

Research Assistant Agentic System

Multi-Agent Workflow using CrewAI

Nithin Yash Menezes

Northeastern University – Master of Science in Information Systems

Project Overview

This project introduces an **automated multi-agent system** designed specifically for academic research workflows. Built on the CrewAI Studio framework, it represents a cutting-edge approach to intelligent task orchestration.

The system executes a complete **end-to-end research pipeline**, from initial query to final deliverable, with minimal human intervention.



3 Specialized Agents

Each with distinct roles and expertise

4 Integrated Tools

3 built-in + 1 custom-developed

Markdown Reports

Professional, structured output

Core Objectives

01

Multi-Agent Orchestration

Design sophisticated coordination mechanisms for complex, sequential research tasks requiring specialized expertise at each stage

02

Tool Integration

Seamlessly integrate web search APIs, intelligent scraping capabilities, and advanced LLM analysis into a unified workflow

03

Custom Development

Develop a specialized tool for generating professional-grade, academically-formatted research reports with consistent structure

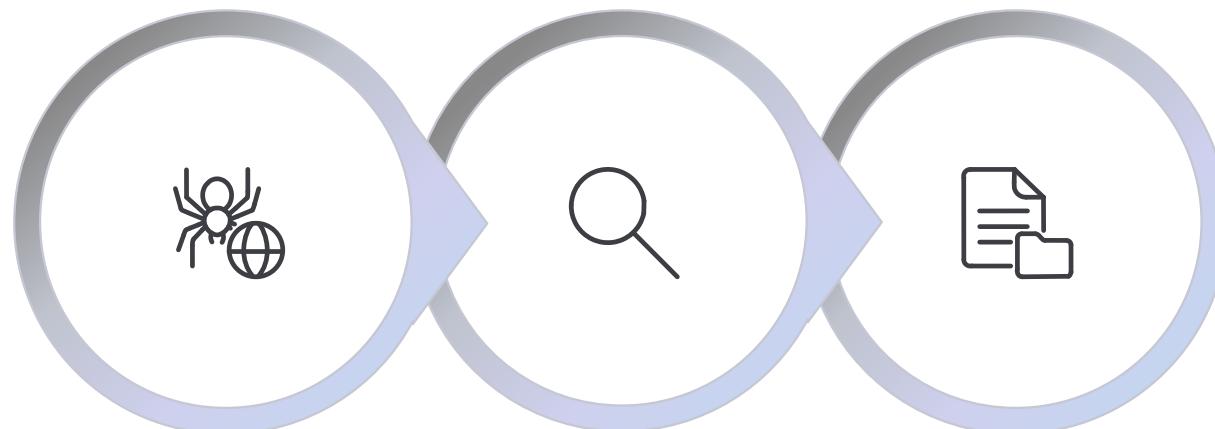
04

Operational Excellence

Ensure system reliability, output clarity, and optimal performance throughout the entire execution pipeline



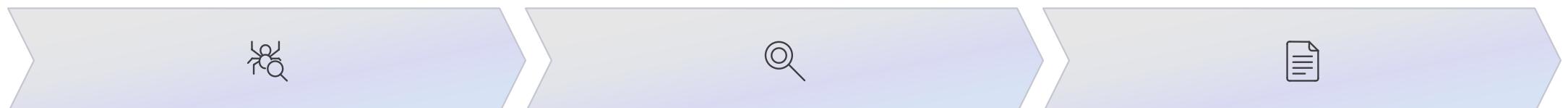
System Architecture



Search

Analysis

Report



Search Phase

Query formulation and source discovery

Analysis Phase

Content extraction and insight generation

Report Phase

Structured synthesis and formatting

The **controller orchestrates** the entire workflow, managing inter-agent communication, maintaining shared memory, and coordinating data flow through structured JSON. The output is delivered as a complete, publication-ready research report.

Agent Specialization

Three intelligent agents work collaboratively, each bringing domain-specific expertise to the research pipeline:



Research Specialist

Primary Function: Information Discovery

- Executes sophisticated web searches using optimized queries
- Identifies and evaluates credible, authoritative sources
- Filters results based on relevance and academic rigor



Content Analyst

Primary Function: Deep Analysis

- Scrapes and parses web pages with precision
- Extracts key insights, statistics, and supporting evidence
- Synthesizes information across multiple sources



Research Report Writer

Primary Function: Document Generation

- Generates final structured research reports
- Ensures academic writing standards and clarity
- Applies consistent formatting and citation styles

Integrated Toolset

Built-In Tools

SerperDevTool

Purpose: Web Search API

Provides access to real-time search results with advanced filtering and relevance ranking capabilities

ScrapeWebsiteTool

Purpose: Intelligent Web Scraping

Extracts structured content from web pages while handling dynamic content and diverse HTML structures

LLM Processing Tool

Purpose: Natural Language Analysis

Powers semantic understanding, summarization, and intelligent content synthesis across all agents

