
Solver-Guided Reasoning for Mixed-Equilibrium Strategies

Anonymous Author(s)

Affiliation

Address

email

Abstract

Reasoning in large language models (LLMs) is often grounded in human text, human demonstrations, and human-generated rationales. For equilibrium reasoning in complex games, however, relying on human data can be suboptimal. In fact, humans play games mostly based on intuitions and heuristic reasoning, which deviates largely from the game equilibrium. This discrepancy is amplified in games with mixed-strategy equilibria, where human data is heavily biased toward pure strategies. Consequently, conditioning LLMs on this data yields weak game strategies. To grant LLMs the reasoning capacity in games, in this work, we study how to elicit equilibrium play using solver output. We propose Mixed-Strategy Decision Tree (MDT), which articulates the silent optimality of the equilibrium into sparse strategic rules that both humans and LLMs could understand. Using solver output rather than by human annotation allows us to extend the input to arbitrarily new states and continuations. We instantiate this study on No-Limit Texas Hold'em by querying a solver oracle for over **250 million mixed-strategy decisions**, where MDT together with other techniques **reduces the ℓ_1 distance to the equilibrium by 52.6%** across 8 different LLM configurations.

1 Introduction

Large language models are trained and evaluated with an extensive amount of human data, including demonstrations, solutions, rationales, and reasoning traces Wei et al. [2022], Achiam et al. [2023]. For reasoning in complex games, however, learning from human data is fundamentally limited. On the one hand, human play and commentary are selective. Many human strategies are exploitative, largely deviating from game-theoretically optimal equilibrium strategies. Even players considered “good” by human standards utilize strategies that are mostly effective against weaker humans but underperform against AI solvers Brown and Sandholm [2019], Silver et al. [2018, 2017]. On the other hand, human data is mostly presented as pure strategies. It is intuitive for humans and LLMs to treat decisions as prediction tasks, attempting to figure out which single action yields the best outcome. But for games with imperfect information, the equilibrium strategy is often mixed. It is difficult to find such mixed strategies in human data, making it equally difficult for LLMs to acquire them.

Recent evaluations of LLMs on No-Limit Texas Hold'em (NLH) poker make this gap concrete. The GTO Wizard Benchmark reports that the strongest evaluated model, GPT-5.3 Extra High reasoning, performs substantially below their approximate equilibrium strategy, at roughly -16 ± 3.0 bb/100 under their luck-adjusted evaluation [Provost et al., 2026]. LLMs also suffer from fundamental mistakes in card representation, hand-strength evaluation and strategy mistakes such as range-level mixing and strategic frequency allocation. Earlier, PokerBench has similar observations that LLMs fail to predict the highest-frequency actions on certain decision points even after task-specific fine-tuning [Zhuang et al., 2025]. These failures are not simply a lack of poker vocabulary. LLMs can

37 often discuss pot odds, blockers, equity, and bluffing, yet still miss the equilibrium logic that couples
38 those concepts across hidden states and action frequencies.

39 This fundamental gap stems from a mismatch in the underlying objectives. LLMs are optimized
40 for linguistic predictions and reasoning, whereas game solvers compute policies by minimizing
41 exploitability across complex, hidden-state game trees [Zinkevich et al., 2007, Moravčík et al., 2017,
42 Brown and Sandholm, 2018]. Consequently, an LLM can produce fluent, conceptually accurate
43 poker commentary while entirely failing to execute the precise frequency allocations required for
44 equilibrium. Relying on human-generated examples or LLM self-rationalization cannot bridge this
45 gap, as neither reliably captures full mixed-strategy policies [Lin et al., 2026]. Conversely, while AI
46 solvers naturally generate optimal mixed strategies, their outputs consist of raw numerical distributions
47 rather than generalized, textual reasoning. To endow LLMs with game-theoretic optimality, we must
48 translate these raw numbers into a linguistic format. We define this task as *solver articulation*:

49 *How can we extract articulate, verifiable reasoning from*
50 *the silent optimality of solver-generated data?*

51 We study this problem in imperfect-information extensive-form games, using No-Limit Texas
52 Hold’em (NLH) decision points where two players are left in the game. We introduce Mixed-
53 Strategy Decision Tree (MDT), which converts solver-implied decision logic into an inspectable
54 and readable form. MDT represents each decision point with solver-derived public-state and its
55 range-level and hand-level summaries. The decision process went through a sparse hierarchical
56 routing to assign probability mass to pure-action leaf prototypes. The hierarchy avoids forcing all
57 strategic interactions into a single dense mapping. In this way, coarse public and range conditions
58 select a local strategic regime, while sparse node-level summaries expose the hand-specific boundary
59 that changes the action mixture.

60 Motivated by counterfactual methods for game solving [Wachter et al., 2017, Agarwal et al., 2021],
61 we propose Scenario-Constrained Counterfactual Sampling (SCCS) that provides additional reasoning
62 paths on top of MDT. SCCS selects shadow hands that share the same public context but exhibit
63 clear solver-policy divergence, route to different MDT leaves, and differ along a salient summary. By
64 isolating such contrastive pairs, SCCS exposes the local boundary at which a hand changes its role
65 inside the mixed strategy. In this way, it converts an inspectable tree route into a transferable strategic
66 statement.

67 Our method is applied to two-player NLH postflop game play with over 250 million decision points
68 accessed. To ensure the quality of the articulation, we used one of the best available commercial
69 poker solvers in the world. The solver enjoys a Nash Distance less than 0.3% of the current pot and
70 gives solutions to arbitrary spots in the game tree. The stream of data therefore spans all 1,755 NLH
71 flops and their turn and river continuations. The obtained MDT is tested across 8 different LLM
72 configurations, where the sparse rules by the MDT are given to LLMs before they are asked to reason
73 the equilibrium strategy. The LLMs have their ℓ_1 distance to the solver target reduced from 0.211 to
74 0.100. The argmax-action agreement, defined by the highest-probability action in each distribution,
75 improves from 57.2% to 76.1%.

76 Beyond the immediate improvements in game-theoretic reasoning, our framework sheds light on
77 a fundamentally new regime in artificial intelligence: endowing LLMs with complex reasoning
78 capabilities entirely through synthetic, AI-generated data. This paradigm is especially appealing in
79 the current landscape, where the supply of high-quality human data is rapidly becoming depleted.
80 By demonstrating that the implicit optimality of an AI solver can be systematically extracted and
81 translated into readable linguistic rules, we provide a concrete pathway to bypass the human data
82 bottleneck. As envisioned by Silver and Sutton [2025], the future of artificial intelligence relies
83 not on mimicking human demonstrations, but on learning directly from ground-truth interactions
84 and solver-backed experience. **Our work represents an effort toward realizing this new era of**
85 **experiences.**

86 2 Related Work

87 **Solver-based game reasoning.** CFR, subgame solving, and self-play search have enabled strong
88 imperfect-information game agents, including superhuman poker systems [Zinkevich et al., 2007,
89 Brown and Sandholm, 2018, 2019, Moravčík et al., 2017, Brown et al., 2020]. These systems

90 compute mixed policies and values, but their outputs are primarily numerical prescriptions rather
 91 than communicable reasoning.

92 **Interpretable policy distillation.** Prior work distills learned policies into trees, programs, or
 93 concept-based representations [Bastani et al., 2018, Verma et al., 2018, Frosst and Hinton, 2017,
 94 McGrath et al., 2022]. Our setting differs because the target is a mixed equilibrium policy in an
 95 imperfect-information game, where local decisions depend on range-level coupling and action-
 96 frequency balance.

97 **LLMs and poker reasoning.** Recent poker benchmarks show that LLMs struggle with solver-
 98 level poker decisions despite fluent strategic language [Gupta, 2023, Zhuang et al., 2025, Provost
 99 et al., 2026]. Rather than training an LLM to play poker directly, we study how solver-derived
 100 mixed-strategy distinctions can be converted into contrastive rules that independent LLMs can use.
 101 Additional related work is discussed in Appendix A.

102 3 Preliminaries

103 **Imperfect-information extensive-form games.** No-Limit Texas Hold’em (NLH) is a zero-sum
 104 extensive-form game with imperfect information; Appendix D summarizes the rule structure and
 105 domain vocabulary used throughout. Postflop decisions occur after public community cards are
 106 revealed: the flop is the betting round after three public cards, and the turn is the round after the fourth
 107 public card. An extensive-form game is defined by $\mathcal{G} = (\mathcal{N}, \mathcal{H}, \mathcal{Z}, \mathcal{A}, P, u, \mathcal{I})$, where $\mathcal{N} = \{1, 2\}$
 108 is the player set, \mathcal{H} is the set of finite histories, $\mathcal{Z} \subset \mathcal{H}$ is the set of terminal histories, $\mathcal{A}(h)$ is
 109 the set of legal actions after a non-terminal history h , $P(h) \in \mathcal{N} \cup \{c\}$ specifies whether a player
 110 or chance acts at h , and $u_i(z)$ is player i ’s payoff at terminal history z . The game is zero-sum, so
 111 $u_1(z) + u_2(z) = 0$. In poker, a history contains public events such as betting actions and community
 112 cards, together with private cards dealt by chance.

113 Imperfect information is represented by information sets. For player i , \mathcal{I}_i partitions the decision
 114 histories at which i acts. Histories $h, h' \in I \in \mathcal{I}_i$ are indistinguishable to player i : they share the
 115 same public betting/card history and the same private hand for i , but may differ in the opponent’s
 116 private hand. A behavioral strategy is therefore a distribution over actions at each information
 117 set, $\pi_i(\cdot | I) \in \Delta(\mathcal{A}(I))$. Throughout the paper, h denotes a generic non-terminal decision history,
 118 while z is reserved for terminal histories.

119 **Nash equilibrium and GTO strategies.** A strategy profile $\pi = (\pi_1, \pi_2)$ induces an expected utility
 120 $u_i(\pi)$ by integrating terminal utilities over chance outcomes and both players’ randomized actions. A
 121 Nash equilibrium (NE) is a profile π^* such that no player can improve by unilateral deviation [Nash,
 122 1951]:

$$u_i(\pi_i^*, \pi_{-i}^*) \geq u_i(\pi_i', \pi_{-i}^*) \quad \forall i \in \mathcal{N}, \forall \pi_i'. \quad (1)$$

123 In poker terminology, a Game-Theoretic Optimal (GTO) strategy is an approximate NE strategy
 124 computed by a solver. Modern poker solvers are commonly based on regret minimization, search,
 125 public-belief reasoning, and subgame solving [Zinkevich et al., 2007, Bowling et al., 2015, Moravčík
 126 et al., 2017, Brown and Sandholm, 2018]. Solver outputs are commonly reported as mixed action
 127 distributions together with action values or counterfactual values. For a decision history h , $\pi^*(\cdot | h)$
 128 denotes the computed action distribution, and $Q^*(h, a)$ denotes the value associated with legal
 129 action $a \in \mathcal{A}(h)$. The notation $Q^*(I, a)$ is used when discussing the corresponding information-set
 130 formalism.

131 **Public belief states and ranges.** Solvers do not reason about a single fully observed state. Given
 132 a public state s consisting of the public board and betting sequence, there are many private-card
 133 assignments consistent with what has been observed. The public belief state (PBS) [Brown et al.,
 134 2020], also called a range representation in poker, is the conditional distribution over these private
 135 assignments: $\beta_s(c_1, c_2) = \Pr(c_1, c_2 | s)$, where c_i denotes player i ’s private hand. A player’s range
 136 is the corresponding marginal distribution over that player’s possible private hands. The solver’s
 137 equilibrium policy can be viewed as a high-dimensional mapping from this public belief state and
 138 a particular private hand to a mixed action distribution: $f^* : (s, c_i, \beta_s) \mapsto \pi_i^*(\cdot | I_i(s, c_i))$. This
 139 mapping is high-dimensional because the action distribution for one private hand can depend on the

140 public state, the player’s range, the opponent’s possible range, and continuation values induced by
 141 future play.

142 **Indifference and mixed strategies in Nash equilibrium.** Mixed strategies are often necessary in
 143 imperfect-information games because deterministic action patterns can reveal exploitable information.
 144 An observed action changes what an opponent can infer about hidden private states. If an action is
 145 associated too strongly with a narrow class of private states, the opponent may be able to respond
 146 profitably to that revealed structure. Equilibrium mixing helps prevent such profitable deviations by
 147 distributing probability mass across actions so that no player can improve unilaterally.

148 Indifference is a consequence of this equilibrium condition, not an independent assumption. If two
 149 actions are both used with positive probability at an information set, then neither can have strictly
 150 higher counterfactual value against the opponent’s equilibrium strategy; otherwise probability could
 151 be shifted toward the better action. Thus supported actions have equal value up to approximation error,
 152 while unsupported actions have no higher value [Nash, 1951, Osborne and Rubinstein, 1994]. In
 153 games with perfect recall, behavioral strategies represent a player’s randomized choices at information
 154 sets [Kuhn, 1953], and CFR-style algorithms approach equilibrium by minimizing counterfactual
 155 regret [Zinkevich et al., 2007]. Formally, for an information set I and any action a in the support of
 156 the equilibrium strategy,

$$Q_i^*(I, a) \approx V_i^*(I) \quad \text{for } a \text{ with } \pi_i^*(a \mid I) > 0, \quad (2)$$

157 with $Q_i^*(I, a) \leq V_i^*(I)$ for actions outside the support.

158 **Global dependence in imperfect-information games.** The preceding definitions still leave the
 159 central difficulty: in imperfect-information games, a local decision generally cannot be interpreted as
 160 an isolated choice at a fully observed state. The value of a local strategy can depend on the belief
 161 over hidden states and on constraints imposed by the full-game strategy, rather than only on the
 162 visible public state. This is a standard obstacle in imperfect-information subgame solving: unlike
 163 perfect-information games, an optimal strategy for a reached subgame may depend on strategies in
 164 other, unreached parts of the game, so the subgame cannot be solved independently of the full-game
 165 strategy [Brown and Sandholm, 2017].

166 4 Why Equilibrium Mixing in Poker Is Difficult

167 NLH is difficult for language models not only because the game tree is large, but because the strategic
 168 object to be learned is a *mixed equilibrium policy* over imperfect-information states. The relevant
 169 target is not a single best action for a visible hand. It is a range-level allocation of action frequencies
 170 that remains hard to exploit after the opponent updates beliefs from the observed betting line. This
 171 distinction is central to modern poker AI: superhuman systems such as DeepStack, Libratus, and
 172 Pluribus rely on equilibrium-oriented search, self-play, abstraction, and subgame reasoning in hidden-
 173 information games rather than on human explanations alone [Moravčík et al., 2017, Brown and
 174 Sandholm, 2018, 2019]. These results suggest that solver-generated equilibrium behavior is a more
 175 appropriate source of strategic targets than human verbal heuristics alone.

176 **Publicly available poker discourse provides language, not equilibrium logic.** Publicly available
 177 poker text is abundant but structurally mismatched to the object we need to learn. Forum posts,
 178 coaching examples, and hand histories are selective: they usually explain memorable or exploitative
 179 decisions, not the full support of a balanced range at an information set. This is consistent with recent
 180 LLM poker evaluations. PokerBench reports that strong pretrained LLMs substantially underperform
 181 on curated GTO decision spots, with GPT-4 reaching only 53.55% accuracy before task-specific
 182 fine-tuning [Zhuang et al., 2025]. Earlier work similarly finds that ChatGPT and GPT-4 can discuss
 183 starting-hand value, position, and GTO concepts while still failing to play game-theoretic optimal
 184 poker [Gupta, 2023]. This gap does not mean that verbal heuristics are useless: conservative advice
 185 such as calling rather than raising in marginal aggressive nodes may reduce immediate losses. The
 186 limitation is that such advice is local, whereas equilibrium requires solver-level frequency allocation
 187 across the whole range. Our own pilot experiments with scraped Two Plus Two (2+2) forum text
 188 and SFT on an open Qwen model point in the same direction: next-token imitation can teach poker
 189 vocabulary and local heuristics, but it does not reliably recover basic range-level decision logic. We
 190 therefore treat publicly available poker discourse as a useful source of terminology and surface-level

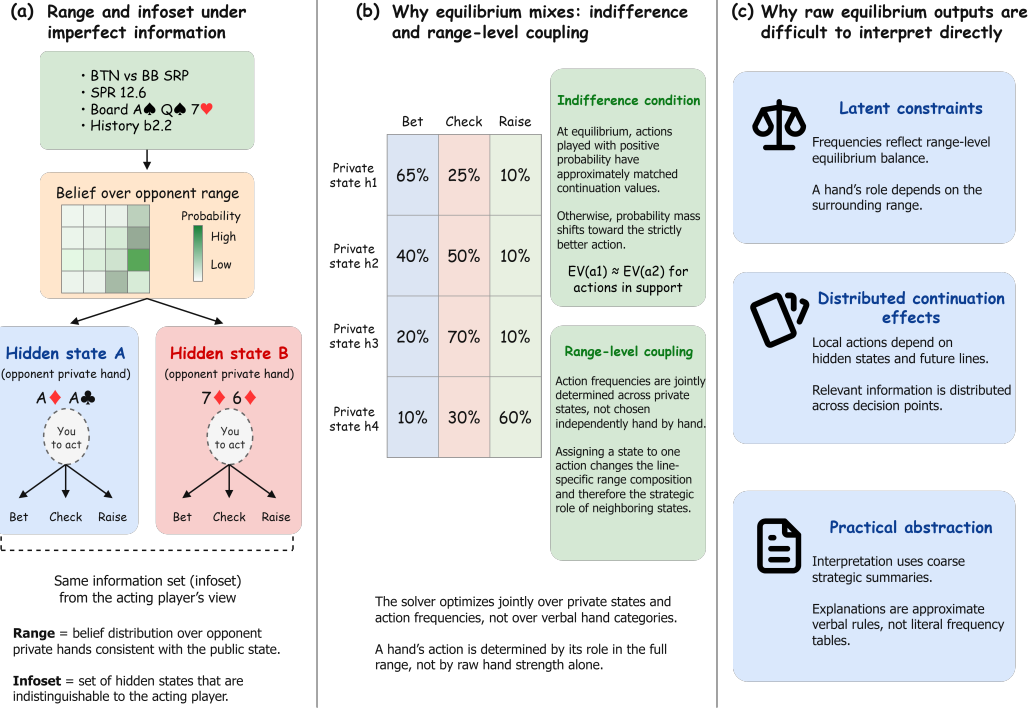


Figure 1: **Why equilibrium strategies in imperfect-information games are hard to interpret.** (a) The same public situation corresponds to many hidden private-card states inside one information set, so a visible action cannot be explained from public context alone. (b) Equilibrium mixing is constrained by indifference and belief-dependent continuation values, so action frequencies cannot be reduced to a single best action. (c) A solver table exposes the numerical policy and continuation values, but not the contrastive rule that makes strategically similar hands diverge; our goal is to compress this latent continuation logic into communicable strategic summaries.

