Systematic analysis and design of the kaggle competition “20 questions”

Muñoz Ipus Laura Daniela[1] Bautista Solano Esteban Alexander[1] Guerrero Ordoñez Luisa Fernanda[1]

School of Computer Engineering[1]
Universidad Distrital Francisco José de Caldas

Introduction

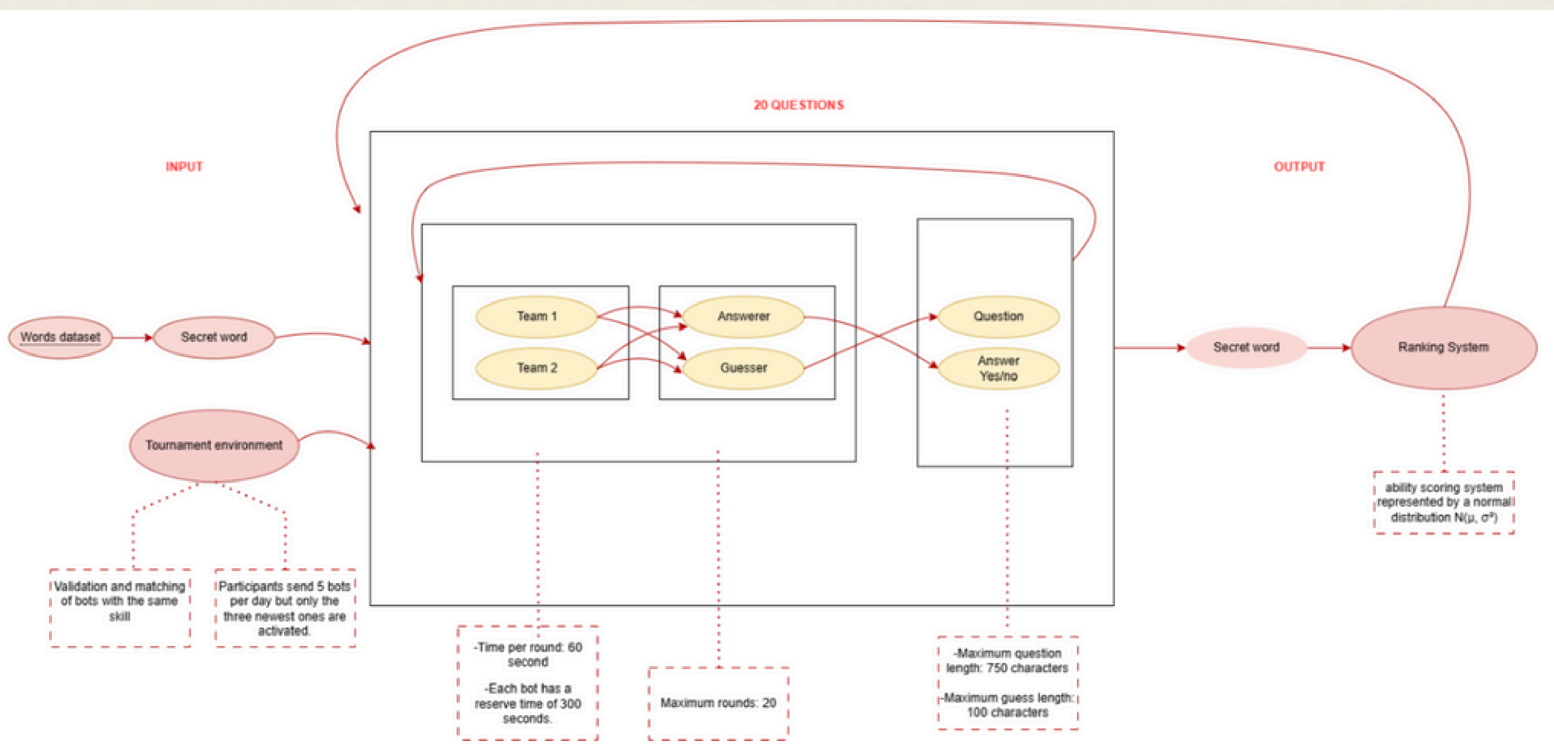
Kaggle's “20 Questions” competition poses the challenge of designing systems where language models (LLMs) guess a secret word by asking yes/no questions in an environment with uncertainty and limited time. Although there are rule-based or supervised learning approaches, challenges remain, such as effective question formulation and managing chaos in interaction. This project proposes a solution from a systems engineering perspective, integrating a modular architecture, simulations with real data, and the Gemma 2B-IT model to automate decisions through prompt engineering.