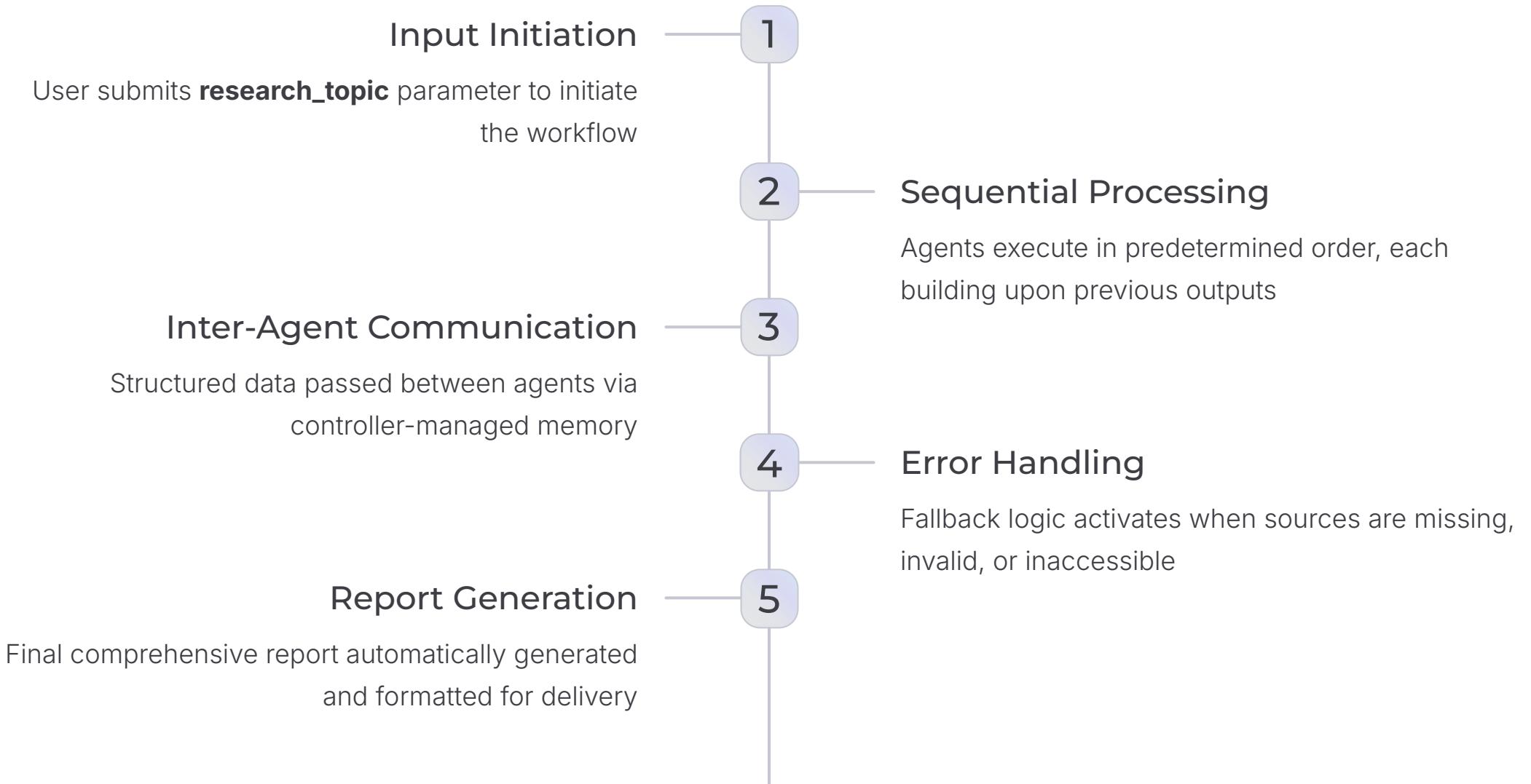
Custom-Built Tool



Markdown Report Generator

Proprietary tool that produces **academic-style structured documents** with consistent formatting, proper citations, and professional presentation standards

