CatanBench: Investigating Spatial-Reasoning and Long-Term Planning in Large-Language Models

Marco Yang
Caltech
Pasadena, CA 91126
myang2@caltech.edu

Justin Luo
Caltech
Pasadena, CA 91126
jluo2@caltech.edu

Rahel Joshi
Caltech
Pasadena, CA 91126
raheljoshi@caltech.edu

Shan Gupta
Caltech
Pasadena, CA 91126
srgupta@caltech.edu

Abstract

Large Language Models (LLMs) have quickly improved in tasks like generating code, yet they still struggle with spatial-reasoning, long-term planning, and negotiation. We introduce CatanBench, a novel benchmark built around the board game Settlers of Catan to understand the limits of LLMs. CatanBench simulates Catan games with an API meant to be be used with LLMs, including stochastic resource yields, partially observable environment, and two-party trading. Our benchmark includes LLM agents with various prompting strategies, like chain-of-thought, rule-based, and tree-search. We evaluate these models on a miniature board using a Bradley-Terry rating model. We show that LLMs struggle to spatially reason about settings that are explicitly described to them, and that they demonstrate a greedy bias towards action that indicates that they are unable to recognize when long-term planning should take priority over short-term gain.

1 Introduction

LLMs dominate other forms of machine learning (ML) in classifying texts, generating code, and answer open-ended questions, yet their strategic planning is not fully understood. GameBench demonstrated that LLMs are weak in this sort of planning, as GPT4, the at the time best model available only won 45% of the time against humans across 9 games, despite well-crafted prompts (Costarelli et al., 2024). Followup research such as GTBench on canonical game-theory tasks (Duan et al., 2024), LLM-Deliberation on negotiation (Abdelnabi et al., 2023), and even for chess with ChessGPT (Noever et. al 2024) show that LLMs are weak at long-term planning and strategy.

These game benchmarks have have two critical limits:

1. Low-dimensionality of action space: at any turn, the number of actions is less than 10.
2. Short-term: Games often end after only a few moves

Settlers of Catan breaks these limits. A typical game of Catan consists of 50 to 100 turns, random resource yields simulating dice rolls, spatial reasoning over a hex grid, two-party trades, and partial information with hidden opponent hands and development cards. These characteristics better relate to real-world situations.

This paper offers:

Preprint.

1. CatanBench, a sim for Catan games for LLMs
2. Modular LLM agents that are interchangeable and various prompting strategies
3. Evaluating protocols with game logs, and rating strategies by win-rate with Bradley-Terry models
4. An exploration of information asymmetry and prompt design for LLM agent games.

2 Links to Code

`https://github.com/mcrco/catanllm`

3 Background and Previous Work

Short-term strategy benchmarks.

GameBench created a suite of nine simple games and showed that human players outperformed LLMs even when LLMs had chain-of-thought prompting (Costarelli et al., 2024). GTBench added more games like Kuhn Poker and Prisoner's Dilemma, and demonstrated that LLMs struggled to understand how the opponents would play, particularly for partially observable games (Duan et al., 2024). Regarding negotiation, LLM-Deliberation found that LLMs struggled to make strategically strong trades (Abdelnabi et al., 2023). These papers show that for games with many moves, LLMs tend to struggle and fail to create strong strategies.

Complex games

Recently, LLMs have been tested on more complicated games. Bakhtin et al. (2022) managed to get close to human performance by combining LLMs with search-based planning. In Koch et al. (2024), it was shown that in even simple simultaneous-move games, pure LLMs without search or simulation tools struggled to perform well. Regarding chess, Noever et. al (2024) showed that an instructed-tuned GPT model played at an intermediate skill level but lacked long-term strategic thinking. And Zhou et al. (2025) added new benchmarks for Monopoly, Codenames, and Risk.

LLMs with tools

Agents that have access to tools like in ReAct and AutoChain had improvements in succeeding tasks by adding reasoning steps, external simulation, or memory updates. But this requires a single LLM focused on fully observable problems and do not extend well to environments where there is a lot of hidden information or multiple players sharing common goals.

Our Benchmark

CatanBench creates a unique benchmark when it comes to LLMs used for games. It has a large number of actions, there is hidden information from the players, there are multi-agent interactions, and there is also negotiating and trading. Other games like Diplomacy focuses too much on alliances and games like chess have no hidden information; instead Catan involves spatial planning and randomness, which other game benchmarks lack. These characteristics better represent real-world circumstances of multi-agent planning than other LLM game benchmarks, since it tests LLMs on long-term strategic planning with incomplete information and with randomness. CatanBench allows for directly comparing prompting strategies, reasoning policies, and LLM models under a high-complexity benchmark.

4 Research Question and Approach

We introduce CatanBench, a benchmark for language models that tests their ability to plan, strategize, and reason when playing Settlers of Catan, where spatial thinking, long-term planning, and negotiation are essential to success. While our methodology can be used to compare across model architectures, we, in this study, focus on how different prompting strategies affect LLM players' performance against botss and each other on the same model architecture. We give one player access to the Catan rulebook and another access to a list of possible strategies to use; we also use one model endowed with greater thinking capabilities and instruct yet another to use tree-search reasoning. In this way,

we explore how information asymmetry can affect reasoning in LLMs and evaluate their performance against non-reasoning models.

At its core, this study drives at three main questions:

1. How does an LLM with preferential information asymmetry (e.g. access to the rules of the game or possible strategies to employ during the game) reason in a long-term planning environment compared to its naive opponent?
2. Can reasoning models, placed in a setting where strategy over the course of multiple turns can be important, develop one without explicit prompting?
3. How do different reasoning methods (e.g. chain-of-thought or tree-search) affect players' performance in the game?

To that end, we pit players against each other in head-to-head Catan games using the `Catanatron` engine, which allows code-mediated gameplay. We use reference bot players, as well as our own specially prompted reasoning players on Gemini-2.5-Flash and Qwen3-235B-A22B. We then evaluate all players' performance according to the Bradley-Terry scheme described in Costarelli et al., 2024.

Our approach was carefully selected. In this way, we isolate the effect of the choice of prompt, as we evaluate each prompting strategy against reference models with and without reasoning. We also evaluate models against each other to investigate how LLMs with different prompts interact with one another. We adopt an objective and universal metric that ranks players, which enables us to make more generalized comments on prompting strategies' effectiveness. We finally save replays of each game containing the reasoning tokens of the models to more carefully investigate some emergent behaviors of LLM players.

5 Experiment

We have several players face off against one another in Settlers of Catan.

5.1 Experimental Setup

We use the `Catanatron` engine to mediate our games. Both to speed up the games and to make computation easier, we use a miniature version of the game board: we play with only seven hexes rather than the standard nineteen in the usual Settlers of Catan. See Figure 1 for an example board. While this setup does introduce misalignment between provided rules and strategies (see Rules/Strategy vs. Basic), it provides a baseline implementation to test models' performance while not exhausting computational power.

We below describe in-depth each player that we use. Each player plays against every other five times, and win rates, total scores, and rankings are subsequently derived from these games.

5.2 Player Descriptions

We use a variety of players that are either provided box-standard by `Catanatron` or are differentially prompted Gemini-2.5-Flash/Qwen3-235B-A22B models.

5.2.1 Box-Standard Bots

`Catanatron` provides several box-standard players that can be used to evaluate our models against. We choose:

- `Random`: This player makes a random move at every turn. Since it is intentionally non-reasoning, we expect that its performance is poor against all other models.
- `AlphaBeta`: This player maximizes its position at every turn through the minimax algorithm. The name, "AlphaBeta," refers to the alpha-beta pruning algorithm that refines the minimax, providing a more efficient way of reaching the same minimax conclusion. This player lacks long-term strategy, but it plays the optimal decision for every turn.
- `ValueFunction`: This player maximizes a predefined heuristic value function at every turn.

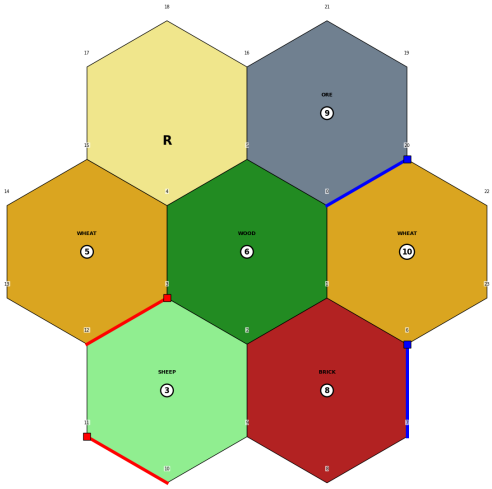

Figure 1: Sample miniature board used in experiments. Note that there are only seven hexes rather than the usual 19.

5.2.2 Prompted Players

We also prompt our own versions of Gemini-2.5-Flash and Qwen3-235B-A22B to test how different prompts affect players' performance. Both of these model architectures have "Thinking" capabilities, where the model is explicitly enhanced for chain-of-thought reasoning. We test models with and without this Thinking capability enabled. See the Appendix for the verbatim prompt for each player. We feature:

- `Basic`: This player is instructed to use chain-of-thought reasoning and then to return an action. This player has Thinking disabled.
- `Rules`: Just like `Basic`, this player is instructed to use chain-of-thought reasoning and then to return an action. However, after the same `Basic` prompt, we give the model access to the Settlers of Catan rule book to introduce information asymmetry between it and other models. This player has Thinking disabled.
- `Strategy`: Just like `Basic`, this player is instructed to use chain-of-thought reasoning and then to return an action. However, after the same `Basic` prompt, we give the model access to a Medium post that describes various Settlers of Catan strategies, again to introduce information asymmetry (Miller 2025). This player has Thinking disabled.
- `Search`: This player is instructed to use chain-of-thought reasoning but is explicitly told to have its reasoning simulate a tree search, where it considers the consequence of each possible action. This player has Thinking disabled.
- `Thinking`: This player is given the same prompt as `Basic`; however, it uses Gemini-2.5-Flash and Qwen3-235B-A22B with its thinking feature enabled.

