POKÉLLMON: A Human-Parity Agent for Pokémon Battles with Large Language Models

Sihao Hu, Tiansheng Huang, Ling Liu Georgia Institute of Technology Atlanta, GA 30332, United States(2024)

Background

Why Pokémon is a Complex Testbed for AI

1. Tactical Complexity: Turn-based system, vast move/switch options, dynamic state.

2. Information Asymmetry: Unknown opponent sets (moves, EVs, nature).

3. LLM Limitations: Existing LLM agents suffer from two main issues:

a) Hallucination: Stating incorrect type matchups or move effects.

b) Panic Switching: Consecutive switches to avoid powerful opponents.

Figure 1. At each turn, the player is requested to decide which action to perform, *i.e.*, whether to let *Dragonite* to take a move or switch to another Pokémon off the field.

Figure 2. Two representative Pokémon: *Charizard* and *Venusaur*. Each Pokémon has type(s), ability, stats and four battle moves.

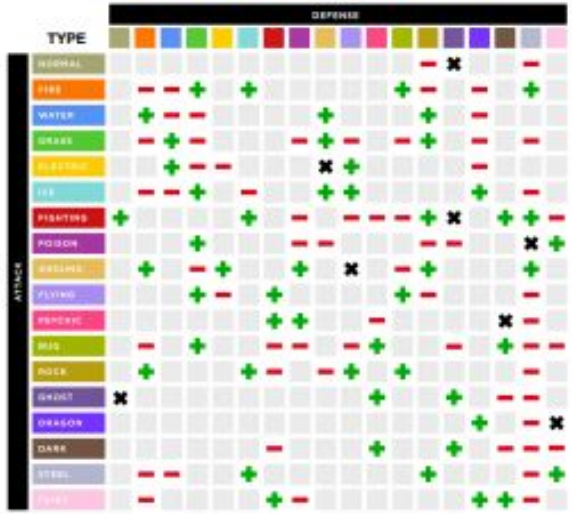

Figure 3. Type advantage/weakness relationship. "+" denotes super-effective (2x damage); "−" denotes ineffective (0.5x damage); "×" denotes no effect (0x damage). Unmarked is standard (1x) damage.

