

Interim Report

Name: Ayman Arif Khan

Project Title in Project Plan: A real-time multiplayer narrative-building game using AI

Project Title: LLM-as-Judge: Designing a Multiplayer Game with AI-Driven Scoring Mechanics

Supervisor: Lewis Griffin

Progress Made to Date

Project Overview

- Multiplayer web game (10-20 players, 60-90 minutes)
- Players submit "headlines from the future" evaluated by an AI Juror
- Dice roll selects from 5 headline variants at different plausibility levels
- Multi-component scoring: plausibility, story connections, planet bonus

Tech Stack

- **Frontend:** React 18, TypeScript, Vite, Tailwind CSS, Socket.IO Client
- **Backend:** Node.js, Express, TypeScript, Socket.IO, PostgreSQL
- **AI:** OpenAI gpt-4o-mini via Responses API
- **Testing:** 188 tests across 11 suites (Jest)

Session & Lobby

- Host creates session with unique 6-char join code
- Players join via code, get player ID and nickname
- Socket.IO rooms for real-time broadcasts
- LocalStorage persistence for auto-rejoin on refresh

Game Loop

- State machine: WAITING → PLAYING → BREAK → FINISHED
- Configurable durations (default 15 min play + 5 min break)
- Accelerated in-game timeline (60x speed)
- Automatic phase transitions synced to database

Headline Submission Pipeline

- Player submits story direction → rate limit (1/min) → LLM evaluation → dice roll → band selection → scoring → broadcast

- AI Juror returns: plausibility grade (1-5), top 3 planets, top 3 connections (STRONG/WEAK), 5 headline variants
- Full LLM request/response stored for auditing
- Server-side dice roll with weighted bands (3:5:8:3:1 ratio)

Scoring

- **Baseline:** Fixed points per submission
- **Plausibility:** AI grades story direction; grade 3 → +2, grade 2/4 → +1, grade 1/5 → 0
- **Connection:** Discard weak connections; others' headline → +3, own → +1, none → 0 (mutually exclusive)
- **Planet bonus:** Awarded when headline matches player's priority planet

Planet Priority System

- 12 Greek-god-themed planets classify headlines thematically
 - Per-player tally of planet frequency in top-3 evaluations
 - Priority planet picked randomly from bottom half of tally (excluding previous)
 - Incentivises thematic variety across submissions
-

Currently Working On

- **End-to-end round summarisation:** Automating PLAYING → BREAK → next round transitions, storing round snapshots, generating AI break summaries, and handling game end with final leaderboard
 - **Priority planet frontend display:** Showing each player their current priority planet in the game UI so they can aim for the bonus
-

Remaining Work Before Final Report Deadline

Deployment (Target: early February)

- Host frontend and backend online (accessible via public URL)
- Configure production database and environment variables
- Ensure WebSocket connections work through hosting infrastructure

Playtesting (Target: mid-February)

- Run playtest session with 10-20 participants
- Gather feedback on scoring balance, headline quality, player experience

- Fix usability issues and edge cases found during live play
- Document findings and apply improvements

Leaderboard & Replay (Target: early March)

- Live rank updates during gameplay
- Per-round data stored for post-game replay
- Replay viewer UI
- Input filtering and moderation logs

Stretch Features (Target: mid-March)

- **Real Headlines Warmup:** Actual 2022-2025 headlines play during a 3-minute clock advance from 2022 to present; submit buttons activate when clock reaches today
- **AI Player:** LLM-powered competitor with same rules/cooldowns, host toggle
- **Audio Summaries:** TTS for break summaries with narrative chapter structure