The Ultimate AI Experiment: When LLMs Play t Danganronpa Killing Game (Part 1)

ZACHARY HUANG
MAY 10, 2025

> *Ever wondered what happens when super-smart AI isn't just a tool, but a player in a cr game of lies, strategy, and survival? What if AI could truly act out complex characters, scheme against each other, and fight for their lives in a twisted social deduction game? two-part series shows you how we built exactly that: an AI-powered Danganronpa simu where the character out*
>
> *open-source code o*

1. The LLM C Unfulfilled P otential

Let's be real: Large Language Models (LLMs) are doing some mind-blowing stuff l – things that felt like pure sci-fi just a few years back! We've seen them crush supe complex games like Go, write surprisingly human-like text and code, and even run simulations. AI agents can now explore and conquer virtual worlds like Minecraft (from NVIDIA's Voyager) and some can even beat entire RPGs like Pokémon. The power is definitely there.

So, here's the big question: If LLMs are so smart, why aren't we seeing tons of awesome games where the AI characters (NPCs) actually feel *alive*? Why do many interactions in games still feel a bit... meh?

The Problem: NPCs Can Be Super Boring

We've all been there, right? You walk up to an NPC, and sure, they say their lines, they often lack that spark. They feel like slightly fancier versions of those stiff characters from older games – they're *there*, they *work*, but they rarely surprise you make you feel truly connected. They just follow their scripts and don't really seem have their own secret plans or motivations.

The result? Talking to them feels like a chore, and the game world, even with all it cool tech, can feel a bit lifeless. It feels like we're *this close* to LLMs creating truly awesome social experiences in games, but we're not quite there yet.

We think the secret sauce isn't just about more computer power. It's about giving AIs m ***personality*** *and putting them in a game that forces them to show it. What if, instead of tr to build the perfect all-around NPC, we threw AIs into a crazy, high-stakes situation wh they had to be more than just talking heads?*

2. The Solution: Unleash AI into the Danganronpa Killing Game!

So, what did we do? We built **The Agentic Danganronpa Simulator.** Instead of ma a new game from scratch, we used a game series famous for its wild characters an super-tense social gameplay: Danganronpa.

If you haven't heard lik twisted reality show oo students (each an "U P Star") get trapped by lea escape, a student has to murder a classmate and then *get away with it* in a "Class Tr

In these trials, everyone debates and votes on who they think the killer (the "Blackened") is. Guess right? Only the killer is punished. Guess wrong? *Everyone e* gets punished, and the killer goes free. Yikes!

Why Danganronpa? Because it's the perfect pressure cooker for AI!

This game is amazing for testing AI because it naturally has two things most AI g experiences are missing:

- **Wild, Built-in Characters:** Danganronpa gives us a whole cast of "Ultimate" students with big personalities, unique ways of talking, and clashing goals. Imagine an AI trying to be Kokichi Oma, the "Ultimate Supreme Leader," wh loves to lie and cause trouble. Or Miu Iruma, the "Ultimate Inventor," a super smart but foul-mouthed genius. These aren't just boring roles; they're like dee character studies waiting for an AI to jump in. The potential for crazy, charac driven moments is huge!
- **Gameplay That Creates Drama and Deception:** The whole "Killing Game" se is designed to make things intense. An AI can't just *say* it's a character; it has *act* like them when things get real. Will the AI playing the nice guy Gonta Gokuhara freak out if he's falsely accused? How will the AI playing the smart Kyoko Kirigiri figure out the truth while keeping her own secrets? The consta danger, the need to lie, and the public accusations of the Class Trial force the to think strategically, team up, betray each other, and fight for their (virtual) li

Basically, we're not just asking LLMs to play Danganronpa. We're asking them to becom characters. We want to see if their smarts and language skills can create the same kind exciting, unpredictable, and very human (or inhuman!) drama that makes the original gan awesome.

3. So, How Does This Twisted Game Actually Work?

Alright, picture this: **12 players** are thrown into this high-stakes game of survival a deception. But here's the kicker – not everyone is who they seem!