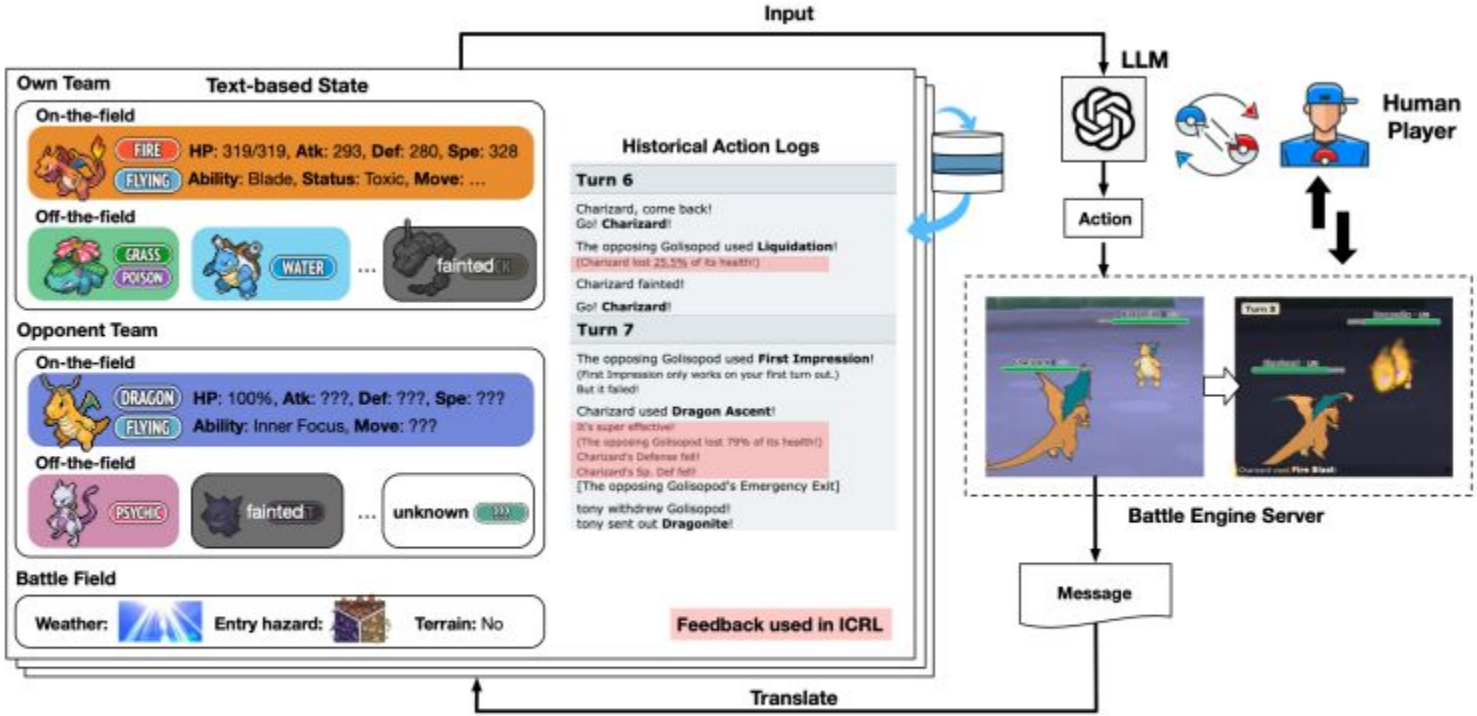

Figure 4. The framework that enables LLMs to battle with human players: It parses the messages received from the battle server and translates state logs into text. LLMs take these state descriptions and historical turn logs as input and generates an action for the next step. The action is then sent to the battle server and executed alongside the action chosen by the opponent player.

Table 1. Performance of LLMs in battles against the bot.

Player	Win rate ↑	Score ↑	Turn #	Battle #
Human	59.84%	6.75	18.74	254
Random	1.2%	2.34	22.37	200
MaxPower	10.40%	3.79	18.11	200
LLaMA-2	8.00%	3.47	20.98	200
GPT-3.5	4.00%	2.61	20.09	100
GPT-4	26.00%	4.65	19.46	100

-MaxPower is always selecting the move with the highest base power available to the current Pokémon.
-LLMs is used in a state of unimproved.
-The battle score is defined as the sum of the numbers of the opponent's fainted Pokemon and the player's unfainted Pokemon.

Test of Hallucination

Table 2. Confusion matrices for type advantage prediction.

Model	LLaMA-2				GPT-3.5				GPT-4			
Class	A	B	C	D	A	B	C	D	A	B	C	D
A	5	46	0	0	0	0	49	2	37	8	5	1
B	25	179	0	0	2	6	185	11	0	185	17	2
C	15	46	0	0	0	2	57	2	3	24	32	2
D	1	7	0	0	0	0	7	1	0	0	0	8

A (Super Effective): The predicted or actual result is twice the damage.
B (Standard): The predicted or actual result is normal damage.
C (Not Very Effective): The predicted or actual result is half the damage.
D (Ineffective): The predicted or actual result is no damage (the move has no effect).
Row: The Predicted result by the LLM
-(e.g., the LLM thought, "This is Super Effective!").
Column: The Actual correct type effectiveness
-(e.g., the reality was, "It was Super Effective.").
Cell Value: The percentage (or frequency) of that specific pattern occurring.

