Autonomous Research Agent (Agentic AI + RAG + MCP)

1. Project Overview

This project is an Autonomous Research Agent that performs structured, multi-step research automatically.
Instead of giving a single chatbot response, the agent:
- Breaks down the topic into sub-questions
- Searches the web using tools (via MCP server)
- Extracts and summarizes key information
- Stores knowledge in vector memory (RAG)
- Reflects on missing gaps
- Generates a structured research report with citations

---

2. Core Architecture

User → Orchestrator Agent → Planner → Research Tool (MCP) → Extractor →
Vector Database (RAG Memory) → Reflection Agent → Report Generator

Each component has a specific role in creating a reliable, multi-step reasoning system.

---

3. Components Explained

Orchestrator Agent:
Controls the overall workflow and coordinates tasks.

Planner Agent:
Breaks the main topic into structured sub-questions.

Research Tool (via MCP Server):
Allows secure external tool access such as:
- Web Search
- PDF Reader
- Citation Tool
- Article Extractor

Extractor Agent:
Cleans raw content and extracts meaningful insights.

RAG Memory (Vector DB):
Stores embedded summaries for future retrieval and reduces hallucination.

Reflection Agent:
Checks for missing information and triggers additional research if needed.

Report Generator:
Produces a clean, structured report with headings and citations.

---

4. RAG Implementation

Steps:
1. Chunk extracted content
2. Generate embeddings
3. Store in vector database (Chroma/Pinecone)
4. Retrieve relevant chunks during reasoning
5. Ground responses in retrieved content

Benefits:
- Reduces hallucinations
- Enables memory persistence
- Supports follow-up queries

---

5. MCP Server Role

The MCP server acts as a structured tool registry.
The agent does not directly access external systems.
Instead:
Agent → Requests tool → MCP executes → Returns structured result

Benefits:
- Secure architecture
- Modular design
- Production-ready tool orchestration

---

6. Suggested Tech Stack

Backend: FastAPI
Agent Framework: LangGraph
LLM: GPT / Claude / Open-source model
Embeddings: OpenAI / BGE
Vector DB: Chroma / Pinecone
Frontend: Streamlit
Evaluation: RAGAS

---

7. Advanced Features (Optional)

- Hybrid Retrieval (BM25 + Embeddings)
- Source credibility scoring
- Reflection loop improvements
- Research depth control
- PDF export
- Research history tracking

---

8. Resume Bullet Example

Designed and deployed a multi-agent autonomous research system using RAG and MCP tool orchestration.
Implemented structured planning, hybrid retrieval, citation validation, and reflection-based reasoning
to improve factual grounding and reduce hallucinations.
