Conversation Games and a Strategic View of the Turing Test

Kaveh Aryan
King's College London, UK
kaveh.aryan@kcl.ac.uk

Mudit Jain
Maharaja Agrasen Institute of Technology, India
mudit.jain2303@gmail.com

Abstract

Although many game-theoretic models replicate real interactions that often rely on natural language, explicit study of games where language is central to strategic interaction remains limited. This paper introduces the *conversation game*, a multi-stage, extensive-form game based on linguistic strategic interaction. We focus on a subset of the games, called verdict games. In a verdict game, two players alternate to contribute to a conversation, which is evaluated at each stage by a non-strategic judge who may render a conclusive binary verdict, or a decision to continue the dialogue. The game ends once a limit is reached or a verdict is given. We show many familiar processes, such as interrogation or a court process fall under this category. We also, show that the Turing test is an instance of verdict game, and discuss the significance of a strategic view of the Turing test in the age of advanced AI deception. We show the practical relevance of the proposed concepts by simulation experiments, and show that a strategic agent outperforms a naive agent by a high margin.

1 Introduction

Language functions not only as a channel for communication but also as a strategic instrument, whether in legal disputes, everyday conversations, or high-stakes negotiations. However, it has as not traditionally been studied this way. In this paper, we focus on a class of dialogue-driven interactions that we call *conversation games*. Our goal is to provide a formal framework that captures the strategic essence of these interactions, allowing us to analyse how participants exchange information and manipulate discourse to achieve specific objectives. After motivating examples, we introduce a general definition of conversation games and then specialize to the subset of games, called *verdict games*, in which a conversation leads to an external judgment, such as a court ruling or a verdict of authenticity, while players have private types and distinct incentives. We show that the Turing test [Oppy and Dowe, 2021] is an instance of the verdict game and discuss its strategic aspects. Furthermore, our simulation experiments demonstrate that strategic agents outperform naive ones and Demon-

strates the relevance of the proposed concept. The rest of this paper is structured as follows: Sections 2 and 3 introduce the notions of *conversation games* and *verdict games*, highlighting their significance through various examples. Section 4 discusses our experiments, Section 5 reviews related works, and Section 6 concludes the paper.

2 Conversation Game

Language is a ubiquitous medium for strategic interactions in real-world settings, from negotiations and debates to customer support and collaborative problem-solving. Unlike traditional games, where moves are tangible actions, conversation games involve contributions to an evolving dialogue. Understanding these dynamics enables researchers to model and predict behaviour in domains like diplomacy, AI-human interactions, and social decision-making.

Example 1. *We begin by highlighting how people use language strategically, leveraging nuances, ambiguity, and calculated phrasing to influence outcomes and manage relationships. Whether it is in workplace ("To deliver the best results, it might be helpful to revisit the timeline?"), friendship ("I've heard journaling can be really helpful."), or dating ("That new exhibit sounds interesting!"), language is used strategically as tool for navigating social interactions and achieving goals.*

Example 2. *(Court) One of the most notable examples of a conversation game is a courtroom trial. Here, the players are the prosecution and the defense, each strategically interacting to achieve their respective goals (conviction or exoneration). Although physical or forensic evidence may be presented, the dialogue surrounding these pieces of evidence often drives the judge or jury's perception. Each utterance aims to persuade the fact-finders and shape their interpretation of the evidence.*

Conversation games can involve hidden roles and Bayesian reasoning, as the following couple of examples illustrate:

Example 3. *(Interrogation) In an interrogation setting, the key players are the interrogator and the suspect. In many legal contexts, the ultimate goal is not merely to gather intelligence but to do so in a manner admissible and convincing in court. Consequently, the interrogator must guide the dialogue to elicit information in a form that will be deemed*

credible and legally defensible, while the suspect may strategically reveal or conceal details to influence legal outcomes.