5.3 Rating System

We adopt the rating system as described by Costarelli et al., 2024. Using the sample $S = \{m_1, m_2, ..., m_N\}$ of matches that we observe, we perform bootstrapping $B = 1000$ times, where we sample uniformly from S each time. Every bootstrapping yields a S_b^* resampled dataset for $b \leq B$. We then use maximum likelihood to fit to the exponential Bradley-Terry model:

$$P(i > j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}$$

with parameters $\theta_b = \{\beta_{Random}, \beta_{AlphaBeta}, ...\beta_{Thinking}\}$. We finally average over all bootstrapped samples to give a final rating.

In this way, we obtain a global ranking of all players across the observed sampled matches. We also employ a methodology that can easily be generalized to featuring more models, more games (other than Catan), and more matches, where not all players need to play against each other.

6 Results

We refer to Figures 2 and 3 for a presentation of our results. We also refer to the reasoning tokens outputted by Gemini-2.5-Flash models, located in the Github repository.

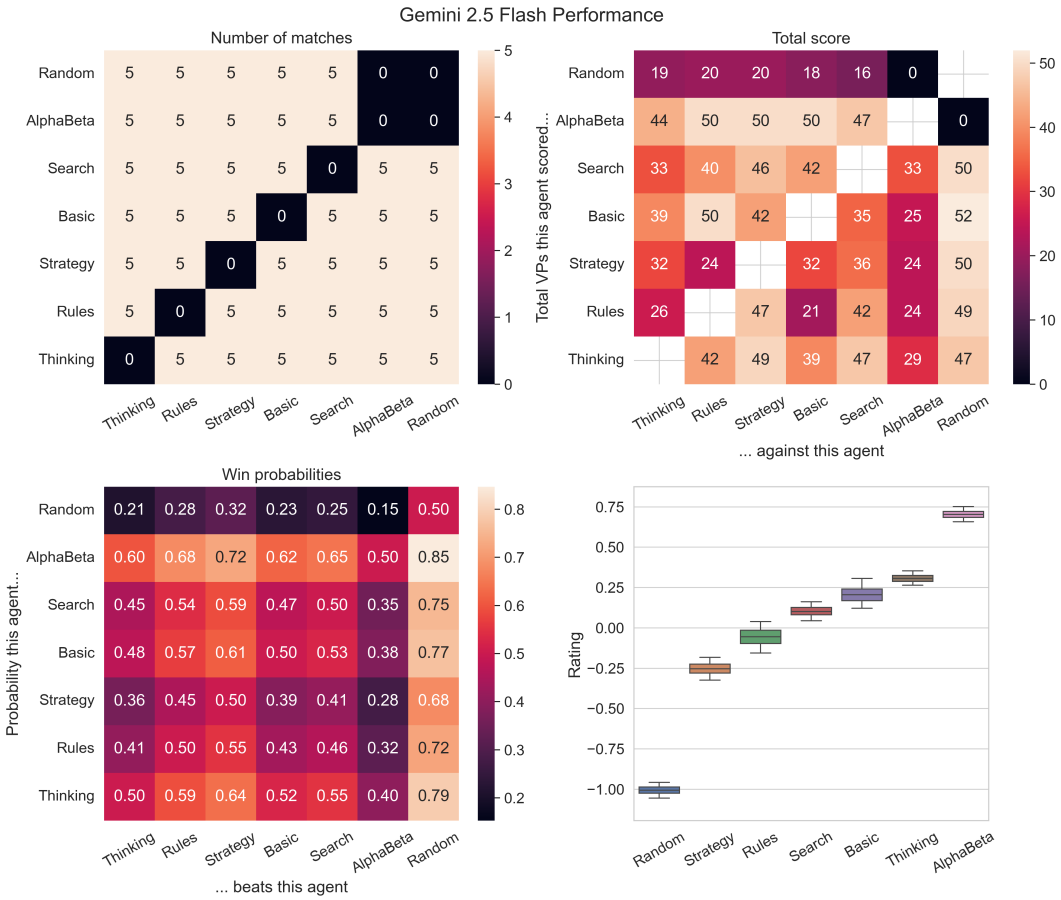

Figure 2: Results for Gemini-2.5-Flash. The top right panel shows the total number of victory points that each player scored against the other, aggregated over all five games that were played against each other. The bottom left panel shows extrapolated win probabilities, bootstrapped over many samples. The bottom right panel shows the rating calculations, implemented as described in Rating System.

We note that, due to computational limits, we only ran the Qwen3-235B-A22B model against the ValueFunction bot, which we selected because it was at least somewhat competitive with our players. Figures 2 and 3 should therefore not be compared against one another (as they are different models playing against different bot opponents); rather, results should be discussed within the context of their own trials. The combined rating calculation of all agents we studied can be found in the appendix.

6.1 Bot Performance against LLM Players

The AlphaBeta bot routinely outperformed the Gemini-2.5-Flash models, and the ValueFunction bot did the same with the Qwen3-235B-A22B models. These bots lack long-term planning strategies, and instead optimize for the best move at every turn as defined by some heuristic. The fact that the bots that optimize for short-term gain did better than the reasoning models, which, at least in theory, should have an advantage in long-term planning, indicates that the reasoning models may

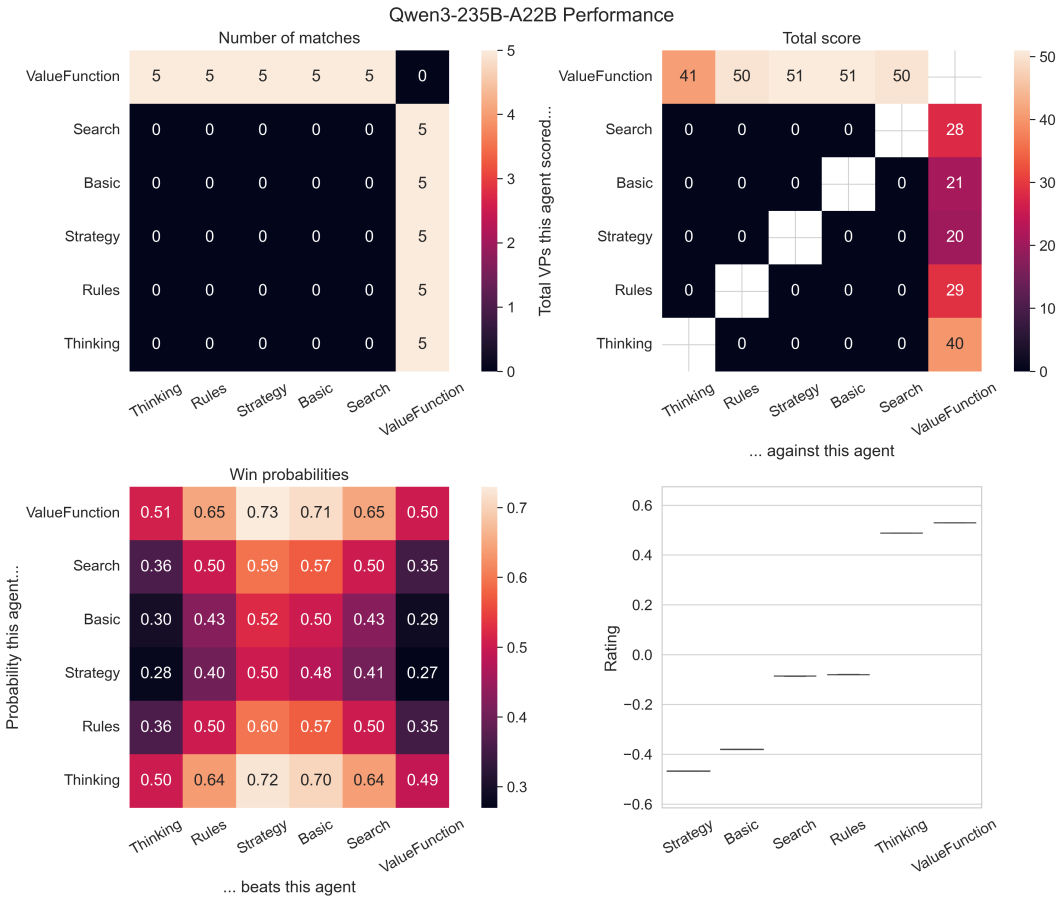

Figure 3: Results for Qwen3-235B-A22B. For the sake of computational efficiency, we only ran Qwen3 against the ValueFunction bot. The top right panel shows the total number of victory points that each player scored against ValueFunction, aggregated over all five games that were played against each other. The bottom left panel shows extrapolated win probabilities, bootstrapped over many samples. The bottom right panel shows the rating calculations, implemented as described in Rating System.

have failed to perform this long-term strategy. As a result, when put on even footing with a model that specifically optimizes for the short-term, the reasoning model underperforms.

We investigate this possible failure to perform long-term planning in Lack of Future-Planning.