191 heuristics, but not as ground-truth evidence for equilibrium decision logic. Appendix E provides
192 additional details.

193 **Equilibrium mixing is functional, not incidental.** In imperfect-information games, randomization
194 is not noise around an underlying pure decision. It is a mechanism for controlling information leakage.
195 A bet must often contain value hands, bluffs, protection hands, and blocker-driven candidates in
196 proportions that prevent profitable counter-strategies. Consequently, a hand may bet not because it is
197 locally strong, but because it occupies a necessary role inside the betting range; conversely, a stronger
198 hand may check because its showdown value realizes well and the betting line needs weaker bluff
199 candidates for balance. The decision boundary is therefore determined jointly by private cards, public
200 board texture, both players' ranges, blocker effects, and continuation values. This is exactly the kind
201 of coupled hidden-state reasoning that CFR-style and search-based poker solvers are designed to
202 approximate [Zinkevich et al., 2007, Brown and Sandholm, 2017, Brown et al., 2020].

203 **Solver outputs are precise but not communicable.** Solver outputs provide the desired equilibrium
204 target, but not in a directly communicable form. A solver table gives action frequencies and continua-
205 tion values for many private hands under a fixed public state; it does not state the compact contrastive
206 rule explaining why two similar hands diverge. An LLM can therefore produce a plausible one-hand
207 rationale while still missing the frequency allocation that makes the whole range balanced. The core
208 task in this paper is to bridge this gap: first distill solver behavior into a sparse strategic representation,
209 and then articulate local counterfactual distinctions that an independent LLM can transfer to unseen
210 hands. This motivates our use of MDT as a solver-grounded intermediate representation and SCCS
211 as the rule-extraction mechanism.

212 5 Methodology

213 Our goal is to achieve solver-guided articulation, which is a conditional prediction problem with an
 214 explicit intermediate representation. At a decision point h , let $\pi^*(\cdot | h)$ be the computed mixed action
 215 distribution, let $\{Q^*(h, a)\}_{a \in \mathcal{A}(h)}$ be the corresponding action-value summaries, and let \mathbf{x} denote a
 216 compact representation of public context and continuation summaries. The solver-side object is

$$\mathcal{O}(h) = (\pi^*(\cdot | h), \{Q^*(h, a)\}_{a \in \mathcal{A}(h)}, \mathbf{x}), \quad (3)$$

217 and the articulation procedure produces a rule $r = A(\mathcal{O}(h), \mathcal{D})$, where \mathcal{D} is a reference collection
 218 used to locate matched public contexts and policy-divergent comparisons. This collection can be
 219 expanded by querying additional solver states, for example by extending action continuations or
 220 adding private-hand assignments. The rule works as an intermediate representation supplied to a
 221 downstream predictor.

222 The rule is constrained to use a small set of quantities from \mathbf{x} and comparisons drawn from matched
 223 public contexts in \mathcal{D} . This rules out explanations that simply restate the full mixed policy table. It
 224 also distinguishes articulation from dominant-action labeling: the target remains the full distribution
 225 $\pi^*(\cdot | h)$, and the intermediate rule must preserve a local mixed-policy distinction rather than only
 226 identify the largest-probability action.

227 For evaluation, the target hand h_{test} is held out from the rule construction process. A predictor
 228 receives the public scenario, the target hand, and optionally the rule r , then outputs a distribution
 229 $\tilde{\pi}(\cdot | h_{\text{test}})$. The primary metric is the distance between $\tilde{\pi}$ and the masked solver target $\pi^*(\cdot | h_{\text{test}})$,
 230 compared against direct prompting and prompting with raw summaries alone. Under this formulation,
 231 a rule is useful only if it improves distributional prediction on an unseen target rather than copying a
 232 displayed policy.

233 5.1 Mixed-Strategy Decision Tree

234 The main technique is a sparse mixed-strategy distillation model, for which we call Mixed-strategy
 235 decision tree (MDT; Figure 2). For each solver-labeled decision point, we write the MDT input as
 236 \mathbf{x} . It contains public context (board, action history, position, stacks, and pot information) together
 237 with line-conditioned range-level and hand-level continuation summaries derived from solver outputs.
 238 These summaries include EV and EQ quantities and action-gap quantities under available lines. In
 239 this way, \mathbf{x} captures decision-relevant continuation information in a compact form for solver-policy
 240 articulation.

241 Given strategic-summary input \mathbf{x} , each leaf $l \in \mathcal{L}$ stores a pure action $a_l \in \mathcal{A}$. The router induces a
 242 probability mass over leaves,

$$\rho_l(\mathbf{x}) = \prod_{(n,c) \in \text{Path}(l)} p_{n,c}(\mathbf{x}), \quad (4)$$

243 and the distilled mixed policy is obtained by aggregating the mass of leaves assigned to each action:

$$\hat{\pi}(a | \mathbf{x}) = \sum_{l \in \mathcal{L}} \rho_l(\mathbf{x}) \mathbf{1}[a_l = a]. \quad (5)$$

244 Thus, MDT represents mixed strategies through probabilistic routing over pure-action leaves, rather
 245 than by storing a mixed action distribution inside each leaf. A leaf never carries a full action-
 246 frequency vector; it names one action prototype. The leaves remain intentionally simple, while the
 247 routing structure encodes when each pure strategic action should receive probability mass. The final
 248 model uses hard sparse local routers; implementation details and router ablations are provided in
 249 Appendix F.1.

250 We train the MDT against solver policy labels using a composite objective. Let $\pi^*(\cdot | h)$ denote the
 251 oracle action distribution and let $Q^*(h, a)$ denote solver-provided action values. We use $\mathcal{L}_{\text{task}} =$
 252 $\lambda_{\pi} \mathcal{L}_{L_1} + \lambda_{\text{ev}} \mathcal{L}_{\text{EV}}$, with

$$\mathcal{L}_{L_1} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} |\pi^*(a | h) - \hat{\pi}(a | h)|, \quad (6)$$

$$\mathcal{L}_{\text{EV}} = V^*(h) - \sum_{a \in \mathcal{A}} \hat{\pi}(a | h) Q^*(h, a), \quad V^*(h) = \sum_{a \in \mathcal{A}} \pi^*(a | h) Q^*(h, a). \quad (7)$$

Here \mathcal{L}_{L_1} measures fidelity to the oracle mixing frequencies, while \mathcal{L}_{EV} measures the oracle-conditioned EV gap under solver-provided action values. We treat this strictly as a local fidelity measure, not as a full-game exploitability estimate.

There are two optimization strategies. The first strategy keeps the router differentiable and adds sparsity-inducing regularization: $\mathcal{L}_{\text{soft}} = \mathcal{L}_{\text{task}} + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \mathcal{H}(\mathbf{p}) + \lambda_3 \mathcal{L}_{\text{ortho}}$. This regime generally achieves better fidelity, but in practice it leaves a long tail of weak summaries with non-zero influence. As a result, the visible top summaries do not fully explain the final decision. The second strategy progressively converts the router into a strict top- K sparse structure. We use a teacher-assisted curriculum: a dense teacher first absorbs the raw mapping, then a structured student inherits topology and is finally locked into hard summary selection with $K = 5$ active local summaries per node. The goal is not maximal fidelity, but explicitness: the displayed sparse summaries are the variables used to compute the router probabilities. We compare both approaches in the experiments.

5.2 Scenario-Constrained Counterfactual Sampling

The MDT exposes which sparse summaries are used along a route, but a routed path alone is not yet a communicable rule. A path can indicate that a hand depends on draw strength, kicker quality, hand EV, or range-relative equity, but it does not identify which local change would move the hand across a strategic boundary. This distinction matters in poker because strategically nearby hands can share the same public context and similar raw equity while occupying different roles inside the equilibrium range.

SCCS addresses this gap by explaining a target hand contrastively. Instead of describing the hand in isolation, it selects a shadow hand from a reference collection of solver queries in the same public scenario whose solver policy diverges and whose MDT route crosses a critical branching decision. For example, under the same board and betting line, a weak-kicker draw may be used as a semi-bluff, while a higher-showdown-value version may check because it realizes enough equity without building the pot. By fixing the public scenario and varying only the private hand, SCCS isolates the strategic quantity that changes the hand’s role.

For a target state h , SCCS identifies its trained-tree route, finds policy-divergent shadow hands under the same public context, localizes the routing boundary where the target and shadow diverge, and converts the active-summary contrast into a natural-language rule. The procedure is targeted at policy-divergent samples rather than nearest visual or lexical neighbors; details are given in Algorithm 1 and Appendix F.2. The resulting rule also defines a communicability test: an independent reasoner should be able to apply the extracted local distinction to an unseen target state with matched public context.

6 Experiments and Results

Our evaluation aims to answer three questions. First, can the proposed MDT articulate the solver output to improve LLM reasoning? Second, do the rules extracted by SCCS help independent LLMs use solver-derived distinctions on unseen target hands in matched public contexts? Third, what kind of strategic distinction does the rule expose in an individual case?

6.1 Solver Oracle Interface and Evaluation Samples

Our experiments use over 250 million solver-labeled postflop decision samples obtained by querying a commercial, high-end NLH solver oracle, including mixed policies and continuation-value quantities. These samples comprise approximately 16M flop decisions and 235M turn decisions under a 6-player 100BB no-rake NLH configuration. Decision points involve postflop spots where two players are left in the game. We include five preflop configurations: SB vs BB single-raised pot and 3-bet pot, BTN vs BB single-raised pot and 3-bet pot, and BTN vs SB 3-bet pot. The same pipeline can continue to generate additional labeled states for new board textures, action branches, private-hand assignments, and configuration choices.

To ensure coverage of strategically distinct public states, flop data is sampled from 1,755 strategic board textures. For each board we record multiple canonical action nodes, including the root, check line, bet line, bet-call line, and bet-bet line. Turn data is generated by extending representative flop branches such as check-check, bet-call, and check-bet-call, then sampling five turn cards per

Table 1: **Training loss and distillation fidelity on the NLH solver-labeled evaluation set.** Lower L_1 and oracle-conditioned EV gap are better. Oracle EV Gap is a local fidelity metric under solver-provided action values, not full-game exploitability. The final hard MDT is the model used for communicable rule extraction.

REPRESENTATION	MODEL	L_1 LOSS	ORACLE EV GAP, % OF POT	ARCHITECTURE / APPROX. PARAMS	VISIBILITY
-	<i>Oracle</i>	-	<i>0.04%</i>	-	<i>N/A</i>
RAW PBS	2-LAYER MLP	0.068 ± 0.002	$0.55 \pm 0.04\%$	512 HIDDEN; $\sim 1.02\text{M}$	BLACK-BOX POLICY
	8-LAYER RESNET	0.036 ± 0.001	$0.18 \pm 0.02\%$	512 HIDDEN; $\sim 2.60\text{M}$	BLACK-BOX POLICY
STRATEGIC SUMMARIES	1-LAYER MLP [†]	0.098 ± 0.010	$0.72 \pm 0.05\%$	256 HIDDEN; $\sim 41\text{K}$	DENSE SUMMARY MODEL
	2-LAYER MLP [†]	0.053 ± 0.005	$0.37 \pm 0.03\%$	256 HIDDEN; $\sim 107\text{K}$	DENSE SUMMARY MODEL
	4-LAYER MLP [†]	0.045 ± 0.005	$0.26 \pm 0.03\%$	256 HIDDEN; $\sim 238\text{K}$	DENSE SUMMARY MODEL
	8-LAYER RESNET	0.021 ± 0.003	$0.05 \pm 0.01\%$	256 HIDDEN; $\sim 502\text{K}$	DENSE SUMMARY MODEL
TREE STUDENTS (ABLATIONS)	DENSE-ROUTER TREE [†]	0.031 ± 0.004	$0.12 \pm 0.03\%$	$\sim 19\text{K}$	TREE, DENSE ROUTING
	SOFT-SPARSE TREE [†]	0.057 ± 0.005	$0.21 \pm 0.02\%$	SPARSITY REG.; $\sim 19\text{K}$	TREE, REGULARIZED ROUTING
HARD CURRICULUM	TEACHER 1 (GLOBAL SPARSE)	0.028 ± 0.002	$0.07 \pm 0.01\%$	8-LAYER RESNET; $\sim 502\text{K}$	GLOBAL SUMMARY GATE
	TEACHER 2 (LOCAL SPARSE)	0.044 ± 0.005	$0.23 \pm 0.03\%$	2-LAYER MLP ROUTERS; $\sim 1.64\text{M}$	LOCAL TREE ROUTING
	FINAL HARD MDT [†]	0.087 ± 0.007	$0.44 \pm 0.04\%$	TOP-K/NODE MASK; $\sim 0.7\text{K}$	

branch. Bet sizes follow the solver’s abstraction; all-in actions are folded into the generic “bet” action category.

For communicability, we construct matched-context tests with unseen target hands. Each test fixes the public context and asks an independent LLM to predict the solver-equilibrium mixed strategy for an unseen target hand. We compare three prompting conditions: **Direct**, which provides only the public state and hand; **Direct+Summaries**, which additionally provides strategic summary quantities; and **SCCS Rule**, which provides an abbreviated contrastive trace extracted from policy-divergent shadow hands in the trained MDT. To prevent the SCCS rule from serving as a near-label lookup, the unseen target hand is never included in the SCCS trace, and its solver policy must differ from the policy of every rule-displayed hand by at least 0.20 under the same action-averaged L_1 metric. Thus, the rule cannot be applied by copying a displayed strategy from a near-duplicate hand; improvement requires transferring the extracted strategic distinction to the held-out target. We report L_1 both to the solver target and to the distilled MDT policy. Additional construction details, including the SCCS matching criteria and prompt format, are reported in the appendix I.

6.2 Training Loss and Distillation Fidelity

The distillation results motivate MDT as an articulation layer rather than only a predictor. Within the same strategic-summary input, tree-structured routing fits the solver policy substantially better than flat MLPs at comparable or smaller parameter counts in our experiments, suggesting that mixed-equilibrium decisions benefit from a hierarchical representation: coarse public-state and range-level conditions first select a strategic region, while hand-level summaries define local action-frequency boundaries. The hard sparse MDT sacrifices some fidelity by allowing only a small set of summaries at each node, but this constraint makes the routing boundaries explicit enough for SCCS rule extraction. Full training-loss comparisons and EV-gap analysis are reported in Appendix B.

6.3 Communicability on Unseen Target Hands

Table 2 is the main quantitative result. Across eight LLM configurations, SCCS rules reduce average L_1 to the solver target from 0.211 to 0.100, a 52.6% relative improvement over direct prompting. The distance to the distilled MDT policy falls from 0.204 to 0.114, a 44.0% relative improvement. The SCCS columns also show low variation across the evaluated LLM configurations, suggesting that the effect is not specific to one model setting. Argmax-action agreement, defined as whether the highest-probability predicted action matches the highest-probability solver action, also rises from 57.2% to 76.1%.

The summaries-only condition is intentionally included as a negative control. Direct+Summaries has worse average L_1 to the solver target (0.256) than the direct prompt, suggesting that raw continuation quantities are not automatically communicable to an LLM. Without a contrastive rule, the model may not know which summary changes are decision-relevant in the current scenario; the extra quantities can be treated as noise or can reinforce a locally conservative interpretation. SCCS improves performance because it organizes those quantities around a policy-divergent boundary under the same public context.

Table 2: **Communicability on unseen target hands in matched public contexts.** Lower L_1 is better. Each model entry reports mean \pm standard error over unseen target cases. The Mean row reports mean \pm sample standard deviation across eight LLM configurations. Bold numerical entries mark the lowest L_1 within each column; the bold model name marks GPT-5.5 high, the strongest LLM configuration evaluated in our experiments.

