WordPlay: An Agent Framework for Language Learning Games
Suma Bailis, Lara McConnaughey, Jane Friedhoff, Feiyang Chen, Chase Adams, Jacob Moon

1 WordPlay Framework

WordPlay Framework generates engaging, adaptive mini-games using LLMs and text-to-image models to address the challenge of balancing engaging and effective language practice.

2 Motivation

- Allows content creators to quickly author bite-sized, puzzles that cater to various CEFR proficiency levels, and uses generated images to aid comprehension.

3 Prompts

1. Setup

   I need help crossing the road [...]

2. Critic

   CONTEXT: You are the game engine in which a player is trying to help a chicken cross the road. Your job is to evaluate if their suggestion would be valid in a cartoon physics environment. I will provide you some examples and then you must evaluate the player input.

   EXAMPLES:

   INPUT: Use a rocketship
   status: VALID
   critique: NONE

   INPUT: Hey, how’s it going?
   status: NOT VALID
   critique: This is a standard greeting, not a valid suggestion for how a chicken can cross the road.

   INPUT: (@user_input)
   Respond using the following JSON format, matching the TypeScript interface:

   ```json
   { "status": string // NOT VALID | VALID,
     "critique": string // one sentence explanation if the suggestion is not valid }
   ```

   - Wordplay consists of only three prompts (or agents) we use PaLM for language generation and Imagen for image generation. All output is in JSON format and the game’s world-state is updated accordingly.

3. Image

   CONTEXT: You are the game engine in which a player is trying to help a cartoon chicken cross the road. The player suggested: (@user_input)

   Respond using the following JSON format, matching the TypeScript interface:

   ```json
   { "solution": string //Summarize the solution using only the key words
     "image_description": string //An image prompt that describes a white 2-D cartoon chicken crossing the road using the solution.
   }
   ```

5 Evaluation

- To assess our puzzles’ alignment to CEFR standards, we conducted experimental sessions with native Hindi-speaking individuals in India learning English.

- The CEF level of responses from both participants and the model underwent analysis through a custom CEF classification model.

- Puzzles evaluated:
  - Wedding - beginner puzzle asking users for suitable wedding attire
  - Finish the Story - intermediate puzzle asks users to take turns with the tutor to construct a narrative
  - Room Prepositions - intermediate puzzle to practice prepositions

- The CEF level of utterances from the tutor varied, with tutor utterances consistently hitting the B2 level, evidenced by the narrow interquartile ranges.

6 Conclusion & Next Steps

- WorldPlay allows content creators to author engaging and level-appropriate puzzles by customizing three simple prompts.
- The framework design prioritizes structured conversation over freeform dialogue, emphasizing safety in the tutor’s output.
- Future work aims to automatically generate puzzles from higher-level descriptions and also explore puzzles that use generated images on a turn-by-turn basis.