Example 4. *(Turing test) Another notable example is the Turing test [Oppy and Dowe, 2021]. Here, the players are a human interrogator and a witness who may be either human or machine. The interrogator updates their beliefs (in a Bayesian fashion) about the witness's identity based on the responses received. The witness, in turn, strategically produces language that minimises the chances of being identified as .*

It should be noted that the utility of players is not necessarily determined by the conversation itself, but rather by some external effect or outcome of the conversation:

Example 5. *(Psychoanalysis) In a psychoanalysis or cognitive behavioral therapy (CBT) session, the therapist and the patient engage in a conversation aimed at improving the patient's mental well-being. Although the dialogue is central to the therapeutic process, the eventual measure of success lies in the patient's external life improvements or changes in mental health, rather than the conversation itself.*

Example 6. *(Teaching) In a classroom setting, a teacher and students engage in conversational exchanges. The teacher's ultimate objective is effective learning as evidenced by students' grasp of the material, rather than simply guiding a smooth dialogue.*

One can readily identify other examples that are subsumed under the overarching concept of the conversation game, such as diplomacy, bargaining, negotiation, debate, and persuasion. A formal definition of the conversation game is as follow:

Definition 1. *Conversation Game. Formally, a conversation game is a multistage, extensive-form game with the following characteristics:*

1. Players and roles. *There can be multiple players, each potentially holding a private type.*

2. Actions as utterances. *At each stage, a player's action is an utterance, which contributes to the evolving dialogue.*

3. Information structure. *Players may have private information about their type or objectives, and they update their beliefs based on the content of the conversation.*

4. Payoffs. *The utility of each player is determined by the ultimate outcome, which may be influenced but not fully determined by the conversation's contents.*

In this paper, we primarily focus on a type of conversation game illustrated by Examples 2, 3, and 4. These are games where players guide the conversation so that a non-strategic judge (or evaluator) can make a binary judgment solely based on the dialogue's content. The utility of the players depends on their private types and on the outcome of this external judgment. We refer to this subset of conversation games as *verdict games*, and they will be formalized in detail in the following section.
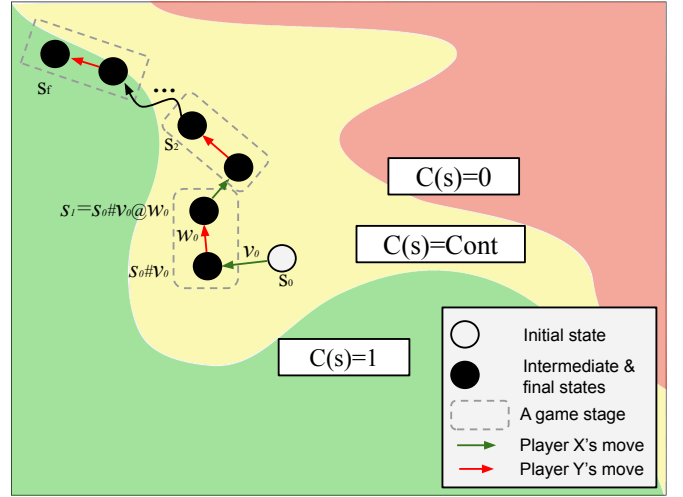

Figure 1: An illustration of the verdict game.

3 Verdict Games

A *verdict game* is a conversation game in which players' utilities are determined by the outcome of a binary sequence classifier on the conversation history.

3.1 Problem Setup

The *verdict game* (Fig. 1) game needs to be general to accommodate different games and player goals. The game involves two players, player X (she/her/her) and player Y (he/his/him). Each player has a type drawn from their respective set, T_X for X and T_Y for Y, which might or might not be revealed to the other. The game starts at state s_0, which is a string (possibly empty) from an alphabet Σ. The game is played in multiple stages. Each stage starts with the X appending a string $v \in \Sigma^+$ with length at most l to the current state s_t, resulting in the string $s_t \# v$, to which Y adds another string $w \in \Sigma^+$ ($|w| \leq l$), resulting to $s_{t+1} = s_t \# v @ w$ ("#" and "@" are X and Y's delimiters, respectively. In reality, they are replaced by speakers designators, such as "X: ..."). This string is then fed to a string classifier, $\mathcal{C} : (\Sigma^* \# \Sigma^+ @ \Sigma^+)^+ \rightarrow \{0, 1, Cont\}$. The classifier's outcome, $c_t = \mathcal{C}(s_t)$, determines whether the game needs to stop (if it is 0 or 1) or continue to the next stage (if it is $Cont$, unless a maximum number of stages, d, is reached). In practice, $Cont$ represents a case that a conclusive verdict cannot be made. In the extended-form formalism, $\mathcal{C}(s_t)$ determines if a decision node, s_t, in the game tree is terminal or not. To keep the game general, we propose using separate utility functions, one for each type of the players, in the form of $u : T_X \times T_Y \times \{0, 1, Cont\} \rightarrow \mathbb{R}$. This determines the utility of the players in the terminal nodes.