6.2 Rules/Strategy vs. Basic

We had initially designed the experiment such that two players – `Rules` and `Strategy` – have access to information that the other players do not. We had initially hypothesized that these players would differentially perform better than their peers.

However, we note that `Rules` and `Strategy` were the two lowest performing models, with performances even worse than the `Basic` prompt which formed the beginning of their prompt. We then note that the addition of the rule book and the strategy guide actively harmed these models. Then, there are two possible interpretations about the results: either the models were unable to reason properly about the additional information that they were presented with, or that information itself was actively unhelpful.

We turn to the reasoning token outputs to offer some insight. We note that the `Strategy` model was the only one out of all non-thinking models (including `Basic`, `Search`, `Rules`, and `Strategy`) to focus on reducing its "hand size," as the `Strategy` model decided to "Trade 4 Ore for 1 Wheat to

reduce hand size" and to "Buy a development card to... reduce hand size." Given that the `Strategy` prompt included a comment discouraging staggered production, as "you are more likely to hold more than 7 cards... in your hand," it is possible that the model attempted to put into effect the suggestions that the prompt gave it. However, the model took this suggestion out of context: the listed strategies also included a clause about how it is better to not play some cards "even if [playing them] helps get you to 7 or less cards." We then note that the underperformance of `Rules` and `Strategy` may derive from misunderstanding the provided context.

Another possible reason for this underperformance is a fundamental misalignment between the presented information and what might be useful for the `Rules` and `Strategy` models in our experiment. The Catan rule book and the provided strategies are meant for a game played on a 19 hex board; we, in this study, use a 7 hex board. It is possible that the optimal strategies to employ are different when in a smaller board when compared to the full boards. While the decision to use a miniature board was intentional (using the full board resulted in games taking too long and expensive computation), it introduces the possibility that the asymmetrical information that the models are provided is actually actively detrimental rather than preferential. Similarly, many of the strategies presented in the Medium post are geared towards humans playing the board game while interacting at the same table (e.g. giving advice on trade negotiations, keeping cards close to the body, knowing the competition, etc.), which, of course, are unhelpful for the LLMs and may only serve to confuse it (Miller 2025). The fundamental misalignment between the in-person game that the information was prepared for and the online version of the game that the models played may have misled the models.

6.3 Thinking vs. Search vs. Basic

We then compare the models across reasoning types: the `Thinking` capability model, the tree `Search` reasoning, and the `Basic` reasoning. We note that the `Basic` and `Search` models did approximately equally well in both Figures 2 and 3 (we ignore the ratings in Figure 3 because it is so extrapolated, instead focusing on the total score metric which is based on objective data). In both cases, the `Thinking` model was the highest-performing model of all of them; in Figure 3, its performance nearly equaled that of ValueFunction. We conclude that the `Thinking` capabilities enhance models' performance. This finding is nontrivial. Our results show that default reasoning for hybrid models definitely works for generalizable instruction-following and reasoning in complicated and unseen domains. While these default reasoning models are in general benchmarked on mathematical reasoning or abilities to code, the success of these `Thinking` models to Catan indicate generalization to unseen domains.

While the `Basic` and `Search` models did approximately similarly in terms of hard numbers, the actual gameplay that they employed were very different. When tracing through the reasoning and analyzing the game boards, we note that the `Search` model was unconcerned with future-planning, as shown in Figure 4, where it traced out a hexagon around tile 4 rather than simply building a road between nodes 3 and 12. While the model recognized that it eventually might get Longest Route, this was merely incidental in its thinking, which was more focused on the desire to reach a better settlement location.

While we cover the models' failure to form long-term plans in Lack of Future-Planning, we here simply note that the `Basic` game board features much more strategic moves than does the `Search` game board, which indicates that the tree search reasoning structure that emphasizes finding the optimal move (which may have been interpreted in the context of a single turn) may be unsuited for a game where long-term strategy – i.e. planning across multiple turns – is critical.

6.4 Models' Spatial Failures & Hallucinations

When tracing through reasoning tokens, we find that the models were incredibly susceptible to spatial reasoning failures, where they erroneously believed that a node connected to resources that it, in fact, did not.

We collected the reasoning tokens for one game from each of the model types. We note that *every model* at least once demonstrated spatial reasoning failures of this kind. For example, in Figure 4 above, the `Search` model said it was "Building road 9-10 to set up a strong settlement at node 10 with Wheat, Ore, and Brick"; however, node 10 does not touch Ore or Brick (while they are close by, they are not touching). Given that every other model made a similar mistake, where it either placed a

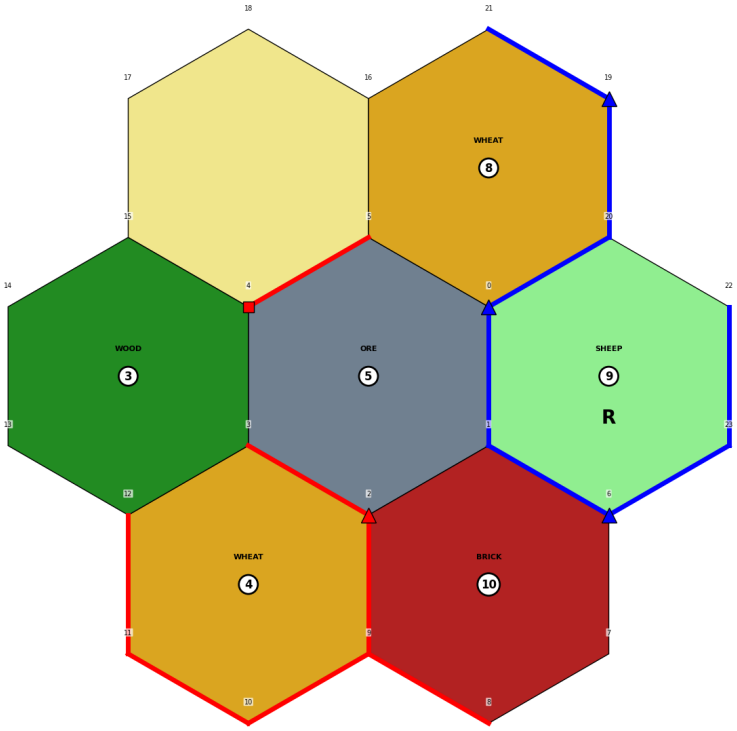

Figure 4: Sample game board for Search (RED). When tracing through the reasoning tokens, the model repeatedly emphasized its desire to get to a better settlement location, which is why it kept circling around tile 4; however, it failed to recognize that it could have simply placed a road from nodes 3 to 12. This decision may indicate that the model fails to develop a long-term strategy and instead greedily optimizes over single moves.

settlement on a node believing that it would gain access to resources that the node does not touch or placed roads with similar justifications, we note that language models may fundamentally be unable to fully comprehend a game that is inherently spatial. This inability to spatially reason, of course, harms models' performance in the game, as the decision to place a settlement or build a road wastes resources that could be used elsewhere and may be hard to replenish.

We therefore conclude that the most advanced models available currently display fundamental inabilities to spatially reason even when presented with explicit descriptions of the setting in a coordinate grid – a finding that has implications far beyond Catan.

We also note that the enhanced reasoning capabilities of some models may actually make this issue worse. We note that the `Thinking` model, which is endowed with superior reasoning capabilities, not only erroneously believed that nodes were touching where they were not; it actively hallucinated nonexistent squares. In Figure 5, when deciding to build a road "from node 15 to node 17," the model justified itself because "node 17 is an excellent settlement spot, providing access to Wheat (3) and two Sheep tiles (3 and 11)." However, we see that, while there is access to Wheat (3) in Figure 5, there are no sheep tiles attached to the node, and, what's more, there are no such Sheep (3 and 11) tiles anywhere on the board. We therefore conclude that enhanced reasoning models are not only susceptible to the same spatial reasoning problems, but their tendency to reason results in hallucination that may make this problem even worse.

6.5 Lack of Future-Planning

The underperformance of the reasoning models may also be attributed to their greedy bias towards action on a turn rather than employing long-term planning. Given that we did not explicitly instruct the models to use long-term planning, we argue that this suggests that reasoning models may be

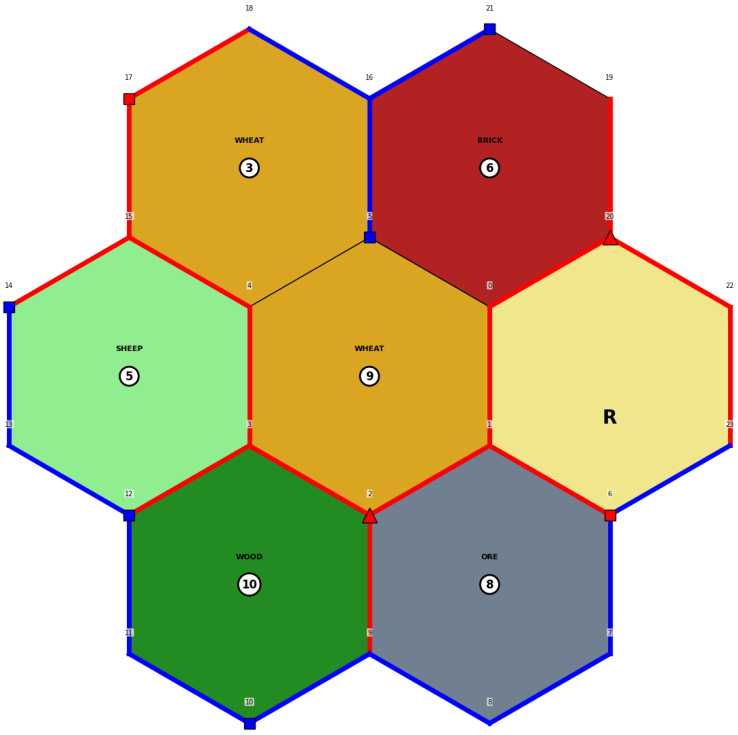

Figure 5: Sample game board for Thinking (RED).

incapable of recognizing when short-term gain should be subordinated to long-term benefits. We then argue that this demonstrates that models, unless explictly instructed, currently struggle to put together coherent long-term strategies when in a competitive environment.

