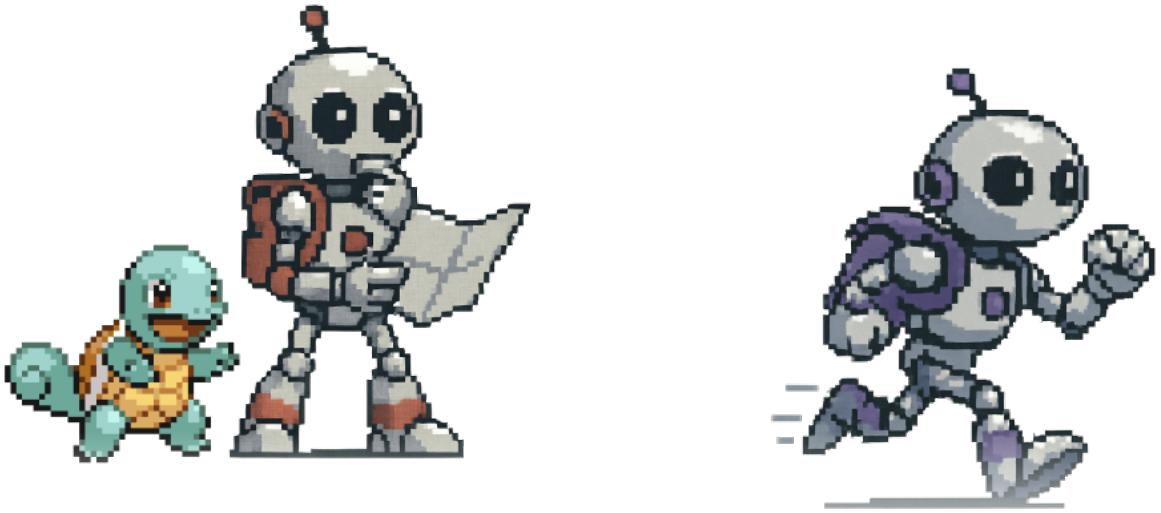
Track 2: Speedrunning

Long-Horizon RPG Gameplay



This track challenges agents to complete a full Pokémon role-playing game (Pokémon Emerald) as quickly and efficiently as possible, navigating a massive, partially observable world with hundreds of NPCs and thousands of possible actions.

Long-horizon planning, efficient exploration, and strategic resource management are critical to succeeding in this track. Agents must learn to balance immediate objectives with long-term strategic goals, making decisions that span thousands of timesteps while adapting to the unpredictable nature of RPG gameplay.

algorithms and efficient resource management to achieve optimal completion times in complex, open-world environments.

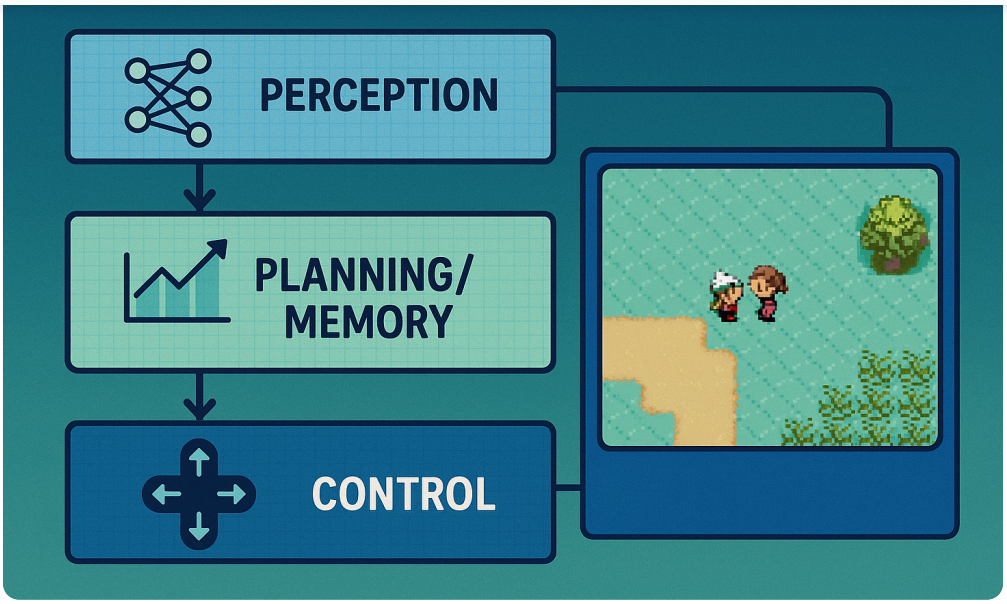
⏰ COMPETITION ENDS IN

37	13	58	18
DAYS	HOURS	MINUTES	SECONDS

Track 2 Submissions end November 15th, 2025

Starter Kits

Our starter kit provides a real-time agent loop with modular components for perception (game frame recognition), planning & memory (long term vs. short term goals, knowledge storage), and control (gameboy emulator



View Starter Kit Repository

What's Included

- **Agent Scaffolding:** Modular framework for building Pokémon Emerald speedrunning agents
- **Pokémon Emerald Wrapper:** Custom emulator API for real-time game interaction
- **Baseline Implementation:** Reference agent with VLM setup and basic planning
- **Evaluation Tools:** Automated testing and performance measurement

Compute Credits

Application closed and credits awarded. All compute credits have been distributed to approved teams.

