Cover Page – NeurIPS 2025 Education Resource Submission

Title:
Application of Multi-Agent Systems for Essay Scoring

Authors:

- **Ali Keramati**, PhD Student, Digital Learning Lab, University of California, Irvine
- **Mark Warschauer**, **Distinguished Professor of Education and Informatics**, *and Director of the Digital Learning Lab*, School of Education, University of California, Irvine

Target Audience:

This resource is designed primarily for educators in K–12 and higher education, interested in AI for learning and assessment, and professionals in assessment, testing, and essay evaluation, as well as education researchers and policymakers exploring AI in learning contexts. It may also be useful for high school and undergraduate students from non-CS backgrounds, and for general audiences curious about AI's real-world applications.

Format & Length:

Recorded video of a slide presentation with live narration, estimated viewing time of 15 minutes.

Description:
This educational resource introduces the concept of AI agents and then expands to multi-agent systems, showing how different architectures can work together to improve performance.

The presentation begins with simple examples that highlight both the power and the limitations of AI, making the idea of an "AI agent" concrete and accessible. It then transitions into a real-world challenge in education: essay scoring, where teachers often face overwhelming workloads and students need faster, more actionable feedback. From there, the slides explain how a single-agent system tackles this task, followed by its limitations. The core of the presentation introduces and compares several multi-agent architectures, supervisor, collaboration, and debate, each illustrating different ways agents can share work, critique each other, or argue opposing perspectives to reach stronger outcomes. Finally, results from experiments on the ASAP[1] dataset demonstrate how **Multi-Agent Debate Essay Scoring Triangulation (MADEST)** can even outperform average human-to-human agreement. The conclusion highlights the trade-offs: single-agent systems are simple and fast, while multi-agent systems are more complex but yield richer, more reliable scoring, closer to how real teachers evaluate essays.

This resource is designed to make advanced AI ideas like multi-agent architectures accessible and engaging for educators, assessment professionals, and education researchers, with a clear focus on real-world application and responsible use.

---

[1] https://www.kaggle.com/datasets/lburleigh/asap-2-0