

# 説明可能な麻雀 AI に向けた 2 段階 Dual-LoRA を用いたニューロシンボリック指示学習

Zhaohao Fang<sup>1</sup> 李吉屹<sup>2</sup> 吉岡真治<sup>2</sup>

<sup>1</sup> 北海道大学 大学院情報科学院 <sup>2</sup> 北海道大学 大学院情報科学研究院

fangzhaohao@elms.hokudai.ac.jp

{jyli,yoshioka}@ist.hokudai.ac.jp

## 概要

近年の麻雀 AI は超人的な強さを実現したが、その判断過程はブラックボックスであり教育的価値が低い。本研究では、麻雀の「何切る」問題に対し、判断根拠を言語化する説明可能なニューロシンボリック AI NeSy-Mahjong を提案する。本手法では、意思決定と推論を分離する 2 段階 Dual-LoRA と、客観的指標を注入するシンボリック計算機の導入により、LLM における幻覚と破滅的忘却を抑制した。天鳳の高段位データを用いた実験の結果、7B 規模のモデルで商用 LLM を凌駕する説明品質と決定精度を達成した。

## 1 はじめに

不完全情報ゲームの代表格である麻雀において、Suphx [1] をはじめ、NAGA [2] や MAKA [3] 等の近年の AI は人間を凌駕する性能を達成している。しかし、これらのシステムは主に行動の確率分布や価値を出力するブラックボックスとして動作するため、なぜその打牌が最適なのかという戦略的根拠を人間に教授することは困難である。特に、攻守のバランスが問われる「何切る」問題において、数理的根拠に基づく説明可能な AI (XAI) [4] が求められている。

大規模言語モデル (LLM) は流暢な説明生成において有望であるが、麻雀特有の厳密な計算 (シャンテン数や有効牌等) において事実と異なる幻覚 [5] を生じやすい。また、戦略的な意思決定能力と、その言語的な説明能力を単一のモデルで同時に学習させると、相互干渉により性能が低下する破滅的忘却が発生するという課題がある。

そこで本研究では、説明可能な麻雀 AI NeSy-Mahjong を提案する。提案手法は以下の 2 つの特徴

を有する。第一に、意思決定を担う直感 LoRA [6] と推論 LoRA を分離した 2 段階 Dual-LoRA を採用し、干渉を回避する。第二に、確定的なシンボリック計算機を導入し、客観的指標を注入することで、LLM の推論を数理的事実に接地させる [7]。

天鳳の鳳凰卓における高段位プレイヤーの対局データを用いた評価実験の結果、提案手法を適用した 7B 規模のオープンソースモデルは、決定精度および説明の質の両面において、DeepSeek-V3 や Qwen-Plus などの大規模商用モデルを上回る性能を示した。

## 2 提案手法

本研究では、不完全情報ゲームにおける意思決定の正確性と、その判断根拠を言語的に説明する能力を同時に満たすことを目的とし、ニューロシンボリック・フレームワーク NeSy-Mahjong を提案する。提案手法の全体構成を図 1 に示す。本節では、問題設定、シンボリック知識の導入、および学習アーキテクチャについて順に述べる。

### 2.1 問題定式化とシンボリック知識

「何切る」問題は、不完全情報下において観測可能な局面情報  $s$  に基づき、手牌から一枚の打牌  $a$  を選択する意思決定タスクである。この判断は、単に手牌効率を最大化する問題ではなく、将来的な放銃リスクや局面進行を考慮した総合的な戦略判断を必要とする。実際、上級者においても、同一局面に対して複数の打牌が合理的と見なされることが多い。

一方、LLM は自然言語による説明生成に優れる反面、麻雀特有の厳密な数理計算を内在的に正確に扱うことが難しい。特に、シャンテン数や有効牌枚数といった組合せの計算や、安全牌判定のような論理制約を伴う判断において、事実と異なる説明を生