Definition 2. *A verdict game, is a conversation game parametrised by $(T_X, T_Y, \Sigma, l, d, \mathcal{C}, \{u_i\}_{i \in T_X \cup T_Y})$ that is played as described in the beginning of this section.*

Example 7. *Court (formal). This is a formalised version of Example 2. Players X and Y are the prosecution and the defence, respectively. The classifier is a judge, where the*

verdicts of 0 and 1, represent verdicts of "not guilty" and "guilty", respectively. As long as the players are concerned, the true type of the defence (actually guilty or not) is irrelevant. Therefore, both players are of a single types which are omitted from u function for brevity. The utility functions are therefore defined as below. $u_X(Cont)$ reflects the doctrine of presumption of innocence.

$$u_X(0) = -1, u_X(1) = 1, u_X(Count) = -1$$
$$u_Y(0) = 1, u_Y(1) = -1, u_Y(Cont) = 1$$
(1)

Example 8. *Interrogation (formal).* *This is a formalised version of Example 3. Players X and Y are the interrogator and the suspect, respectively. As the suspect could be guilty or non-guilty, $t_Y = \{Guilty, Non\text{-}Guilty\}$. At first glance, the classifier might seem to reflect the interrogator's internal judgment about the suspect's status. However, upon closer examination, it becomes clear that the relevant judgment for an effective interrogation is that of an impartial observer—akin to an indifferent judge assessing the interrogation's progression in a courtroom. This perspective is captured by the non-strategic classifier in our model. Again, the verdicts of 0 and 1, represent verdicts of "guilty" and "not guilty", respectively. However, the goal of the interrogator is to establish the suspect's true type. The respective utility functions are shown in Table 1.*

Player	t_Y	$c = 0$	$c = 1$	$c = Cont$
X	Non-Guilty	1	-1	0
X	Guilty	-1	1	0
Y (Non-Guilty)	Non-Guilty	1	-1	1
Y (Guilty)	Guilty	1	-1	1

Table 1: Utility outcomes for players X and Y under different types and actions in an interrogation game.

It can be observed that for the interrogator, this is just the usual (unweighted) classification cost function, where false positives and false negatives are penalised while true positives and true negatives are rewarded. Alternative utility functions, that represent various preferences are of course possible. The goal of the suspect is to be exoneration, no matter what his type is.

Example 9. *Turing test (formal).* *The usual Turing test (Example 4) is isomorphic to the interrogation example, where the witness types, i.e., "Non-Guilty" and "Guilty" can correspond to "Human" and "Machine", respectively. Classifier is an AI-detection system, that decides, based on the history of the conversation at each state, whether the witness is Human (0) or Machine (1).*

We conclude this section by putting this approach to the Turing test in context. Historically, during the era of ELIZA [Weizenbaum, 1966], the Turing Test was viewed primarily as a benchmark for intelligence, often involving simple, non-strategic interactions. However, the landscape has changed significantly; advanced Large Language Models (LLMs) are now sometimes judged as "more human than humans" [Rathi *et al.*, 2024]. Current statistical and machine learning models perform well at detecting AI-generated text [Mitchell *et*

al., 2023; Mireshghallah *et al.*, 2024], but their effectiveness is limited to texts directly produced by LLMs as they are not designed to detect adversarially manipulated texts designed to obscure their machine origins. As deceptive techniques advance, it is likely that more sophisticated detection methods will be necessary—methods that strategically interact with agents to reveal their machine nature, akin to the sequential and interactive *Voight-Kampff* Test from the science-fiction classic *Blade Runner*.

3.2 Considerations and Limitations

The Verdict game, despite its versatility, is constrained by several limitations. One major challenge is the enormous branching factor at each step, as it includes all possible utterances that could occur, such as topic changes or tangential remarks. A potential workaround is to narrow down the options to only the most likely and natural utterances for that moment. Alternatively, players can adopt high-level strategies—such as focusing on eliciting contradictions, pressing for details, or establishing inconsistencies—and restrict their utterances to those aligned with their chosen strategy.

Another limitation arises in cases, e.g., where the interrogator seeks specific information, such as the name of a person or non-binary details, which are harder to model within the game's structure. Additionally, the importance of impartial judgment poses a challenge, as the players' personal judgments might influence outcomes, for example, in scenarios where judges themselves ask questions (true that even here the judge's decision might be judged by a higher-level court or by the public or even the "history", but we might equally decide not to delve into this complexity) or when a human suspects that the other side might be a chatbot and it is not easy for them to use AI-detector tools to verify the fact.

