

ChatGPT and Competitive Pokémon

Jack Frampton, Kelly Xu

Introduction

The goal of our experiment is to assess ChatGPT's capabilities in pokemon competitive single battles.

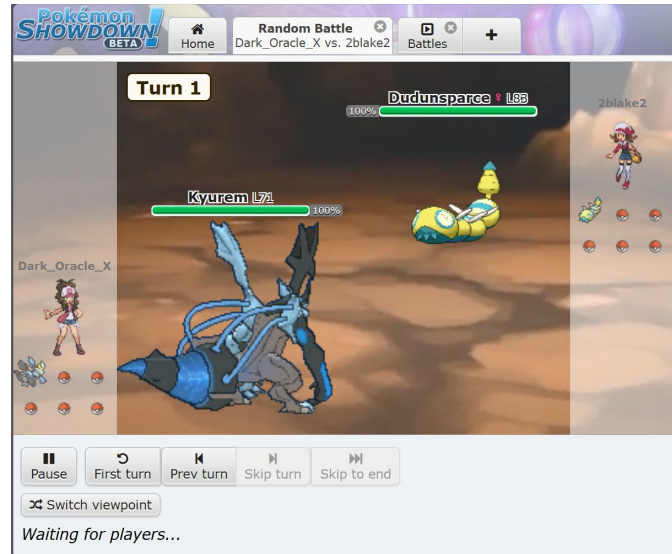
To be a proficient player you must have a good understanding of:

- Pokemon type advantages
- Pokemon Abilities
- Speed control
- Pokemon stats
- General strategies
- Meta-game knowledge



Pokemon Showdown

We used pokemon showdown, a online pokemon simulator for our test battles. We had ChatGPT play Gen9 random single battles.



Methods

We conducted two distinct experimental approaches to assess ChatGPT-5's strategic capabilities: manual interactive battles and autonomous API-driven battles.

In the manual testing phase, we acted as an interface between ChatGPT-5 and Pokemon Showdown. We provided ChatGPT with an initial prompt containing it's random team information and objective and after each turn, we provided ChatGPT with all knowledge that would be provided to a regular player.

Methods

We wrote a python script to have GPT play autonomously through the pokemon showdown API. Each turn we prompted with the current state of battle and history about what has happened so far. ChatGPT was instructed to respond with a single command as required by the API, such as “switch 3” or “move 2.”

We ran it for about 90 battles and saved each battle into a json log that had all of the battle information. Finally, we parsed the data for analysis.

```
|t:|1761250676
|switch|p2a: Morpeko|Morpeko, L88, M|100/100
|move|p1a: Honchkrow|Brave Bird|p2a: Morpeko
|-resisted|p2a: Morpeko
|-damage|p2a: Morpeko|61/100
|-damage|p1a: Honchkrow|280/312|[from] Recoil
|
|-heal|p2a: Morpeko|67/100|[from] item: Leftovers
|-formchange|p2a: Morpeko|Morpeko-Hangry|[from] ability: Hunger Switch
|upkeep
|turn|2
```

Results

Manual Testing

50%

Win Rate

Human-like performance with strategic intent

Autonomous Testing

18%

Win Rate

Well below expected 50% baseline

Analysis

What It Does Well

- Understands type matchups and often selects super-effective moves
- Uses setup moves like Calm Mind and Shell Smash at smart times
- Manages hazards effectively (Stealth Rock, Spikes, Whirlwind)
- Shows awareness of abilities (such as Storm Drain and Swift Swim)
- Can plan ahead and maintain defensive positioning (Skarmory sequences)

Analysis

What It Struggles With

- Makes illegal moves (Choice-lock and double Terastallize)
- Sometimes attacks into faster opponents without priority or protection
- Uses ineffective or immune moves in key moments (Toxic on Weezing)
- Sends invalid commands the server ignores
- Relies on low-accuracy moves at critical times, losing winning positions
- Lacks consistency (Reasoning is strong, but the moves it decides are not reliable)

Conclusion

- ChatGPT can think strategically, but it struggles to execute reliably without guidance.
- Human-guided interaction brings out its best strategic thinking