LLM run	L_1 to solver			L_1 to MDT		
	Direct	+Summaries	+SCCS Rule	Direct	+Summaries	+SCCS Rule
Gemini-3.1 Flash	0.211 \pm 0.015	0.223 \pm 0.016	0.099 \pm 0.009	0.204 \pm 0.014	0.224 \pm 0.015	0.114 \pm 0.009
Gemini-3.1 Pro low	0.235 \pm 0.020	0.299 \pm 0.021	0.106 \pm 0.010	0.235 \pm 0.020	0.300 \pm 0.020	0.124 \pm 0.010
Gemini-3.1 Pro high	0.268 \pm 0.019	0.303 \pm 0.015	0.089 \pm 0.012	0.261 \pm 0.019	0.314 \pm 0.012	0.109 \pm 0.013
DeepSeek-V4 Flash	0.185 \pm 0.015	0.263 \pm 0.019	0.116 \pm 0.013	0.173 \pm 0.013	0.256 \pm 0.018	0.130 \pm 0.012
DeepSeek-V4 Pro	0.254 \pm 0.016	0.255 \pm 0.017	0.112 \pm 0.011	0.252 \pm 0.014	0.259 \pm 0.014	0.125 \pm 0.010
GPT-5.4	0.187 \pm 0.013	0.216 \pm 0.014	0.095 \pm 0.010	0.175 \pm 0.011	0.210 \pm 0.014	0.104 \pm 0.008
GPT-5.5 low	0.158 \pm 0.013	0.241 \pm 0.016	0.096 \pm 0.009	0.155 \pm 0.011	0.241 \pm 0.016	0.109 \pm 0.009
GPT-5.5 high	0.186 \pm 0.007	0.248 \pm 0.017	0.086 \pm 0.010	0.173 \pm 0.008	0.260 \pm 0.015	0.097 \pm 0.008
Mean	0.211 \pm 0.038	0.256 \pm 0.032	0.100 \pm 0.011	0.204 \pm 0.041	0.258 \pm 0.035	0.114 \pm 0.011

The table supports a specific interpretation of the result. SCCS does not make an LLM a standalone poker agent; it makes a local solver distinction communicable enough for an independent model to transfer the extracted rule to an unseen target hand in a similar strategic neighborhood.

6.4 Case Study: From Over-Folding to Draw-Aware Mixing

To qualitatively inspect the strategy acquired by the LLM, we inspect one of the most difficult spots for humans: SB vs BB single raised pot. Table 3 in the appendix reports the case where the public state is a 8s6h5d board after the bet-raise actions. Now the out-of-position player (SB) is facing pressure with Td9c.

Directly prompting an LLM identifies the hand as an offsuit gutshot with poor equity realization and assigns most mass to Fold. Despite that the LLM is able to correctly identify the hand bucket, the solver disagrees with its action and assigns no fold mass and mixes between Call and the Raise action. A studied human player will be able to understand that at a relatively high stack depth, a hand like T9 can hit the absolute nuts on a 8 turn or river. At the same time, with the wide range of SB vs BB, the overcards J and T provides additional equity even without the backdoor flush draw. SCCS makes this distinction communicable by contrasting a folding prototype (Jc4c) with the same gut shot bucket. Their differences in draw strength, hand EV, and range-relative equity move Td9c into a continuing draw class with some aggressive raise frequency.

This example illustrates why conservative heuristics are not enough. The direct and summaries-only prompts over-fold, while the MDT rule exposes the local boundary between weak folding draws and continuing draw candidates. With that boundary exposed, GPT-5.5 high shifts toward the solver’s Call/Bet allocation. Appendix C provides the full case details.

7 Conclusion

This paper studies how to elicit equilibrium reasoning in LLMs by translating solver-computed mixed strategies into communicable rules. Human play, commentary, and self-rationalization provide weak supervision for complex games because they are selective, heuristic, and biased toward pure actions, whereas equilibrium play in imperfect-information games requires precise mixed-strategy frequency allocation. Solver outputs provide the desired optimality signal, but only as numerical policy distributions and continuation values.

We introduced Mixed-Strategy Decision Tree (MDT) to articulate solver-implied decision logic into sparse strategic rules, and Scenario-Constrained Counterfactual Sampling (SCCS) to expose local contrastive boundaries between hands in the same public context. In No-Limit Texas Hold’em, using over 250 million solver-labeled mixed-strategy decisions, the resulting rules reduce LLM prediction distance to the solver equilibrium by 52.6% across eight LLM configurations. These results suggest that solver-generated data can serve not only as supervision for policy prediction, but also as a source of readable reasoning traces for LLMs.

References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems*, 34:4699–4711, 2021.
- Osbert Bastani, Yewen Pu, and Armando Solar-Lezama. Verifiable reinforcement learning via policy extraction. *Advances in neural information processing systems*, 31, 2018.
- Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold’em poker is solved. *Science*, 347(6218):145–149, 2015.
- Noam Brown and Tuomas Sandholm. Safe and nested subgame solving for imperfect-information games. *Advances in neural information processing systems*, 30, 2017.
- Noam Brown and Tuomas Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.
- Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.
- Noam Brown, Anton Bakhtin, Adam Lerer, and Qucheng Gong. Combining deep reinforcement learning and search for imperfect-information games. *Advances in neural information processing systems*, 33:17057–17069, 2020.
- Nicholas Frosst and Geoffrey Hinton. Distilling a neural network into a soft decision tree. *arXiv preprint arXiv:1711.09784*, 2017.
- Sam Ganzfried and Tuomas Sandholm. Potential-aware imperfect-recall abstraction with earth mover’s distance in imperfect-information games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.
- Suriya Gunasekar, Yi Zhang, Jyoti Aneja, Caio César Teodoro Mendes, Allie Del Giorno, Sivakanth Gopi, Mojan Javaheripi, Piero Kauffmann, Gustavo de Rosa, Olli Saarikivi, et al. Textbooks are all you need. *arXiv preprint arXiv:2306.11644*, 2023.
- Akshat Gupta. Are chatgpt and gpt-4 good poker players? a pre-flop analysis. *arXiv preprint arXiv:2308.12466*, 2023.
- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- Chenghao Huang, Yanbo Cao, Yinlong Wen, Tao Zhou, and Yanru Zhang. Pokergpt: An end-to-end lightweight solver for multi-player texas hold’em via large language model. *arXiv preprint arXiv:2401.06781*, 2024.
- Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.
- Pang Wei Koh, Thao Nguyen, Yew Siang Tang, Stephen Mussmann, Emma Pierson, Been Kim, and Percy Liang. Concept bottleneck models. In *International Conference on Machine Learning*, pages 5338–5348, 2020.
- Harold W. Kuhn. Extensive games and the problem of information. In Harold W. Kuhn and Albert W. Tucker, editors, *Contributions to the Theory of Games II*, volume 28 of *Annals of Mathematics Studies*, pages 193–216. Princeton University Press, 1953.
- Boning Li and Longbo Huang. Efficient online pruning and abstraction for imperfect information extensive-form games. In *The Thirteenth International Conference on Learning Representations*, 2025.

Minhua Lin, Enyan Dai, Hui Liu, Xianfeng Tang, Yuliang Yan, Zhenwei Dai, Jingying Zeng, Zhiwei Zhang, Fali Wang, Hongcheng Gao, Chen Luo, Xiang Zhang, Qi He, and Suhang Wang. How far are LLMs from professional poker players? revisiting game-theoretic reasoning with agentic tool use. In *The Fourteenth International Conference on Learning Representations*, 2026. URL <https://openreview.net/forum?id=vV54ShHvGi>.

Thomas McGrath, Andrei Kapishnikov, Nenad Tomašev, Adam Pearce, Martin Wattenberg, Demis Hassabis, Been Kim, Ulrich Paquet, and Vladimir Kramnik. Acquisition of chess knowledge in alphazero. *Proceedings of the National Academy of Sciences*, 119(47):e2206625119, 2022.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

Marc-Antoine Provost, Nejc Ilenic, Christopher Solinas, and Philippe Beardsell. Gto wizard benchmark, 2026. URL <https://arxiv.org/abs/2603.23660>.

Andrei A Rusu, Sergio Gomez Colmenarejo, Caglar Gulcehre, Guillaume Desjardins, James Kirkpatrick, Razvan Pascanu, Volodymyr Mnih, Koray Kavukcuoglu, and Raia Hadsell. Policy distillation. *arXiv preprint arXiv:1511.06295*, 2015.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarczyk, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

David Silver and Richard S Sutton. Welcome to the era of experience. *Google AI*, 1:11, 2025.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

Abhinav Verma, Vijayaraghavan Murali, Rishabh Singh, Pushmeet Kohli, and Swarat Chaudhuri. Programmatically interpretable reinforcement learning. In *International Conference on Machine Learning*, pages 5045–5054, 2018.

Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841, 2017.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

Richard Zhuang, Akshat Gupta, Richard Yang, Aniket Rahane, Zhengyu Li, and Gopala Anumanchipalli. Pokerbench: Training large language models to become professional poker players. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 26175–26182, 2025.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

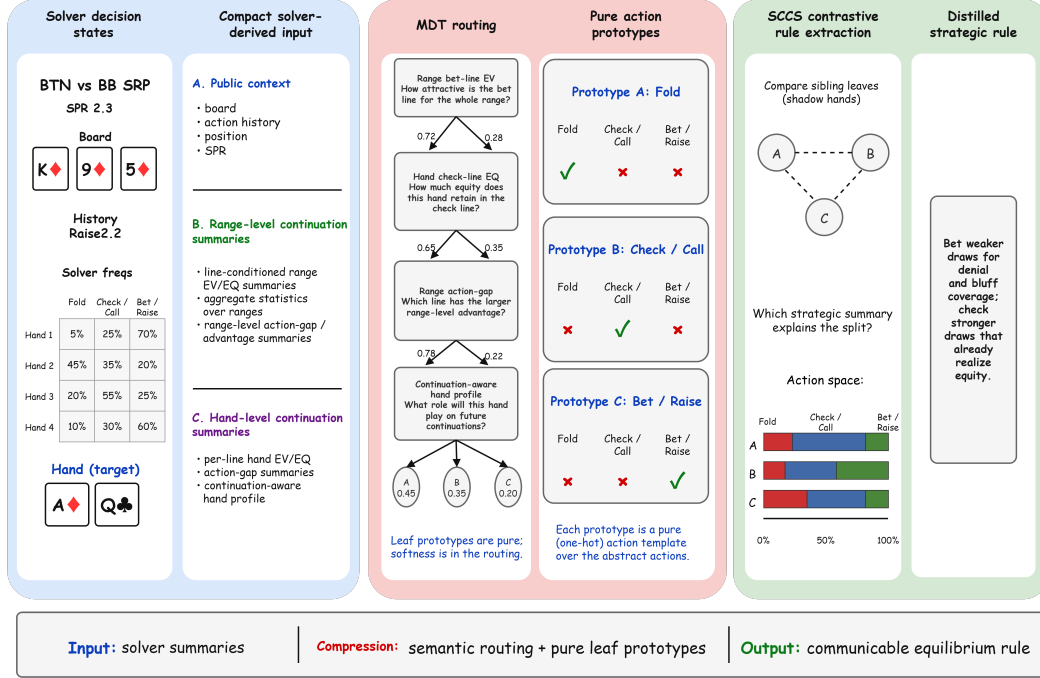


Figure 2: **Distilling solver policies into communicable strategic rules.** Solver policy and continuation outputs are first converted into a compact input x containing public context, range-level continuation summaries, and hand-level continuation summaries. MDT then uses sparse routing over these summaries to assign probability mass to pure-action leaf prototypes; the mixed strategy arises from the routing distribution, not from mixed leaves. SCCS compares shadow hands with matched public context but clear solver-policy divergence, then converts the routing contrast into a rule that an independent reasoner can apply to unseen target hands.

A Additional Related Work

Superhuman solvers as silent oracles. The resolution of imperfect-information games has been driven by equilibrium-finding algorithms like Counterfactual Regret Minimization (CFR) and its variants [Zinkevich et al., 2007, Bowling et al., 2015, Brown and Sandholm, 2018, 2019, Moravčík et al., 2017, Brown et al., 2020]. In other strategic domains, superhuman systems have become more than competitors: AlphaGo and AlphaZero changed how strong players and researchers study Go and chess [Silver et al., 2016, 2017], and later analyses recovered human-understandable chess concepts from AlphaZero play [McGrath et al., 2022]. Poker has undergone a parallel shift toward solver- and GTO-guided study, as reflected by solver-based benchmarks and training resources [Zhuang et al., 2025, Provost et al., 2026]. Yet poker solvers remain largely silent oracles: they provide exact frequency prescriptions (e.g., “bet 33.4%”) but not communicable rationales for those frequencies.

Concept discovery and interpretable policy distillation. A growing body of work aims to expose structure inside learned or optimized policies. Concept-based analysis has been effective for perfect-information agents such as AlphaZero [McGrath et al., 2022], while policy distillation compresses cumbersome teacher models into lightweight students [Hinton et al., 2015, Rusu et al., 2015]. Interpretable variants [Koh et al., 2020, Jacobs et al., 1991, Shazeer et al., 2017], including Tree Imitation Learning, VIPER, and Programmatically Interpretable RL, project policies into trees or programs [Bastani et al., 2018, Verma et al., 2018]; soft decision trees provide another differentiable route to tree-structured explanations [Frosst and Hinton, 2017]. In imperfect-information games, abstraction and bucketing methods cluster strategically similar hands for computational tractability [Ganzfried and Sandholm, 2014, Li and Huang, 2025]. These methods expose useful structure, but their objectives are usually concept probing, compression, or efficient solving; they do not directly

address the range-level coupling and action-frequency mixing that make poker equilibrium policies hard to communicate.

LLMs, poker benchmarks, and synthetic reasoning. LLMs have recently been tested as poker decision makers, but existing results show substantial gaps. Gupta [2023] evaluate ChatGPT and GPT-4 on preflop poker decisions, while Huang et al. [2024] explore LLM-based poker agents trained from online poker data; both lines highlight the difficulty of obtaining reliable strategic behavior from language models alone. PokerBench provides a broader benchmark and training set over curated preflop and postflop spots, using dominant-action and bet-size labels with action-accuracy and exact-match metrics [Zhuang et al., 2025]. The GTO Wizard Benchmark evaluates frontier LLMs against a superhuman poker agent and finds persistent failures in card representation, range construction, and solver-level action selection [Provost et al., 2026]. These works primarily evaluate or train LLM poker play. By contrast, structured reasoning and synthetic-data work suggests that explicit rationales can improve model behavior [Wei et al., 2022, Gunasekar et al., 2023, Zelikman et al., 2022], but poker requires such rationales to be grounded in solver-computed mixed strategies rather than in fluent human commentary alone.

B Distillation Fidelity Details

Table 1 reports the training objective components used to evaluate solver-policy distillation: action-distribution L_1 and oracle-conditioned EV gap. Dense high-capacity models obtain the lowest numerical loss, but their decisions are not directly inspectable. The raw-PBS baselines use a 1482-dimensional input and 512-wide hidden layers, whereas the summary-input baselines use a 156-dimensional input and 256-wide hidden layers. In the tree variants, the soft-sparse model keeps dense routers over the 156 summaries and relies on regularization rather than hard feature selection; Teacher 2 uses higher-capacity two-layer MLP routers; and the final hard MDT selects its top five summaries at each node through a mask over the original 156-dimensional input. Within the summary-input comparison, tree-structured routing improves substantially over the flat MLP, supporting the use of MDT for context-dependent policy structure. The final hard MDT has higher loss because hard sparsification removes small corrective effects used by dense routers. This fidelity cost is intentional: the displayed node-local summaries are exactly the variables used by the model, which avoids explanations that omit many low-weight contributors.

C Case Study Details

This appendix expands the draw-aware transfer case study from the main text using the GPT-5.5 high visible-rationale diagnostic run. The example is chosen because all methods receive the same public state and target hand, and the main error is strategic rather than notational: Td9c is recognized as an offsuit gutshot, but the question is whether it belongs to the folding part of the range or to a continuing draw class. We report action distributions in the order shown in the public context, and report the average L_1 distance to the solver target.

Table 3: **Draw-aware rule transfer on an unseen out-of-position target hand.** The SCCS rule exposes the boundary between weak folding draws and continuing semi-bluff candidates.

Public context	SB vs BB single-raised pot, board 8s6h5d, history b4, b11, OOP decision, actions {Call, Fold, Bet(1.79 \times)}.
Contrastive trace	Fold prototype Jc4c: model [0.02, 0.97, 0.00], solver [0.00, 1.00, 0.00]. Policy-divergent continuing prototypes, including Qh9h and Jc7c, route to Call/Bet mixtures with solver Fold = 0. The contrastive split is governed by draw strength, hand EV, and range-relative equity.
Unseen target hand	Td9c. Key SCCS values: draw strength 0.30 vs average 0.13; hand EV 0.03; hand equity 0.32; range-relative equity -0.14.
Solver target	[Call : 0.63, Fold : 0.00, Bet : 0.37].
GPT-5.5 direct	[0.12, 0.88, 0.00], average L_1 to solver = 0.585.
GPT-5.5 + summaries	[0.07, 0.91, 0.02], average L_1 to solver = 0.605.
GPT-5.5 + SCCS rule	[0.59, 0.00, 0.41], average L_1 to solver = 0.025.

530 The visible rationales clarify the failure mode. Direct prompting describes Td9c as a gutshot with
 531 overcards but poor out-of-position realization, and therefore assigns most mass to Fold. Adding
 532 summaries does not change the qualitative decision: the model notes the 9 blocker, but still treats
 533 the hand as low-ranked air facing pressure and views the large aggressive action as too ambitious.
 534 SCCS changes the evidence structure by presenting a matched folding prototype together with
 535 continuing prototypes in the same public context. With this contrast, the model instead anchors Td9c
 536 to the nearby T9 gutshot profile: low showdown value but enough straight equity and playability to
 537 continue, with suit differences secondary. It therefore removes the fold mass and recovers the solver’s
 538 Call/Bet mixture up to small error. This is the type of local range-boundary transfer measured in the
 539 communicability experiment; it is not an evaluation of live exploitability.