Goal

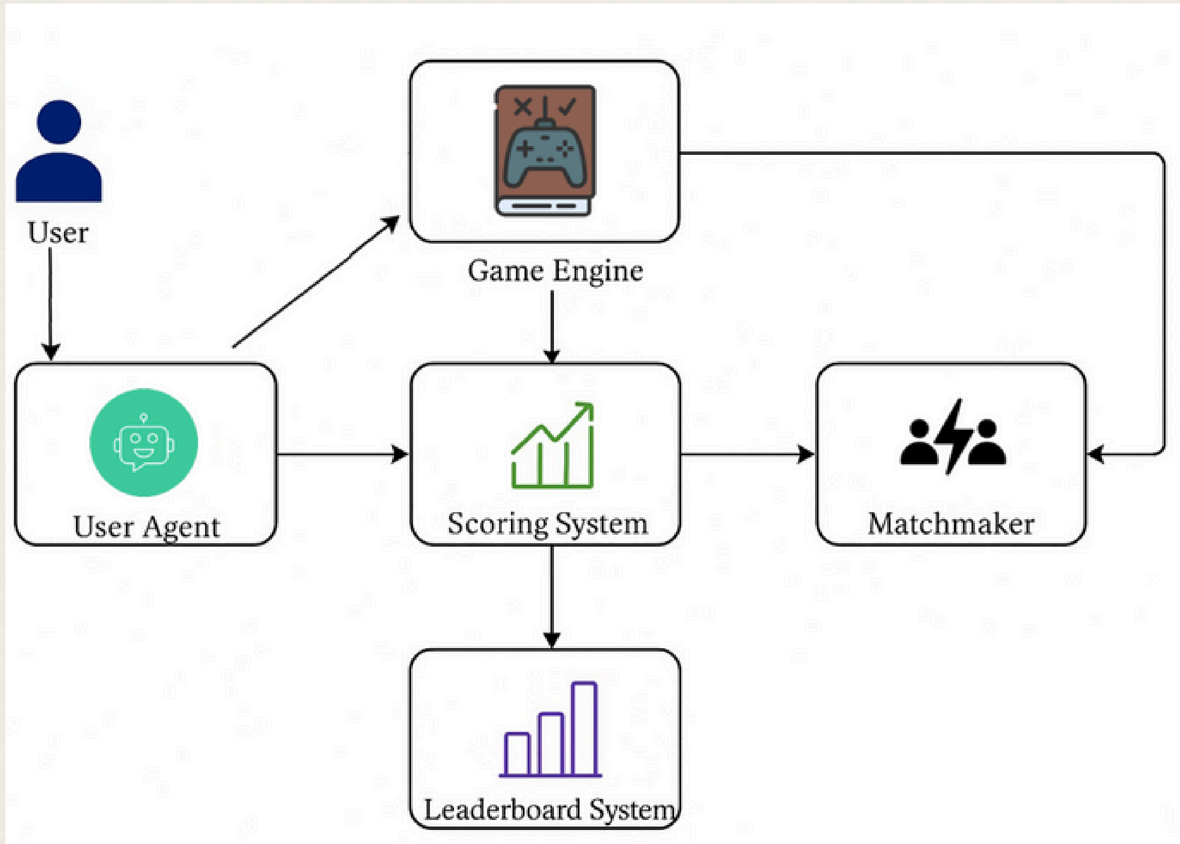
- Research question: How can the Kaggle “20 Questions” competition be analyzed, simulated, and enhanced by applying systems engineering principles?
- Expected final product: We aim to develop a modular architecture capable of simulating real competition dynamics and integrating the Gemma 2B-IT model, enabling automated interaction between agents and allowing for performance evaluation.

Proposed solution

The system models interactions using feedback loops, enabling homeostasis and a life cycle of creation, competition, and evaluation. It incorporates core systems analysis concepts such as structure, behavior, sensitivity, and self-regulation.



A four-layer modular design—interaction, control, evaluation, and management—ensures scalable, traceable, and adaptive system behavior under uncertainty.



Experiments

We simulated matches using real Kaggle data under three scenarios: balanced, chaotic, and skilled. The system was later extended with the Gemma 2B-IT model to automate question-answering, allowing us to compare LLM-driven performance under controlled conditions.

Results

The system solved the games in an average of 4.1 rounds, demonstrating its deductive efficiency.

The chaotic and balanced scenarios had similar success rates, indicating low variability.

Greater skill meant shorter game duration, with predictable patterns between rounds and success.

Conclusions

- The system successfully simulated the dynamics of competition, reflecting consistent behavior among agents.
- Modular architecture proved to be efficient, scalable, and capable of adapting to different scenarios.
- The use of real data allowed us to validate the robustness of the system under real conditions

Bibliografía

- Waechter, T. (2024). LLM 20 Questions Games Dataset. Kaggle. https://www.kaggle.com/datasets/waechter/llm-20-questions-games?select=EpisodeAgents.csv
- Muñoz, L. D. (2025). Systems Analysis and Design. GitHub. https://github.com/LauraDanielaa/systems-analysis-and-design-
- Sierra, C. A. (2025). Systems Analysis and Design – Course Materials. GitHub. https://github.com/EngAndres/ud-public/tree/main/courses/systems-analysis