Bic Rouge: Revolutionizing Theoretical Exam Grading with GPT-3.5 Turbo and Llama-3 for Enhanced Educational Accessibility

MUHIGIRI Ashuza Albin[1],
Prof. BARAKA MUSHAGE Olivier[1],
Ms. KRAME KADURHA David[1]

Introduction

The Challenge in Education

High student-to-teacher ratios across Africa, reaching up to **1/200 in the DRC**, create immense challenges. Manual grading of theoretical exams is not only time-consuming but also subjective, leading to teacher burnout and hindering personalized student feedback.

Figure 1: The high student-teacher ratio is a major hurdle in providing quality education.

"Our goal is to leverage AI to empower educators, making grading more efficient, consistent, and fair for every student."

BicRouge is an intelligent system leveraging Large Language Models (LLMs) to assist teachers in grading theoretical exams.

Key Features

- Automated correction generation
- Teacher validation and adjustment interface
- Immediate, personalized student feedback
- Modular and scalable architecture

Methodology & Architecture

LLM Integration

We evaluated and compared two powerful models for this task:

- **GPT-3.5 Turbo:** A proprietary model known for its reliability and high performance.
- **Llama 3 8B - Instruct:** A leading open-source model, prized for its accessibility and adaptability.

System Workflow

The system is built with Python, FastAPI, and ReactJS, deployed on Azure AI. The workflow is designed for seamless interaction:

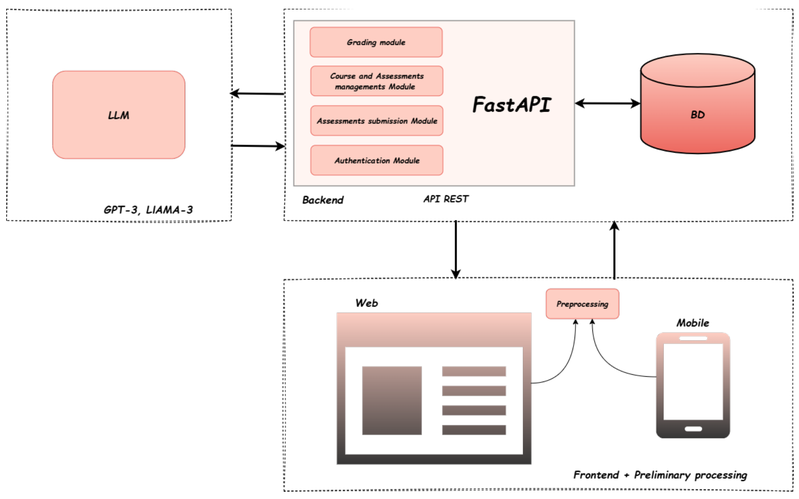

Figure 2: Bic Rouge System Architecture: From student submission to teacher-validated feedback.

90%

The potential time-saving for educators using an automated grading assistant. This allows more focus on teaching and less on manual grading. Based on recent studies showing 80-90% time savings with AI grading tools.

Experimental Results

Our evaluation focused on coherence, fluency, and semantic similarity with human-graded responses.

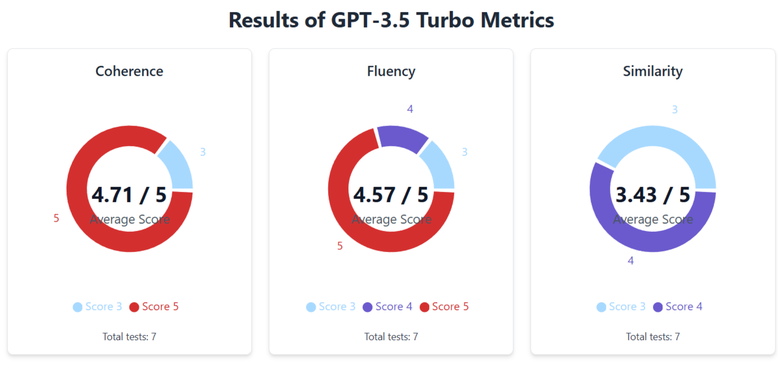

Figure 3: GPT-3.5 Turbo showed strong alignment with human evaluations across all metrics.

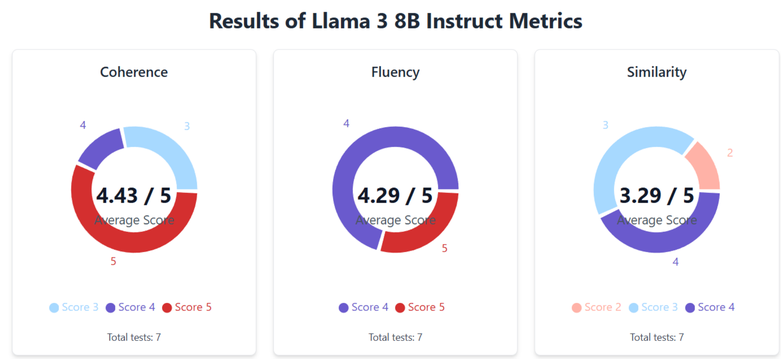

Figure 4: Llama 3 8B - Instruct proved to be a powerful open-source alternative.

Conclusion & Future Directions

Bic Rouge successfully demonstrates the feasibility of LLM-powered automated grading, providing a robust architecture that improves efficiency and offers immediate, personalized feedback.

Future Work

- **Enhancing Accuracy:** Fine-tuning models and integrating Retrieval-Augmented Generation (RAG) for context-aware evaluations.
- **Broader Accessibility:** Developing a mobile application and integrating vision models to grade handwritten exams, addressing infrastructure limitations.
- **Exploring New Models:** Continuously integrating more advanced LLMs as they become available.

More Information

Affiliations

[1] Université Libre des Pays des Grands Lacs, Democratic Republic of the Congo (DRC)

Acknowledgements

Deep Learning Indaba 2025 | #DLIndaba2025

Scan to explore the full project (code, video, paper).
Or visit: `https://tinyurl.com/bicrouge`