540 D NLH Rules and Poker Terminology

541 This appendix provides a compact reference for the NLH rule structure and poker terminology used
 542 throughout the paper. The descriptions are intended to fix notation and vocabulary for the experiments,
 543 not to introduce new modeling assumptions.

544 D.1 Rules of No-Limit Texas Hold’em

545 **Game format.** No-Limit Texas Hold’em (NLH) is a 6-player version of Texas Hold’em. Each
 546 player receives two private cards, usually called *hole cards*. Up to five public *community cards* are
 547 then revealed on the board. At showdown, each remaining player forms the best five-card poker hand
 548 using any combination of their two private cards and the five community cards.

549 **Blinds and positions.** Each hand begins with forced bets called blinds. In play, the button posts the
 550 small blind (SB) and acts first before the flop; the big blind (BB) posts the larger forced bet. After the
 551 flop, the BB acts first and is therefore out of position (OOP), while the button/SB acts second and is
 552 in position (IP). The button alternates between players across hands.

553 **Betting streets.** A hand proceeds through four betting rounds, also called *streets*. The *preflop* round
 554 occurs after private cards are dealt and before any community card appears. The *flop* reveals three
 555 community cards, the *turn* reveals a fourth community card, and the *river* reveals the fifth and final
 556 community card. In this paper, the dataset and evaluations focus on postflop decisions, especially
 557 flop and turn states.

558 **Legal actions.** At a decision point, the legal actions depend on the previous betting sequence. A
 559 player may *check* if no bet is currently faced, *bet* to put chips into the pot, *call* to match an opponent’s
 560 bet, *fold* to surrender the pot, or *raise* to increase an existing bet. In no-limit poker, a bet or raise can
 561 be any legal size up to the player’s remaining stack; an all-in action commits the full remaining stack.
 562 Our action abstraction groups available aggressive actions under the generic bet/raise category when
 563 reporting solver mixtures.

564 **Pots, stacks, and bet sizes.** The *pot* is the number of chips currently contested. A player’s *stack* is
 565 their remaining chips. Stack-to-pot ratio (SPR) is the remaining effective stack divided by the pot
 566 and measures how much future betting leverage remains. Bet sizes are often written as fractions or
 567 multiples of the pot, e.g., a $0.5\times$ pot bet or a $1.79\times$ pot raise. The experiment configuration uses
 568 100BB starting stacks and no rake.

569 **Hand ranking.** Texas Hold’em uses the standard poker hand order: high card, one pair, two pair,
 570 three of a kind, straight, flush, full house, four of a kind, and straight flush. A *kicker* is a side card
 571 used to break ties between otherwise similar made hands, such as top pair with an ace kicker versus
 572 top pair with a weaker kicker.

573 D.2 Common Poker Terms Used in the Paper

Term	Meaning in this paper
Action line / history	The sequence of previous betting actions and public card events leading to the current decision point.
All-in	A bet or raise that commits a player's entire remaining stack.
Air	A hand with little or no current showdown value and limited immediate equity.
Backdoor draw	A draw that needs favorable cards on both later streets to complete, such as needing both turn and river to make a flush.
Bet size	The amount placed into the pot, often normalized by the current pot size.
Blocker	A card in a player's hand that removes combinations from the opponent's possible range, often reducing the chance that the opponent holds strong continuing hands.
Board	The public community cards visible to both players.
Board texture	Strategic properties of the board, such as pairedness, connectedness, straight potential, and flush potential.
Bluff	An aggressive action with a hand that is unlikely to be best if called, used to make better hands fold.
Call	Matching the current bet to continue in the hand.
Check	Passing the action when no bet is faced.
Continuation bet	A postflop bet made by the player who was the previous aggressor, commonly abbreviated as c-bet.
Draw	A hand that is not currently strong but can improve to a strong hand on later community cards.
Equity (EQ)	The probability, or solver-computed share, that a hand or range wins at showdown under the relevant future-card distribution.
Expected value (EV)	The expected payoff of a hand, range, or action under the solver's continuation strategy.
Fold equity	The value gained from the probability that an opponent folds to an aggressive action.
Flush draw	A draw to five cards of the same suit.
Gutshot	An inside straight draw that can complete with one specific rank.
GTO	Game-Theoretic Optimal; in this paper, an approximate Nash-equilibrium poker strategy computed by a solver.
Hand	Usually the player's two private cards, and sometimes the resulting best five-card category depending on context.
In position (IP)	The player who acts second on postflop streets.
MDF	Minimum defense frequency, a pot-odds-derived threshold describing how often a range must continue to avoid being immediately exploitable by a bet.
Mixed strategy	A probability distribution over legal actions at a decision point.
Nuts	The strongest possible hand, or class of strongest hands, for the current board.
Nut advantage	A range-level advantage in the frequency or equity of nut-class hands.
Offsuit	A two-card private hand whose cards have different suits.
Open-ended straight draw (OESD)	A straight draw that can complete with a card on either end of the sequence.
Out of position (OOP)	The player who acts first on postflop streets.
Overcard	A private card higher than every card on the board.
Postflop	Any decision after the flop has been dealt; includes flop, turn, and river.
Pot odds	The price offered by the pot for calling a bet, usually expressed as a required equity threshold.
Protection bet	A bet with a vulnerable made hand or semi-made hand intended to deny equity to hands that can improve.
Range	The probability distribution over private hands a player can hold after conditioning on public cards and betting history.
Range advantage	A range-level equity or EV edge for one player over the other in the current public state.

Term	Meaning in this paper
River	The fifth community card and final betting street.
Semi-bluff	A bluffing bet or raise with a hand that can improve to a strong hand on later streets.
Set	Three of a kind made with a pocket pair and one matching board card.
Showdown value	The ability of a hand to win if betting stops and the hand reaches showdown.
Single-raised pot	A pot where the preflop action contains one raise and no 3-bet.
Suited	A two-card private hand whose cards share the same suit.
Thin value	A value bet with a hand that is ahead of some calling hands but not strong enough to be clearly dominant.
Three-bet pot / 3-bet pot	A pot where the preflop action contains a raise and then a re-raise.
Trap	A passive action with a very strong hand, used to keep weaker hands or bluffs in the opponent’s range.
Turn	The fourth community card and the betting street after it is dealt.
Unblocker	A card property that leaves the opponent’s folding range relatively intact, which can improve bluff quality in some contexts.
Value bet	A bet made with a hand expected to be called by worse hands often enough to profit.

574 E Pilot Study on Public Poker Discourse

575 This appendix gives additional context for the pilot study mentioned in Section 4. Two Plus Two
576 (2+2) refers to the Two Plus Two poker forum¹, a long-running public discussion forum for poker
577 strategy, theory, and community discussion. Such forum text is useful for exposing a language model
578 to poker vocabulary, common strategic concepts, and the informal reasoning style used by human
579 players.

580 In our pilot experiment, we scraped public 2+2 forum discussions and used them to perform supervised
581 fine-tuning (SFT) on an open Qwen model with the standard next-token prediction objective. The
582 goal was not to build the final system in this paper, but to test whether imitation of naturally occurring
583 poker discourse could by itself teach equilibrium-relevant poker reasoning. The resulting model
584 learned to use many terms that appeared frequently in the forum text, such as range, blocker, equity,
585 pot odds, bluff, value bet, and GTO. It also produced more fluent local hand explanations than the
586 base model.

587 However, the improvement was mostly linguistic and heuristic. In qualitative evaluations, the fine-
588 tuned model still showed shallow understanding of many poker-theoretic concepts that appeared
589 in the training text. It could often repeat the vocabulary of range advantage, blockers, or mixed
590 strategy, but it did not reliably apply these concepts to make decisions close to solver-computed
591 GTO strategies. In particular, it frequently reduced mixed-equilibrium decisions to single-hand
592 narratives, over-relied on visible hand strength, and failed to reason about how one private hand’s
593 action frequency is constrained by the rest of the range.

594 These observations support the distinction made in the main text. Public poker discourse is valuable
595 as a source of terminology and human-readable explanation style, but it does not provide ground-truth
596 equilibrium targets. Therefore, our main pipeline uses solver-derived mixed policies and continuation
597 summaries as the strategic target, while using language only as the medium for articulation.

598 F Additional Method Details

599 F.1 Router Parameterization

600 Each internal node n computes a branch distribution from a small set of active summaries. Let
601 $\mathbf{m}_n \in \{0, 1\}^D$ denote the node-local hard mask and let $\|\mathbf{m}_n\|_0 \leq K$. The branch logits are

$$\mathbf{g}_n(\mathbf{x}) = r_n(\mathbf{x} \odot \mathbf{m}_n), \quad \mathbf{p}_n(\mathbf{x}) = \text{softmax}(\mathbf{g}_n(\mathbf{x})), \quad (8)$$

¹<https://forumserver.twoplustwo.com/>

where r_n is a small local router. In the ablations in Table 1, r_n can be either a sparse linear map,

$$\mathbf{g}_n(\mathbf{x}) = W_n(\mathbf{x} \odot \mathbf{m}_n) + \mathbf{b}_n, \quad (9)$$

or an additive one-dimensional router,

$$\mathbf{g}_n(\mathbf{x}) = \sum_{j:m_{n,j}=1} \mathbf{h}_{n,j}(x_j), \quad (10)$$

with each $\mathbf{h}_{n,j}$ implemented as a small scalar-to-logit network. These choices are implementation variants for fitting the tree; the final hard MDT exposes the same object in either case: a node-local set of at most K summaries and the resulting branch probabilities.

F.2 SCCS Sampling and Verification

SCCS is designed to extract a rule for a held-out target without showing the target policy to the downstream LLM. The sampling procedure first fixes the public context, then looks for private-hand changes that both cross an MDT routing boundary and induce a clear solver-policy change. The resulting prompt displays only the contrastive hands used to form the rule; the evaluation target is required to be policy-separated from every displayed hand.

Algorithm 1 SCCS: Scenario-Constrained Contrastive Rule Extraction

Require: Target hand h_{test} , trained MDT \mathcal{M} , reference solver set \mathcal{D} , policy-divergence threshold τ_π , prompt-separation threshold τ_{sep}

Ensure: SCCS prompt containing contrastive rule evidence, with the target policy masked

- 1: Fix the public scenario $c = \text{Context}(h_{\text{test}})$.
 - 2: Compute target summaries \mathbf{x}_{test} , target route $P_{\text{test}} = \text{Route}_{\mathcal{M}}(\mathbf{x}_{\text{test}})$, and target policy $\pi^*(\cdot \mid h_{\text{test}})$.
 - 3: Initialize candidate set $\mathcal{C} \leftarrow \emptyset$.
 - 4: **Scenario-constrained candidate sampling**
 - 5: For each solver-labeled hand $h' \in \mathcal{D}$ with $\text{Context}(h') = c$:
 - 6: compute $\mathbf{x}', P' = \text{Route}_{\mathcal{M}}(\mathbf{x}')$, and $\pi^*(\cdot \mid h')$.
 - 7: discard h' if $P' = P_{\text{test}}$.
 - 8: discard h' if $\bar{L}_1(\pi^*(\cdot \mid h'), \pi^*(\cdot \mid h_{\text{test}})) < \tau_\pi$.
 - 9: otherwise add h' to \mathcal{C} .
 - 10: **Boundary localization**
 - 11: For each candidate $h' \in \mathcal{C}$, identify the earliest node $n(h')$ where P' and P_{test} diverge.
 - 12: At $n(h')$, collect the active summaries selected by the hard MDT router.
 - 13: Rank candidates by policy divergence, routing-boundary clarity, and sparsity of the active-summary contrast.
 - 14: Select one or more shadow hands \mathcal{S} from the top-ranked candidates.
 - 15: **Rule construction**
 - 16: For each shadow hand $h_s \in \mathcal{S}$, extract the diverging node, route branch, active summaries, summary values, MDT policy, and solver policy.
 - 17: Convert these contrasts into a short natural-language rule that names the public scenario, the shadow-hand contrast, and the active summary boundary.
 - 18: **Held-out verification prompt**
 - 19: Hide $\pi^*(\cdot \mid h_{\text{test}})$ from the downstream LLM.
 - 20: Require $\bar{L}_1(\pi^*(\cdot \mid h_{\text{test}}), \pi^*(\cdot \mid h_s)) \geq \tau_{\text{sep}}$ for every displayed hand $h_s \in \mathcal{S}$.
 - 21: Query the LLM with the public scenario, target hand, target summaries, and SCCS rule evidence; evaluate its predicted distribution against the masked target policy.
-

Here \bar{L}_1 denotes the same action-averaged L_1 distance used in the main evaluation. In the communicability experiments, we use $\tau_{\text{sep}} = 0.20$ to prevent the displayed SCCS examples from serving as near-label lookups for the held-out target.

G Detailed Optimization Curriculum

Here we detail the three-phase curriculum for the **Hierarchical Hard Distillation**. This curriculum is designed to distill knowledge from a dense Oracle into a strict symbolic structure without performance collapse.

Input (SCCS Prompt - Abbreviated)

Scenario: BTN vs BB 3-Bet. **Board:** $4\spadesuit 3\heartsuit 3\diamondsuit$. **Hand:** $T\heartsuit 8\heartsuit$ (Hero).

Decision Trace: Node #0 (Filter) \rightarrow Node #21 (Split) \rightarrow Leaf #X (Strategy: Bet 100%).

SCCS Comparison: Shadow Hand $K\heartsuit 9\heartsuit$ (Shadow1) diverges at Node #21 to **Check**.

Critical Summary Contrast: Hero Phys:Kicker = 0.57 (T -High). Shadow1 Phys:Kicker = 0.64 (K -High).

Hero EQ - Nut MDF = 0.03. Shadow1 EQ - Nut MDF = 0.20.

Model Output (Hard MDT Logic - Generated by Gemini-3.1-Pro)

1. Strategy Summary: The model executes a Polarized Continuation Bet strategy. It aggressively semi-bluffs hands with low Showdown Value but high potential (T -high flush draws) to deny equity, while checking back stronger K -high hands to realize equity as (S.D. Value).

2. Path Logic & Divergence:

- **Node #21 (The Polarization Split):** This is the pivot point separating Semi-Bluffs from S.D. Value.
- **Hero ($T\heartsuit 8\heartsuit$) \rightarrow Bet:** Hero has a weaker Phys:Kicker (0.57). The model identifies this hand as having insufficient Showdown Value to check-call. It bets to fold out better hands (like Q -high) and deny their equity realization.
- **Shadow1 ($K\heartsuit 9\heartsuit$) \rightarrow Check:** Shadow1 possesses a stronger Phys:Kicker (0.64). Despite similar raw equity, the K -high acts as a blocker and has enough Showdown Value to beat missed draws. Checking maximizes EV by avoiding isolation against the villain’s calling range.

3. Test Case Solution (Unseen Hand: $K\diamondsuit 9\diamondsuit$)

- **Predicted Action:** Check (Frequency $\sim 60\%$).
 - **Reasoning:** The target hand follows the same routed strategic pattern as Shadow1 ($K\heartsuit 9\heartsuit$). With a Kicker of 0.64 (K -High), it falls into the S.D. Value bucket. It is too strong to turn into a pure bluff (like the $T\heartsuit 8\heartsuit$) but too weak to value bet, confirming the model’s preference for Range Protection over Polarization for this specific kicker class.
-

Figure 3: **Example of SCCS Rule Extraction.** The SCCS pipeline successfully distills the subtle strategic boundary between a *Semi-Bluff* and *Showdown Value*. The Hard MDT uses the summary Phys: Kicker as a precise cut-off: on a $4\spadesuit 3\heartsuit 3\diamondsuit$ board, K -High (Kicker 0.64) is strong enough to Check, whereas T -High (Kicker 0.57) must Bet to deny equity. The LLM correctly generalizes this rule to the unseen target hand $K\diamondsuit 9\diamondsuit$.

620 **Phase 1: The Global Teacher (Denoising).** We first train a “Global-Gated ResNet” (Teacher 1). It
621 is restricted to use only $K_g = 50$ summaries globally but allows unrestricted non-linear interactions.
622 This phase acts as a filter, removing strictly irrelevant summaries while preserving high-dimensional
623 correlations inherent in the solver data.

624 **Phase 2: The Structural Teacher (Topology).** We then distill Teacher 1 into a “Deep Tree”
625 (Teacher 2). This model adopts the target tree topology but uses higher-capacity local routers. This
626 step establishes the correct decision hierarchy (e.g., branching on Board Texture before Kickers)
627 without being constrained by limited routing capacity.

Phase 3: Hard Student Locking (Logic Extraction). Finally, we distill Teacher 2 into the target
“Hard MDT”. Crucially, we switch from soft gating to “Hard Gating” using the Straight-Through
Estimator (STE). We physically lock the summary set to the Top- K (e.g., 5) at each node:

$$\mathbf{m}_{hard} = \text{TopK}(\mathbf{m}_{logits}, k = 5)$$

628 This forces the student to find the optimal strategy that exists *strictly within* the 5-summary subspace,
629 mathematically eliminating the residue problem found in soft optimization.

630 H Limitations and Future Work

631 While our framework successfully distills articulate reasoning from solver data, we identify two
632 primary limitations regarding its scope and current deployment.

633 **Dependence on Mixed-Strategy Equilibria** The Mixed-Strategy Decision Tree (MDT) is architec-
634 turally specialized for imperfect-information games characterized by mixed Nash equilibria. The

model’s inductive bias, specifically its decomposition of strategy into probabilistic routing over pure-action leaf prototypes, is designed to capture the delicate frequency balancing required in games like No-Limit Texas Hold’em. Consequently, this approach may yield diminishing returns in perfect-information domains (e.g., chess) or games dominated by pure strategies, where such complex mixing specifications are unnecessary.