6.5.1 Bias towards Action

We again refer to Figure 4. During the game, the `Search` model built a "road 9-10 to set up a strong settlement at node 10"; then, even though it repeated its goal of building a "settlement at node 10," it stated that "The decision to buy a development card is the most impactful action I can take right now, as it's the only one I can afford..." Rather than ending its turn and saving up for that settlement, the model chose to buy whatever it could with the resources in hand, which rendered it unable to buy that settlement later on. This bias towards action and refusal to save up – even when it can be beneficial – is seen across models.

We again refer to Figure 5. The `Thinking` model repeatedly built roads – even when it said in its reasoning tokens that it wanted to focus on building settlements. The model noted "Building a road is the only action possible with current resources" and, separately, that "I have the necessary resources (1 Wood, 1 Brick) to build a road." Even though that 1 Wood and 1 Brick could be used to eventually make a settlement (and the road was unnecessary), the model chose to spend whatever it had to build a road.

We also see this behavior when models begin to realize that they are losing. The `Rules` model, for instance, recognizes early in the game that it is losing to Blue. Rather than sticking with its strategy of building a city, the model says that "I need to buy a development card to try and get a VP or a Knight to disrupt Blue." For the next *fifty turns*, the model focuses on buying development cards, hoping to pick up a hidden VP. Rather than sticking to a strategy where waiting is inherent, the model opts for immediate action, gambling on the possibility of picking up a development card that *may have* a VP instead of waiting for the city that would be a *surefire* extra VP. The fact that this behavior is seen across models indicates that the models have a bias towards short-term action rather than saving for long-term benefit.

In fact, when reviewing the playable actions across all models, unless they were physically unable to make any moves, they rarely (if ever) forfeited the ability at least do *something* on their turn (including building something or playing a development card). This emergent behavior indicates that LLMs, when placed in these competitive situations without explicit instruction to strategize over the long-term, opt to optimize for short-term action rather than plan ahead, even though longer-term strategy is more beneficial. The fact that the models are unable to realize this – especially the `Strategy` model that is explicitly aware of the long-term planning inherent to the game – indicates that reasoning models currently cannot recognize when strategy is necessary, instead repeatedly choosing to take short-term action.

7 Discussion and Conclusion

We present CatanBench, where we pit Gemini-2.5-Flash and Qwen3-235B-A22B models with different prompts against each other as they play Settlers of Catan. While we intended to answer questions about information asymmetry in these models, we, in the process, uncovered emergent behaviors about reasoning models that are exposed only in a competitive, long-term planning environment. We first show that default reasoning in Gemini-2.5-Flash and Qwen3-235B-A22B do much better than chain-of-thought prompting and demonstrate generalization to unseen domains. However, our most important results are described in Models' Spatial Failures and Hallucinations and Lack of Future-Planning. First, we find that current reasoning models demonstrate gross inabilities to spatially reason about a setting (even when provided with explicit coordinate grids), and that models endowed with `Thinking` capabilities are perhaps even more susceptible to this phenomenon, where they even hallucinate about the surroundings. This finding has implications for LLMs' use in robotics and other domains where accurate spatial reasoning is inherent to success. Second, we find that models, if not explicitly instructed to, refuse to engage in long-term planning, instead opting for short-term action. This bias for action suggests that they may struggle to perform well in decision-making where delayed gratification is key, which hints at important safety concerns about models' abilities to suppress the instinct for short-term success in favor of abiding by their long-term safeguards.

We note several limitations of this work and propose future extensions:

7.1 Board State Representation

Obviously, natural language is not the original modality of board games, and thus we acknowledge that there are a few shortcomings to our methodology. The complexity of the board representation in natural language definitely impacted the performance of models against bot benchmarks. We observed multiple instances of the LLMs hallucinating characteristics of the board that did not exist, such as wrong resource production for a give node.

We note that images of the board more accurately represent the spatial relationships. Thus, it would definitely be a worthwhile endeavor to benchmark vision-language models on images of the board state (which we have already provided code for). We ourselves were unable to do so given the time and financial constraints.

7.2 Further Experimentation

Due to financial constraints, we were only able to run a small amount of matches of a limited number of matchup pairings. It is definitely worth exploring the impact of the prompt on the reasoning variants of Gemini and Qwen in addition to the non-thinking variants. 5 matches is also not enough to generate high-confidence results.

Another project idea stemming from the board-state representation issue that we thought of but did not have time to pursue was to compare the performance of an LLM fine-tuned (or with in-context learning) on a verified question-answer dataset about spatial relationships on the board, such as whether or not two nodes are adjacent or if two settlements are connected by any road, and plot the validation accuracy of the LLM on the the board QA dataset against the Bradley-Terry performance rating. Finally, having humans play these games against bots or LLMs may be informative, benchmarking all players against human intelligence, reasoning, and strategy.

8 References

Costarelli, A., Allen, M., Hauksson, R., Sodunke, G., Hariharan, S., Cheng, C., … Yadav, A. (2024). GameBench: Evaluating Strategic Reasoning Abilities of LLM Agents. arXiv [Cs.CL]. Retrieved from http://arxiv.org/abs/2406.06613

Duan, J., Zhang, R., Diffenderfer, J., Kailkhura, B., Sun, L., Stengel-Eskin, E., … Xu, K. (2024). GTBench: Uncovering the Strategic Reasoning Limitations of LLMs via Game-Theoretic Evaluations. arXiv [Cs.CL]. Retrieved from http://arxiv.org/abs/2402.12348

Abdelnabi, S., Gomaa, A., Sivaprasad, S., Schönherr, L., & Fritz, M. (2024). Cooperation, Competition, and Maliciousness: LLM-Stakeholders Interactive Negotiation. arXiv [Cs.CL]. Retrieved from http://arxiv.org/abs/2309.17234

Feng, X., Luo, Y., Wang, Z., Tang, H., Yang, M., Shao, K., … Wang, J. (2023). ChessGPT: Bridging Policy Learning and Language Modeling. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2306.09200

Miller, N. (2025, March 25). 17 advanced tactics to win more settlers of Catan. Medium. https://medium.com/@noahmiller400/17-advanced-tactics-to-win-more-settlers-of-catan-2d9280b7fa4b

Park, D., Kim, M., Choi, B., Kim, J., Lee, K., Lee, J., … Cho, J. (2025). Orak: A Foundational Benchmark for Training and Evaluating LLM Agents on Diverse Video Games. arXiv [Cs.AI]. Retrieved from http://arxiv.org/abs/2506.03610

(fair)†, M. F. A. I. R. D. T., Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., … Zijlstra, M. (2022). Human-level play in the game of Diplomacy by combining language models with strategic reasoning. Science, 378(6624), 1067–1074. doi:10.1126/science.ade9097

9 Appendix

9.1 Combined Ratings

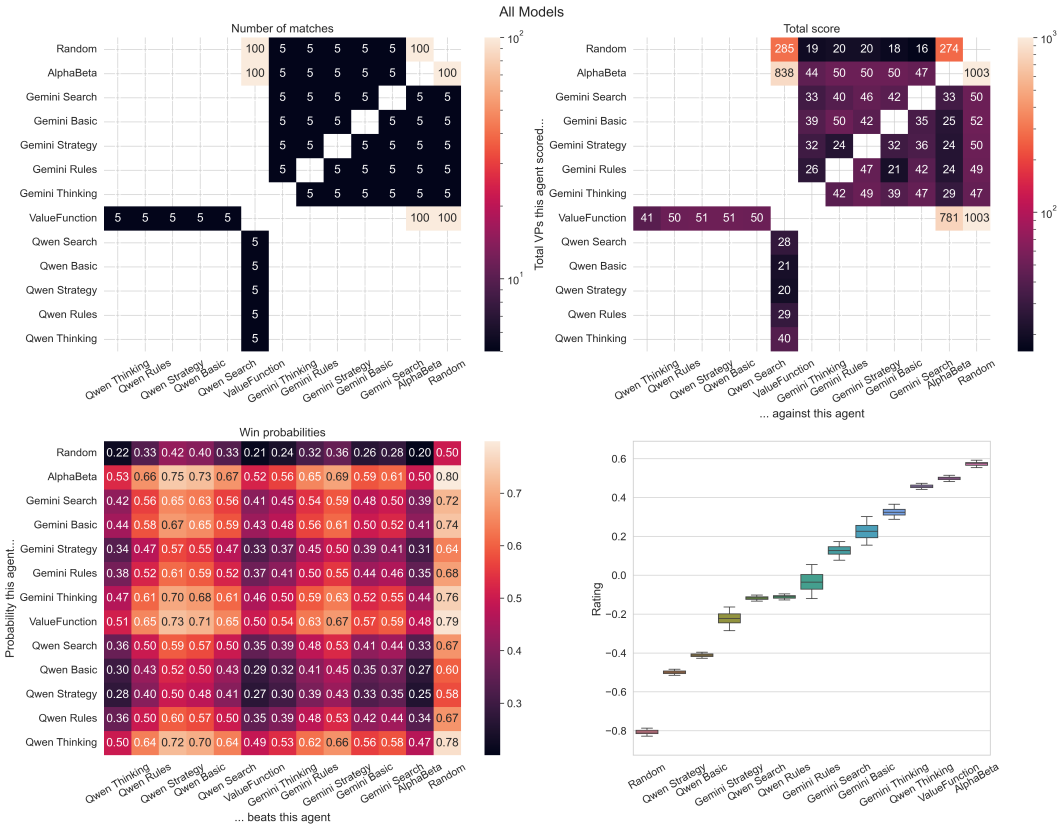

Figure 6: Results for all models combined. In order to rank the non-LLM players accurately relative to each other, we played 100 games between each of them. Matches between Qwen and Gemini models were not recorded; as such, their ratings are calculated according to their performance relative to ValueFunction and AlphaBeta, respectively. We therefore do not claim to conclude *anything* about the relative rating of Qwen models with respect to Gemini models.