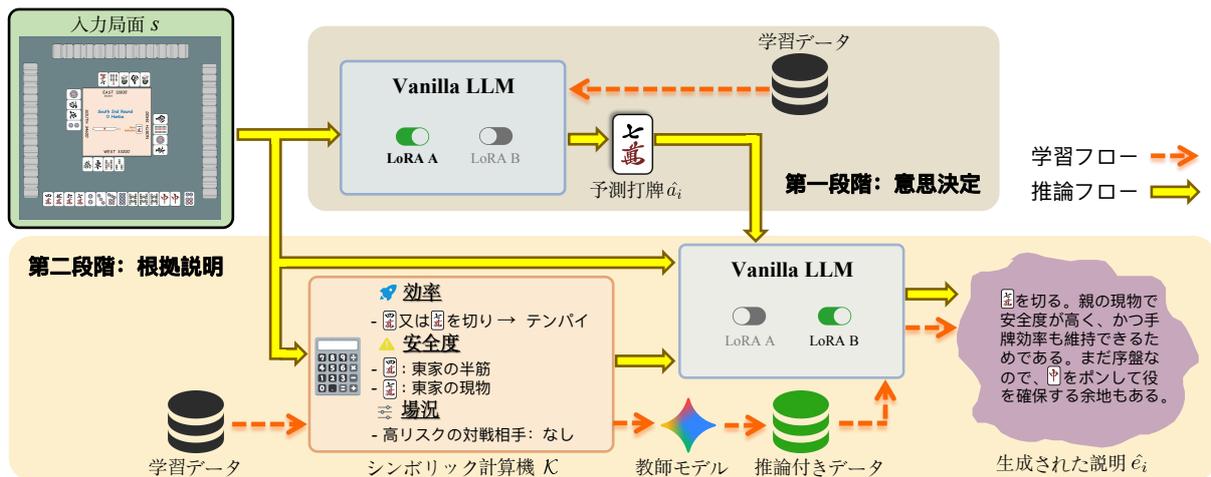


図1 NeSy-Mahjong のアーキテクチャ.

成する傾向がある。この性質は、説明可能な麻雀 AI を構築する上で大きな障害となる。

そこで本研究では、人間プレイヤーが実際に参照している判断材料をモデル化するため、確定的なアルゴリズムに基づくシンボリック計算機  $\mathcal{K}$  を導入する。計算機は局面  $s$  を入力とし、打牌判断に関わる客観的指標集合  $k = \mathcal{K}(s)$  を算出する。本研究では、これらの指標を以下の三つの観点に分類する。

**効率** 効率指標は、手牌が和了にどれだけ近づいているかを定量的に評価するものである。代表的な指標として、シャンテン数および有効牌の種類と残り枚数を用いる。これらは攻撃的判断の基盤となる情報であり、多くの人間プレイヤーが打牌選択時に最初に考慮する要素である。

**安全度** 効率の高い打牌であっても、放銃の危険性が高ければ合理的とは言えない。安全度指標は、各他家に対する放銃リスクを、公開情報のみに基づいて評価する。本研究では、現物に代表される論理的に安全が保証される牌に加え、スジ構造、壁、生牌といった経験的に用いられてきた守備指標を扱う。これらは絶対的な保証を与えるものではないが、実戦で広く共有されている判断基準である。

**場況** 最適な打牌は、局所的な手牌情報だけでなく、局面全体の進行状況にも依存する。例えば、他家の副露状況や巡目の深さによって、攻撃を優先すべきか、安全重視に切り替えるべきかが変化する。本研究では、他家の副露数、捨て牌推移、危険牌の出現などを統合したヒューリスティックな指標により、局面の緊張度を段階的に評価する。

これらの指標は、いずれも人間上級者が暗黙的に用いている判断材料を形式化したものであり、シン

ボリック計算機によって一貫した形で算出される。本研究では、これらの客観的指標  $k$  を LLM への入力として明示的に与えることで、生成される説明を数理的・論理的事実に基づいたものとし、判断根拠の一貫性を確保する。

## 2.2 ニューロシンボリック蒸留によるデータ構築

「何切る」問題に対する高品質な説明データは、対局ログや既存のデータセットから直接得ることが困難である。そのため、本研究では教師モデルを用いた蒸留により、説明文データを構築する。

具体的には、強力な推論能力を有する Gemini-2.0-Flash [8] を教師モデルとして用い、局面  $s$  に対する説明文  $e$  を生成させる。この際、単に局面のみを入力するのではなく、シンボリック計算機によって算出された指標  $k$  と、専門家が選択した正解打牌  $a$  を同時に与える。これにより、教師モデルは計算結果  $k$  に整合した形で、打牌  $a$  を正当化する説明を生成するよう制約される。

この手法により、数理的整合性を保った説明文からなる推論データセット  $\mathcal{D}_{\text{reason}} = \{(s, a, k, e)\}$  を構築することができる。

## 2.3 2 段階 Dual-LoRA アーキテクチャ

意思決定と説明生成は、必要とされる表現や最適化目標が異なる。これらを単一のモデルで同時に学習させると、相互干渉により性能低下が生じやすい。そこで本研究では、事前学習済みモデル  $\Theta$  を凍結した上で、二つの LoRA アダプタを段階的に学習する 2 段階 Dual-LoRA アーキテクチャを採用する。