Dependence on Solver-Derived Summaries Currently, our system operates as an offline analytical agent rather than a standalone poker agent. Because the MDT relies on solver-derived strategic summaries, it cannot yet function in a live setting where such ground-truth solver information is unavailable. Bridging this gap, potentially by training a separate state-estimation module to approximate these summaries from raw history, remains a critical direction for future work to enable live-agent deployment.

I Prompt Templates

We use a structured prompt to ground the LLM’s generation in the routed logic exposed by MDT. The template below is populated dynamically by the SCCS engine. We also provide example output of Gemini-3.1-Pro.

Template 1:

```
# Role
You are an elite Poker AI Strategist and GTO Solver Analyst. Your task is to reverse-engineer the decision-making
process exposed by a Mixed-Strategy Decision Tree (MDT) model.

# Task
Analyze the provided trace logic for the specific hand provided below. You must explain *why* the AI chose this
specific path over others, using Poker Theory concepts (Equity Realization, Blockers, Range Morphology).

# Semantic Definitions (Crucial)
The trace uses specific tags based on Equity (EQ) vs. Opponent Range:
- **Value**: Aggressive action with High EQ (>65%).
- **Bluff**: Aggressive action with Low EQ (<35%).
- **Thin Value/Protect**: Aggressive action with Moderate EQ.
- **S.D. Value**: Passive action (Check/Call) with enough EQ to win at showdown but not enough to bet for value.
- **Give Up**: Passive action with near-zero EQ.
- **Trap**: Passive action with Nut-class EQ.

# Input Data
=== SCENARIO CONTEXT ===
Scenario : BTN_vs_BB_3B
Board : 4s3h3d
History : k
Player : IP
Hand : Th8h
Hand Wgt : 0.0039 (Prob in Range)
Pot : 26.5
Actions : [Check, Fold, Bet(0.48x)]

=== SUMMARY GLOSSARY (Definitions for current path) ===

--- Strategic Adv ---
- **Hero EQ Adv** (Avg: -0.03): **RELATIVE SUMMARY (Hero - Villain)**. Difference in Range Equity. Positive (>0) = Hero
is Ahead. Negative (<0) = Villain is Ahead. (Automatically adjusted from Global OOP-IP value based on Hero's
position).
- **Hero EV Adv** (Avg: -0.19): **RELATIVE SUMMARY (Hero - Villain)**. Difference in EV Pot Share. Positive = Favors
Hero.
- **Hand EQ** (Avg: 0.49): Raw Equity (0.0-1.0) against opponent's current range.
- **EQ - Nut MDF** (Avg: 0.25): Hand Equity minus Nut MDF. Positive = Hand is strong enough to play for stacks.
- **EQ - Nut Range Avg** (Avg: 0.27): Hand EQ minus the Average EQ of the Hero's Nut Range. Are you Top Nut or Bottom
Nut?
- **Rank in Range (EQ)** (Avg: 0.50): Percentile of Hand Equity (0-1).

--- Hand Physics ---
- **Phys: Kicker** (Avg: 0.65): Normalized Kicker Strength (Rank / 14.0). Ace=1.0, King=0.92, ..., 2=0.14. Crucial for
domination issues (e.g., distinguishing Top Pair Top Kicker from Top Pair Weak Kicker).

--- Range Buckets ---
- **OOP Rng EV** (Avg: 0.59): OOP's EV expressed as a percentage of the total pot.

--- Other ---
- **Bluff Efficiency** (Avg: 0.61): Measure of how well our 'Air' range blocks opponent's calling range. High
Efficiency = We have 'Natural Bluffs' (Low Equity + High Blocker Score). Low Efficiency = Our bluffs have poor
removal (random trash).
- **Hand Blocker** (Avg: 3.62): Score representing how much this hand blocks opponent's continuing range.
- **Hand EV** (Avg: 0.41): Expected Value of the hand normalized by Pot.
- **MDF (Decision)** (Avg: 0.65): Decision-based MDF = Pot / (Pot + Bet). The break-even equity required to call.
```

651

```

- **Vuln - Range Avg** (Avg: 0.00): Hand Vulnerability minus Range Average. Positive = More vulnerable than average (
  needs protection).
- **Regret: Fold-Call** (Avg: 0.89): EV(Fold) - EV(Call). Positive = Fold is better. Diff in EV (Action A - Action B)
  normalized by Pot.

=== DECISION LOGIC TRACE ===
Legend:
- [Path]: format is 'Role [Hand] -> PathID -> IntendedAction'. Indicates which internal branch was taken and the final
  action intent.
- (Reach): The probability of the hand actually reaching this node vs. surviving to the next node.
- R: [0:xx 1:xx ...]: Router Probabilities. The internal neural network's confidence distribution across Branch 0, 1,
  and 2. Shows how 'split' or 'certain' the decision was.
- [Summary Impact]: Shows which summaries pushed the router towards specific branches.
+-- NODE #0
  [Path] Hero [Th8h] -> P0->Bet (Reach: 100%->96%) | R: [0:95 1:00 2:04]
  [Path] Shadow1 [Kh9h] -> P0->Check (Reach: 100%->74%) | R: [0:73 1:00 2:26]
  [Path] Shadow2 [Tc9c] -> P0->Check (Reach: 100%->100%) | R: [0:99 1:00 2:00]
  [Path] Shadow3 [Ks6s] -> P0->Check (Reach: 100%->63%) | R: [0:62 1:00 2:37]
  [Summary Impact Analysis - Node #0]
  | Summary | Avg | Hero Impact (Chk-Bet) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | Rank in Range (EQ) | 0.50 | **0.06** (+29.1) | 0.24 | 0.02 | 0.25
  | Hand EQ | 0.49 | **0.26** (+15.1) | 0.43 | 0.22 | 0.44
  | MDF (Decision) | 0.65 | **-1.00** (-6.3) | -1.00 | -1.00 |
  | Phys: Kicker | 0.65 | **0.57** (-2.1) | 0.64 | 0.64 | 0.43
  | Hand EV | 0.41 | **0.32** (-0.1) | 0.34 | 0.24 | 0.38
+-- NODE #1
  [Path] Hero [Th8h] -> P2->Bet (Reach: 96%->75%) | R: [0:24 1:00 2:75]
  [Path] Shadow1 [Kh9h] -> P2->Check (Reach: 74%->100%) | R: [0:00 1:00 2:99]
  [Path] Shadow2 [Tc9c] -> P2->Check (Reach: 100%->62%) | R: [0:37 1:00 2:62]
  [Path] Shadow3 [Ks6s] -> P2->Check (Reach: 63%->100%) | R: [0:00 1:00 2:99]
  [Summary Impact Analysis - Node #1]
  | Summary | Avg | Hero Impact (Bet-Chk) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | Hand EV | 0.41 | **0.32** (+29.1) | 0.34 | 0.24 |
  | Hero EV Adv | -0.19 | **0.12** (-27.5) | 0.12 | 0.12 |
  | Vuln - Range Avg | 0.00 | **-0.16** (+15.8) | -0.04 | -0.13 |
  | Hand Blocker | 3.62 | **1.00** (+1.7) | 3.00 | 2.00 |
  | Hand EQ | 0.49 | **0.26** (-0.0) | 0.43 | 0.22 |
+-- NODE #6
  [Path] Hero [Th8h] -> P2->Bet (Reach: 72%->100%) | R: [0:00 1:00 2:99]
  [Path] Shadow1 [Kh9h] -> P2->Check (Reach: 74%->100%) | R: [0:00 1:00 2:99]
  [Path] Shadow2 [Tc9c] -> P2->Check (Reach: 62%->100%) | R: [0:00 1:00 2:99]
  [Path] Shadow3 [Ks6s] -> P2->Check (Reach: 63%->100%) | R: [0:00 1:00 2:99]
  [Summary Impact Analysis - Node #6]
  | Summary | Avg | Hero Impact (Bet-(Chk)) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | Hand EV | 0.41 | **0.32** (+51.2) | 0.34 | 0.24 |
  | Hero EQ Adv | -0.03 | **0.09** (-20.7) | 0.09 | 0.09 |
  | Phys: Kicker | 0.65 | **0.57** (+7.4) | 0.64 | 0.64 |
  | Bluff Efficiency | 0.61 | **0.42** (+7.2) | 0.42 | 0.42 |
  | EQ - Nut Range Avg | 0.27 | **0.03** (+4.7) | 0.20 | -0.02 |
+-- NODE #21
  [Path] Hero [Th8h] -> P2->Bet (Reach: 72%->100%) | R: [0:00 1:00 2:99]
  [Path] Shadow1 [Kh9h] -> P0->Check (Reach: 74%->81%) | R: [0:81 1:00 2:18]
  [Path] Shadow2 [Tc9c] -> P2->Check (Reach: 62%->95%) | R: [0:02 1:02 2:95]
  [Path] Shadow3 [Ks6s] -> P2->Check (Reach: 63%->97%) | R: [0:02 1:00 2:97]
  [Summary Impact Analysis - Node #21]
  | Summary | Avg | Hero Impact (Bet-(Chk)) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | EQ - Nut MDF | 0.25 | **0.03** (+5.6) | 0.20 | -0.01 |
  | Vuln - Range Avg | 0.00 | **-0.16** (+4.0) | -0.04 | -0.13 |
  | Phys: Kicker | 0.65 | **0.57** (+1.6) | 0.64 | 0.64 |

```

```

| Regret: Fold-Call | 0.89 | **1.02** (+1.0) | 0.99 | 1.12
| 0.90 |
| OOP Rng EV | 0.59 | **0.44** (+0.3) | 0.44 | 0.44
| 0.44 |
+-- LEAF #64 (Static Prototype)
| -> Strat: [Check:99%] | Samples: [Shadow1:Kh9h] (Reach:59.6%)
+-- LEAF #66 (Static Prototype)
| -> Strat: [Check:99%] | Samples: [Hero:Th8h] (Reach:71.8%), [Shadow2:Tc9c] (Reach:58.8%), [
Shadow3:Ks6s] (Reach:60.8%)

=== FINAL STRATEGY vs GTO COMPARISON (Training Data) ===
| Hand | Role | Final Model Strat | GTO Target Strat |
| :-----: | :-----: | :-----: | :-----: |
| Th8h | Hero | C:0.40 F:0.00 B:0.60 | C:0.40 F:0.00 B:0.60 |
| Kh9h | Shadow1 | C:0.61 F:0.00 B:0.39 | C:0.59 F:0.00 B:0.41 |
| Tc9c | Shadow2 | C:0.66 F:0.00 B:0.34 | C:0.35 F:0.00 B:0.65 |
| Ks6s | Shadow3 | C:0.52 F:0.00 B:0.48 | C:0.41 F:0.00 B:0.59 |

## 3. Test Case: The Unseen Hand
Consider a new hand in the same scenario:
- **Hand**: Kd9d
- **Key Difference**: (Inspect the summaries yourself compared to Hero)
**Question**: Based on the logic learned above, what is the optimal action for this hand? Explain why using the model's
decision boundaries.

# Analysis Instructions
1. **Layer-by-Layer Review**: For each NODE, explain how the specific **Impact values** determined the branch choice.
2. **Impact Comparison**: Compare Hero's summary values vs. Global Avg and Shadows.
3. **Consistency Check**: Use Section 3 (Test Case) to verify if the logic you reverse-engineered applies to an unseen
hand.
4. **Solve Test Case**: Provide your answer for the Unseen Hand in Section 3.
- Key Summary Values of Test Case (Hierarchical):
--- **Depth 0** - Top 5 Drivers:
- **Rank in Range (EQ)**: 0.24 (Avg: 0.50) | Impact: +18.28
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -6.30
- **Phys: Kicker**: 0.64 (Avg: 0.65) | Impact: -0.95
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: -0.50
- **Hand EQ**: 0.43 (Avg: 0.49) | Impact: +0.20
--- **Depth 1** - Top 4 Drivers:
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +31.56
- **Hero EV Adv**: 0.12 (Avg: -0.19) | Impact: -27.52
- **Vuln - Range Avg**: -0.04 (Avg: 0.00) | Impact: +13.43
- **Hand Blocker**: 3.00 (Avg: 3.62) | Impact: +12.95
--- **Depth 2** - Top 3 Drivers:
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +55.78
- **Hero EQ Adv**: 0.09 (Avg: -0.03) | Impact: -20.67
- **Bluff Efficiency**: 0.42 (Avg: 0.61) | Impact: +7.21
--- **Depth 3** - Top 2 Drivers:
- **Phys: Kicker**: 0.64 (Avg: 0.65) | Impact: +3.12
- **EQ - Nut MDF**: 0.20 (Avg: 0.25) | Impact: +1.43

# Output Format (Markdown)
## 1. Strategy Summary
(One concise sentence)

## 2. Path Logic & Divergence
- **Node #0**: ...

## 3. Test Case Solution
- **Predicted Mixed Strategy**: ...
- **Reasoning**: ...

```

653

654 Answer 1:

```

# Test Case Solution (Sample 5181)
Hand: Kd9d
Scenario: BTN_vs_BB_3B

## 1. Ground Truth Strategies
- **GTO Strategy**: C:0.59 F:0.00 B:0.41
- **Model Strategy**: C:0.61 F:0.00 B:0.39

## 2. Actual Decision Trace
+-- NODE #0
[Path] TestHand [Kd9d] -> P0->Check (Reach: 100%>73%) | R: [0:73 1:00 2:26]
[Summary Impact Analysis - Node #0 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Chk-Bet) |
| :-----: | :-----: | :-----: |
| Rank in Range (EQ) | 0.50 | **0.24** (+18.3) |
| MDF (Decision) | 0.65 | **-1.00** (-6.3) |
| Phys: Kicker | 0.65 | **0.64** (-0.9) |
| Hand EV | 0.41 | **0.34** (-0.5) |
| Hand EQ | 0.49 | **0.43** (+0.2) |
+-- NODE #1
[Path] TestHand [Kd9d] -> P2->Check (Reach: 73%>100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #1 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-(Chk)) |
| :-----: | :-----: | :-----: |

```

655

```

| Hand EV | 0.41 | **0.34** (+31.6) |
| Hero EV Adv | -0.19 | **0.12** (-27.5) |
| Vuln - Range Avg | 0.00 | **-0.04** (+13.4) |
| Hand Blocker | 3.62 | **3.00** (+12.9) |
| Hand EQ | 0.49 | **0.43** (+4.4) |
+-- NODE #6
[Path] TestHand [Kd9d] -> P2->Check (Reach: 73%>100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #6 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-(Chk)) |
| :-----: | :-----: | :-----: |
| Hand EV | 0.41 | **0.34** (+55.8) |
| Hero EQ Adv | -0.03 | **0.09** (-20.7) |
| Bluff Efficiency | 0.61 | **0.42** (+7.2) |
| Phys: Kicker | 0.65 | **0.64** (+7.1) |
| EQ - Nut Range Avg | 0.27 | **0.20** (+6.4) |
+-- NODE #21
[Path] TestHand [Kd9d] -> P0->Check (Reach: 73%>81%) | R: [0:81 1:00 2:18]
[Summary Impact Analysis - Node #21 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Chk-Bet) |
| :-----: | :-----: | :-----: |
| Phys: Kicker | 0.65 | **0.64** (+3.1) |
| EQ - Nut MDF | 0.25 | **0.20** (+1.4) |
| Regret: Fold-Call | 0.89 | **0.99** (-1.1) |
| Vuln - Range Avg | 0.00 | **-0.04** (-0.9) |
| OOP Rng EV | 0.59 | **0.44** (-0.3) |
+-- LEAF #64 (Static Prototype)
-> Strat: [Check:99%] | Samples: [TestHand:Kd9d] (Reach:59.1%)

```

3. Key Summary Drivers (Path-Dependent)

```

### ** (Depth 0) ** - Top 5 Drivers:
- **Rank in Range (EQ)**: 0.24 (Avg: 0.50) | Impact: +18.28
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -6.30
- **Phys: Kicker**: 0.64 (Avg: 0.65) | Impact: -0.95
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: -0.50
- **Hand EQ**: 0.43 (Avg: 0.49) | Impact: +0.20

### ** (Depth 1) ** - Top 4 Drivers:
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +31.56
- **Hero EV Adv**: 0.12 (Avg: -0.19) | Impact: -27.52
- **Vuln - Range Avg**: -0.04 (Avg: 0.00) | Impact: +13.43
- **Hand Blocker**: 3.00 (Avg: 3.62) | Impact: +12.95

### ** (Depth 2) ** - Top 3 Drivers:
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +55.78
- **Hero EQ Adv**: 0.09 (Avg: -0.03) | Impact: -20.67
- **Bluff Efficiency**: 0.42 (Avg: 0.61) | Impact: +7.21

### ** (Depth 3) ** - Top 2 Drivers:
- **Phys: Kicker**: 0.64 (Avg: 0.65) | Impact: +3.12
- **EQ - Nut MDF**: 0.20 (Avg: 0.25) | Impact: +1.43

```

656

657 LLM Response 1:

```

## 1. Strategy Summary
The model executes a **Polarized Continuation Bet** strategy where hands with low Showdown Value but high potential (
Ten-high flush draws) aggressively semi-bluff to deny equity, while stronger King-high hands (S.D. Value) check
back to realize equity and protect the checking range.

## 2. Path Logic & Divergence

### Node #0 (The Equity Filter)
* **Role**: Preliminary Router.
* **Logic**: This node acts as a low-frequency filter rather than a strategic decision maker. While 'Rank in Range'
is the primary summary (Hero Impact +29.1), the router sends the vast majority of all analyzed hands (Hero 96%,
Shadows 74-100%) down the same pipe (Path 0 -> Node #1).
* **Divergence**: There is no meaningful divergence here; the model has determined that all these IP hands (Th8h,
Kh9h, etc.) belong in the active decision tree rather than an immediate "Give Up" bucket.

### Node #21 (The Polarization Split)
* **Role**: The primary strategic pivot point separating **Semi-Bluffs** from **S.D. Value**.
* **Logic**: This node splits Hero (Th8h) from Shadow1 (Kh9h).
  * **Hero (Th8h) -> Bet**: Hero has a weaker 'Phys: Kicker' (0.57) and very low 'EQ - Nut MDF' (0.03). The model
identifies this as a hand with insufficient Showdown Value to check-call but excellent "Natural Bluff" properties
(Backdoor Flush Draw + Overcards). It bets to fold out better hands (Q-high, J-high) and deny their equity
realization.
  * **Shadow1 (Kh9h) -> Check**: Shadow1 has a stronger 'Phys: Kicker' (0.64) and higher 'EQ - Nut MDF' (0.20).
Despite having similar raw equity, the King-high acts as a blocker to the opponent's continuing range and has
enough Showdown Value to beat missed draws or lower high-cards. Checking maximizes EV by avoiding isolation
against the villain's calling range (which is pair-heavy).

## 3. Test Case Solution

* **Hand**: **Kd9d** (King-Nine of Diamonds)
* **Predicted Mixed Strategy**: **Check** (Frequency ~60% Check / 40% Bet)
* **Reasoning**:

```

658

1. ****Morphological Twin****: Kd9d is effectively identical to Shadow1 (Kh9h). Both are King-high hands with a backdoor flush draw on a paired low board (4s3h3d). In MDT models, identical summary vectors result in identical routing.
2. ****Showdown Value Threshold****: The key decision boundary at ****Node #21**** is 'Phys: Kicker'. The Test Case has a Kicker of 0.64 (King), placing it firmly in the ****S.D. Value**** bucket alongside Kh9h. It is too strong to turn into a pure bluff (like the Th8h) because it beats the villain's air range, but it is too weak to value bet for three streets.
3. ****Impact Interpretation****: While the Test Case shows positive impact for 'Hand EV' and 'Kicker', these summaries in the context of King-high hands on this board promote ***Equity Realization*** (Checking) over ***Equity Denial*** (Betting). The hand serves to protect the checking range rather than polarizing the betting range.

