Veris AI Candidate Task: Build an AI Agent

Your task is to build an AI agent for the financeQA benchmark presented on https://www.financearena.ai/. The benchmark currently ranks LLMs in a non-agentic approach, where OpenAI is leading with only 54.1% overall score. The goal is to introduce an agentic approach to improve on this baseline. You can decide to use any tools that you wish including a RAG system to improve on the benchmark evaluation dataset. The task's main focus is on your thought process and implementation rather than the actual score you get on the benchmark.

Requirements

- Language: Python or TypeScript.
- Frameworks & Models: Any LLM of your choice (e.g., OpenAI, open-source LLMs). You may use commercial APIs or open-source models. You may also choose to use an agentic framework of your choice or not.
- Tools: You can choose any API, MCP, or existing tool library/framework. You can decide to use RAG as a tool.
- You are free to use any coding assistant or environment for the project.

Deliverables

1. Working code for your agent
2. Documentation explaining:
    a. An agent card explaining what the agent does
    b. Your overall design approach
    c. Key trade-offs and decisions
    d. Evaluation over the benchmark dataset
    e. Suggestions for future improvements
3. (Optional) Containerized deployment (e.g., Docker) or host on a cloud environment

Evaluation Criteria

- Code quality & organization (clarity, modularity, best practices)
- Functionality (agent should work reliably and meet the task objective)
- Documentation quality (clear explanation of design and reasoning)
- (Bonus but not required) Benchmark score and improvement over the current non-agentic baselines

Expected Effort

This task is designed to take 4 to 8 hours of work. We expect a submission within a week.

Submission

Email us a GitHub repository link or a zip file containing:

- All source code
- Any configuration or environment files (e.g., requirements.txt or package.json)
- Documentation
- (Optional) Deployment files (Dockerfile, scripts, etc.)

If you have any questions, please reach out to **hello@veris.ai**