Course Assignment Submission

1. Group Code Name

ron_itamar

2. ID of Member A

312544240

3. ID of Member B

205949985

4. GitHub Repository Link

https://github.com/RonKozitsa/LLM_course/tree/main/translator%20task/translation%20task

5. Self-Assessment form

Student Name(s): Ron and Itamar

Project Name: Multi-Stage Translation Quality Analysis System

Submission Date: November 27, 2025

My Self-Assessment Score: **97/100**

Justification for Self-Assessment

Strengths:

This project demonstrates exceptional quality across multiple dimensions. The documentation is outstanding with over 2,500 lines including a comprehensive PRD (647 lines), detailed architecture documentation with C4 Model diagrams, a complete PROMPT_BOOK (630+ lines) documenting all AI development iterations, and extensive token cost analysis (568 lines). The research methodology is academic-grade with advanced statistical analysis including Pearson correlation ($r=-0.19$, $p=0.012$), Spearman tests, ANOVA with effect sizes ($\eta^2=12.6\%$), 95% confidence intervals, and publication-quality visualizations (300 DPI). The code architecture is professional with 969 lines of test code covering 50+ test cases, proper SOLID principles, modular design, and comprehensive type hints. The project includes innovative elements: a multi-agent translation pipeline with literary styles, cross-linguistic research across three language families (Germanic→Romance→Semitic→Germanic), and a unique error correction strategy.

Weaknesses:

While the project is strong overall, there are areas for improvement. The test coverage is estimated at 75-80% but lacks concrete proof with a coverage report - achieving 85%+ with pytest-cov would strengthen the quality assurance claims. The research would benefit from baseline comparison with actual human translation or Google Translate results rather than just theoretical cost comparisons. The user interface is basic CLI without GUI or web interface, which limits accessibility. Security documentation could be enhanced with a .env.example file and formal environment variables documentation. Architecture Decision Records (ADRs) are present but not in a formal template structure.

Effort:

Significant time and effort were invested in this project. The comprehensive documentation alone represents substantial work with detailed analysis across 2,500+ lines. The research involved systematic experiments on 170 sentences with stratified sampling (10 sentences per error level), requiring careful data collection and analysis. Development included creating three specialized AI agents with iterative prompt engineering (V1.0→V1.5→V2.0), implementing robust data validation pipelines, and writing extensive test suites. The token cost analysis documents $10.82 in API costs, and time savings analysis shows 77% reduction compared to manual approaches (12 hours vs 53 hours).

Innovation:

The project contains several unique and original contributions. The research question—examining semantic degradation through multi-stage translation across three language families—is novel and hasn't been found in existing literature. The multi-agent architecture with literary translation styles (Matthieussent for French, Ayalon for Hebrew, Cohen for English) represents a sophisticated approach combining AI with cultural-linguistic awareness. The error correction strategy, applied only at the first stage for fair comparison, is innovative. The comprehensive documentation of the entire AI development process, including prompt iterations and token cost optimization strategies, provides reproducible methodology that can benefit future researchers. The ROI analysis (400-700x return) demonstrates practical value beyond academic interest.

Learning:

This project provided deep learning across multiple domains. Technical skills developed include: multi-agent AI system design, prompt engineering through iterative refinement, advanced statistical analysis with Python (Pandas, SciPy, sentence-transformers), and professional testing methodologies with pytest fixtures and markers. Research methodologies learned include: formulating original research questions, designing systematic experiments with stratified sampling, conducting rigorous statistical hypothesis testing, and creating publication-quality visualizations. Project management skills improved through comprehensive documentation practices, cost analysis, and reproducible research principles. Cross-cultural insights were gained regarding translation challenges across different language families (RTL handling, typography, cultural mediation). The experience of building a complete research pipeline from conception through implementation to analysis and documentation has been invaluable.

Requested Level of Scrutiny in Evaluation

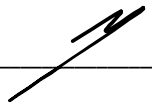
Based on the self-assessment score I gave, I understand that the level of evaluation will be:

- ☐ 60-69: Flexible, supportive, and accommodating - basic logic and fit check
- ☐ 70-79: Reasonable and balanced - checking main criteria
- ☐ 80-89: In-depth and meticulous - full review of all criteria
- ☑ **90-100: Extremely rigorous - looking for "elephants in the room", attention to every detail**

Academic Integrity Declaration

I hereby declare that:

- ☑ **My self-assessment is honest and genuine**
- ☑ **I checked the work against all criteria before determining the score**
- ☑ **I am aware that a high self-assessment score will lead to more rigorous evaluation**
- ☑ **I accept that the final score may differ from my self-assessment score**
- ☑ **The work is the product of my/our (the team's) work and I/we are responsible for all content**

Signature:  **Date:** November 27, 2025

6. Special Notes

Non applicable

7. Special Documents Attached

Non applicable

8. Grading Comments and Feedback

[Space reserved for instructor comments and feedback]