3.3 Equilibria

All conversation games, including the verdict game, are finite, because the number of stages and the length of player actions are finite. In particular, the court game (Example 7) is a complete-information, zero-sum game, in principle, solvable by an standard minimax algorithm to a Subgame Perfect Equilibrium [Tadelis, 2013, Chapter 8]. Admittedly, the action space is too large to allow for an exhaustive search. Therefore, approaches such as Monte Carlo Tree Search [Świechowski *et al.*, 2023] need to be employed to approximately solve the game. The interrogation games, e.g., Examples 8, 4, are Bayesian [Tadelis, 2013, Chapter 15]. The appropriate solution concept is Perfect Bayesian Equilibrium [Tadelis, 2013, Chapter 16]. Again, the action space is too large, calling for a surrogate methods, such as Information Set Monte Carlo Tree Search [Cowling *et al.*, 2012]. In the next section, we present the results obtained under a severely limited branching factor as a proof of concept.

4 Experiments

This section serves as a proof of concept illustration of the ideas proposed in this paper, and we leave a through exploration to future work.

4.1 Court Process

To simulate the court example (Example 7), we used LLM agents [Huang *et al.*, 2024; Xi *et al.*, 2023], to act as prosecution, defence, and judge. To illustrate the efficacy of strategic planning, we compared a naive prosecutor with a strategic one. While the naive prosecutor selects questions using the LLM's default temperature, the strategic prosecutor performs a shallow search to introspect and choose the question most likely to convince the judge (Table 2). The results of the experiments show that the strategic agent wins 64% of the times, while the naive agent wins 27% of the times. The difference is statistically highly significant with a p-value of $\leq 1e - 5$.

Detail of experiments. LLM: OpenAI GPT-4o [OpenAI, 2024] (`gpt-4o-2024-11-20`), depth of introspection: 1, breadth of introspection: 10, number of experiments for each type of prosecutor: 100.

Component	Description
Context (Case Details)	*Victim*: Emily Harper (34, journalist) *Suspect*: Ryan Carter (ex-boyfriend) *Evidence*: fingerprints, torn jacket, text messages, being seen by witnesses, signs of struggle in victim's apartment.
Play X (Prosecutor, Naive)	Generates a response based on default temperature settings, asking simple and direct questions without leveraging psychological or strategic patterns.
Play X (Prosecutor, Naive)	Introspects using 10 simulated conversations to generate the most effective question.
Player Y (Defence)	Represents Ryan Carter (suspect), responding with plausible denials, evasions, or misleading statements designed to conceal guilt.
Classifier	Analyzes the conversation history and assigns one of three verdicts: Guilty, Innocent, or Non-conclusive, with an emphasis on identifying guilt if vague responses occur to the questins regarding the suspect being seen by witnesses.

Table 2: Description of components in the court experiment.

4.2 Debate Tournament

A structured debate between two language model agents represents a natural instance of a verdict game. In this setting, the participants assume fixed roles—proposition and opposition—and alternately contribute concise argumentative utterances in response to a stated motion. The interaction unfolds in discrete turns, typically bounded by a predetermined number of rounds. At the end of the exchange, a non-strategic evaluator—implemented as a language model prompted to act as a neutral judge—renders a binary verdict for one side or declares a draw, based entirely on the content of the dialogue. This setup encapsulates the defining features of verdict games: players engage through language, and their utterances influence the judgment of an external evaluator whose decision determines the players' payoffs. Unlike fact-finding contexts such as interrogations, the emphasis here lies in rhetorical strength and persuasive effectiveness. Each agent's objective is not to assert a universal truth but to steer the conversation strategically within the constraints of its assigned stance, in order to maximize its likelihood of success under evaluative scrutiny.

Proof of Concept

To demonstrate the practical viability of the debate setting within the verdict game framework, we developed a simulated tournament in which language model agents participated in multi-round debates under varying strategic assumptions. In the baseline condition, agents generated responses turn by turn without forward planning. In contrast, a strategic variant evaluated multiple candidate utterances per turn, selecting the one most likely to influence the judge's final assessment. Each debate proceeded for a fixed number of rounds, after which the entire conversation was submitted to a neutral language model acting as the judge. The judge evaluated the dialogue using criteria such as clarity, coherence, and alignment with the assigned stance, ultimately producing a structured verdict. While outcomes varied across motions and matchups, the overall pattern was clear: strategically guided agents regularly outperformed their non-strategic counterparts, confirming the core intuition of the verdict game model. This result illustrates the value of planning in language-based interactions and supports the broader premise that strategic use of discourse can measurably impact externally judged outcomes.