659

660 Template 2:

```
# Role
You are an elite Poker AI Strategist and GTO Solver Analyst. Your task is to reverse-engineer the decision-making
process exposed by a Mixed-Strategy Decision Tree (MDT) model.

# Task
Analyze the provided trace logic for the specific hand provided below. You must explain *why* the AI chose this
specific path over others, using Poker Theory concepts (Equity Realization, Blockers, Range Morphology).

# Semantic Definitions (Crucial)
The trace uses specific tags based on Equity (EQ) vs. Opponent Range:
- **Value**: Aggressive action with High EQ (>65%).
- **Bluff**: Aggressive action with Low EQ (<35%).
- **Thin Value/Protect**: Aggressive action with Moderate EQ.
- **S.D. Value**: Passive action (Check/Call) with enough EQ to win at showdown but not enough to bet for value.
- **Give Up**: Passive action with near-zero EQ.
- **Trap**: Passive action with Nut-class EQ.

# Input Data
=== SCENARIO CONTEXT ===
Scenario : BTN_vs_BB_3B
Board : As7s6h
History : k
Player : IP
Hand : 8h8c
Hand Wgt : 0.0068 (Prob in Range)
Pot : 26.5
Actions : [Check, Fold, Bet(0.48x)]

=== SUMMARY GLOSSARY (Definitions for current path) ===

--- Strategic Adv ---
- **Nut EQ Adv (Hero-Vill)** (Avg: 0.05): Difference in Felting Range Equity. Positive = Hero has stronger nuts.
- **Hand EQ** (Avg: 0.49): Raw Equity (0.0-1.0) against opponent's current range.
- **EQ - MDF** (Avg: -0.20): Hand Equity minus Decision MDF. >0 implies raw equity is sufficient to call.
- **EQ - Range Avg** (Avg: -0.00): Hand EQ minus Range Average EQ. Relative Strength.
- **Rank in Range (EQ)** (Avg: 0.50): Percentile of Hand Equity (0-1).

--- Hand Physics ---
- **Phys: Kicker** (Avg: 0.65): Normalized Kicker Strength (Rank / 14.0). Ace=1.0, King=0.92, ..., 2=0.14. Crucial for
domination issues (e.g., distinguishing Top Pair Top Kicker from Top Pair Weak Kicker).

--- Range Buckets ---
- **IP Rng: Set** (Avg: 0.03): Density (0.0-1.0) of Set (ID 12) in IP's range. Sum of all Made Hand densities is
approximately 1.0. High value means range is concentrated on this hand type.
- **OOP Rng EV** (Avg: 0.59): OOP's EV expressed as a percentage of the total pot.

--- Other ---
- **Bluff Efficiency** (Avg: 0.61): Measure of how well our 'Air' range blocks opponent's calling range. High
Efficiency = We have 'Natural Bluffs' (Low Equity + High Blocker Score). Low Efficiency = Our bluffs have poor
removal (random trash).
- **Hand: AceHigh** (Avg: 0.21): Is the current hand a AceHigh? (1.0 = Yes, 0.0 = No).
- **Hand Unblocker** (Avg: 3.62): Score representing how much this hand unblocks opponent's folding range.
- **Opp Card Scarcity (Avg)** (Avg: 0.06): Average Scarcity Effect of our hole cards. Measures the 'Scarcity Effect' we
inflict on the opponent by holding cards they need. Calculated based on card frequency in their range. High
Value = The opponent's range heavily relies on this card. By holding it, we create a severe shortage (High
Blocker Power). Low Value = The opponent's range rarely contains this card. Holding it creates minimal shortage (
Low Blocker Power).
- **Opp Card Scarcity (C1)** (Avg: 0.07): Scarcity Effect of Card 1. High Value = We block a key card for the opponent
(e.g. holding an Ace vs an Ace-heavy range).
- **Hand EV** (Avg: 0.41): Expected Value of the hand normalized by Pot.
- **Hand Vuln** (Avg: 0.02): Probability (0-1) that hand is currently ahead but will lose by river.
- **MDF (Decision)** (Avg: 0.65): Decision-based MDF = Pot / (Pot + Bet). The break-even equity required to call.
- **percentage of hand in own range** (Avg: 0.01): Percentage of hand in own range
- **Regret: Call-Raise** (Avg: 0.34): EV(Call) - EV(Raise). Positive = Call is better. Diff in EV (Action A - Action B)
normalized by Pot.

=== DECISION LOGIC TRACE ===
Legend:
- [Path]: format is 'Role [Hand] -> PathID -> IntendedAction'. Indicates which internal branch was taken and the final
action intent.
- (Reach): The probability of the hand actually reaching this node vs. surviving to the next node.
- R: [0:xx 1:xx ...]: Router Probabilities. The internal neural network's confidence distribution across Branch 0, 1,
and 2. Shows how 'split' or 'certain' the decision was.
```

661

```

- [Summary Impact]: Shows which summaries pushed the router towards specific branches.
+-- NODE #0
[Path] Hero [8h8c] -> P2->Check (Reach: 100%->52%) | R: [0:48 1:00 2:51]
[Path] Shadow1 [JsTs] -> P2->Bet (Reach: 100%->80%) | R: [0:20 1:00 2:79]
[Path] Shadow2 [KsQh] -> P0->Check (Reach: 100%->83%) | R: [0:82 1:00 2:17]
[Path] Shadow3 [QsJs] -> P2->Bet (Reach: 100%->84%) | R: [0:16 1:00 2:83]
[Summary Impact Analysis - Node #0]
| Summary | Avg | Hero Impact (Bet-Chk) | Shadow1 | Shadow2 |
|-----|-----|-----|-----|-----|
| Shadow3 | | | | |
| :-----: | :-----: | :-----: | :-----: | :-----: |
| :-----: | | | | |
| MDF (Decision) | 0.65 | ** -1.00** (+6.3) | -1.00 | -1.00 |
| -1.00 | | | | |
| Rank in Range (EQ) | 0.50 | ** 0.43** (-6.0) | 0.55 | 0.38 | 0.59
| | | | |
| Phys: Kicker | 0.65 | ** 0.57** (+2.1) | 0.71 | 0.86 | 0.79
| | | | |
| Hand EQ | 0.49 | ** 0.41** (-1.6) | 0.48 | 0.40 | 0.52
| | | | |
| Hand EV | 0.41 | ** 0.29** (-0.2) | 0.71 | 0.29 | 0.75
| | | | |

+-- NODE #1
[Path] Shadow2 [KsQh] -> P2->Check (Reach: 83%->100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #1 (Analysis Focus: Shadow2 - Hero diverged)]
| Summary | Avg | Shadow2 Impact (Bet-(Chk)) |
|-----|-----|-----|
| Hand Blocker | 3.62 | ** 4.00** (+37.9) |
| Hero EV Adv | -0.19 | ** 0.17** (-33.3) |
| Hand EV | 0.41 | ** 0.29** (+26.8) |
| Vuln - Range Avg | 0.00 | ** -0.06** (+13.8) |
| Hand EQ | 0.49 | ** 0.40** (+3.3) |

+-- NODE #6
[Path] Shadow2 [KsQh] -> P2->Check (Reach: 83%->100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #6 (Analysis Focus: Shadow2 - Hero diverged)]
| Summary | Avg | Shadow2 Impact (Bet-(Chk)) |
|-----|-----|-----|
| Hand EV | 0.41 | ** 0.29** (+46.8) |
| Hero EQ Adv | -0.03 | ** 0.04** (-12.5) |
| Phys: Kicker | 0.65 | ** 0.86** (+7.0) |
| Bluff Efficiency | 0.61 | ** 0.56** (+6.8) |
| EQ - Nut Range Avg | 0.27 | ** 0.16** (+5.9) |

+-- NODE #21
[Path] Shadow2 [KsQh] -> P0->Check (Reach: 83%->99%) | R: [0:99 1:00 2:00]
[Summary Impact Analysis - Node #21 (Analysis Focus: Shadow2 - Hero diverged)]
| Summary | Avg | Shadow2 Impact (Chk-Bet) |
|-----|-----|-----|
| Phys: Kicker | 0.65 | ** 0.86** (+17.9) |
| Vuln - Range Avg | 0.00 | ** -0.06** (-1.2) |
| EQ - Nut MDF | 0.25 | ** 0.17** (-0.7) |
| OOP Rng EV | 0.59 | ** 0.42** (+0.6) |
| Regret: Fold-Call | 0.89 | ** 1.16** (+0.4) |

+-- LEAF #64 (Static Prototype)
-> Strat: [Check:99%] | Samples: [Shadow2:KsQh] (Reach:82.3%)

+-- NODE #3
[Path] Hero [8h8c] -> P0->Check (Reach: 52%->79%) | R: [0:78 1:00 2:21]
[Path] Shadow1 [JsTs] -> P0->Bet (Reach: 80%->80%) | R: [0:79 1:00 2:20]
[Path] Shadow3 [QsJs] -> P0->Bet (Reach: 84%->70%) | R: [0:70 1:00 2:29]
[Summary Impact Analysis - Node #3]
| Summary | Avg | Hero Impact (Chk-Bet) | Shadow1 | Shadow3 |
|-----|-----|-----|-----|-----|
| Regret: Call-Raise | 0.34 | ** 0.33** (+6.1) | 0.33 | 0.27 |
| EQ - MDF | -0.20 | ** -1.00** (-2.0) | -1.00 | -1.00 |
| IP Rng: Set | 0.03 | ** 0.05** (+1.2) | 0.05 | 0.05 |
| Hand: AceHigh | 0.21 | ** 0.00** (-0.6) | 0.00 | 0.00 |
| Bluff Efficiency | 0.61 | ** 0.56** (+0.3) | 0.56 | 0.56 |

+-- NODE #10
[Path] Hero [8h8c] -> P2->Check (Reach: 41%->100%) | R: [0:00 1:00 2:99]
[Path] Shadow1 [JsTs] -> P0->Bet (Reach: 63%->100%) | R: [0:99 1:00 2:00]
[Path] Shadow3 [QsJs] -> P0->Bet (Reach: 59%->100%) | R: [0:99 1:00 2:00]
[Summary Impact Analysis - Node #10]
| Summary | Avg | Hero Impact (Bet-(Chk)) | Shadow1 | Shadow3 |
|-----|-----|-----|-----|-----|
| :-----: | :-----: | :-----: | :-----: | :-----: |
| EQ - Range Avg | -0.00 | ** -0.11** (+14.0) | -0.04 | 0.00 |
| Hand Vuln | 0.02 | ** 0.18** (+5.0) | -0.36 | -0.40 |
| Nut EQ Adv (Hero-Vill) | 0.05 | ** -0.04** (-1.2) | -0.04 | -0.04 |
| EQ - MDF | -0.20 | ** -1.00** (-0.4) | -1.00 | -1.00 |
| Hand Unblocker | 3.62 | ** 1.00** (-0.1) | 3.00 | 3.00 |

+-- NODE #31
[Path] Shadow1 [JsTs] -> P2->Bet (Reach: 63%->90%) | R: [0:09 1:00 2:89]
[Path] Shadow3 [QsJs] -> P2->Bet (Reach: 59%->98%) | R: [0:02 1:00 2:97]
[Summary Impact Analysis - Node #31 (Analysis Focus: Shadow1 - Hero diverged)]
| Summary | Avg | Shadow1 Impact (Bet-Chk) | Shadow3 |
|-----|-----|-----|-----|

```



```

| | Regret: Fold-Call | 0.89 | **0.47** (+8.4) | 0.40 |
| | Block Opp EQ | 0.51 | **0.50** (-7.8) | 0.53 |
| | Bet Morph (EQ Diff) | 0.30 | **0.43** (+1.6) | 0.43 |
| | Nut EQ Adv (Hero-Vill) | 0.05 | **0.04** (+1.5) | -0.04 |
| | Hand: NoDraw | 0.52 | **0.00** (-0.4) | 0.00 |
+-- LEAF #96 (Static Prototype)
| -> Strat: [Bet:99%] | Samples: [Shadow1:JsTs] (Reach:56.8%), [Shadow3:QsJs] (Reach:57.8%)
+-- NODE #33
[Path] Hero [8h8c] -> P2->Check (Reach: 41%>99%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #33]
| Summary | Avg | Hero Impact (Bet-Chk) |
| :-----: | :-----: | :-----: |
| Opp Card Scarcity (Avg) | 0.06 | **0.04** (+12.1) |
| Opp Card Scarcity (C1) | 0.07 | **0.04** (-2.1) |
| MDF (Decision) | 0.65 | **1.00** (-1.8) |
| OOP Rng EV | 0.59 | **0.42** (-0.9) |
| percentage of hand in own range | 0.01 | **0.01** (+0.4) |
+-- LEAF #102 (Static Prototype)
| -> Strat: [Bet:99%] | Samples: [Hero:8h8c] (Reach:40.7%)

=== FINAL STRATEGY vs GTO COMPARISON (Training Data) ===
| Hand | Role | Final Model Strat | GTO Target Strat |
| :-----: | :-----: | :-----: | :-----: |
| 8h8c | Hero | C:0.65 F:0.00 B:0.35 | C:0.63 F:0.00 B:0.37 |
| JsTs | Shadow1 | C:0.48 F:0.00 B:0.52 | C:0.45 F:0.00 B:0.55 |
| KsQh | Shadow2 | C:0.60 F:0.00 B:0.40 | C:0.79 F:0.00 B:0.21 |
| QsJs | Shadow3 | C:0.46 F:0.00 B:0.54 | C:0.47 F:0.00 B:0.53 |

## 3. Test Case: The Unseen Hand
Consider a new hand in the same scenario:
- **Hand**: 8s8c
- **Key Difference**: (Inspect the summaries yourself compared to Hero)
**Question**: Based on the logic learned above, what is the optimal action for this hand? Explain why using the model's decision boundaries.

# Analysis Instructions
1. **Layer-by-Layer Review**: For each NODE, explain how the specific **Impact values** determined the branch choice.
2. **Impact Comparison**: Compare Hero's summary values vs. Global Avg and Shadows.
3. **Consistency Check**: Use Section 3 (Test Case) to verify if the logic you reverse-engineered applies to an unseen hand.
4. **Solve Test Case**: Provide your answer for the Unseen Hand in Section 3.
- Key Summary Values of Test Case (Hierarchical):
--- **Depth 0** - Top 5 Drivers:
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: +6.30
- **Rank in Range (EQ)**: 0.45 (Avg: 0.50) | Impact: -4.42
- **Phys: Kicker**: 0.57 (Avg: 0.65) | Impact: +2.13
- **Hand EV**: 0.33 (Avg: 0.41) | Impact: +0.35
- **Hand EQ**: 0.44 (Avg: 0.49) | Impact: +0.09
--- **Depth 1** - Top 4 Drivers:
- **Regret: Call-Raise**: 0.39 (Avg: 0.34) | Impact: +8.67
- **EQ - MDF**: -1.00 (Avg: -0.20) | Impact: -2.04
- **IP Rng: Set**: 0.05 (Avg: 0.03) | Impact: +1.19
- **Hand: AceHigh**: 0.00 (Avg: 0.21) | Impact: -0.61
--- **Depth 2** - Top 3 Drivers:
- **EQ - Range Avg**: -0.08 (Avg: -0.00) | Impact: +12.29
- **Hand Vuln**: 0.16 (Avg: 0.02) | Impact: +2.25
- **Nut EQ Adv (Hero-Vill)**: -0.04 (Avg: 0.05) | Impact: -1.16
--- **Depth 3** - Top 2 Drivers:
- **Opp Card Scarcity (Avg)**: 0.03 (Avg: 0.06) | Impact: +18.36
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -1.85

# Output Format (Markdown)
## 1. Strategy Summary
(One concise sentence)

## 2. Path Logic & Divergence
- **Node #0**: ...

## 3. Test Case Solution
- **Predicted Mixed Strategy**: ...
- **Reasoning**: ...