Workflow Execution Pipeline



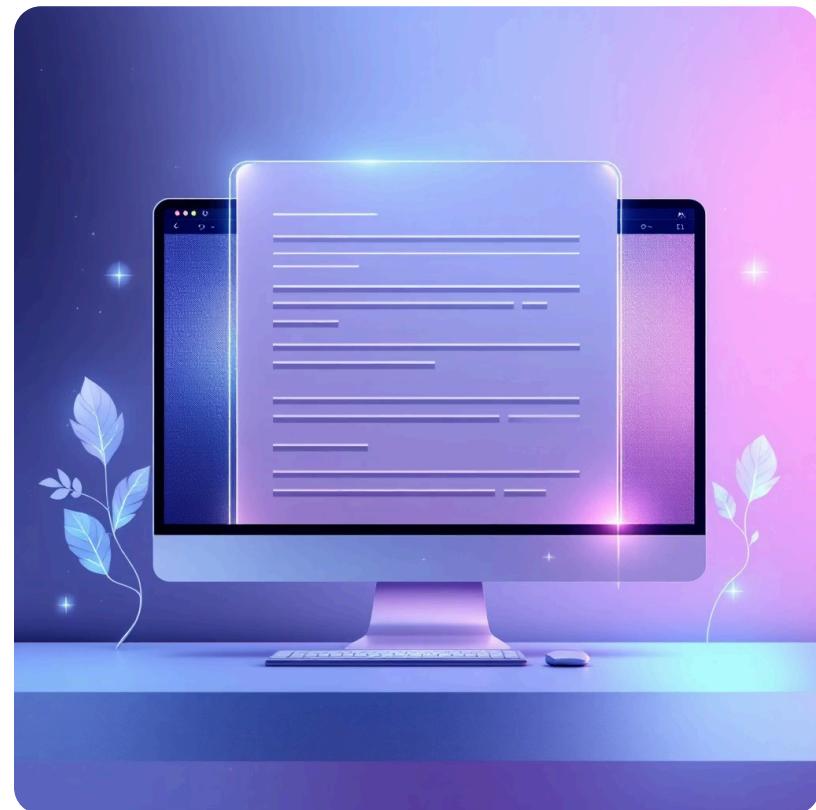
- ❑ **Key Feature:** The system operates autonomously from start to finish, requiring human intervention only for topic specification and final review.

Sample Output Structure

The system generates comprehensive research reports with standardized academic formatting:

Executive Summary High-level overview and key takeaways	Introduction Context, scope, and research objectives
Key Findings Primary discoveries and data points	Analysis & Discussion Deep interpretation and implications
Conclusion Synthesis and future directions	References Properly formatted citations

"Each section is meticulously structured to meet academic standards, with clear hierarchies, logical flow, and evidence-based argumentation throughout the document."



Performance Evaluation

Strengths

Full Automation

Complete pipeline operates autonomously from query to final report with zero manual intervention

Quality Output

Generates high-quality, academically structured documents that meet professional standards

Seamless Integration

Tools and agents work cohesively through well-designed interfaces and data contracts

Scalable Design

Architecture supports expansion with additional agents, tools, or workflow stages

Limitations

Structure Dependency

Scraping effectiveness varies based on website HTML structure and consistency

Access Barriers

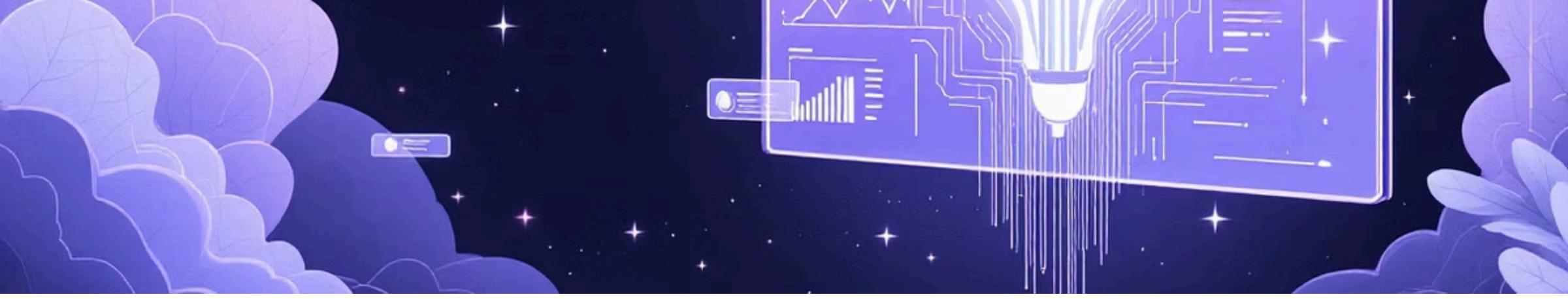
Cannot extract content from paywalled, login-protected, or dynamically rendered sites

Source Limitations

Quality depends on available online sources and search API coverage

Context Bounds

LLM token limits may constrain analysis of extremely long documents



Conclusion & Impact



Successful Demonstration

Proves viability of multi-agent AI systems for complex academic workflows



Operational Excellence

Delivers efficient, reliable, and consistent research outputs



Expandable Foundation

Provides robust architecture for advanced agentic workflows

This project **successfully meets all assignment requirements** while demonstrating practical applications of CrewAI's multi-agent orchestration capabilities. The system represents an ideal foundation for building more sophisticated research automation tools, with clear pathways for enhancement including additional specialized agents, expanded tool integration, and advanced reasoning capabilities.

- ❑ **Future Directions:** The architecture supports integration with citation databases, advanced NLP models, and collaborative multi-user research environments.