⚠ Prize Eligibility Notice

Exact clones of organizer-hosted baselines are not eligible for prizes. Submissions must demonstrate novel approaches, meaningful modifications, or original implementations. Simple repackaging or minimal changes to existing baseline code will be disqualified from prize consideration.

How to Submit for Track 2

Submissions for this track focus on achieving maximum game completion under time constraints. Your agent must interact exclusively through our custom Pokémon Emerald emulator API. Use any method, as long as the final action comes from a neural network.

Important: All submissions will undergo anti-cheat verification to ensure fair competition. This includes validation of agent behavior, action logs, and verification that submissions follow the competition rules.

Submit Your Entry

Submission Requirements

- **Code Archive:** Your agent implementation as a ZIP or TAR.GZ file including all dependencies and README
- **Action & State Logs (Anti-Cheat):** The `submission.log` and detailed logs generated by the starter kit during your agent's run. These logs validate that your agent followed competition rules and provide action/state information for evaluation.
- **Methodology Description:** A brief document (1-2 paragraphs via the google form) describing your approach and detailing your scaffolding components across the five dimensions (State Information, Tools, Memory, Feedback, Fine-tuning). This is required for calculating your Adjusted Performance score.

> **completely rewrite the starter kit code** to implement your approach. The only requirement is that your submission includes the valid logs (including `submission.log`) generated by the starter kit's logging system, which verifies your agent interacted with the game through the official API and followed competition rules.

Final Ranking Criteria

Final rankings are determined by **Adjusted Performance**, which accounts for both your agent's raw performance and the complexity of its external scaffolding:

> $$\text{Adjusted Performance} = \text{Raw Performance} \times \lambda(\text{Scaffolding})$$
>
> where **λ(Scaffolding)** is a penalty factor (≤1) based on the external support your agent uses

Raw Performance Components

- **Milestone Completion:** Percentage of game milestones accomplished (e.g., gym badges, story progression)
- **Completion Efficiency:** Time and action count to achieve milestones
- **Reproducibility:** Clear documentation and verifiable results

Scaffolding Penalty Components

The scaffolding penalty λ evaluates your agent's support structure across five dimensions:

- **State Information (S):** What information your agent receives (raw pixels vs. parsed game state vs. privileged information)
- **Tools (T):** External tools available during gameplay (web search, calculators, planning utilities, etc.)
- **Memory (M):** Memory mechanisms beyond immediate context (vector databases, knowledge graphs, external storage)
- **Feedback (F):** Human or automated feedback during runs (human-in-the-loop,

> **Important:** Agents with more external support receive larger penalties. Pure end-to-end learned systems with minimal scaffolding are favored in rankings, all else being equal. This encourages development of generalizable AI capabilities rather than task-specific engineering.

Teams must document their scaffolding components in detail during submission. The organizing committee will review and assign scaffolding penalty factors based on the submitted information.

Timeline

June 11th, 2025

Competition Website Launch

Official competition website goes live with preliminary documentation.

June 25th, 2025

Formal Competition

starter code with a baseline RPG agent (scaffolding and VLM setup) and emulator API available for beta testers.

July 7th, 2025

Competition Begins

Track 2 Competition Begins. Submit runs of your Pokémon Emerald agent to the leaderboard.

November 15th, 2025

Results Announcement

Final submission deadline.

December 2025

NeurIPS 2025 Presentation

Winners announced at NeurIPS 2025.

Prizes

🏆 Speedrun Rankings

Top performing agents in the RPG speedrunning challenge will be awarded **$4,500** and **1000 GCP** total:

- **1st Place:** $1,500 + 700 GCP
- **2nd Place:** $1,000 + 300 GCP
- **3rd-4th Place:** $500
- **5th-8th Place:** $250

★ Judge's Choice Award

Senior organizers will award at least **two projects** with **$400** or **500 GCP** to help continue their work after the competition ends.

This project does not necessarily have to place highly in the speedrun rankings but should propose a novel approach or demonstrate interesting capabilities in long-horizon planning or RPG navigation.

A NeurIPS 2025 competition advancing AI decision-making through the complex environments of Pokémon battles and gameplay.

Home

Track 1

Track 2

Leaderboard

Rules

Hackathon

FAQ

For questions or support, please contact:

Email: pokeagentchallenge@gmail.com

Join our Discord community for direct assistance and discussions.