POKÉLLMON Overview: The Three Pillars

POKÉLLMON: The First Human-Parity LLM Agent

— Built on three core strategies to overcome LLM limitations—

1. In-context Reinforcement Learning (ICRL)

2. Knowledge-Augmented Generation (KAG)

3. Consistent Action Generation (via Self-Consistency)

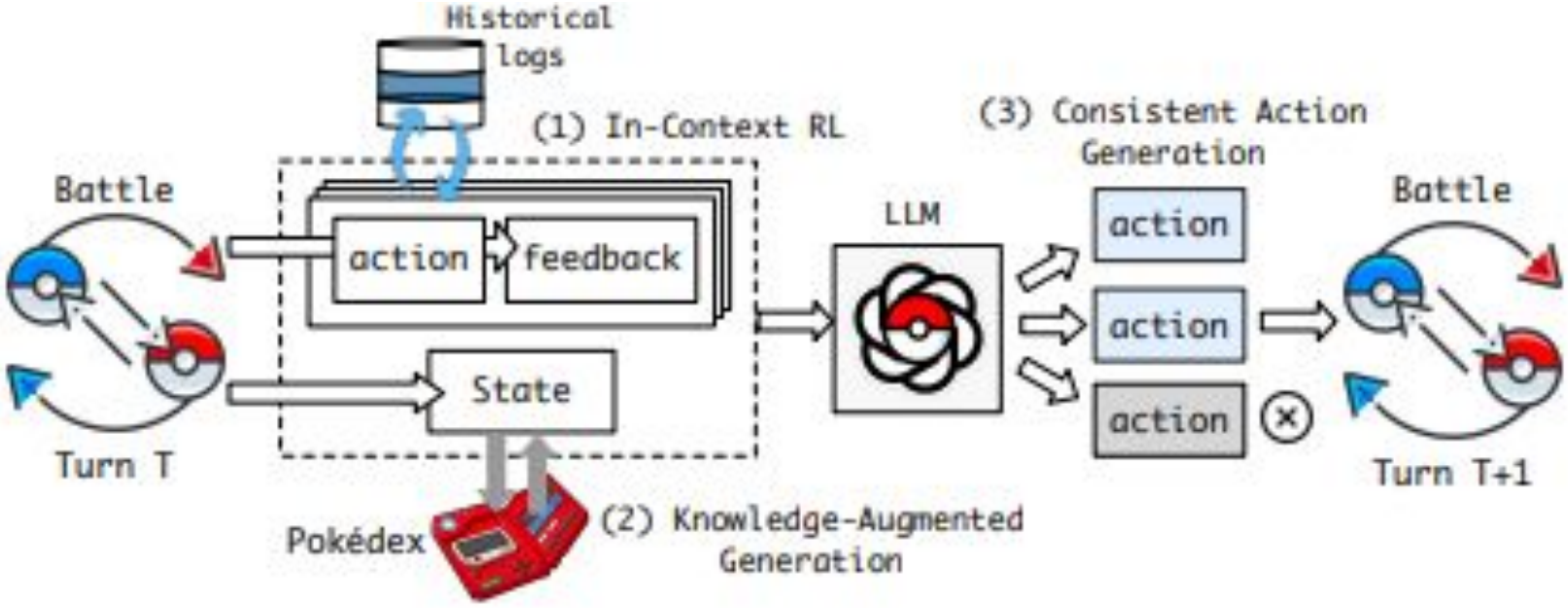

Figure 5. POKÉLLMON is equipped with three strategies: (1) ICRL that leverages instant feedbacks from the battle to iteratively refine generation; (2) KAG that retrieves external knowledge to combat hallucination and to act timely and properly; (3) Consistent Action Generation to prevent the panic switching problem.

Pillar 1: In-context Reinforcement Learning (ICRL)

Learning "Just-in-Time" from Textual Feedback

1. Goal: Instantly consume battle outcomes as feedback to refine the policy without re-training.

2. Feedback Types (Reward Signals): Damage taken/dealt, move effectiveness (Super Effective/Not Very Effective), estimated Speed/Priority, actual move effects (stat changes).

3. Impact: Improved baseline GPT-4 win rate from 26% to 36%.

Figure 6. The agent repeatedly uses the same attack move but has zero effect to the opposing Pokémon due to its ability "Dry Skin."

Figure 7. In turn 3, the agent uses "Psyshock", which cause zero damage to the opposing Pokémon. With ICRL, the agent switch to another Pokémon.

Table 3. Performance of ICRL in battles against the bot.

Player	Win rate ↑	Score ↑	Turn #	Battle #
Human	59.84%	6.75	18.74	254
Origin	26.00%	4.65	19.46	100
ICRL	36.00%	5.25	20.64	100

Table 3 shows the improvement brought by ICRL. Compared to the original performance of GPT-4, the win rate is boosted by 10%, and the battle score increases by 12.9%. During the battles, we observe that the agent begins to change its action if the moves in previous turns do not meet the expectation, as shown in **Figure** 7: After observing that the opposing Pokémon is immune to the attack, it switches to another Pokémon.