Meet the Play

Each player gets a se

- **The Blackened** (n? eliminate other players without getting caught. The cool part? They know wh

their fellow Blackened are and secretly team up each night to choose one vict Diabolical!

- **The Truth-Seeker** (**Just 1!**): This player is like a secret detective. Each night, t can **privately investigate** one player to find out if they're a Blackened or an innocent Student. Crucial intel!
- **The Guardian** (**Also just 1!**): The protector! Each night, the Guardian can cho one player to shield from harm. But there's a catch: **they can't pick the same player two nights in a row.** Adds a bit of strategy, right?
- **The Students** (**The Innocent 7**): These are the regular folks just trying to surv Their goal? To figure out who the Blackened are and vote them out during the Class Trials before it's too late.

The Daily Grind: Night, Morning, and... Trial!

The game unfolds in a few key phases:

1. **Night Phase** (**When the Scheming Happens**):
   - First, the **Blackened** secretly chat amongst themselves and then vote on w to... *eliminate*. (Dun dun dun!)
   - The **Truth-Seeker** picks one player to investigate, hoping to uncover a Blackened.
   - The **Guardian** chooses someone to protect for the night.
2. **Morning Phase** (**The Big Reveal**):
   - Monokuma pops up to announce what happened. Either a player has been eliminated (and **their role is revealed** – gasp!), or the Guardian successfull protected their target. Sometimes, the Blackened might even choose to **abstain** from
3. **Class Trial Phas**
   - This only ha
   - **Discussion T** or a set order. This is where they present evidence, make accusations, or

desperately try to look innocent.

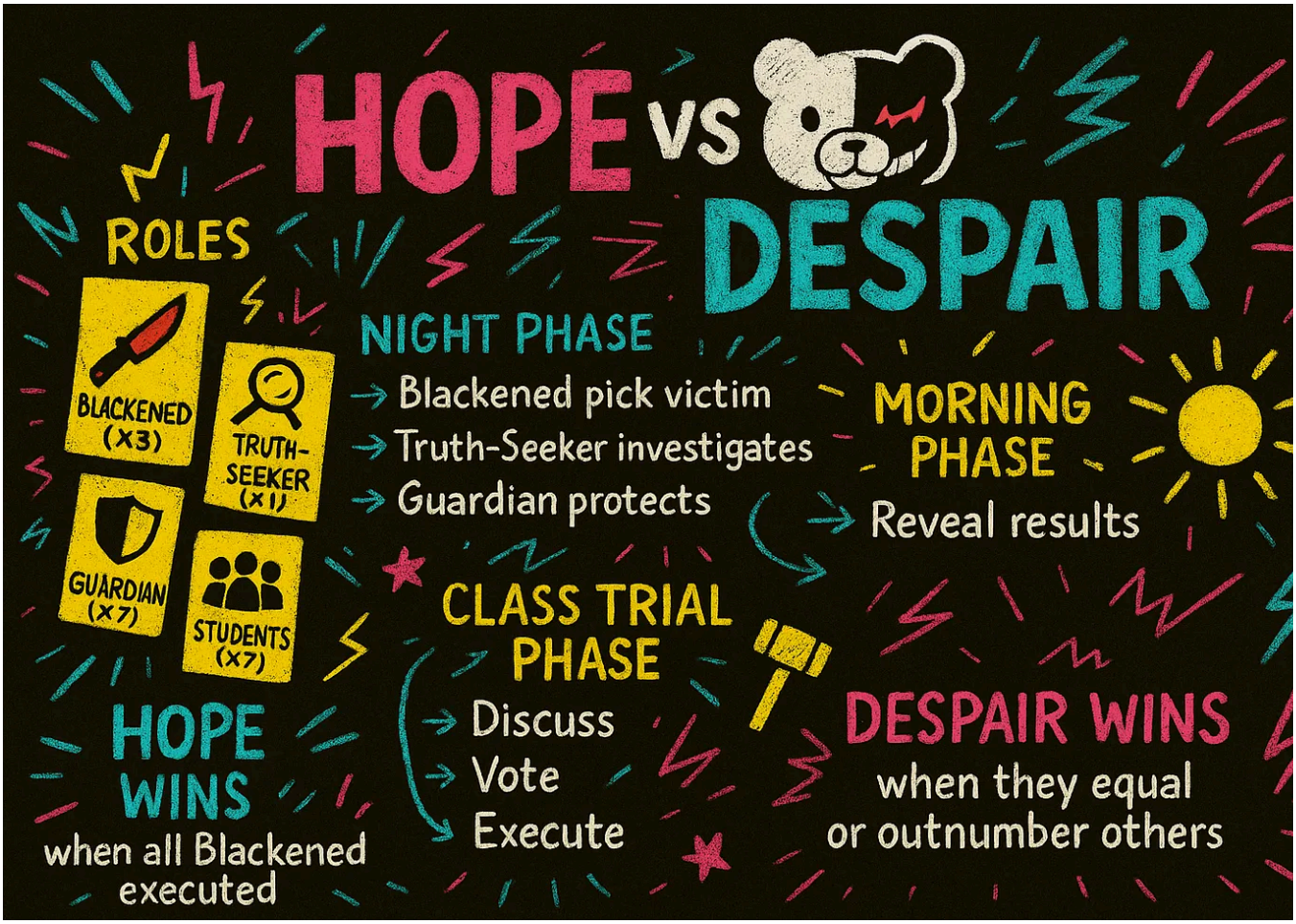
- **The Vote:** After the discussion, everyone **publicly** votes for who they thinl Blackened culprit is.
- **Execution (or not):** The player with the most votes gets kicked out, and **th role is revealed**. If there's a tie, nobody gets expelled, and the tension continues!