**Stage 1: 意思決定 (LoRA-A)**

第一段階では、打牌選択に特化した意思決定アダプタ  $\Delta\theta_A$  を学習する。天鳳の高段位対局データ  $\mathcal{D}_{\text{decision}}$  を用い、局面  $s$  に対する正解打牌  $a$  の予測確率を最大化することで、人間上級者の直感的判断をモデルに獲得させる。この段階では説明生成を行わず、簡潔な行動出力に限定することで、幻覚の発生と推論遅延を抑制する。

$$\mathcal{L}_{\text{stage1}} = - \sum_{(s,a)} \log P(a | s; \Theta, \Delta\theta_A) \quad (1)$$

### Stage 2: 根拠説明 (LoRA-B)

第二段階では、第一段階で獲得した意思決定能力を保持したまま、説明生成に特化した推論アダプタ  $\Delta\theta_B$  を学習する。ここでは推論データセット  $\mathcal{D}_{\text{reason}}$  を用い、シンボリック指標  $k$  と打牌  $a$  を条件として説明文  $e$  を生成するよう最適化を行う。

$$\mathcal{L}_{\text{stage2}} = - \sum_{(s,a,k,e)} \log P(e | s, k, a; \Theta, \Delta\theta_B) \quad (2)$$

推論時には、まず LoRA-A を用いて打牌  $\hat{a}$  を決定し、続いて LoRA-B に切り替えた上で、シンボリック指標  $k$  と予測打牌  $\hat{a}$  を入力として説明  $\hat{e}$  を生成する。この段階的処理により、高い意思決定精度と数理的に一貫した説明生成を両立する。

## 3 評価実験

本節では、以下の 3 つのリサーチクエスチョン (RQ) を通じて提案手法を評価する。

- **RQ1 (決定効用)**: 提案手法は正確な打牌決定が可能か？
- **RQ2 (説明品質)**: 生成される説明は論理的かつ正確か？
- **RQ3 (消融研究)**: Dual-LoRA とシンボリック接地は各々機能しているか？

### 3.1 実験設定

**データセット** 実験には、オンライン麻雀プラットフォーム「天鳳」より公開されている対局ログを用いた。全ての対局は四人南赤ルールの鳳凰卓（全員七段以上）に限定し、各対局から上位 2 名の自発的な打牌行動（立直後を除外）を抽出して「何切る」問題として整形した。

データの偏りを防ぐため、ゲーム進行 (Early/Mid/Late) と戦略状況 (Offense/Defense) による 6 分割で層化抽出し、最小カテゴリに合わせてダウンサンプリングした。その結果、Stage 1 (意

思決定) 学習用に 44,034 件、検証用に 4,416 件、テスト用に 4,896 件を用意した。また、説明生成用の Stage 2 には、教師モデルによって生成された 5,544 件の推論付きデータを使用した。

**教師モデルによる推論データ生成** Stage 2 の学習データ構築には、Gemini-2.0-Flash を教師モデルとして用いた。各局面に対して、正解打牌とシンボリック計算機により算出された客観的指標を制約条件として与え、説明文を生成させた。この設計により、数理的事実と矛盾しない推論データセットを構築している。

**比較手法** 提案手法は、7B 規模の DeepSeek [9], Llama-3.1 [10], Qwen2.5 [11], Yi [12] をベースとして評価した。比較対象として、各 Vanilla 版に加え、Qwen-Plus [13], DeepSeek-V3 [14], GLM-4 [15], Moonshot-Kimi [16], Gemini-2.0-Flash [8] の商用 API モデルを用いた。

**評価指標** 意思決定の正確さは、テストデータにおけるプレイヤーの実際の打牌とモデルの予測 (貪欲法による Top-1) が一致した割合 (**Top-1 Accuracy**) で評価した。説明文の品質については、GPT-4o [17] を用いた LLM-as-a-Judge による自動評価を採用した。評価観点は以下の 3 点とし、それぞれ 1 から 10 のスコアで採点させた。

- **論理性 (Logic)**: 説明の文章構成や主張の一貫性。
- **正確性 (Precision)**: シャンテン数や安全牌（現物・スジ）の判定など、客観的な数理的事実との整合性。
- **戦術性 (Strategy)**: 麻雀のセオリーに照らし合わせ、その状況において妥当な判断根拠を述べているか。

