Project #	Team Members	Github Repo	Industry Judge Grade (/10) - 20%	System Design & Implementation (/10) - 20%	Evaluation Rigor (/10) - 15%	Task Performance (/10) - 20%	Safety & Ethics (/10) - 5%	Poster & Demo (/10) - 10%	Report & Documentation (/10) - 10%	Final Grade
				Architecture quality (single vs multi-agent rationale) Clear roles, scratchpad, memory, state management Tools integration (≥3) Cost tracking and latency considerations Observability (logs, traces, dashboards) Reliability (timeouts, retries, schema validation) Proper use of software tools	Construction of task suite and baselines Measurement of accuracy, reliability, latency, and cost Automated evaluation harness Bias checks and human spot-checks Completeness and reproducibility of results	Completion of ≥70 percent of tasks Ability to generalize to held-out test cases Quality of outputs Correct orchestration Measured improvement over baselines	Domain-appropriate disclaimers Human-in-the-loop strategies for risky tasks PII redaction Prompt-injection defenses Use of safe workflows and clear fallback paths	Problem & users Tasks & success criteria Architecture diagram Evaluation plan Safety & ethics Pilot evidence Roadmap Plus demo clarity, timing, and professionalism	GitHub repo structure Reproducibility (install steps, configs, seeds) Clear documentation of experiments, evaluation, and design decisions Cost log.csv and dashboards Code readability and comments	
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										