Pillar 2: Knowledge-Augmented Generation (KAG)

Overcoming Hallucination with External Knowledge

1. Goal: Retrieve accurate, explicit knowledge to prevent fundamental mistakes (hallucination) before they lead to defeat.

2. Implementation: Integrate a Pokédex (external database from Bulbapedia) into the prompt context.

3. Key Information: Precise Type Matchups, Move Effects, Ability Effects.

4. Impact: Combined with ICRL, win rate jumped from 36% to 58% (against heuristic bots).

Table 4. Performance of KAG in battles against the bot.

Player	Win rate ↑	Score ↑	Turn #	Battle #
Human	59.84%	6.75	18.74	254
Origin	36.00%	5.25	20.64	100
KAG[Type]	55.00%	6.09	19.28	100
KAG[Effect]	40.00%	5.64	20.73	100
KAG	58.00%	6.53	18.84	100

Figure 8. The agent understands the move effect and uses it properly: *Klefki* is vulnerable to the ground-type attack of *Rhydon*. Instead of switching, the agent uses "Magnet Rise", a move that protects itself from the ground-type attack for five turns, invalidating the ground-type attack "Earthquake" of the opposing *Rhydon*.

Pillar 3: Consistent Action Generation

Mitigating "Panic Switching" with Self-Consistency

1. The Problem: When faced with a strong threat, LLM agents using Chain-of-Thought (CoT) often enter a loop of "Panic Switching" to avoid combat.

2. Observation: CoT actually decreased performance in our tests.

3. Solution (Self-Consistency / SC): Generate multiple action candidates independently, then vote for the most consistent action.

4. Impact: Increased win rate (against heuristic bots) up to 64%, effectively suppressing panic behavior.

Table 5. Performance of prompting approaches in battles against the bot.

Player	Win rate ↑	Score ↑	Turn #	Battle #
Human	59.84%	6.75	18.74	254
Origin	58.00%	6.53	18.84	100
CoT	54.00%	5.78	19.60	100
SC (k=3)	**64.00%**	6.63	18.86	100
ToT (k=3)	60.00%	6.42	20.24	100

Figure 9. When facing a powerful Pokémon, the agent with CoT switches different Pokémon in three consecutive to elude the battle. This gives the opponent three free turns to quadruple its attack stats and quickly defeat the agent's entire team.

Table 6. Statistic analysis of panic switching

Player	Win rate ↑	Switch rate	CS1 rate	CS2 rate
Origin	58.00%	17.05%	6.21%	22.98%
CoT	54.00%	26.15%	10.77%	34.23%
SC (k=3)	64.00%	16.00%	1.99%	19.86%
ToT (k=3)	60.00%	19.70%	5.88%	23.08%

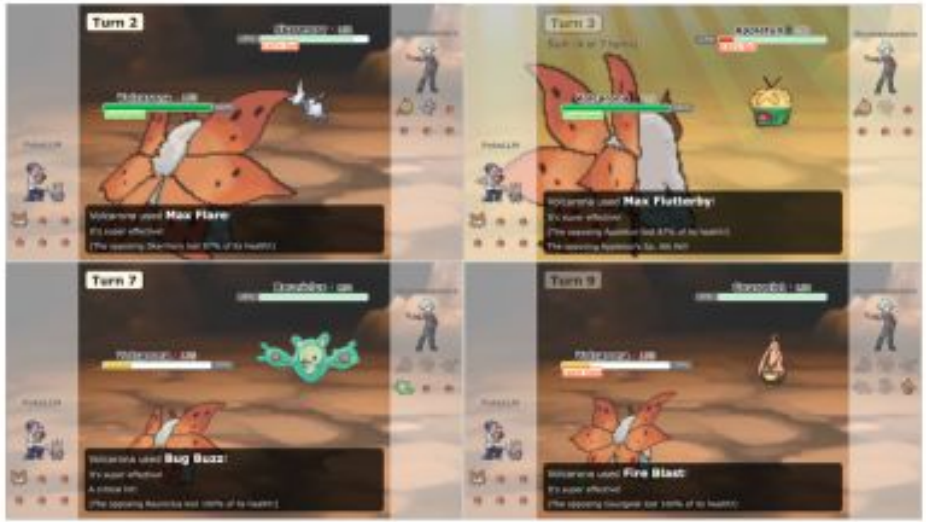

Figure 10. POKÉLLMON selects effective moves in every turn, causing the opponent's entire team to faint using one Pokémon.