### 3.2 RQ1：打牌決定精度

表 1 の左列に、テストデータにおける打牌決定精度を示す。

**ベースモデルの限界**: Vanilla モデルはいずれも Top-1 精度が 10% 未満であり、手牌に存在しない牌を出力するケースも多く確認された。これは、事前学習のみでは麻雀の複雑なルールや役の構造を理解できていないことを示している。

**商用モデルとの比較**: 商用 LLM は一定の改善を示すものの、最良の DeepSeek-V3 でも 32.1% に留まった。これに対し、提案手法は全てのバックボーンに

表 1 全体評価結果. 意思決定精度 (Top-1) と GPT-4o による説明品質の比較.

モデル	決定精度 Top-1 Acc.	説明品質		
		論理性	正確性	戦術性
ベースモデル (Vanilla)				
DeepSeek-LLM-7B	3.3%	2.09	1.46	1.44
Llama-3.1-8B	3.0%	2.26	1.38	1.40
Qwen2.5-7B	8.8%	2.93	2.14	2.21
Yi-1.5-6B	3.6%	2.37	1.61	1.60
商用 API モデル				
Qwen-Plus	22.5%	7.61	7.38	7.07
DeepSeek-V3	32.1%	7.78	7.44	7.35
GLM-4	25.4%	7.05	6.72	6.47
Moonshot-Kimi	24.1%	7.32	7.19	6.82
Gemini-2.0-Flash	19.9%	7.18	6.93	6.51
提案手法 (NeSy-Mahjong)				
NeSy-DeepSeek	65.4%	8.09	8.39	7.75
NeSy-Llama	66.5%	<b>8.12</b>	<b>8.48</b>	<b>7.82</b>
NeSy-Qwen	<b>67.0%</b>	8.08	8.43	7.81
NeSy-Yi	66.2%	8.01	8.33	7.78

において 65–67% の精度を達成し、商用モデルを大きく上回った。大規模な自己対戦学習を行った Suphx (76.7%) と比較しても、軽量な学習規模で高い競争力を示している。この結果は、汎用的な推論能力よりも、ドメイン特化の Instruction Tuning が「何切る」問題において支配的であることを示唆する。

### 3.3 RQ2：説明品質評価

表 1 の右列に、説明品質の評価結果を示す。

**商用モデルの幻覚:** 商用モデルは論理性 (Logic) において高いスコア (7.0–7.8) を記録し、流暢な言語生成能力を示した。しかし、正確性 (Precision) は相対的に低く、牌効率や安全度に関する具体的な数値計算において誤り (幻覚) が散見された。

**提案手法の優位性:** 一方、提案手法は全ての指標において 8.0 前後の高いスコアを達成した。特に正確性 (Precision) においては商用モデルを 1.0 ポイント以上引き離しており、これはシンボリック計算機から注入された客観的事実が、LLM の推論を強力に接地させている効果である。また、戦術性 (Strategy) の高さは、モデルが単に事実を述べるだけでなく、攻守のバランスを考慮した説得力のある説明を構成できていることを示している。

### 3.4 RQ3：消融研究とメカニズムの検証

提案手法の構成要素である「シンボリック接地」と「Dual-LoRA」の有効性を検証するため、NeSy-Llama をベースとした消融実験を行った。各バリエーションの比較結果を図 2 に示す。

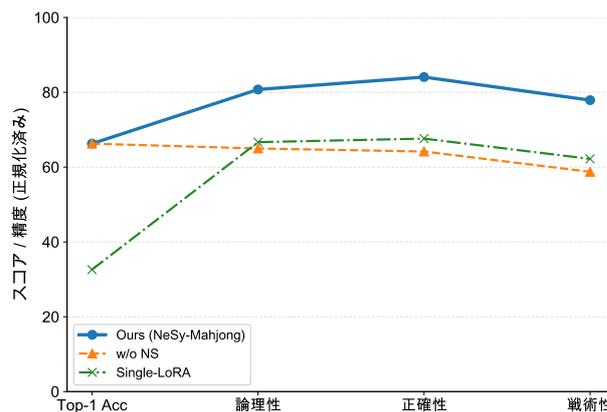


図 2 消融実験の結果. 4つのバックボーンモデルにおける平均性能を比較したものである。なお、説明品質のスコア (1–10) は、決定精度 (0–100%) との比較を容易にするため、10倍に正規化してプロットしている。