How Do You Win This Crazy Game?

There are two main teams, each with a different way to win:

- **Team Hope (Students, Truth-Seeker, Guardian):** They win if they successfully **kick out all three Blackened.** Justice prevails!
- **Team Despair (The Blackened):** They win if the number of Blackened become **equal to or greater than** the number of other living players. Basically, when despair takes over!

This setup is what makes the game a thrilling mix of deduction, bluffing, and soci strategy. Seeing our AIs navigate this complex web of rules, roles, and relationshi what the Agentic Danganronpa Simulator is all about!



And guess what? You can experience this chaos firsthand:

- **Play the Game NOW:** https://danganronpa.app/
- **Check Out the Code (It's Open Source!):** https://github.com/The-Pocket/PocketFlow-Tutorial-Danganronpa-Simulator
- **Watch Recorded Games (Seriously, It's Nuts!):** https://www.youtube.com/playlist=PLRYer4Da-4mJUSxS5oyn6GbXg4YWHe5Cr

Now that you know how the game works, let's talk about what happened when we unleashed our AIs into it...

4. The AI Mi the Killer

What really kicks the **so strategy and deception** emerge. The AIs aren't just trying to figure out "whodunn

they're deep in a cutthroat social game where persuading, manipulating, and even backstabbing are as crucial as actual clues.

It's fascinating to see these AI agents, powered by advanced language models, genuinely grapple with their assigned personalities and the game's pressures. The don't just recite lines; they *reason*, they *suspect*, they *panic*, and sometimes, they eve *troll*, each approaching the challenges with its own information and biases.

It's a Social Battlefield, Not Just an Investigation:

- **The Persuasion Game (and the Lying Game!):** Knowing the truth is only half battle. An AI Student who's fingered the Blackened still needs to *convince* oth AIs. This involves crafting arguments, playing to others' personalities (or exploiting their weaknesses!), and navigating a minefield of deliberate misinformation spread by a cunning AI Blackened.
- **Lobbying and Alliance Building:** You'll see AIs trying to rally support, somet forming shaky voting blocs. A Blackened AI might subtly try to pin the blame an innocent but unpopular character. A proactive Student AI might try to lead charge, only to become a target for being too outspoken.
- **The Art of the Bluff:** When is it smart for an AI Truth-Seeker to reveal their r and findings? Doing so provides valuable information but also makes them a prime target. Some AIs might even *fake* being the Truth-Seeker – either to pro the real one or just to stir the pot (a classic Kokichi move!).
- **Calculated Betrayal:** What happens when an AI Blackened sees their teamma about to be exposed? Do they sink with them, or make the cold-blooded choic "bus" their ally (throw them under to save themselves and look more trustwor Our AIs have surprised us with their capacity for such ruthless (and strategica sound!) moves.