9.2 Prompts for Each Player

As a note, all contents under `=== BOARD COORDINATE SYSTEM EXPLANATION ===` are identical across prompts, so they are not replicated after the `Basic` prompt below. Instead, `...` will refer to its contents.

9.2.1 Basic & Thinking

```
You are an expert Settlers of Catan player. Your goal is to win the game by
     reaching
10 victory points first. You should think about which action to take every
    single time.
Explain your reasoning, step by step. However, your output should always
    end with these
two lines:

A brief summary of the reasoning behind your move.
    You will use these as notes when making future moves.
    Output these notes in one line with no header.
```

```
The action number (The line should only be the number.)

Do not put a hyphen, bullet point, or any other formatting before these two
     final lines.

=== BOARD COORDINATE SYSTEM EXPLANATION ===
This game is played on a 7 tile board with no ports. The 7 tiles are
    arranged in a hexagon
around a central tile.

SPATIAL RELATIONSHIPS:
- Tiles: Identified by cube coordinates (x,y,z) where x+y+z=0
- Nodes: Intersection points between tiles, identified by integer IDs
    (0-53)
- Edges/Roads: Connections between two nodes, represented as (node_id1,
    node_id2)
- Each tile has 6 nodes at its corners and 6 edges connecting those nodes
- Adjacent tiles share nodes and edges
- Ports are located on specific edges facing the ocean

=== BOARD LAYOUT ===

--- HEXAGON TILES ---
Tile (-1, 0, 1) (ID 3): ORE with number 4 | corner nodes: [3, 4, 12, 13, 14,
     15]
Tile (-1, 1, 0) (ID 4): WOOD with number 10 | corner nodes: [4, 5, 15, 16,
    17, 18]
Tile (0, -1, 1) (ID 2): WHEAT with number 9 | corner nodes: [2, 3, 9, 10,
    11, 12]
Tile (0, 0, 0) (ID 0): WHEAT with number 8 | corner nodes: [0, 1, 2, 3, 4,
    5]
Tile (0, 1, -1) (ID 5): BRICK with number 6 | corner nodes: [0, 5, 16, 19,
    20, 21]
Tile (1, -1, 0) (ID 1): DESERT (no resources) | corner nodes: [1, 2, 6, 7,
    8, 9]
Tile (1, 0, -1) (ID 6): SHEEP with number 3 | corner nodes: [0, 1, 6, 20,
    22, 23]

--- PORTS ---
This map has no ports.

--- NODES (Intersection Points) ---
Node 0: neighbors: [1, 5, 20] | adjacent to tile IDs: 0, 5, 6
Node 1: neighbors: [0, 2, 6] | adjacent to tile IDs: 0, 1, 6
Node 2: neighbors: [1, 3, 9] | adjacent to tile IDs: 0, 1, 2
Node 3: neighbors: [2, 4, 12] | adjacent to tile IDs: 0, 2, 3
Node 4: neighbors: [3, 5, 15] | adjacent to tile IDs: 0, 3, 4
Node 5: neighbors: [0, 4, 16] | adjacent to tile IDs: 0, 4, 5
Node 6: neighbors: [1, 7, 23] | adjacent to tile IDs: 1, 6
Node 7: neighbors: [6, 8, 24] | adjacent to tile IDs: 1
Node 8: neighbors: [7, 9, 27] | adjacent to tile IDs: 1
Node 9: neighbors: [2, 8, 10] | adjacent to tile IDs: 1, 2
Node 10: neighbors: [9, 11, 29] | adjacent to tile IDs: 2
Node 11: neighbors: [10, 12, 32] | adjacent to tile IDs: 2
Node 12: neighbors: [3, 11, 13] | adjacent to tile IDs: 2, 3
Node 13: neighbors: [12, 14, 34] | adjacent to tile IDs: 3
Node 14: neighbors: [13, 15, 37] | adjacent to tile IDs: 3
Node 15: neighbors: [4, 14, 17] | adjacent to tile IDs: 3, 4
Node 16: neighbors: [5, 18, 21] | adjacent to tile IDs: 4, 5
```

```
Node 17: neighbors: [15, 18, 39] | adjacent to tile IDs: 4
Node 18: neighbors: [16, 17, 40] | adjacent to tile IDs: 4
Node 19: neighbors: [20, 21, 46] | adjacent to tile IDs: 5
Node 20: neighbors: [0, 19, 22] | adjacent to tile IDs: 5, 6
Node 21: neighbors: [16, 19, 43] | adjacent to tile IDs: 5
Node 22: neighbors: [20, 23, 49] | adjacent to tile IDs: 6
Node 23: neighbors: [6, 22, 52] | adjacent to tile IDs: 6
```