**シンボリック接地の効果** 図 2 のオレンジ色の折れ線 (w/o NS) は、推論時のシンボリック注入のみを無効化した設定である。意思決定を行う Stage 1 には変更がないため、Top-1 Accuracy は提案手法 (Ours) と一致する。しかし、正確性および戦術性のスコアが大幅に低下しており、LLM 単体ではシャッテン数や安全度の厳密な計算が困難であり、外部知識による制約が不可欠であることを示している。

**Dual-LoRA による忘却の回避** 緑色の折れ線 (Single-LoRA) は、意思決定と説明生成を単一のアダプタで同時学習させた設定である。グラフ左端が示す通り、Top-1 Accuracy が劇的に低下している (平均で約 20 ポイント以上の下落)。これは、戦略的最適化と説明生成の言語的最適化が同一パラメータ空間で競合し、破滅的忘却を引き起こしたことを示唆しており、Dual-LoRA とシンボリック計算機の導入が本タスクに不可欠であることが分かる。

## 4 おわりに

本研究では、麻雀の「何切る」問題において、正確な意思決定と論理的な説明を両立するニューロシンボリック AI NeSy-Mahjong を提案した。提案手法は、意思決定と推論を分離する 2 段階 Dual-LoRA アーキテクチャと、確定的なシンボリック知識の注入により、LLM の課題である幻覚と破滅的忘却を効果的に抑制した。評価実験の結果、7B 規模のオープンソースモデルにおいて、DeepSeek-V3 等の商用 LLM を凌駕する性能を確認した。今後の展望として、副露 (鳴き) を含む全体局への拡張や、マルチエージェント相互作用の分析が挙げられる。

## 謝辞

本研究の遂行にあたり、清華大学の Xu Junhuai 氏, Yu Jiawei 氏, Chen Shuotian 氏, ならびに西安電子科技大学の Li Hanjie 氏には、有益な助言と多大なる協力をいただいた。ここに記して謝意を表する。

## 参考文献

- [1] Junjie Li, Mykel J Kochenderfer, Yang Gao, Hangiu Liu, Xiaotian Liu, Cheng Wu, and Haobo Qin. Suphx: Mastering mahjong with deep reinforcement learning. **arXiv preprint arXiv:2003.13590**, 2020.
- [2] DWANGO Co., Ltd. Naga: Deep learning mahjong ai, 2019. Accessed: 2025-12-01.
- [3] Catfood Studio. Maka: Ai-assisted log analysis tool for mahjong soul (beta release), 2025. Official announcement; accessed 2025-12-01.
- [4] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. XAI: Explainable artificial intelligence. **Science Robotics**, 4(37), 2019.
- [5] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. **ACM Comput. Surv.**, 55(12), March 2023.
- [6] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In **ICLR 2022**, April 2022.
- [7] Timo Schick, Jane Dwivedi-Yu, Roberto Dessí, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: language models can teach themselves to use tools. In **Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23**, Red Hook, NY, USA, 2023. Curran Associates Inc.
- [8] Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, 2024. We utilize the Gemini-2.0-Flash version.
- [9] DeepSeek-AI. Deepseek llm: Scaling open-source language models with longtermism, 2024.
- [10] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, et al. The llama 3 herd of models, 2024.
- [11] Qwen. Qwen2.5 technical report, 2025.
- [12] 01. AI. Yi: Open foundation models by 01.ai, 2025.
- [13] Alibaba Cloud Intelligence. Qwen-Plus. <https://qwen-ai.chat/models/qwen-plus/>, 2025. Accessed: 2025-12-01.
- [14] DeepSeek-AI. Deepseek-v3 technical report, 2025.
- [15] 5 Team. Glm-4.5: Agentic, reasoning, and coding (arc) foundation models, 2025. We utilize the GLM-4.6 version.
- [16] Kimi Team. Kimi k2: Open agentic intelligence, 2025.
- [17] Hui Huang, Xingyuan Bu, Hongli Zhou, Yingqi Qu, Jing Liu, Muyun Yang, Bing Xu, and Tiejun Zhao. An empirical study of LLM-as-a-judge for LLM evaluation: Fine-tuned judge model is not a general substitute for GPT-4.

In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Findings of the Association for Computational Linguistics: ACL 2025**, pages 5880–5895, Vienna, Austria, July 2025. Association for Computational Linguistics.