```

663

664 Answer 2:

```

# Test Case Solution (Sample 220974)
Hand: 8s8c
Scenario: BTN_vs_BB_3B

## 1. Ground Truth Strategies
- **GTO Strategy**: C:0.56 F:0.00 B:0.44
- **Model Strategy**: C:0.63 F:0.00 B:0.37
## 2. Actual Decision Trace
+-- NODE #0
[Path] TestHand [8s8c] -> P2->Check (Reach: 100%>66%) | R: [0:34 1:00 2:65]
[Summary Impact Analysis - Node #0 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-Chk) |
| :-----: | :-----: | :-----: |

```

665

```

| MDF (Decision)          | 0.65 | ** -1.00** (+6.3) |
| Rank in Range (EQ)      | 0.50 | ** 0.45** (-4.4)  |
| Phys: Kicker            | 0.65 | ** 0.57** (+2.1)  |
| Hand EV                 | 0.41 | ** 0.33** (+0.4)  |
| Hand EQ                 | 0.49 | ** 0.44** (+0.1)  |
+-- NODE #3
[Path] TestHand [8s8c] -> P0->Check (Reach: 66%->87%) | R: [0:87 1:00 2:12]
[Summary Impact Analysis - Node #3 (Analysis Focus: TestHand - Hero diverged)]
| Summary          | Avg | TestHand Impact (Chk-Bet) |
| :-----:        | :---: | :-----: |
| Regret: Call-Raise | 0.34 | ** 0.39** (+8.7) |
| EQ - MDF          | -0.20 | ** -1.00** (-2.0) |
| IP Rng: Set       | 0.03 | ** 0.05** (+1.2) |
| Hand: AceHigh     | 0.21 | ** 0.00** (-0.6) |
| Bluff Efficiency  | 0.61 | ** 0.56** (+0.3) |
+-- NODE #10
[Path] TestHand [8s8c] -> P2->Check (Reach: 57%->100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #10 (Analysis Focus: TestHand - Hero diverged)]
| Summary          | Avg | TestHand Impact (Bet-(Chk)) |
| :-----:        | :---: | :-----: |
| EQ - Range Avg    | -0.00 | ** -0.08** (+12.3) |
| Hand Vuln         | 0.02 | ** 0.16** (+2.2) |
| Nut EQ Adv (Hero-Vill) | 0.05 | ** -0.04** (-1.2) |
| EQ - MDF          | -0.20 | ** -1.00** (-0.4) |
| Hand Unblocker    | 3.62 | ** 1.00** (-0.1) |
+-- NODE #33
[Path] TestHand [8s8c] -> P2->Check (Reach: 57%->100%) | R: [0:00 1:00 2:99]
[Summary Impact Analysis - Node #33 (Analysis Focus: TestHand - Hero diverged)]
| Summary          | Avg | TestHand Impact (Bet-(Chk)) |
| :-----:        | :---: | :-----: |
| Opp Card Scarcity (Avg) | 0.06 | ** 0.03** (+18.4) |
| MDF (Decision)        | 0.65 | ** -1.00** (-1.8) |
| Opp Card Scarcity (C1) | 0.07 | ** 0.03** (-1.2) |
| OOP Rng EV            | 0.59 | ** 0.42** (-0.9) |
| percentage of hand in own range | 0.01 | ** 0.01** (+0.4) |
+-- LEAF #102 (Static Prototype)
-> Strat: [Bet:99%] | Samples: [TestHand:8s8c] (Reach:57.4%)

## 3. Key Summary Drivers (Path-Dependent)

### ** (Depth 0) ** - Top 5 Drivers:
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: +6.30
- **Rank in Range (EQ)**: 0.45 (Avg: 0.50) | Impact: -4.42
- **Phys: Kicker**: 0.57 (Avg: 0.65) | Impact: +2.13
- **Hand EV**: 0.33 (Avg: 0.41) | Impact: +0.35
- **Hand EQ**: 0.44 (Avg: 0.49) | Impact: +0.09

### ** (Depth 1) ** - Top 4 Drivers:
- **Regret: Call-Raise**: 0.39 (Avg: 0.34) | Impact: +8.67
- **EQ - MDF**: -1.00 (Avg: -0.20) | Impact: -2.04
- **IP Rng: Set**: 0.05 (Avg: 0.03) | Impact: +1.19
- **Hand: AceHigh**: 0.00 (Avg: 0.21) | Impact: -0.61

### ** (Depth 2) ** - Top 3 Drivers:
- **EQ - Range Avg**: -0.08 (Avg: -0.00) | Impact: +12.29
- **Hand Vuln**: 0.16 (Avg: 0.02) | Impact: +2.25
- **Nut EQ Adv (Hero-Vill)**: -0.04 (Avg: 0.05) | Impact: -1.16

### ** (Depth 3) ** - Top 2 Drivers:
- **Opp Card Scarcity (Avg)**: 0.03 (Avg: 0.06) | Impact: +18.36
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -1.85

```

666

667 LLM Response 2:

```

## 1. Strategy Summary

The AI plays as a mixed strategy (Check ~65% / Bet ~35%), primarily checking to realize equity with marginal
showdown value, but betting occasionally to deny equity (Protection) from overcards when blocker configurations
allow.

---

## 2. Path Logic & Divergence

### Node #0: The Strategic Split (Protection vs. Pot Control)

* Logic: The AI is torn between two competing incentives.
* Pull to Check (Path P0): Driven by Rank in Range (0.43) and Hand EQ (0.49). is a mediocre hand (
percentile) on an board. It cannot withstand heavy heat (indicated by negative MDF impact), favoring a low-
variance line.
* Pull to Bet (Path P2): Driven by Phys: Kicker (+2.1 Impact) and MDF Decision (+6.3 Impact). Despite being
marginal, is highly Vulnerable to overcards. The positive MDF impact here suggests that checking gives up
too much equity to random live cards, incentivizing a "merging" bet to fold out overcards (KQ, KJ, QJ).

```

668

```

* Divergence: Shadow hands like (Draws) commit fully to Path P2 (Bet) because they have high Hand EV and Bluff Efficiency. Hero () lacks the "Natural Bluff" properties, causing the split.

Node #33: The Blocker Confirmation (Why Bet?)

* Path: Inside the "Bet" branch (P2), Hero reaches Node #33.
* Logic: The decision to finalize the Bet is driven by Opp Card Scarcity (Avg) (+12.1 Impact) and Hand Vuln.

* The AI recognizes that is incredibly vulnerable (Impact: +5.0 in previous Node #10).
* The Scarcity metric indicates that blocks key portions of the opponent's range (likely straights like 98, 85, or T987 combos). By blocking these static hands, Hero increases the likelihood the opponent holds "Air" or dominatable pairs, making the Protection Bet profitable.

Node #3: The Showdown Realization (Why Check?)

* Path: Inside the "Check" branch (P0).
* Logic: Here, Regret: Call-Raise (+6.1 Impact) dominates. The AI calculates that if it Bets and gets Raised, it faces a catastrophic EV loss. By Checking, it avoids this "Regret" scenario, successfully realizing its equity against the opponent's garbage range without paying to see the turn.

---

3. Test Case Solution

Test Hand: (Holding the Spade Blocker)
Context: Same Board ()

Predicted Mixed Strategy

Bet: 55% / Check: 45% (Significant increase in Betting frequency compared to )

Reasoning

The decision hinges on the Opp Card Scarcity and Hand Unblocker mechanics acting on the Flush Draw board ().

1. Increased Fold Equity (Blocker Effect):
* In the provided Test Case data, Opp Card Scarcity has a massive impact of +18.36 towards Betting (Depth 3), significantly higher than the +12.1 seen for .
* Why? Holding the blocks the opponent's "Continuing Range" (Flush Draws like ).
* By blocking the hands that naturally call, the opponent's remaining range becomes more polarized between "Air" (which folds) and "Ax" (which crushes us).

2. Protection Incentive:
* Since is a vulnerable hand, it benefits immensely from immediate folds (denying equity to random overcards like ).
* With the , the opponent folds more frequently (because we block their calling draws). This makes the Protection Bet more efficient than with .

3. Conclusion: The logic shifts from "Check to avoid draws" (Red) to "Bet to capitalize on fold equity" (Blue). The transforms the hand from a passive showdown hand into a more effective semi-bluff/protection candidate.

```

669

670 Template 3:

```

# Role
You are an elite Poker AI Strategist and GTO Solver Analyst. Your task is to reverse-engineer the decision-making process exposed by a Mixed-Strategy Decision Tree (MDT) model.

# Task
Analyze the provided trace logic for the specific hand provided below. You must explain why the AI chose this specific path over others, using Poker Theory concepts (Equity Realization, Blockers, Range Morphology).

# Semantic Definitions (Crucial)
The trace uses specific tags based on Equity (EQ) vs. Opponent Range:
- (Value): Aggressive action with High EQ (>65%).
- (Bluff): Aggressive action with Low EQ (<35%).
- (Thin Value/Protect): Aggressive action with Moderate EQ.
- (S.D. Value): Passive action (Check/Call) with enough EQ to win at showdown but not enough to bet for value.
- (Give Up): Passive action with near-zero EQ.
- (Trap): Passive action with Nut-class EQ.

# Input Data
=== SCENARIO CONTEXT ===
Scenario : BTN_vs_BB_3B
Board : QsJh5d
History : k
Player : IP
Hand : 7c6c
Hand Wgt : 0.0048 (Prob in Range)
Pot : 26.5
Actions : [Check, Fold, Bet(0.67x)]

=== SUMMARY GLOSSARY (Definitions for current path) ===

```

671

```

--- Strategic Adv ---
- **Hero EV Adv** (Avg: -0.19): **RELATIVE SUMMARY (Hero - Villain)**. Difference in EV Pot Share. Positive = Favors Hero.
- **Hand EQ** (Avg: 0.49): Raw Equity (0.0-1.0) against opponent's current range.
- **EQ if BetBet** (Avg: 0.35): Hypothetical Equity if the game line goes BetBet.
- **EQ - MDF** (Avg: -0.20): Hand Equity minus Decision MDF. >0 implies raw equity is sufficient to call.
- **Rank in Range (EQ)** (Avg: 0.50): Percentile of Hand Equity (0-1).

--- Hand Physics ---
- **Phys: Kicker** (Avg: 0.65): Normalized Kicker Strength (Rank / 14.0). Ace=1.0, King=0.92, ..., 2=0.14. Crucial for domination issues (e.g., distinguishing Top Pair Top Kicker from Top Pair Weak Kicker).

--- Other ---
- **Hand: NoDraw** (Avg: 0.52): Does the hand have a NoDraw? (1.0 = Yes, 0.0 = No).
- **Hand Blocker** (Avg: 3.62): Score representing how much this hand blocks opponent's continuing range.
- **Block Self Realiz** (Avg: 0.49): Self-Blocker: Negative effect where our cards block opponent's folding range (bad for bluffs).
- **Opp Card Scarcity (C1)** (Avg: 0.07): Scarcity Effect of Card 1. High Value = We block a key card for the opponent (e.g. holding an Ace vs an Ace-heavy range).
- **Hand EV** (Avg: 0.41): Expected Value of the hand normalized by Pot.
- **MDF (Decision)** (Avg: 0.65): Decision-based MDF = Pot / (Pot + Bet). The break-even equity required to call.
- **Blocker - Range Avg** (Avg: 0.00): Hand Blocker Score minus Range Average. Positive = Better than average blockers.
- **Vuln - Range Avg** (Avg: 0.00): Hand Vulnerability minus Range Average. Positive = More vulnerable than average (needs protection).
- **percentage of hand in own range** (Avg: 0.01): Percentage of hand in own range
- **Regret: Fold-Call** (Avg: 0.89): EV(Fold) - EV(Call). Positive = Fold is better. Diff in EV (Action A - Action B) normalized by Pot.

=== DECISION LOGIC TRACE ===
Legend:
- [Path]: format is 'Role [Hand] -> PathID -> IntendedAction'. Indicates which internal branch was taken and the final action intent.
- (Reach): The probability of the hand actually reaching this node vs. surviving to the next node.
- R: [0:xx 1:xx ...]: Router Probabilities. The internal neural network's confidence distribution across Branch 0, 1, and 2. Shows how 'split' or 'certain' the decision was.
- [Summary Impact]: Shows which summaries pushed the router towards specific branches.
+-- NODE #0
  [Path] Hero [7c6c] -> P0->Check (Reach: 100%->99%) | R: [0:98 1:00 2:01]
  [Path] Shadow1 [8s7s] -> P0->Check (Reach: 100%->95%) | R: [0:95 1:00 2:04]
  [Path] Shadow2 [7s6s] -> P0->Bet (Reach: 100%->96%) | R: [0:95 1:00 2:04]
  [Path] Shadow3 [8h7h] -> P0->Bet (Reach: 100%->95%) | R: [0:94 1:00 2:05]
  [Summary Impact Analysis - Node #0]
  | Summary | Avg | Hero Impact (Chk-Bet) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | Rank in Range (EQ) | 0.50 | **0.00** (+31.9) | 0.03 | 0.01 | 0.04
  | Hand EQ | 0.49 | **0.16** (+25.4) | 0.24 | 0.20 | 0.24
  | MDF (Decision) | 0.65 | **-1.00** (-6.3) | -1.00 | -1.00 |
  | Phys: Kicker | 0.65 | **0.43** (-4.1) | 0.50 | 0.43 | 0.50
  | Hand EV | 0.41 | **0.29** (+0.4) | 0.34 | 0.35 | 0.34
+-- NODE #1
  [Path] Hero [7c6c] -> P0->Check (Reach: 99%->54%) | R: [0:54 1:00 2:45]
  [Path] Shadow1 [8s7s] -> P2->Check (Reach: 95%->52%) | R: [0:47 1:00 2:52]
  [Path] Shadow2 [7s6s] -> P2->Bet (Reach: 96%->52%) | R: [0:47 1:00 2:52]
  [Path] Shadow3 [8h7h] -> P2->Bet (Reach: 95%->52%) | R: [0:47 1:00 2:52]
  [Summary Impact Analysis - Node #1]
  | Summary | Avg | Hero Impact (Chk-Bet) | Shadow1 | Shadow2 |
  | :-----: | :-----: | :-----: | :-----: | :-----: |
  | Hero EV Adv | -0.19 | **0.27** (+44.9) | 0.27 | 0.27 |
  | Hand EV | 0.41 | **0.29** (-25.8) | 0.34 | 0.35 |
  | Vuln - Range Avg | 0.00 | **-0.08** (-14.4) | -0.15 | -0.11 |
  | Hand EQ | 0.49 | **0.16** (+1.9) | 0.24 | 0.20 |
  | Hand Blocker | 3.62 | **1.00** (-1.7) | 1.00 | 1.00 |
+-- NODE #4
  [Path] Hero [7c6c] -> P2->Check (Reach: 54%->100%) | R: [0:00 1:00 2:99]
  [Summary Impact Analysis - Node #4]
  | Summary | Avg | Hero Impact (Bet-(Chk)) |
  | :-----: | :-----: | :-----: |
  | percentage of hand in own range | 0.01 | **0.00** (+68.5) |
  | Block Self Realiz | 0.49 | **0.39** (-4.7) |
  | Blocker - Range Avg | 0.00 | **-0.27** (+2.0) |
  | EQ if BetBet | 0.35 | **0.12** (-1.1) |
  | EQ - MDF | -0.20 | **-1.00** (-0.6) |
+-- NODE #15
  [Path] Hero [7c6c] -> P0->Check (Reach: 54%->100%) | R: [0:99 1:00 2:00]
  [Summary Impact Analysis - Node #15]

