Protocol Competence of Small Language Models in Interactive Language Game Environments

Suchir Salhan, Paula Buttery

Department of Computer Science & Technology, University of Cambridge, U.K. · ALTA Institute, Cambridge, U.K.

The LM Playschool Workshop @ EMNLP 2026

Before AI can interact, it must learn to play.

Can demonstrations from a larger model teach an SLM how to participate in Interactive Language Learning Games?

Once an SLM can participate in an interactive language game, can it learn from its own experience without a teacher?

How much successful experience is needed for this self-learning to work?

Teaching models how to participate unlocks learning from their own experience.

Methods

We evaluate Qwen, Llama and Gemma (0.8B–7B) on five CLEMBench Playpen games using ClemScore, then compare protocol repair, teacher demonstrations and self-ReST on Qwen3.5-0.8B under matched compute budgets.

Teacher Demonstrations: Learn from supervised examples of valid interactions and successful solutions.

Self-ReST: Generate, filter, and learn from the model's own successful game trajectories.

- Generate: Sample 4 trajectories per game.
- Filter: Keep valid wins; balance across games.
- Fine-tune: Train on successful self-generated experience.
- Repeat: Generate better data with the improved model.

Alternative Strategies

- **ZPD Selection: Train on games with 25–75% win rates.**
- **Credit-Weighted ReST:** Prioritise informative trajectories and progress-making actions.
- Decoding Control: Vary temperature and sampling to explore alternative training data.

Results

IMAGEGAME — ABORTED EPISODE (QWEN2.5-3B-INSTRUCT)

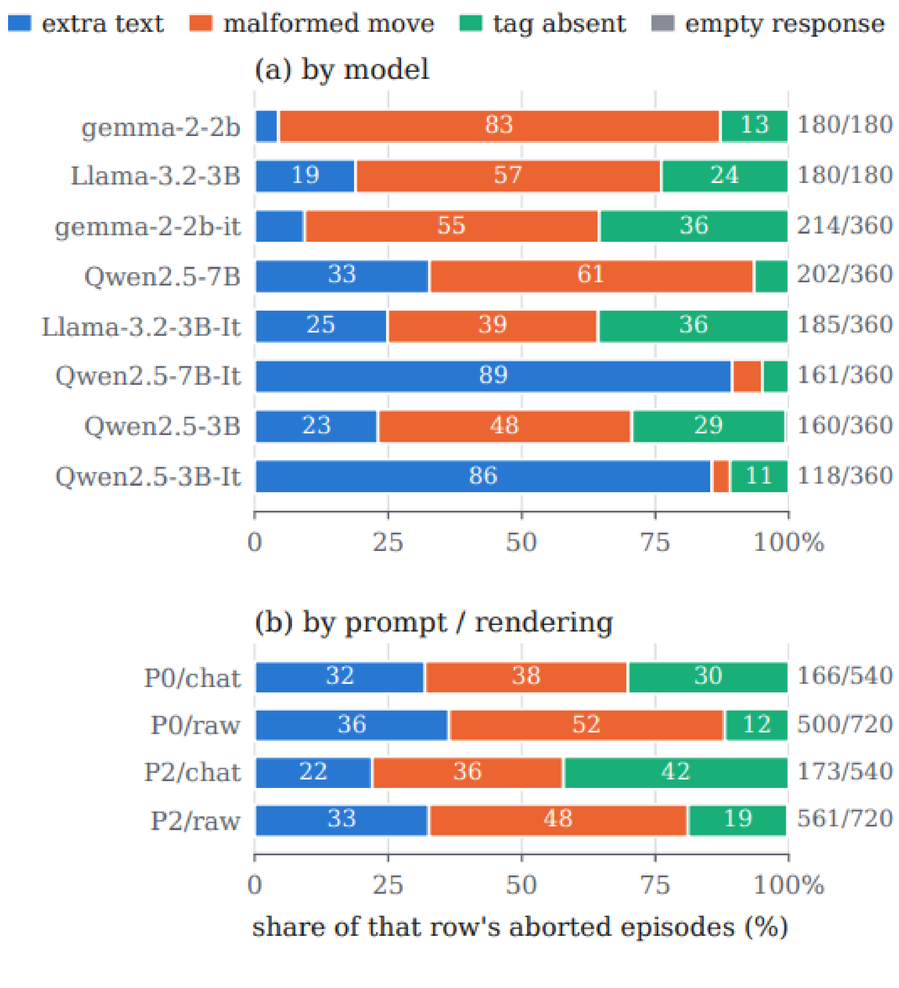
[A⟨GM]
Let us do the following task. The goal is to fill an empty grid that looks like this:
□ □ □ □ □
□ □ □ □ □
□ □ □ □ □
□ □ □ □ □
□ □ □ □ □
A filled grid below is 5 by 5 and can look like this:
□ □ □ □ □
□ □ E □ □
□ □ □ □ □
□ □ □ □ □
X X X [. . .]

[A⟩GM]
Command: Put an F in second row fourth column
What is your next command?

[GM|GM]
Invalid instruction format

[GM|GM]
Game aborted.

The player emits a well-formed command and then writes the co-player's turn as well; the game master rejects the turn and the episode aborts.



Small language models (Qwen, Llama and Gemma; 0.8B–7B) often fail at interacting correctly, not solving the task.

- *Smaller can outperform larger models in valid participation.*
- *3× CLEMSCORE improvement with self-ReST: 6.67 → 20.27.*

Model	Size	Play-rate	CLEMSCORE
Qwen3.5-0.8B	0.8B	0.59	6.67
Qwen2.5-1.5B	1.5B	0.00	0.00
Qwen2.5-3B	3B	0.38	7.04
Qwen3.5-4B	4B	0.00	0.00
Qwen2.5-7B	7B	0.17	0.00

Self-generated data is only useful when models can reliably interact with their environment.

Further work might include whether interaction and self-training can improve smll models in non-interactive language game environments.

DEMONSTRATIONS UNLOCK SELF-IMPROVEMENT

Model: Qwen3.5-0.8B; protocol repair, teacher demonstrations and self-ReST.

3× CLEMSCORE gain: 6.67 → 14.97 → 20.27 across protocol repair, demonstrations and self-ReST.

- **Credit assignment:** Credit-weighted ReST achieves the highest mean (19.17) and lower variance, but gains are not statistically significant.

Get in touch: sas245@cam.ac.uk