9.2.2 Rules

```
You are an expert Settlers of Catan player. Your goal is to win the game by
     reaching
10 victory points first. You should think about which action to take every
    single time.
Explain your reasoning, step by step. However, your output should always
    end with these
two lines:
A brief summary of the reasoning behind your move.
    You will use these as notes when making future moves.
    Output these notes in one line with no header.
The action number (The line should only be the number.)

Do not put a hyphen, bullet point, or any other formatting before these two
     final lines.

=== CATAN RULEBOOK ===

**THE GAME: CATAN**

**INTRODUCTION**
Welcome to the unexplored island of Catan!
As you and your fellow explorers gaze across the uninhabited landscape of
    hexagonal tracts
of terrain, you quickly realize that this unique land arrangement provides
    ample space for
each of you to put down roots.

You race off to establish your respective communities by gathering
    resources from nearby
terrain, building your infrastructure, and nurturing trade relationships.
    Watch out for
surprises-- your rivals can block your path or send the robber to plunder
    your wealth!

Good luck, explorers!

**OBJECTIVE**
The first player to reach 10 victory points (VPs) on their turn wins. Earn
    VPs by building.
Collect and trade for the resources you need to build.

* Settlement: 1 VP each
* City: 2 VPs each
* Largest Army and Longest Route: 2 VPs each
* Development card with a VP: 1 VP each

**CONTENTS**
* 6 sea frame pieces
```

```
* 19 terrain hexes (3x hills, 4x forests, 4x pastures, 4x fields, 3x
    mountains, 1x desert)
* 2 bonus victory point (VP) tiles: Longest Route and Largest Army
* 4 player aids.
* 18 number discs (2: 1x, 3: 2x, 4: 2x, 5: 2x, 6: 2x, 8: 2x, 9: 2x, 10: 2x,
     11: 2x, 12: 1x)
* 25 development cards (2x Monopoly, 2x Road Building, 2x Invention, 14x
    Knight, 5x Victory
Point)
* 95 resource cards (19x brick, 19x wood, 19x wool, 19x wheat, 19x ore)
* 2 card trays
* 2 dice
* 1 robber
* 20 settlements (5x each color)
* 16 cities (4x each color)
* 60 roads (15x each color)

Each player aid reads the following:
**PLAYER AID**
**TURN OVERVIEW**
**Production Phase**
* Play a development card (optional)
* Roll dice
* Collect resources or Resolve a 7

**Action Phase**
* In any order (optional):
* Trade
* Build roads, settlements, cities, and development cards
* Play a development card (if you did not play one before the dice roll)

**PLAY DEVELOPMENT CARDS**
* Not on the turn you bought it
* Only 1 development card per turn
* Before rolling dice or during the Action phase
* VP card exception: You may play multiple VP cards (even on the turn you
    buy them) in order
to win the game.

If this is your first time playing CATAN, we recommend using the setup
    shown here. For
additional variety, check out the Variable Setup described on page 11.

**1. Assemble the Frame**
Match the numbers at the puzzle piece ends of the sea frame pieces to
    assemble
the coast of Catan.

**2. Place the Hexes and Number Discs**
Place the terrain hexes and number discs inside the frame as shown.

**3. Create the Supply**
Sort the resource cards by type and place them in five faceup stacks in the
     card trays.
Shuffle the development cards and place them in a facedown stack in the
    empty card tray slot.
Place the Longest Route and Largest Army tiles near the game board.

**4. Place the Robber and Player Pieces**
```

```
* Place the robber on the desert hex
* Each player selects a color and takes the roads and buildings (
    settlements and cities) in
that color along with a player aid. In a 3-player game, do not use the
    white pieces.
* Place 2 starting settlements and roads for each player as shown.

**5. Collect Your Starting Resources**
Each player takes the resource cards from the supply that match the hexes
    adjacent to
their second settlement, highlighted in black. Keep these cards hidden in
    your hand.

**6. Choose the First Player**
Each player rolls the dice. The player with the highest roll is the first
    player.

**TURN OVERVIEW**
CATAN is played over a series of turns, starting with the first player, and
     moving clockwise
around the table. A turn consists of two phases, which occur in this order:
1. Production phase
2. Action phase
Once you have finished the Action phase, if you have not won, pass the dice
     to the player on
your left. They begin their turn with the Production phase.

**PRODUCTION PHASE**
**PLAY A DEVELOPMENT CARD**
If you have development cards, you may play one of them before rolling the
    dice. See
"Development Cards" on page 9.
**ROLL DICE**
Roll both dice and add them together. This total determines which hexes
    produce resources
this turn.
**COLLECT RESOURCES**
Hexes with a number disc matching the dice roll produce this turn. Each
    player with a
settlement on a producing hex receives 1 resource card of the corresponding
     type from the
supply.

If a player has 2 or 3 settlements on that hex, they receive 1 resource
    card for
each settlement.

Similarly, a player receives 2 resource cards for each of their cities on
    that hex.

If there are not enough resource cards in the supply to fulfill everyone's
    production, then
no one receives any of that resource. However, if only one player is
    affected, give that
player as many of those resource cards as remain in the supply.

* Forests produce wood
* Hills produce brick
* Pastures produce wool
```

```
* Mountains produce ore
* Fields produce wheat
* The desert produces nothing

**Example:**
The White player rolls an 8. Blue has a settlement on the mountain hex with
     the number 8 disc,
so Blue receives 1 ore resource card. Orange receives 2 ore for their city
    on that hex.

**RESOLVE A 7**
When you roll a 7, hexes do not produce any resources. Instead:
**1. DISCARD RESOURCES**
Each player who has more than 7 resource cards must choose half (rounded
    down) of their
resource cards and return them to the supply.

**Example:**
White has 9 resource cards in hand. When a 7 is rolled, they must discard 4
     cards.

**2. ACTIVATE THE ROBBER**
You must move the robber to a new hex.

Steal 1 random resource card from a player who has a building on that hex.
    If multiple players
have buildings on the new hex, you choose one player to rob.
**Important:**
A hex with the robber does not produce resources when its number is rolled.

**Example:**
Orange rolls a 7. They move the robber to the pasture hex with a red
    settlement and a blue city. Orange chooses to steal a card from Blue
    and takes a random card without looking.

**ACTION PHASE**
You may take actions in this phase as often as you like and in any order,
    as long as you have
the resources to do so.

**TRADE**
You may trade freely with other players and the supply to get the resources
     you need to build.
During your turn, other players may only trade with you, not with each
    other or with the supply.
There are three types of trades you may perform:

**TRADE WITH OTHER PLAYERS**
To trade with other players, announce which resource(s) you want and which
    resource(s) you are
willing to trade. Other players may accept your proposal, make
    counteroffers, or make their
own proposals.
**Important:**
You may not give away cards in any way, which includes trading matching
    resource cards
(for example, trying to trade 3 ore for 1 ore is not allowed).
**Example:**
```

```
It is Blues turn, and they want 1 wheat. Orange will trade 1 wheat and
    wants 1 wool in return. Blue agrees and the players exchange cards.
    White also wants to trade for 1 wool, but Blue no longer has any. White
     may not trade with Orange because it is not Orange's turn.

**GENERAL TRADE WITH THE SUPPLY (4:1)**
To trade with the supply, put 4 of the same resource cards into the supply
    and take 1 card of
a different resource from the supply.
**Example:**
It is Blue's turn. They choose to trade 4 wood for 1 ore.

**PORT TRADE WITH THE SUPPLY (3:1 OR 2:1)**
If you have a building on a 3:1 port, you may put 3 of the same resource
    cards into the supply
and take 1 card of a different resource from the supply.

If you have a building on a 2:1 port, you may put 2 of the resource cards
    shown on the port
into the supply and take 1 card of a different resource from the supply.

**Example:**
White wants 1 ore. On their turn, they may not trade 2 wheat for 1 ore
    because they do not
have a building on the 2:1 wheat port. However, they may trade 3 wood for 1
     ore because they
have a settlement built on a 3:1 port.

**BUILD**
Roads, buildings (settlements and cities), and development cards each have
    different costs.
During your turn, you may spend resources (aka "build") to place roads and
    buildings on the
board, as well as to take development cards from the supply. To do so,
    simply return the
required resource cards from your hand to the supply.

**ROADS**
Roads are worth 0 VP
Roads are placed on empty hex edges. A new road must connect to one of your
     existing roads
or buildings. You may not build a road starting on the other side of an
    opponent's building.
Building cost: 1 wood, 1 brick

**Longest Route:** the first player to have 5 continuous roads in play
    receives this tile. If
another player has more continuous roads in play, they immediately receive
    this tile. The
Longest Route tile is worth 2 VPs.
If, when a player's route is broken, they no longer meet the requirements
    for the Longest
Route, the tile is returned to the supply. It remains in the supply until a
     single player
has the longest continuous route of at least 5 roads. That layer
    immediately receives the bonus
tile and the 2 VPs.
**Example:**
```

```
Red has a continuous route of 7 roads and the Longest Route tile. White has
     a continuous
route of 6 roads. On Whites turn, they build a settlement at the
    intersection shown, breaking
Reds route into a four-segment route and a three-segment route. Since White
     now has the
longest route, they receive the tile and the 2 VPs.

**SETTLEMENTS**
Settlements are worth 1 VP.
Settlements are placed on empty intersections. A new settlement must follow
     the Distance Rule
and must connect to at least one of your existing roads. You have 5
    settlement pieces. To
continue building more settlements, you must first upgrade one to a city.
Building cost: 1 brick, 1 wood, 1 wool, 1 wheat

**The Distance Rule:**
When placing a settlement, stay at least two edges away from all other
    buildings!

**CITIES**
Cities are worth 2 VPs.

Cities always replace settlements. To build a city, remove one of your
    settlements from the
board, return it to your player area, and place your city where the
    settlement was located.
You have 4 cities and may not build more.
Building cost: 2 wheat, 3 ore

**DEVELOPMENT CARDS**
When you build a development card, draw the top card of the deck.
    Development cards stay hidden
until used. They do not count toward the number of cards in your hand when
    a 7 is rolled,
and the robber may not steal them.
You may play 1 development card during your turn by placing it face up in
    your player area.
It may not be a card you built this turn. You may play a development card
    before rolling
dice or at any time during the Action phase.
Development cards never go back into the supply. If the supply runs out,
    you may not
build any more cards.
You may not trade or give away development cards.
Building cost: 1 wheat, 1 wool, 1 ore

**Types of development cards:**
**INVENTION**
When you play this card, take any 2 resource cards from the supply. Add
    them to your hand.
They may be 2 of the same resource of 2 different resources.

**MONOPOLY**
When you play this card, announce one type of resource. Each player must
    give you all
their resource cards of that type. You may only request one type of
    resource when you play this
```

```
card, regardless of how many cards you receive.

**ROAD BUILDING**
When you play this card, build 2 roads at no cost (i.e. without spending
    any resource cards).

**VICTORY POINT**
Reveal all your Victory Point cards, including any built this turn, if you
    can reach the number
of VPs needed to win. Otherwise, you must keep Victory Point cards hidden
    in your player area.

**KNIGHT**
Move the robber to a new hex. Steal 1 random resource card from a player
    with a building on that
hex.

**Largest Army:** The first player to have 3 Knight cards in play receives
    this tile.
If another player has more Knight cards in play, they immediately receive
    this tile.
The Largest Army tile is worth 2 VPs.

**WINNING THE GAME**
If you have 10 or more VPs at any point during your turn, the game ends
    immediately and you are
the winner!
Turn over any number of Victory Point cards, including ones built this turn,
     to show that you
have reached 10 VPs.

Congratulations!
**Example:**
Blue ends the game on their turn with 10 VPs (2 cities, 3 settlements, the
    Longest Route tile,
and 1 Victory Point card). Victory!

=== BOARD COORDINATE SYSTEM EXPLANATION ===
...
```

9.2.3 Search

```
You are an expert Settlers of Catan player. Your goal is to win the game by
     reaching
10 victory points first. You should think about which action to take every
    single time.
Explain your reasoning, step by step.
In your reasoning, you should simulate a tree search on moves that you
    think will bring you
the most benefit by considering the consequences of each action.
However, your output should always end with these two lines:
A brief summary of the reasoning behind your move.
    You will use these as notes when making future moves.
    Output these notes in one line with no header.
The action number (The line should only be the number.)
Do not put a hyphen, bullet point, or any other formatting before these two
     final lines.
=== BOARD COORDINATE SYSTEM EXPLANATION ===
....
```