Experimental Results: Online Battles vs. Humans

Achieving Human-Parity Performance

1. Testing Environment: Live online battles against real human players.

2. Results:

- Ladder Players (Random Humans): 49% Win Rate (N=105)

- Invited Players (Experienced Humans): 56% Win Rate (N=50)

3. Conclusion: POKÉLLMON is the first LLM-based agent to demonstrate performance equivalent to human players in this complex tactical game.

Table 7. Performance of POKÉLLMON against human players.

v.s. Player	**Win rate ↑**	**Score ↑**	**Turn #**	**Battle #**
Ladder Player	48.57%	5.76	18.68	105
Invited Player	56.00%	6.52	22.42	50

Battle Analysis: Strengths and Weaknesses

Understanding POKÉLLMON's Strategy

Strengths:

1. Human-like Attrition: Executes complex stalling tactics (e.g., Poisoning, then healing/protecting).

2. Tactical Accuracy: Few mistakes in type-matching and move choice (thanks to KAG).

Weaknesses:

1. Vulnerability to Attrition: Prioritizes short-term gains; easily defeated by human attrition specialists (Win Rate: 18.75%).

2. Deception: Fails against human "tricks" or psychological maneuvers (e.g., baiting an attack and then switching to an immune Pokémon).

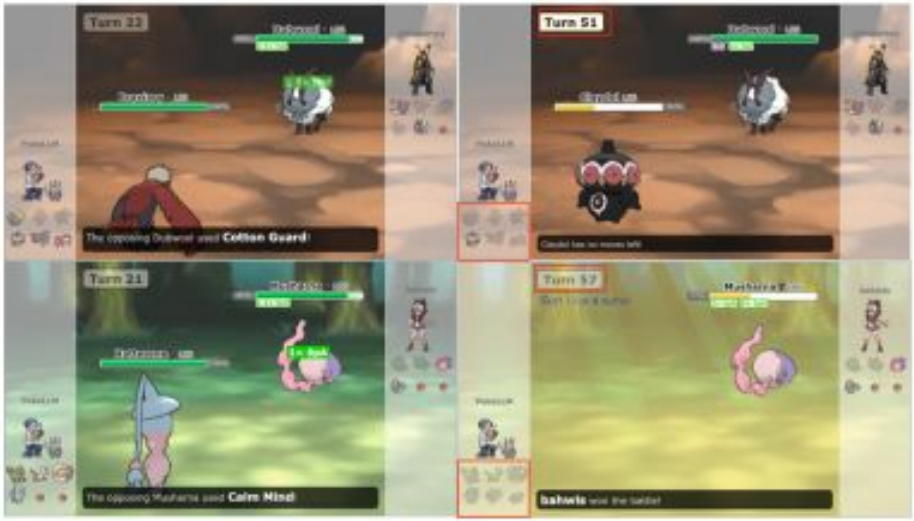

Figure 11. POKÉLLMON suffers from attrition strategies: the opponent players frequently recover high-defense Pokémons. Breaking the dilemma requires joint effects across many turns.

Figure 12. An experienced human player misdirects the agent to use a dragon-type attack by firstly sending out a dragon-type Pokémon and immediately switch to another Pokémon immune to the dragon-type attack.

Table 8. Battle performance impacted by the attrition strategy

Ladder	Win rate ↑	Score ↑	Turn #	Battle #
w. Attrition	18.75%	4.29	33.88	16
w/o Attrition	53.93%	6.02	15.95	89

Conclusion and Future Work

Summary

POKÉLLMON successfully utilized ICRL, KAG, and SC to achieve human-parity in Pokémon battles.

Future Work

1. Integrating Long-Term Planning and deeper strategic reasoning.

2. Enhancing Opponent Prediction and handling deceptive moves.

3. Applying the POKÉLLMON framework to other complex tactical decision-making environments.