Contribution ID: **17** Type: **Poster**

Comparative Analysis of Multi-Agent LLM Systems for Solving Polish Matura in Physics Exams

Thursday 24 April 2025 10:45 (30 minutes)

Large Language Models have gained widespread recognition since OpenAI released their revolutionary model, ChatGPT 3.5. Since then, many new approaches have emerged to improve the capabilities and accuracy of these models for different tasks. One such method involves using multi-agent conversations. This article compares two multi-agent setups designed to solve the Polish standardized high school exam in physics. Comparative benchmarks were performed on several real final exams published by the Polish Central Examination Board (pl. CKE —Centralna Komisja Egzaminacyjna). The study employed ChatGPT-4 Turbo and the AutoGen framework. Benchmarks covered a total of 90 tasks from three Polish Matura physics exams (editions: 2018, 2019, 2023). The simpler multiagent systems achieved an average score of 76.1%, while the more complex systems averaged 85.6%.

Author: WRÓBEL, Piotr (Politechnika Warszawska)

Presenter: WRÓBEL, Piotr (Politechnika Warszawska)

Session Classification: Session B (Poster)