9.2.4 Strategy

```
You are an expert Settlers of Catan player. Your goal is to win the game by
     reaching
10 victory points first. You should think about which action to take every
    single time.
Explain your reasoning, step by step. However, your output should always
    end with these
two lines:

A brief summary of the reasoning behind your move.
    You will use these as notes when making future moves.
    Output these notes in one line with no header.
The action number (The line should only be the number.)

Do not put a hyphen, bullet point, or any other formatting before these two
     final lines.

=== ADVANCED STRATEGY GUIDE TO WIN MORE CATAN ===

Use the below guide when reasoning about your strategy.

Here are 17 TIPS AND TRICKS from the New York City Champion - so you earn
    more bragging
rights.

* *General Strategy*
* *Placements*
* *Trading*
* *Table Presence*
* *Miscellaneous Tactics*

Yes, there really is a competitive scene for Settlers of Catan and we all
    take it *way too
seriously*.

Performing data analysis on hundreds of matches, debating even the most
    insignificant of
moves, and recapping games like what you see in poker or chess  theres no
    end to what some
people like me will do to gain an edge.

Given youre here  I commend you! You and me are alike  we HATE losing in
    Settlers of Catan.
Other games I can come in last place and not care the next minute  this
    onenot so much.

If youre sick of being on the bad end of trades without even realizing it
    and find yourself
winning until its stripped away from you at the very last moment  youve
    come to the right
place.

Reading this guide will legitimately pay dividends for years to come.
    Whether you have family
game night later today or are playing a decade from now, I have no doubt a
    few pointers you are
```

```
about to learn will stick with you.

This post is designed to have each pointer as a stand-alone principle to
    keep in mind. Thus,
feel free to jump around and revert back to this post for some last-minute
    refreshers before
playing a game.

**Who am I you ask?** As someone who has been studying Catan for years and
    even won a
couple tournaments  I have learned a lot of tricks of the trade, inventing
    some myself.
For further background on my story, check out my first post here: my
    strange addiction:
settlers of catan.

Think of this guide as learnings from a thousand games culminating into one
     short list.
Inevitably, sharing my unique approach will take away some of my edge;
    however, candidly,
I want to grow on Medium more  so if you get any value from this, follow my
     Medium for
additional Catan content.

I know writing a piece as lengthy as this one will drop off the reading
    completion rate than
if I wrote something half the size, but ultimately I decided that theres a
    lot to say
and those whom I would really like to help improve their strategy will read
     through anyways.
This took a lot of effort, but I really enjoyed writing. There are a lot of
     gold nuggets in
here *even if you may not agree with everything*  I hope you enjoy!

Alright, lets get into the juicy stuff already. Keep in mind, this is an
    advanced guide so
not ideal for those who just learned the rules and excludes a lot of the
    most basic
strategies. *(If you still need that cheat sheet on what cards to build a
    city  this means
you)* On the flip side, some of these tips are not revolutionary for the
    experienced
player, but worth keeping top of mind.

Several of these principles, especially regarding *table presence* (more on
     that later),
probably dont matter against your 3 friends who are each 7 White Claws deep
     trying to watch
sports at the same time.

These rules of thought are most effective at a table of competent players
    who are dialed-in.
Most often, youll find this at tournaments rather than at any given house
    game with
little-to-no stakes.

Additionally, since some of these tactics are a bit cold-blooded *(doing
    whatever it takes to
```

```
win)*, dont expect everything included herein to keep your competition
    happy  user
discretion is advised. Even though Catan is just a game and you should be
    able to do whats best
to win, people do take moves personally. **Sometimes keeping people happy
    at the expense
of short-term wins builds goodwill that could pay off down the line.**

Anyways, lets get started

**1. Know your competition**

* This is what I would deem to be the most important field of thought
    throughout any match:
**knowing who youre playing against at any point in time**. Specifically,
    understanding
personalities, strategic prowess, and temperament. Each of these vectors
    are subject to change
throughout a game and can drastically shape your path to victory. Everyones
     headspace is
different whether theyre far in the leador not. It is your responsibility
    to recognize
this shift so you can capitalize accordingly. Since Catan is so
    interpersonal, **realizing how
each person may work *with* or *against* everyone else matters a lot.**
* Theres this inherent balancing act of how assertive vs. passive someone
    should be toward who
and when. In the game of life, overbearing people usually lose, but in
    Catan..this is not
always the case. Outspoken players may either influence others not to
    double-cross them or
can inadvertently paint a large target on their backs  it depends.
* Since Catan is so interpersonal, **how you go about proposing,
    negotiating, and rejecting
trades is critical.**

2. Bluffing

* This is the area of Catan that really separates the kids from the adults.
     It is openly framing
to the table that youre going to do something even when you have no plans
    to. *(Not to be
confused with going back on a deal)*

**Here are some examples:**

* Persuading someone to build their road a certain way claiming youre going
     to place a
settlement there  meanwhile you intend to keep it open as an expansion spot.

* Extorting someone to trade with you so that you wont knight them, even
    when you never
planned on putting the robber on them at all. *(note: this may have since
    been outlawed in the
official tournament Catan rules)*
* Declaring you have a specific development card (such as a knight) so that
     someone wont
do an action (like putting the robber on you).
```

```
* Negotiating a more attractive trade to help save the game. Ill explain:
    imagine you need to
steal the longest road to save the match. Even if you have the sufficient
    cards to make this
happen, you can declare that youre not going to build the roads unless you
    get an even
more-than-necessary deal from the table. (Of course, if the others arent
    aware of this  no
need to proclaim it, simply state that you need help) This declaration that
     youre willing to
let the game end, even when youre not, can be quite effective. I would
    recommend going
as far as pretending you are about to hand over the dice (even though you
    wont finish the
motion) to get someone else to fold first to your trade demands.
* Even when you really want a trade, its often helpful to sound indecisive
    of how much it
may benefit you. Coming across eager will often make the other person
    recoil their
willingness to follow through with it. It also may build some goodwill with
     the person if you
say the trade is helping them more than you and to keep it in mind when you
     are offering
a deal next time.
* Further to needing a specific card and acting indifferent, it may be best
     to offer the card(s)
you want to give away in exchange for anything. This can help obscure the
    move you would like
to make when you only accept the trade you want. If someone offers
    something useless to you, correct
yourself that you dont want that resource. If youre unsure of what cards
    are in
circulation as well, it will give you an idea of what people have (or more
    specifically,
what cards people dont want).
* The list goes on

3. Cut people off, but dont let it happen to you

* There are three main stages to any round of Catan: the early-, mid-, and
    end-game. Each one
is defined by its unique priorities, with the early-game focused on
    expansion (i.e. building
roads). Securing the settlement spots left unclaimed from initial
    placements should always be
your first priority (building cities is best to focus on later). The
    rationale is that you will
not only gain more production from *different* hexes, but also
    diversification of both
*resources and numbers* (*see Tip #3*).

**Cut people off**

* Even if you are two or three roads away from cutting someone off, dont
    let this deter you.
If you are fairly confident you may be able to build somewhere before
    someone else,
focus on the contested spot(s) first. You may derail someone elses game who
     positioned
```

```
their settlements assuming they could get it  a great offensive move by you.
     Meanwhile, you
are also gaining a spot that may be accretive to your production. Anywhere
    else that may be
less desirable could still be up for grabs later too.
* When cutting someone off, its
best to build all the required roads on *one single turn*. Otherwise, if
    you build one road
to start (even if it helps get you to 7 or less cards), the other player
    will see what youre
doing and prioritize getting there before you. Thus, you will have wasted
    your resources
building to nowhere (if there is no back-up route).

**Dont get cut off**

* Especially during the early stage of the game, AVOID getting cut off at
    all costs. Pay
close attention to anyone who threatens to steal the spot youre aiming for.
     Keep track
if they are accumulating sufficient lumber/brick, be cognizant of any
    trades that may
help them on this objective, and assume any unturned development card could
     be a year
of plenty, or even worse, a road builder. If there is a mutually beneficial
     trade possible,
you can sweeten the deal for yourself by having their promise not to build
    toward you.
*(While verbal agreements like this may not be binding per Catan rules,
    they are almost
always upheld)*
* Even while initial placements are being made, try to dissuade players
    from pointing
their roads in your direction. Even if its a spot that is unlikely for you
    to get anytime
soon, it helps to give you optionality later on. I will often proclaim that
     they are
building toward a high priority spot of mine that would be better for the
    both of us if we
did not have to compete for it.

4. Avoid being the clear leader

* **Early-to-mid game:** You will get knighted to oblivion and miss out on
    trades if you get
too much of a lead early. This is why those who often start strong arent
    able to maintain it.
If you have access to ample ore and grain, it may be more appropriate to
    buy development
cards than cities. I have found that development cards are often more
    subtle. It doesnt raise
alarm bells while bringing you closer to earning the largest army or
    gaining another
beneficial card.
* **End game:** If the largest army or longest road is still up for grabs
    and someone
else is not about to take either  delay getting it for yourself. For
    example, if you
```

```
have two knights face-up with at least one face down (assuming nobody else
    is in the
same boat), it may be worth holding out on flipping it. (The exception is
    that by not
flipping it, you may encourage someone else to buy development cards in
    order to
catch-up to you or be forced to play a knight at an inopportune time) Being
     close
to 10 VPs places more of a target on you. When possible, get those 2 VPs
    during your
move to win the game.

5. Diversify your resources *and* number production

* **Resource diversity:** Its* typically* better to have access to more (or
     all) of the
five resources. Thereby, you are less reliant on trading a 41 with the bank,
     21 with a
port, or forced into making a deal with someone.
* **Number diversity:** You
want to have your numbers as distributed as possible. For example, while
    the numbers 6
and 8 have equal probabilities of being rolled, it is better to have access
     to a 6
AND 8, rather than two 6s or two 8's.
* *Heres why:* your production is less concentrated on isolated numbers.
    Therefore,
if a needed roll of the dice doesnt come, the effect on you will be less
    drastic and
