *arXiv preprint arXiv:2506.07468*, 2025.
- [174] Jiaqi Chen, Bang Zhang, Ruotian Ma, Peisong Wang, Xiaodan Liang, Zhaopeng Tu, Xiaolong Li, and Kwan-Yee K Wong. Spc: Evolving self-play critic via adversarial games for llm reasoning. *arXiv preprint arXiv:2504.19162*, 2025.
- [175] Runlong Zhou, Simon S Du, and Beibin Li. Reflect-rl: Two-player online rl fine-tuning for lms. *arXiv preprint arXiv:2402.12621*, 2024.
- [176] Xinhong Xie, Tao Li, and Quanyan Zhu. Learning from response not preference: A stackelberg approach for llm detoxification using non-parallel data. *arXiv preprint arXiv:2410.20298*, 2024.

- [177] Pengyu Cheng, Yifan Yang, Jian Li, Yong Dai, Tianhao Hu, Peixin Cao, Nan Du, and Xiaolong Li. Adversarial preference optimization: Enhancing your alignment via rm-llm game. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3705–3716, 2024.
- [178] Han Shen, Zhuoran Yang, and Tianyi Chen. Principled penalty-based methods for bilevel reinforcement learning and rlhf. *arXiv preprint arXiv:2402.06886*, 2024.
- [179] Souradip Chakraborty, Amrit Bedi, Alec Koppel, Huazheng Wang, Dinesh Manocha, Mengdi Wang, and Furong Huang. Parl: A unified framework for policy alignment in reinforcement learning from human feedback. In *Proceedings of the International Conference on Learning Representations 12 (ICLR 2024)*, 2024.
- [180] Vinzenz Thoma, Barna Pásztor, Andreas Krause, Giorgia Ramponi, and Yifan Hu. Contextual bilevel reinforcement learning for incentive alignment. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 127369–127435, 2024.
- [181] Hao Ma, Tianyi Hu, Zhiqiang Pu, Liu Boyin, Xiaolin Ai, Yanyan Liang, and Min Chen. Coevolving with the other you: Fine-tuning llm with sequential cooperative multi-agent reinforcement learning. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 15497–15525, 2024.
- [182] Ian Gemp, Yoram Bachrach, Marc Lanctot, Roma Patel, Vibhavari Dasagi, Luke Marrs, Georgios Piliouras, Siqi Liu, and Karl Tuyls. States as strings as strategies: Steering language models with game-theoretic solvers. *Agentic Markets @ ICML 2024*, 2024.
- [183] Athul Paul Jacob, Yikang Shen, Gabriele Farina, and Jacob Andreas. The consensus game: Language model generation via equilibrium search. In *Proceedings of the International Conference on Learning Representations 12 (ICLR 2024)*, 2024.
- [184] Baiting Chen, Tong Zhu, Jiale Han, Lexin Li, Gang Li, and Xiaowu Dai. Incentivizing truthful language models via peer elicitation games. *arXiv preprint arXiv:2505.13636*, 2025.
- [185] Bangguo Yu, Yuzhen Liu, Lei Han, Hamidreza Kasaei, Tingguang Li, and Ming Cao. Vln-game: Vision-language equilibrium search for zero-shot semantic navigation. *arXiv preprint arXiv:2411.11609*, 2024.
- [186] Weitong Zhang, Chengqi Zang, and Bernhard Kainz. Strategic llm decoding through bayesian games. In *Reasoning and Planning for Large Language Models @ ICLR 2025*, 2025.
- [187] Ian Gemp, Roma Patel, Yoram Bachrach, Marc Lanctot, Vibhavari Dasagi, Luke Marrs, Georgios Piliouras, Siqi Liu, and Karl Tuyls. Steering language models with game-theoretic solvers. In *Agentic Markets @ ICML 2024*, 2024.
- [188] Daniel Sefeni, Michael Johnson, and Joshua Lee. Game-theoretic approaches for stepwise controllable text generation in large language models. *Authorea Preprints*, 2024.
- [189] Guoxi Zhang and Jiuding Duan. Vickreyfeedback: Cost-efficient data construction for reinforcement learning from human feedback. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 351–366. Springer, 2024.
- [190] Xie Yi, Zhanke Zhou, Chentao Cao, Qiyu Niu, Tongliang Liu, and Bo Han. From debate to equilibrium: Belief-driven multi-agent LLM reasoning via bayesian nash equilibrium. In *Proceedings of the International Conference on Machine Learning 42 (ICML 2025)*, 2025. URL <https://openreview.net/forum?id=RQwexjUCxm>.
