**Project Deliverable Evaluation (1–5 scale per parameter)**

|  |  |  |
| --- | --- | --- |
| Parameter | Description | Scoring Guide |
| Correctness (Prompt &  Chaining) (Weight: 30%) | Implements prompt design, LLM chaining, RAG  context injection. | 1 = Not implemented, 2 = Minimal  attempt, 3 = Works partially, 4 = Works  correctly, 5 = Fully correct + thoughtful |
| Code Quality & Structure  (Weight: 25%) | Clean, modular, reusable, tested. | 1 = Poor, 2 = Some structure, 3 = Decent  modularity, 4 = Good structure + some  tests, 5 = Excellent quality + strong tests |
| Resilience & Error Handling  (Weight: 20%) | Handles long jobs, retries, randomness, API  failures. | 1 = Missing, 2 = Minimal, 3 = Partial  handling, 4 = Solid handling, 5 = Robust,  production-ready |
| Documentation & Explanation  (Weight: 15%) | README clarity, setup instructions, trade-off  explanations. | 1 = Missing, 2 = Minimal, 3 = Adequate, 4  = Clear, 5 = Excellent + insightful |
| Creativity / Bonus (Weight: 10%) | Extra features beyond requirements. | 1 = None, 2 = Very basic, 3 = Useful  extras, 4 = Strong enhancements, 5 =  Outstanding creativity |