variable. With staggered production (aka accessing repeat numbers), you are
     more likely
to hold more than 7 cards or few in your hand at any point.

6. How to play your development cards

* **Road builder:** To maximize its impact, use this to cut someone off
    toward a
desirable spot or make an unexpected steal of longest road to win the game.
     Alternatively,
if your production of lumber/brick is minimal so youre having a tough time
    building
roads, it is equally fine to use road builder for acquiring an uncontested
    settlement spot.
* **Year of plenty:** Unless you need to steal the longest road or urgently
     need to acquire
a development card to hold the largest army, use this card for the
    resources you are
producing least. For example, if you have access to ore on your hexes, but
    need one or two
so you can build a city now  dont. Wait until those numbers are rolled
    organically
instead.
* **Monopoly:** This card increases in power as the game progresses. When
    production is
much higher since there are more cities in play, you will get more bang for
     your buck.
* **Knight:** Even without using it, you can threaten others that if the
    robber is
```

```
placed on you, you will use the knight against them as revenge (whether or
    not this
is a bluff  is up to you). Additionally, in times when your hand is stuck (
    you need
another card to make a move on your turn) and the robber is not on you, it
    may be fine to
use the knight simply for taking a card from someone.

7. Robbers can decide games

* New players often underestimate the power of the robber. It can
    drastically slow down
someones game, even if it may just *block one number* or *steal one card*
    at a time.
If its placed on you  okay, fine  but really avoid being stolen from as
    well. The robber
deserves respect and caution.

**Here are some pointers:**

* If theres an obvious placement that does not involve you and the person
    may not realize,
it may be best to bring it to their attention. Otherwise, its better to
    keep a low profile
to let them arrive at the decision and not direct attention to you.
* Direct your attention away from the game itself. This means avoiding eye
    contact with
the player robbing someone and potentially chatting up someone uninvolved.
    And yes, it
really can work. Someone who is indifferent will take a card from the
    person who is more
engaged in whats happening  its more convenient and potentially wont have
    to
interrupt your conversation.
* If the person does choose to take from you nonetheless, you can tell them
     that you
dont have the resource they need *(even naming said resource if you know
    what
they want)*. You can even ask them and then share their odds of actually
    getting it
sharing their better odds with the other person, if youre really paying
    attention.
* Keep your cards close to your body when someone is deciding who to pick
    from.
Remember that note about making it convenient for them? If they want to
    pick from you,
make them reach  sometimes people will actually change their minds during
    this moment.
* Most notably, always try to make a trade if it makes sense to either have
     someone
not knight you or at least take a card from you. Its often better to trade
    a card
*(unless its a resource you really need)* for something you dont want than
    to lose
one.
* When in control of the robber, you must also keep in mind the opposite
    perspective of
everything above. When you may be placing the robber, use these
    opportunities to extort
```

```
specific cards out of other players in exchange for you not knighting/
    taking from them.

8. Always speak up

* Remember: every trade you are not involved in is helping the games of
    others, not yours.
* If there is a trade that may be at least slightly one-sided, bring this
    light so
the deal does not occur. You want to highlight how much the exchange will
    help the aggressor
and only be a marginal benefit for the other. You can even offer a
    different deal
including yourself. While every situation is different, Catan does not
    typically
favor those who remain silent.

9. Think about win-scenarios

* One of the most common scenarios people lose is that they have tunnel
    vision on their
own win-scenarios (the path someone needs to win) while losing sight of
    everyone else.
* In most games, the winner will have the longest road or largest army
    therefore, if
youre too far from acquiring one of them, you need to try to secure the
    other. As you
think about which each person is poised to go after based on their resource
     production
(lumber/brick = road, ore/grain/wool = largest army), dont ever lose focus
    on what it
would take for them to secure it. And if they do, their proximity to
    winning.
* Its difficult for someone to sneak a win with largest army, but the
    longest road is
a different story. The fastest Catan matches usually end with the longest
    road ripe for
the taking at 5-roads or the one holding it having not insulated theirs to
    be long
enough.

10. The person with the most points isnt always winning

* Pay attention to who is most likely to win the game long term (i.e:
    strength of resource
production, open spots for future settlements, ability to build roads or
    buy development
cards). In matches with beginners, players will often incorrectly target
    someone who
may get an early longest road or otherwise get to 5 points quickly. However,
     slowing the
person with the best long-term set up and highest skill level early is
    often best. In
these instances, their games may accelerate beyond what can be stopped.

11. Yes, 11s and 3s roll

* Even at the highest level of Catan, I have found that some people do not
    pay much
```

```
attention to these numbers. They each have a **5.6% probability of rolling
    **
meaningful odds that should not be counted out. They may not be ideal for
    your
initial settlement placements, but could make for great expansion numbers.
    Not to
mention, they will likely never have the robber on them.

12. Table presence

* Your mannerisms at the table matter a lot and are contingent on the field
     of
thought that started this post: knowing your competition. There are times
    it may be better
to keep a low profile while at other moments it may best to be outspoken so
     nobody wants
to point the target on you.
* There are tons of nuances for how you should behave in a game based on
    the unique
circumstances at hand. For example, whether or not you should have urgency
    for
someone to accept a trade should vary based on who you may be playing
    against  people
react very differently.
* Furthermore, how you choose your words matters a lot during negotiations.
     If you
are desperate for a resource, do not make this clear to the other player
    use the word
*want* instead of *need*. If you show weakness, the person will capitalize
    and get
a better deal from you.
* When playing against strangers (with stakes while in a tournament), it
    often pays to
get on their good side. This includes *acting* remorseful when robbing from
     them and
*faking* empathy when their numbers dont roll. Half joking, but also half
    serious.
If you want to be next-level try hard, you can even make some off topic
    conversation
while playing, getting their heads off the game and for them to like you.
* My favorite example of this strategy was when I was able to portray
    whatever table
presence I wanted during the preliminary round of the NYC tournament I won
    in December
2022. Sitting with three strangers, it seemed most advantageous for me to
    act
timid and come across like I barely knew the rules. I kept asking for
    advice on moves
that I should make which helped to keep the target off me the entire game
    even when
I was winning. One guy (who ended up making the semi-finals) even openly
    said
that he felt bad knighting me because I was so nice. Ultimately, I ended up
winning and shared that I had played in a few tournaments before.

13. Card tracking

* **At the highest level of play, tracking is what separates the great
    players from
```

```
the best.** This is the act of keeping a tally of who has what resources.
    This is a
tremendous advantage because you are awake of what trades someone is
    willing to make
so that you can maximize every exchange. It also keeps you informed on what
     move someone
is focused on making next.
* While at first this may seem unrealistic, it becomes easier over time the
     more you
practice. As the game progresses and more trades are made, knowing exactly
    what cards
everyone has is impossible. However, a general idea of what someone has can
     be sufficient.
This is done by remembering the previous rolls that have occurred and what
    someone has
spent their resources on. If they must discard cards from having more than
    7, pay
attention to what they get rid of.
* When first learning to track cards, start with a reasonable goal of
    keeping
track of the last 5 rolls at any given point. Therefore, even if you dont
    remember what
each person has in their hand, you can back solve what cards each person
    has based on
where they are positioned on the board.

14. Involve yourself in others initial placements

* While many people will view the start of the game as independent (where
    there
shouldnt be any dialogue)  this could not be farther from the case.
    Sometimes there
are obvious settlement spots/road directions to start a game; however, most
     of the
time there is a level of subjectivity that you can take advantage of.
* If there are two spots of equal interest to someone, its worth guiding
    them
toward the one that interferes with you less. Additionally, you want
to mitigate situations when a player points their road toward a spot you
    both
may then have to compete for. There could be another spot or road placement
     genuinely
better for them that may not entail taking the same level of risk.

15. Keep a low profile

* When nearing the lead, the less unnecessary attention you can draw to
    yourself,
the better. While these points may seem dramatic, I assure you they can
    make a difference
against people who really care to win.
* If the banks resources are in stacks across the table from you, ask to
    split the
piles in half on each side or have everything moved closer to you.
    Constantly reaching
across the board for cards will accentuate how much youre picking up.
* If youre picking up 3 cards of the same resource, pick all 3 up at once
    instead of 1
at a time to not accentuate all the cards youre getting.
```

```
* When the robber is placed on you and someone is deciding who to pick a
    card from:
always keep your cards in one stack instead of fanned out so that youre not
     flaunting
how many cards you have. If you have more than a few, this may be the
    reason someone
may pick from you.

16. Picking up development cards

* Just like in poker, do not give a read on what development card youve
picked up. Instead, when you buy them, do not look and keep them in front
    of you
facedown. Only look at them at some later point in the cycle. Your
    expression, even
more than you realize, can signal what youve picked up and play a big role
    in
hurting your game.
* Naturally, the exception to this is when youre pulling for a VP to win it
     all.
* On the flip side, closely watch the expressions of opponents when they
    look
at their new development cards.

**If they ...**

* glance at the card quickly without any reaction, then place it down
    maybe a knight.
* have a look of distaste being early in the game  maybe a Victory Point.
* look around to see what cards everyone has  maybe a monopoly.
* pay closer attention to the board  maybe a road builder on where they may
     go.
* look back at their hand to see what they need  maybe a year of plenty.

17. Directing trades

* Cerberus paribus, its better to be on the side offering trades than
    accepting them. Unless
you know exactly whats in someones hand and what they will do or whom they
    may try to
trade with, a move may happen that youre prepared for.
* In Catan, you will regret accepting
someones trade much more often than you will be happy you did it. A
    willingness to walk
away from deals is the highest form of leverage.

=== BOARD COORDINATE SYSTEM EXPLANATION ===
...
```