5 Related Work

This section situates our study within the broader landscape of game theory, linguistic interactions, and AI-related strategic frameworks.

Classic games modelled after linguistic interactions. Many game models in game theory is modelled after games where the language is a medium for actions, e.g., offering a wage in a signalling game [Tadelis, 2013, Chapter 16], and players engaging in *cheap talks* [Tadelis, 2013, Chapter 18]. However, these are oversimplified models where the actions are discretised into a limited set of predefined choices. In other domain, such as auctions [Tadelis, 2013, Chapter 13], the linguistic choices are often irrelevant. This is in contrast with the present work, where the actions as linguistic utterances as such.

Games over language. Several studies consider games that involve alphabets. Two such works consider games where players add letters to the end of shared word [Rosenfeld, 2024; Marcus and Törmä, 2023], with clearly different scope with the present work. Also, there are game theoretic studies of language-based games, such a general study of hidden-role games [Carminati *et al.*, 2024], Werewolf (Mafia) [Wang, 2024], and Secret Hitler [Reinhardt, 2020]. However, these

works either abstract away linguistic interactions or specialise to a specific game [Bertolazzi *et al.*, 2023].

Game theoretic linguistics. There is a branch of studies that studies pragmatics and rational discourse from a game theoretic standpoint [Jacob *et al.*, 2024; Goodman and Frank, 2016; Hintikka, 2012]. This is in contrast with the present study, where the goal is to define a game using language.

Strategic classification. Our verdict game involves strategically taking turns to steer the ongoing conversation toward the boundaries of a sequence classifier. This has similarities with the problem of *strategic classification* [Hardt *et al.*, 2015; Ghalme *et al.*, 2021; Ghalme *et al.*, 2021], where a player plays against a classifier (i.e., the party that controls the classifier): the player strategically games the classifier by changing its observable attributes to achieve a desirable classification outcome (such as a desirable mortgage decision), while the party that controls the classifier tries to oppose such manipulations. This is in contrast with the present work, where the classifier is non-strategic.

Turing test and AI detection. One of our motivating examples is the Turing Test [Oppy and Dowe, 2021]. At the era of large language models, several studies has shown significant improvements advanced Large Language Models (LLMs) are sometimes judged as "more human than humans." [Rathi *et al.*, 2024]. Current statistical and machine learning models achieve high accuracy in detecting AI-generated text [Mitchell *et al.*, 2023; Mireshghallah *et al.*, 2024], but these methods are primarily effective for texts directly produced by LLMs and are not designed for texts adversarially manipulated to obscure their machine origins.

LLM agents. LLM-based agents are AI systems that leverage a large language model (LLM) as their core reasoning engine. A key characteristic of these agents is their ability to engage in multi-turn interactive conversations [Huang *et al.*, 2024; Xi *et al.*, 2023]. We utilized LLM-based agents, in our experiments, in dual roles: as players and as judges, each guided by tailored prompts designed for their respective tasks.

6 Conclusion and Future Work

This paper introduced a formal framework for studying *conversation games*, with a focus on their subset, *verdict games*, where dialogue influences external judgments. By modeling scenarios such as courtroom trials, interrogations, and the Turing test, we showcased how strategic linguistic interactions shape outcomes. Our analysis of the Turing test revealed its strategic depth, emphasizing the importance of robust detection methods in the era of advanced AI systems.

Simulation experiments demonstrated the superiority of strategic agents over naive ones, highlighting the practical relevance of our framework. However, several challenges remain, such as handling the high branching factor in conversational moves and modeling more complex objectives in multi-agent settings.

Future work will focus on exploring trust dynamics in AI-human interactions and developing advanced interrogation strategies. We also aim to leverage advanced search methods, such as Monte Carlo Tree Search, to approximate equilibria in more complex conversation games. This framework provides a foundation for understanding and designing systems that navigate strategic linguistic interactions in both human and AI-driven contexts.

References

[Bertolazzi *et al.*, 2023] Leonardo Bertolazzi, Davide Mazzaccara, Filippo Merlo, and Raffaella Bernardi. ChatGPT's Information Seeking Strategy: Insights from the 20-Questions Game. In C. Maria Keet, Hung-Yi Lee, and Sina Zarrieß, editors, *Proceedings of the 16th International Natural Language Generation Conference*, pages 153–162, Prague, Czechia, September 2023. Association for Computational Linguistics.