## A 学習設定の詳細

Stage 1 (意思決定) と Stage 2 (根拠説明) における詳細な学習設定を表 2 に示す。Stage 2 では、ベースモデルおよび Stage 1 で学習した意思決定アダプタを凍結し、説明生成用のアダプタのみを更新することで、意思決定能力への干渉を防いでいる。

表 2 Stage 1 (意思決定) および Stage 2 (根拠説明) のハイパーパラメータ設定

設定項目	Stage 1: 意思決定	Stage 2: 根拠説明
ベースモデル精度	4-bit (NF4)	bfloat16
LoRA 手法	QLoRA	Standard LoRA
ランク ( $r$ ) / $\alpha$	16 / 32	32 / 64
ドロップアウト	0.05	0.05
対象モジュール	All Linear Layers <sup>†</sup>	
バッチサイズ	32	16
学習率 (LR)	$3 \times 10^{-4}$	$1 \times 10^{-4}$
スケジューラ	Cosine	Cosine
エポック数	3	3
最大系列長	1024	1024
最適化手法	AdamW	AdamW

<sup>†</sup>q, k, v, o, gate, up, down\_proj を含む全層。

## B 代表的な生成事例

提案手法が、ベースラインと比較していかに幻覚を防ぎ、戦術的に妥当な説明を生成できるかを示す。以下に、他家から立直が入った終盤の局面における各モデルの挙動を比較する。なお、本実験ではモデルの出力言語が中国語であったため、本稿ではそれらを日本語に翻訳して掲載している。

**局面概要 (東 2 局・親・19 巡目)** 上家 (北家) が立直しており、自分 (親) は一向聴だが、危険牌を引いて対応を迫られた場面である。

**正解打牌** 2s

**入力データ (プロンプト)**

[Game Details]

Round: East 2, 0 Honba, 1 Riichi Stick

Tiles Remaining: 19

Dealer (Oya): Player 1

Dora Indicator: 8p

[Your Status]

POV: Player 1 (East), Score: 25000

Hand: 4m 5m 5p 6p 9p 2s 4s 6s 7s S S N N 7m

Drawn Tile: 7m

[Table Status]

Player 0 (North, 22400, Riichi):

Discards: W E G Wh R N 9s 8p 6m 5s 6m 1s\* 1m

Player 1 (East, You, 25000):

Discards: 1s 1m 8p 3p W 5m 9s 4m 3p 8m 1s 9s

Player 2 (South, 26600, Melds: Pon Wh):

Discards: 1p 1p 1p 2s 4p E 3m 1m 6s G R 3m 2s

Player 3 (West, 25000):

Discards: 1s 2m E G R 1m 9p Wh W 2p 2s 0m 5s

### 各モデルの出力と比較

#### Qwen2.5-7B (Vanilla モデル) 打 5p

**理由:** 手牌に 5m が含まれているため、5p を切ることで「5m-5p-5s」の順子を構成しやすくなると判断している。また、5p は下家にとって有効牌である可能性があり、他家の手進行を妨げられると考えている。さらに、5p を切っても他家の順子や刻子を助けにくく、テンパイを防げるという認識に基づく選択である。

**分析: 深刻な幻覚.** 麻雀において「5m-5p-5s」という順子は成立しない。基本的な牌種・順子規則を誤って理解しており、ルールレベルでの致命的な知識欠如が見られる。

#### DeepSeek-V3 (商用モデル) 打 9p

**理由:** 手牌には 7m の対子があり雀頭として機能している。9p は周辺形を持たない孤立牌であり、保持価値が低い。また、8p (ドラ表示牌) がすでに場に見えているため、9p は比較的安全であると判断している。順子主体の役 (例: 三色同順, 一气通貫) を目指すことでテンパイ速度を高められ、親番である以上、和了速度を優先すべきだと結論づけている。

**分析: 危険な判断.** 残り巡目が少なく、他家から立直が入っている状況で、無スジかつ山に残っている可能性の高い 9p を切るのは非常にリスクが高い。攻撃効率を過度に重視し、安全性評価が不十分である。

#### NeSy-Llama (Ours) 打 2s

**理由:** 現在は終盤局面であり、上家から立直が入っている。下家も高打点手を進行している可能性が高く、複数人テンパイの確率が高い状況である。7m, 9p, N, S を切っても向聴数は維持できるが、2s は立直者に対するスジ牌であり、場にも多く見えているため比較的安全度が高い。多少の効率低下はあるものの、放銃回避を最優先すべき局面である。

**分析: 戦術的勝利.** シンボリックに計算された「場況 (高脅威)」および「安全性 (スジ・枯れ)」指標を正しく参照し、攻撃ではなくベタオリ (防御) を選択できている。局面理解と意思決定が人間の上級者水準に近い。