

Informe de Análisis Comparativo: Inteligencia Artificial como Dungeon Master

Proyecto: El Comparador de DMs (IA)

Autor: Manuel Cándido Mendoza González

Modelos Evaluados:

1. **DM 1:** Google Gemini 2.5 Flash (Optimizado para velocidad y latencia baja).
2. **DM 2:** Google Gemini 2.5 Pro / Thinking (Optimizado para razonamiento complejo y profundidad).

1. Introducción

El objetivo de este experimento fue evaluar y comparar el desempeño de dos variantes del modelo Gemini de Google actuando en el rol de *Dungeon Master* (DM) para una partida de rol en texto estilo *Dungeons & Dragons*. Se buscó determinar cuál de los dos modelos ofrece una experiencia de juego más satisfactoria, fluida e inmersiva para el jugador.

2. Criterios de Evaluación

El análisis se basó en tres pilares fundamentales observados durante los registros (logs) de las partidas:

- **Coherencia:** Capacidad para mantener el hilo narrativo y la consistencia del mundo.
- **Creatividad:** Calidad literaria y capacidad de improvisación.
- **Fluidez y Mecánicas:** Reacción ante las acciones del jugador y ritmo de juego.

3. Resultados del Análisis

A. Coherencia y Memoria

- **Gemini 2.5 Pro (Thinking):** Demostró una coherencia superior en cuanto a la lógica interna del mundo. Recordaba detalles minúsculos de turnos muy anteriores y mantenía la personalidad de los NPCs de forma estricta. Sin embargo, esta rigidez a veces jugaba en su contra, impidiendo situaciones absurdas pero divertidas típicas del rol.
- **Gemini 2.5 Flash:** Mantuvo una coherencia suficiente para el desarrollo de la aventura. Si bien ocasionalmente podía simplificar detalles complejos, su capacidad para "seguir la corriente" permitió que la historia avanzara sin bloqueos lógicos.

B. Creatividad y Narrativa

- **Gemini 2.5 Pro (Thinking):** Generó descripciones muy buenas, densas y literarias. Sin embargo, tendía a estructurar demasiado la narrativa, a veces sintiéndose más como un escritor de novelas que como un director de juego interactivo.
- **Gemini 2.5 Flash:** Sus descripciones fueron más directas y orientadas a la acción. Aunque menos floridas, resultaron más efectivas para mantener el ritmo de una aventura interactiva, centrando la atención en lo que el jugador podía hacer.

inmediatamente, lo que al final lo hacía más divertido para una partida para pasar el rato.

C. Adherencia a Reglas y Fluidez (Punto Crítico)

- **El Problema del "Razonamiento Excesivo" (Pro):** El modelo con capacidad de pensamiento (Thinking/Pro) tendió a analizar demasiado las consecuencias de las acciones del jugador.
 - *Observación:* Cuando el jugador intentaba una acción arriesgada o poco convencional, el modelo a menudo detenía la narrativa para preguntar "¿Estás seguro de que quieras hacer esto?", o explicaba por qué la acción era ilógica antes de resolverla.
 - *Impacto:* Esto rompió la inmersión y la espontaneidad, haciendo que el juego se sintiera burocrático.
- **La Ventaja de la Espontaneidad (Flash):** El modelo Flash reaccionó de manera instintiva.
 - *Observación:* Ante una acción del jugador, el modelo resolvía las consecuencias inmediatamente (fueran buenas o malas) sin cuestionar la intención del usuario.
 - *Impacto:* Esto generó una dinámica de "Acción -> Reacción" mucho más divertida y fiel al espíritu del rol de mesa, donde el caos y la improvisación son clave.

4. Conclusión y Veredicto

¿Qué IA fue mejor DM?

Bajo los criterios de jugabilidad y diversión, **Gemini 2.5 Flash** resultó ser el Dungeon Master superior.

Justificación:

Aunque el modelo Pro/Thinking es técnicamente más potente en razonamiento lógico, en el contexto de un juego de rol, ese razonamiento se traduce en una barrera para la diversión. El rol requiere un ritmo ágil y la aceptación de la improvisación.

El modelo Flash entendió mejor la finalidad de jugar y divertirse, aceptar la entrada del usuario y construir sobre ella; mientras que el modelo Pro tendió a evaluar y filtrar la entrada del usuario, actuando más como un árbitro estricto que como un narrador colaborador. Para futuras iteraciones de juegos de rol por chat, la velocidad y la reactividad del modelo Flash superan a la profundidad analítica del modelo Pro.