```

```

|      | Summary          | Avg      | Hero Impact (Chk-Fld) |
| :-----: | :-----: | :-----: |
| Hero EV Adv | -0.19 | **0.27** (-22.6) |
| Hand: NoDraw | 0.52 | **0.00** (+15.7) |
| EQ if BetBet | 0.35 | **0.12** (+12.0) |
| Opp Card Scarcity (C1) | 0.07 | **0.04** (-9.1) |
| Regret: Fold-Call | 0.89 | **1.51** (+1.0) |
+-- LEAF #46 (Static Prototype)
   -> Strat: [Check:13%/Bet:86%] | Samples: [Hero:7c6c] (Reach:53.6%)
+-- NODE #6
   [Path] Shadow1 [8s7s] -> P2->Check (Reach: 50%>42%) | R: [0:28 1:28 2:42]
   [Path] Shadow2 [7s6s] -> P2->Bet (Reach: 50%>54%) | R: [0:23 1:23 2:53]
   [Path] Shadow3 [8h7h] -> P2->Bet (Reach: 50%>43%) | R: [0:28 1:28 2:42]
   [Summary Impact Analysis - Node #6 (Analysis Focus: Shadow1 - Hero diverged)]
   | Summary          | Avg      | Shadow1 Impact (Bet-Chk) | Shadow2 | Shadow3
   | :-----: | :-----: | :-----: | :-----: |
+-----+
   | Hand EV          | 0.41 | **0.34** (+54.6) | 0.35 | 0.34
   |
   | Hero EQ Adv      | -0.03 | **0.17** (-35.6) | 0.17 | 0.17
   |
   | Phys: Kicker      | 0.65 | **0.50** (+7.9) | 0.43 | 0.50
   |
   | Bluff Efficiency  | 0.61 | **0.63** (+6.8) | 0.63 | 0.63
   |
   | EQ - Nut Range Avg | 0.27 | **-0.02** (+4.4) | -0.06 | -0.01
   |
+-- NODE #21
   [Path] Shadow1 [8s7s] -> P2->Check (Reach: 21%>97%) | R: [0:03 1:00 2:96]
   [Path] Shadow2 [7s6s] -> P2->Bet (Reach: 27%>98%) | R: [0:01 1:00 2:98]
   [Path] Shadow3 [8h7h] -> P2->Bet (Reach: 21%>97%) | R: [0:02 1:00 2:97]
   [Summary Impact Analysis - Node #21 (Analysis Focus: Shadow1 - Hero diverged)]
   | Summary          | Avg      | Shadow1 Impact (Bet-Chk) | Shadow2 | Shadow3
   | :-----: | :-----: | :-----: | :-----: |
+-----+
   | EQ - Nut MDF      | 0.25 | **0.00** (+6.5) | -0.04 | 0.01
   |
   | Regret: Fold-Call | 0.89 | **1.43** (-6.5) | 1.42 | 1.43
   |
   | Phys: Kicker      | 0.65 | **0.50** (+3.8) | 0.43 | 0.50
   |
   | Vuln - Range Avg  | 0.00 | **-0.15** (+3.6) | -0.11 | -0.15
   |
   | OOP Rng EV        | 0.59 | **0.36** (-2.3) | 0.36 | 0.36
   |
+-- LEAF #66 (Static Prototype)
   -> Strat: [Check:99%] | Samples: [Shadow1:8s7s] (Reach:20.2%), [Shadow2:7s6s] (Reach:26.3%), [Shadow3:8h7h] (Reach:20.7%)

=== FINAL STRATEGY vs GTO COMPARISON (Training Data) ===
| Hand | Role | Final Model Strat | GTO Target Strat |
| :-----: | :-----: | :-----: | :-----: |
| 7c6c | Hero | C:0.71 F:0.00 B:0.29 | C:0.71 F:0.00 B:0.29 |
| 8s7s | Shadow1 | C:0.56 F:0.00 B:0.44 | C:0.59 F:0.00 B:0.41 |
| 7s6s | Shadow2 | C:0.42 F:0.00 B:0.58 | C:0.43 F:0.00 B:0.57 |
| 8h7h | Shadow3 | C:0.48 F:0.00 B:0.52 | C:0.42 F:0.00 B:0.58 |

## 3. Test Case: The Unseen Hand
Consider a new hand in the same scenario:
- **Hand**: 8d7d
- **Key Difference**: (Inspect the summaries yourself compared to Hero)
**Question**: Based on the logic learned above, what is the optimal action for this hand? Explain why using the model's decision boundaries.

# Analysis Instructions
1. **Layer-by-Layer Review**: For each NODE, explain how the specific **Impact values** determined the branch choice.
2. **Impact Comparison**: Compare Hero's summary values vs. Global Avg and Shadows.
3. **Consistency Check**: Use Section 3 (Test Case) to verify if the logic you reverse-engineered applies to an unseen hand.
4. **Solve Test Case**: Provide your answer for the Unseen Hand in Section 3.
   - Key Summary Values of Test Case (Hierarchical):
     --- ** (Depth 0) --- Top 5 Drivers:
     - **Rank in Range (EQ)**: 0.03 (Avg: 0.50) | Impact: +30.35
     - **Hand EQ**: 0.24 (Avg: 0.49) | Impact: +17.70
     - **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -6.30
     - **Phys: Kicker**: 0.50 (Avg: 0.65) | Impact: -3.15
     - **Hand EV**: 0.34 (Avg: 0.41) | Impact: -0.40
     --- ** (Depth 1) --- Top 4 Drivers:
     - **Hero EV Adv**: 0.27 (Avg: -0.19) | Impact: -44.88
     - **Hand EV**: 0.34 (Avg: 0.41) | Impact: +30.94
     - **Vuln - Range Avg**: -0.15 (Avg: 0.00) | Impact: +15.60
     - **Hand Blocker**: 1.00 (Avg: 3.62) | Impact: +1.67
     --- ** (Depth 2) --- Top 3 Drivers:
     - **Hand EV**: 0.34 (Avg: 0.41) | Impact: +54.60
     - **Hero EQ Adv**: 0.17 (Avg: -0.03) | Impact: -35.60
     - **Phys: Kicker**: 0.50 (Avg: 0.65) | Impact: +7.90
     --- ** (Depth 3) --- Top 2 Drivers:
     - **EQ - Nut MDF**: 0.00 (Avg: 0.25) | Impact: +6.50
     - **Regret: Fold-Call**: 1.43 (Avg: 0.89) | Impact: -6.47

```

```

# Output Format (Markdown)
## 1. Strategy Summary
(One concise sentence)

## 2. Path Logic & Divergence
- **Node #0**: ...

## 3. Test Case Solution
- **Predicted Mixed Strategy**: ...
- **Reasoning**: ...

```

674

675 Answer 3:

```

# Test Case Solution (Sample 602381)
Hand: 8d7d
Scenario: BTN_vs_BB_3B

## 1. Ground Truth Strategies
- **GT0 Strategy**: C:0.56 F:0.00 B:0.44
- **Model Strategy**: C:0.56 F:0.00 B:0.44
## 2. Actual Decision Trace
+-- NODE #0
[Path] TestHand [8d7d] -> P0->Check (Reach: 100%>95%) | R: [0:95 1:00 2:04]
[Summary Impact Analysis - Node #0 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Chk-Bet) |
| :-----: | :-----: | :-----: |
| Rank in Range (EQ) | 0.50 | **0.03** (+30.4) |
| Hand EQ | 0.49 | **0.24** (+17.7) |
| MDF (Decision) | 0.65 | **-1.00** (-6.3) |
| Phys: Kicker | 0.65 | **0.50** (-3.1) |
| Hand EV | 0.41 | **0.34** (-0.4) |
+-- NODE #1
[Path] TestHand [8d7d] -> P2->Check (Reach: 95%>52%) | R: [0:47 1:00 2:52]
[Summary Impact Analysis - Node #1 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-Chk) |
| :-----: | :-----: | :-----: |
| Hero EV Adv | -0.19 | **0.27** (-44.9) |
| Hand EV | 0.41 | **0.34** (+30.9) |
| Vuln - Range Avg | 0.00 | **-0.15** (+15.6) |
| Hand Blocker | 3.62 | **1.00** (+1.7) |
| Hand EQ | 0.49 | **0.24** (-0.5) |
+-- NODE #6
[Path] TestHand [8d7d] -> P2->Check (Reach: 50%>42%) | R: [0:28 1:28 2:42]
[Summary Impact Analysis - Node #6 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-Chk) |
| :-----: | :-----: | :-----: |
| Hand EV | 0.41 | **0.34** (+54.6) |
| Hero EQ Adv | -0.03 | **0.17** (-35.6) |
| Phys: Kicker | 0.65 | **0.50** (+7.9) |
| Bluff Efficiency | 0.61 | **0.63** (+6.8) |
| EQ - Nut Range Avg | 0.27 | **-0.01** (+4.4) |
+-- NODE #21
[Path] TestHand [8d7d] -> P2->Check (Reach: 21%>97%) | R: [0:03 1:00 2:96]
[Summary Impact Analysis - Node #21 (Analysis Focus: TestHand - Hero diverged)]
| Summary | Avg | TestHand Impact (Bet-Chk) |
| :-----: | :-----: | :-----: |
| EQ - Nut MDF | 0.25 | **0.00** (+6.5) |
| Regret: Fold-Call | 0.89 | **1.43** (-6.5) |
| Phys: Kicker | 0.65 | **0.50** (+3.8) |
| Vuln - Range Avg | 0.00 | **-0.15** (+3.6) |
| OOP Rng EV | 0.59 | **0.36** (-2.3) |
+-- LEAF #66 (Static Prototype)
-> Strat: [Check:99%] | Samples: [TestHand:8d7d] (Reach:20.3%)

## 3. Key Summary Drivers (Path-Dependent)

### ** (Depth 0) - Top 5 Drivers:
- **Rank in Range (EQ)**: 0.03 (Avg: 0.50) | Impact: +30.35
- **Hand EQ**: 0.24 (Avg: 0.49) | Impact: +17.70
- **MDF (Decision)**: -1.00 (Avg: 0.65) | Impact: -6.30
- **Phys: Kicker**: 0.50 (Avg: 0.65) | Impact: -3.15
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: -0.40

### ** (Depth 1) - Top 4 Drivers:
- **Hero EV Adv**: 0.27 (Avg: -0.19) | Impact: -44.88
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +30.94
- **Vuln - Range Avg**: -0.15 (Avg: 0.00) | Impact: +15.60
- **Hand Blocker**: 1.00 (Avg: 3.62) | Impact: +1.67

### ** (Depth 2) - Top 3 Drivers:
- **Hand EV**: 0.34 (Avg: 0.41) | Impact: +54.60
- **Hero EQ Adv**: 0.17 (Avg: -0.03) | Impact: -35.60
- **Phys: Kicker**: 0.50 (Avg: 0.65) | Impact: +7.90

### ** (Depth 3) - Top 2 Drivers:

```

676

```
- **EQ - Nut MDF**: 0.00 (Avg: 0.25) | Impact: +6.50
- **Regret: Fold-Call**: 1.43 (Avg: 0.89) | Impact: -6.47
```

677

678 LLM Response 3:

1. Strategy Summary

The model applies a polarized bluffing strategy where "Pure Air" (7c6c, 16% EQ) mostly gives up but occasionally bluffs, whereas "Semi-Bluffs" (8d7d, 24% EQ) bet more frequently (40-50%) to realize their higher equity and draw potential.

2. Path Logic & Divergence

The decision tree separates hands based on their potential to improve (Equity) and their utility as bluffs (Blockers/EV).

* **Node #0 (The "Worthlessness" Filter):**

* **Logic:** The router isolates hands with very low **Rank in Range** (<0.05) and negative **MDF** metrics. Both Hero (7c6c) and Shadows (87/76) are identified as the "bottom of range," filtering out all value hands.

* **Result:** All low-equity hands proceed to Node #1.

* **Node #1 (The "Draw vs. Air" Split):**

* **Logic:** This is the critical divergence point. The router uses **Hand EV** and **Hand EQ** to distinguish "Pure Trash" from "Draws."

* **Hero (7c6c):** Has significantly lower **Hand EQ** (0.16) and **Hand EV** (0.29). It is routed primarily to **Path 0** (Node #15), treated as a "Pure Bluff" candidate that mostly checks but takes a high-frequency betting line (Leaf #46) when it *does* bluff to maximize fold equity.

* **Shadows (8s7s, etc.):** Have higher **Hand EQ** (0.24) due to gutshot/backdoor potential. They are routed to **Path 2** (Node #6), entering a "Semi-Bluff" line.

* **Theory:** The model adheres to the GTO principle that *zero-equity hands* (Hero) are often indifferent (mix check/bet or check/fold), while *low-equity draws* (Shadows) are incentivized to bet aggressively to realize equity or fold out better hands.

* **Node #6 & #21 (The Semi-Bluff Mix):**

* **Logic:** The Shadows (and the Test Case) traverse these nodes. Summaries such as **Bluff Efficiency** (+6.8 impact) and **Regret: Fold-Call** drive the decision.

* **Result:** Shadows reach a mix of checking (Leaf #66) and betting (inferred branch), resulting in a higher overall betting frequency (~45-55%) compared to Hero (~29%).

3. Test Case Solution

* **Predicted Mixed Strategy:** **Check: ~54% / Bet: ~46%** (Aggressive Semi-Bluff Mix)

* **Reasoning:**

* **Summary Matching:** The Test Case hand (8d7d) has **Hand EQ (0.24)** and **Hand EV (0.34)**, which are close to the continuing shadow hands (8s7s, 8h7h) and distinct from Hero (0.16 / 0.29).

* **Strategic Classification:** Unlike Hero (7c6c), which is "Pure Air," 8d7d holds a gutshot straight draw (8-7 on Q-J -5 requires 9-T).

* **Path Prediction:** Because its "Depth 1" drivers (Hand EV, Vuln) match the Shadows perfectly, 8d7d will be routed away from Hero's "Pure Bluff" path at Node #1 and into the "Semi-Bluff" path (Node #6).

* **Outcome:** It will adopt the Shadow strategy profile: a balanced mix favoring aggression more than Hero, utilizing its equity to semi-bluff.

679

680 NeurIPS Paper Checklist

681 1. **Claims.** Do the main claims made in the abstract and introduction accurately reflect the
682 paper's contributions and scope?

683 [Yes] The abstract and introduction frame the work as solver-led articulation of reasoning
684 data for imperfect-information games with mixed Nash equilibria. The empirical evidence
685 is presented as an empirical instantiation in NLH, not as a claim of general equilibrium
686 solving or live poker deployment.

687 2. **Limitations.** Does the paper discuss the limitations of the work performed by the authors?

688 [Yes] The appendix discusses the method's dependence on mixed-strategy equilibria and its
689 offline analytical nature. While demonstrated only in NLH-style mixed-equilibrium settings,
690 this suggests a possible path toward solver-grounded articulation in other domains.

691 3. **Theory, assumptions, and proofs.** If you are including theoretical results, did you state the
692 full set of assumptions of all theoretical results, and did you include complete proofs of all
693 theoretical results?

[N/A] The paper formulates a problem setting and gives an empirical instantiation, but does not claim new theoretical guarantees or present formal theorems.

4. **Experimental result reproducibility.** If the contribution is a dataset or model, what steps did you take to make your results reproducible or verifiable?

[Yes] The paper specifies the solver-output interface, reasoning-rule criteria, NLH corpus construction, MDT objectives, SCCS search procedure, evaluation protocol, and prompt templates. The implementation code will be released, but the solver-generated dataset will not be publicly released because it is collected through a commercial API. And exact solver/API version omitted due to license/anonymity.

5. **Open access to data and code.** If you ran experiments, did you include the code, data, and instructions needed to reproduce the main experimental results?

[No] All implementation code, configuration files, prompt templates, data schemas, and reproduction instructions can be made public. The only component that cannot be redistributed is the solver-generated dataset, because it is collected through a commercial solver API. Full reproduction therefore requires regenerating the solver-labeled corpus with API access and rerunning the LLM-based evaluation calls.

6. **Experimental setting/details.** If you ran experiments, did you specify all the training details?

[Yes] The manuscript reports the NLH settings, dataset scale, sampled public states, action abstraction, training objectives, distillation regimes, SCCS matching logic, and LLM evaluation conditions. Model training used a single RTX 5090 GPU; solver data collection used a commercial API; SCCS/communicability evaluation used LLM API calls.

7. **Experiment statistical significance.** Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

[Yes] The communicability table reports means with standard errors over unseen target cases, and the Mean row reports sample standard deviation across LLM configurations, as defined in the table caption.

8. **Experiments compute resource.** For each experiment, does the paper provide sufficient information on the computer resources needed to reproduce the experiments?

[Yes] The main compute sources are identified: a commercial API for solver data collection, a single RTX 5090 GPU for MDT training, and LLM API calls for SCCS/communicability testing.

9. **Code of Ethics.** Have you read the NeurIPS Code of Ethics and ensured that your research conforms to it?

[Yes] The research uses solver-generated game data and LLM evaluations, and does not rely on private personal data, human-subject experiments, or deceptive user interaction.

10. **Broader impacts.** If appropriate for the scope and focus of your paper, did you discuss potential negative societal impacts of your work?

[Yes] The paper discusses that the work articulates solver-derived equilibrium logic rather than providing real-money gambling advice or an autonomous live-play agent. Potential misuse is limited by not releasing the solver-generated dataset and by framing the system as an offline analysis tool.

11. **Safeguards.** Do you have safeguards in place for responsible release of models with a high risk for misuse?

[N/A] The paper does not release a general-purpose model or an autonomous poker-playing agent. The planned release is implementation code for research use, while the solver-generated dataset is not publicly released.

12. **Licenses.** If you are using existing assets, did you cite the creators and respect the license and terms of use?

[Yes] The paper cites relevant benchmark, solver, LLM, and interpretability literature. Public code release will document software dependencies and API requirements; solver data are not redistributed.

13. **Assets.** If you are releasing new assets, did you document them and provide these details alongside the assets?

748 [Yes] The planned public asset is the implementation code, which should include instructions,
749 environment requirements, expected data schema, and limitations. The solver-generated
750 dataset itself will not be released.

751 14. **Crowdsourcing and research with human subjects.** If you used crowdsourcing or con-
752 ducted research with human subjects, did you include the full text of instructions and details
753 about compensation?
754 [N/A] No crowdsourcing or human-subject study is conducted.

755 15. **IRB approvals.** Did you describe any potential participant risks and obtain Institutional
756 Review Board approvals, if applicable?
757 [N/A] The work does not involve human participants or personal data collection.

758 16. **Declaration of LLM usage.** Does the paper describe the usage of LLMs if it is an important,
759 original, or non-standard component of the core methods in this research?
760 [Yes] LLMs are used as independent readers of generated reasoning rules in the communica-
761 bility evaluation. The paper reports the prompting conditions and includes prompt templates
762 in the appendix; the evaluation uses LLM API calls rather than training or releasing an
763 LLM.