- [191] Thomas Kleine Buening, Jiarui Gan, Debmalya Mandal, and Marta Kwiatkowska. Strategyproof reinforcement learning from human feedback. *arXiv preprint arXiv:2503.09561*, 2025.
- [192] Haoran Sun, Yurong Chen, Siwei Wang, Wei Chen, and Xiaotie Deng. Mechanism design for llm fine-tuning with multiple reward models. *Pluralistic Alignment @ NeurIPS 2024*, 2024.
- [193] Shang Liu, Hanzhao Wang, Zhongyao Ma, and Xiaocheng Li. How humans help llms: Assessing and incentivizing human preference annotators. *arXiv preprint arXiv:2502.06387*, 2025.
- [194] Shang Liu, Zhongze Cai, Hanzhao Wang, Zhongyao Ma, and Xiaocheng Li. Incentivizing high-quality human annotations with golden questions. *arXiv preprint arXiv:2505.19134*, 2025.
- [195] Boaz Taitler, Omer Madmon, Moshe Tennenholtz, and Omer Ben-Porat. Data sharing with a generative ai competitor. *arXiv preprint arXiv:2505.12386*, 2025.
- [196] Renzhe Xu, Kang Wang, and Bo Li. Heterogeneous data game: Characterizing the model competition across multiple data sources. In *Proceedings of the International Conference on Machine Learning 42 (ICML 2025)*, 2025.
- [197] Yanxuan Wu, Haihan Duan, Xitong Li, and Xiping Hu. Navigating the deployment dilemma and innovation paradox: Open-source versus closed-source models. In *Proceedings of the ACM on Web Conference 2025, WWW '25*, page 1488–1501, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712746. doi: 10.1145/3696410.3714783. URL <https://doi.org/10.1145/3696410.3714783>.
- [198] Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The Economics of Large Language Models: Token Allocation, Fine-Tuning, and Optimal Pricing. *arXiv preprint arXiv:2502.07736*, 2025.
- [199] Rafid Mahmood. Pricing and competition for generative ai. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.
- [200] Xiang Li, Bing Luo, Jianwei Huang, and Yuan Luo. Strategic prompt pricing for aigc services: A user-centric approach. *arXiv preprint arXiv:2503.18168*, 2025.
- [201] Eden Saig, Ohad Einav, and Inbal Talgam-Cohen. Incentivizing quality text generation via statistical contracts. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.
- [202] Soheil Feizi, MohammadTaghi Hajiaghayi, Keivan Rezaei, and Suho Shin. Online advertisements with llms: Opportunities and challenges. *ACM SIGecom Exchanges*, 22(2):66–81, March 2025.

- [203] Ermis Soumalias, Michael J Curry, and Sven Seuken. Truthful aggregation of llms with an application to online advertising. *Agentic Markets @ ICML 2024*, 2024.
- [204] MohammadTaghi Hajiaghayi, Sébastien Lahaie, Keivan Rezaei, and Suho Shin. Ad auctions for llms via retrieval augmented generation. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.
- [205] Avinava Dubey, Zhe Feng, Rahul Kidambi, Aranyak Mehta, and Di Wang. Auctions with llm summaries. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 713–722, 2024.
- [206] Martino Banchio, Aranyak Mehta, and Andres Perlroth. Ads in conversations. In *Proceedings of the 26th ACM Conference on Economics and Computation*, page 350. Association for Computing Machinery, 2025. ISBN 9798400719431. doi: 10.1145/3736252.3742545. URL <https://doi.org/10.1145/3736252.3742545>.
- [207] Tommy Mordo, Moshe Tennenholtz, and Oren Kurland. Sponsored question answering. In *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 167–173, 2024.
- [208] Ziwei Wu and Junwu Zhu. A review on large model-oriented advertising auction. In *2024 IEEE International Conference on Cognitive Computing and Complex Data (ICCD)*, pages 7–12. IEEE, 2024.
- [209] Menghua Wu and Yujia Bao. Advertising in ai systems: Society must be vigilant. *arXiv preprint arXiv:2505.18425*, 2025.
- [210] Boaz Taitler and Omer Ben-Porat. Selective Response Strategies for GenAI. In *Proceedings of the International Conference on Machine Learning 42 (ICML 2025)*, 2025.
- [211] Sahand Sabour, June M Liu, Siyang Liu, Chris Z Yao, Shiyao Cui, Xuanming Zhang, Wen Zhang, Yaru Cao, Advait Bhat, Jian Guan, et al. Human decision-making is susceptible to ai-driven manipulation. *arXiv preprint arXiv:2502.07663*, 2025.