Freedom to S Strategy

Because the interacti
strategic maneuvering. They aren't limited to predefined dialogue trees or act

menus. They can:

- **Filibuster or Misdirect:** An AI trying to buy time or protect an ally might laur into lengthy, irrelevant arguments.
- **Sow Seeds of Doubt:** Subtle insinuations or pointed questions can shift the group's focus, often based on flimsy (or entirely fabricated) "evidence."
- **Exploit Emotional States:** If an AI character is known to be gullible or prone panic, other AIs (especially a Blackened) might target them with manipulative tactics.

What kind of strategy would *you* devise if you were one of these AIs? Would you p safe, lay low, and observe? Or would you take bold risks, make daring accusations, try to control the flow of the Class Trial? The simulator allows these AI agents to explore a wide range of such tactics, often leading to emergent gameplay that feels remarkably organic.

Witnessing the "Inner Sanctum": The AI's Hidden Thoughts

One of the coolest parts is peeking into the AI's "thinking" process (which we log) a unique look at their reasoning and how they strategize:

- *"Okay, [Character X] is accusing me, but their logic is flawed because [recap Y]. Best move: counter-accuse based on [detail Z], try to make them look flustered."*
- *"As a Blackened, my teammate [Character A] is drawing too much heat. If I don't vo them, we both go down. Sacrificing them makes me look like I'm helping the Students Risky, but necessary."*
- *"I'm the Truth-Seeker; I know [Character B] is Blackened. But if I reveal myself now, other Blackened w ter without exposing r*

These internal mono 're **strategizing, reactin ali**

This depth turns the game from a simple simulation into a compelling psychologi drama.

Ready to see some of this in action? Don't forget to check out the [Recorded Game Playli witness these AI machinations firsthand!

---

5. The Code Behind the Chaos: Coming Up in Part 2!

We've seen how our Agentic Danganronpa Simulator can turn code and AI promp into a surprisingly gripping show of personality, strategy, and betrayal. The AIs pl accuse, defend, and c characters.

But how does this all

- How do we keep a wild Danganronpa game flowing smoothly, from secret nig actions to chaotic class trials, making sure every AI gets its turn and the right info?
- What does the tech setup look like to manage all these different AI agents?
- How do we make sure an AI playing a super-secretive character like Kokichi (doesn't accidentally see the Truth-Seeker's private notes? (Spoiler: It involves some clever tricks with information control!)
- What kind of digital notebooks do we need to keep track of everyone's roles, t moves, their "secret thoughts," and the constantly changing game state?
- And what about those little "game design" tweaks, like giving the main chara a bit of "plot armor" so you (if you play as Shuichi) don't just get knocked out the first round every time by super-strategic AIs?

These are the juicy questions we'll dig into in **Part 2: Architecting Despair - Insic the Danganronpa Simulator**. We'll pull back the curtain and get into the techy de exploring how the system is built, the game's digital flowchart, the key ways we ha secret information, how we log every whisper and accusation, and the tricks that k the despair running smoothly.

If you're curious about building your own complex, AI-driven games, or just want know the engineering secrets behind making AIs convincingly scheme and strateg you won't want to miss it!

For now, why not try to survive the Killing Game yourself, or explore the open-source cod GitHub and see if you can spot the beginnings of the chaos we've talked about? The desp just getting started!

4 Likes

← Previous Next

Discussion about this post

Comments Restacks

Write a comment...