[Carminati *et al.*, 2024] Luca Carminati, Brian Hu Zhang, Gabriele Farina, Nicola Gatti, and Tuomas Sandholm. Hidden-Role Games: Equilibrium Concepts and Computation, July 2024. arXiv:2308.16017 [cs].

[Cowling *et al.*, 2012] Peter I. Cowling, Edward J. Powley, and Daniel Whitehouse. Information Set Monte Carlo Tree Search. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(2):120–143, June 2012. Conference Name: IEEE Transactions on Computational Intelligence and AI in Games.

[Ghalme *et al.*, 2021] Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic Classification in the Dark. In *Proceedings of the 38th International Conference on Machine Learning*, pages 3672–3681. PMLR, July 2021. ISSN: 2640-3498.

[Goodman and Frank, 2016] Noah D. Goodman and Michael C. Frank. Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, 20(11):818–829, November 2016.

[Hardt *et al.*, 2015] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic Classification, November 2015. arXiv:1506.06980 [cs].

[Hintikka, 2012] Jaakko Hintikka. *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*. Springer Science & Business Media, December 2012. Google-Books-ID: Hr19CAAAQBAJ.

[Huang *et al.*, 2024] Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. Understanding the planning of LLM agents: A survey, February 2024. arXiv:2402.02716 [cs].

[Jacob *et al.*, 2024] Athul Jacob, Gabriele Farina, and Jacob Andreas. Regularized Conventions: Equilibrium Computation as a Model of Pragmatic Reasoning. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2944–2955, Mexico City, Mexico, June 2024. Association for Computational Linguistics.

[Marcus and Törmä, 2023] Pierre Marcus and Ilkka Törmä. WINNING SETS OF REGULAR LANGUAGES: DESCRIPTIONAL AND COMPUTATIONAL COMPLEXITY. *Journal of Automata, Languages and Combinatorics*, 28, 2023.

[Mireshghallah *et al.*, 2024] Niloofar Mireshghallah, Justus Mattern, Sicun Gao, Reza Shokri, and Taylor Berg-Kirkpatrick. Smaller Language Models are Better Zero-shot Machine-Generated Text Detectors. In Yvette Graham and Matthew Purver, editors, *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 278–293, St. Julian's, Malta, March 2024. Association for Computational Linguistics.

[Mitchell *et al.*, 2023] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. In *Proceedings of the 40th International Conference on Machine Learning*, pages 24950–24962. PMLR, July 2023. ISSN: 2640-3498.

[OpenAI, 2024] OpenAI. GPT-4o System Card, October 2024.

[Oppy and Dowe, 2021] Graham Oppy and David Dowe. The Turing Test. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2021 edition, 2021.

[Rathi *et al.*, 2024] Ishika Rathi, Sydney Taylor, Benjamin K. Bergen, and Cameron R. Jones. GPT-4 is judged more human than humans in displaced and inverted Turing tests, July 2024. arXiv:2407.08853 [cs] version: 1.

[Reinhardt, 2020] Jack Reinhardt. Competing in a Complex Hidden Role Game with Information Set Monte Carlo Tree Search, May 2020. arXiv:2005.07156 [cs].

[Rosenfeld, 2024] Matthieu Rosenfeld. Ann wins the non-repetitive game over four letters and the erase-repetition game over six letters. *European Journal of Combinatorics*, 118:103924, May 2024.

[Tadelis, 2013] Steven Tadelis. *Game Theory: An Introduction*. Princeton University Press, January 2013. Google-Books-ID: _4OqAaITAWMC.

[Wang, 2024] S. T. Wang. Optimal Strategy in Werewolf Game: A Game Theoretic Perspective, August 2024. arXiv:2408.17177 [econ] version: 1.

[Weizenbaum, 1966] Joseph Weizenbaum. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, January 1966.

[Xi *et al.*, 2023] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. The Rise and Potential of Large Language Model Based Agents: A Survey, September 2023. arXiv:2309.07864 [cs].

[Świechowski *et al.*, 2023] Maciej Świechowski, Konrad Godlewski, Bartosz Sawicki, and Jacek Mańdziuk. Monte Carlo Tree Search: a review of recent modifications and applications. *Artificial Intelligence Review*, 56(3):2497–2562, March 2023.