- [212] Guy Keinan and Omer Ben-Porat. Strategic content creation in the age of genai: To share or not to share? *arXiv preprint arXiv:2505.16358*, 2025.
- [213] Yi Gao, Zhe Wang, and Yan Huang. Pandora box or golden fleece: Economic analysis of generative ai adoption on creation platforms. In Michelle Carter 0001, Kelly J. Fadel, Thomas O. Meservy, Deborah J. Armstrong, Amit Deokar 0001, and Matthew L. Jensen, editors, *30th Americas Conference on Information Systems: Elevating Life through Digital Social Entrepreneurship, AMCIS 2024, Salt Lake City, UT, USA, August 15–17, 2024*. Association for Information Systems, 2024. URL https://aisel.aisnet.org/amcis2024/ai_aa/ai_aa/11.
- [214] Seyed A Esmaeili, Kshipra Bhawalkar, Zhe Feng, Di Wang, and Haifeng Xu. How to strategize human content creation in the era of genai? *arXiv preprint arXiv:2406.05187*, 2024.
- [215] Boaz Taitler and Omer Ben-Porat. Braess's paradox of generative ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14139–14147, 2025.
- [216] Benjamin Laufer, Jon Kleinberg, and Hoda Heidari. The Backfiring Effect of Weak AI Safety Regulation. *arXiv preprint arXiv:2503.20848*, 2025.
- [217] S Nageeb Ali, Nicole Immorlica, Meena Jagadeesan, and Brendan Lucier. Flattening supply chains: When do technology improvements lead to disintermediation? *arXiv preprint arXiv:2502.20783*, 2025.
- [218] Tian Xie, Xuwei Tan, and Xueru Zhang. Algorithmic decision-making under agents with persistent improvement. In *Proceedings of AAAI/ACM Conference on AI, Ethics, and Society 7 (AIES 2024)*, page 1672–1683, 2025.
- [219] Wenhao Li, Yue Lin, Xiangfeng Wang, Bo Jin, Hongyuan Zha, and Baoxiang Wang. Verbalized bayesian persuasion. *arXiv preprint arXiv:2502.01587*, 2025.
- [220] Niclas Boehmer, Sara Fish, and Ariel D Procaccia. Generative social choice: The next generation. In *Proceedings of the International Conference on Machine Learning 42 (ICML 2025)*, 2025.
- [221] Jie Sun, Tianyu Zhang, Houcheng Jiang, Kexin Huang, Chi Luo, Junkang Wu, Jiancan Wu, An Zhang, and Xiang Wang. Large language models empower personalized valuation in auction. *arXiv preprint arXiv:2410.15817*, 2024.
- [222] Nicolas Della Penna. Natural language mechanisms via self-resolution with foundation models. *arXiv preprint arXiv:2407.07845*, 2024.
- [223] Ismail Lotfi, Nouf Alabbasi, and Omar Alhussein. Rethinking strategic mechanism design in the age of large language models: New directions for communication systems. *IEEE Internet of Things Magazine*, pages 1–9, 2025. doi: 10.1109/IOT.2025.3576260.
- [224] David Huang, Francisco Marmolejo-Cossío, Edwin Lock, and David Parkes. Accelerated preference elicitation with llm-based proxies. *arXiv preprint arXiv:2501.14625*, 2025.
- [225] Agnieszka Mensfelt, Kostas Stathis, and Vince Trencsenyi. Autoformalizing and simulating game-theoretic scenarios using llm-augmented agents. *arXiv preprint arXiv:2412.08805*, 2024.
- [226] Shilong Deng, Yongzhao Wang, and Rahul Savani. From natural language to extensive-form game representations. *arXiv preprint*, 2025.
- [227] Agnieszka Mensfelt, Kostas Stathis, and Vince Trencsenyi. Autoformalization of game descriptions using large language models. *arXiv preprint arXiv:2409.12300*, 2024.
- [228] Yue Yin. Too much information? investigating information disclosure in auction systems with llm simulations. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25, New York, NY, USA, 2025*. Association for Computing Machinery. ISBN 9798400713958. doi: 10.1145/3706599.3720022. URL <https://doi.org/10.1145/3706599.3720022>.
- [229] Chanwoo Park, Xiangyu Liu, Asuman E Ozdaglar, and Kaiqing Zhang. Do llm agents have regret? a case study in online learning and games. In *Proceedings of the International Conference on Learning Representations 13 (ICLR 2025)*, 2025.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Manuscript submitted to ACM