**Technical Specification: Deep Research Workflow System**

**1. Overview**

The Deep Research Workflow system is an automated, agent-based pipeline for high-quality research and document generation. It leverages modular tool servers, multiple autonomous agents, and human-in-the-loop review points to ensure accuracy, comprehensiveness, and professional formatting of research outputs.

**2. Architecture**

**2.1. System Components**

**User Prompt Interface**

* Receives initial user research prompt.
* Presents planning results and collects user feedback.

**Planning Agent**

* Analyzes the user prompt and generates a structured multi-step research plan.
* Passes plan to human for review and approval.

**Human Review**

* User approves or requests changes to the plan.
* Ensures plan alignment before resource-intensive steps.

**Research Tool MCP Server**

Gradio MCP Server for Document Processing

* Hosts tool APIs for data gathering, search, and knowledge extraction.
* Tools include:
  + Web Search (DuckDuckGo, Tavily)
  + Web Scrape
  + Google Scholar (SerpAPI)
  + Semantic Scholar
  + arXiv
  + PubMed/NCBI
  + Wikipedia/Wikidata
  + OpenAlex
  + Crossref

**Research Agent**

* Executes the research plan using Research Tools.
* Collects, synthesizes, and documents relevant information.

**Senior Research Agent**

* Validates research quality, checks for coverage and depth.
* Requests additional research or refinements if necessary.
* Research Complete decision point, max 3 iterations before escalation.

**Write Agent**

* Drafts a comprehensive, well-structured report using the gathered research.

**Review Agent**

* Reviews the draft for clarity, completeness, and alignment with the plan.
* Utilizes checklist validator tool.
* Flags required corrections.

**Revision Agent**

* Applies corrections and improvements as per Review Agent feedback.
* Up to 3 iterations allowed; escalation on failure.

**Document Tool MCP Server**

Gradio MCP Server for Document Processing

* Hosts tools for report enhancement, styling, and formatting:
  + Citation Formatter
  + Checklist Validator
  + Template Engine
  + Chart/Table Generator
  + Format Converter

**Formatting Agent**

* Applies templates, converts formats, and inserts charts/tables as needed.

**Summary Agent**

* Generates an executive summary or key takeaways for the report.

**Final Document Output**

* Delivers the finished, formatted report as a PDF file.

**3. Data Flow & Orchestration**

1. User submits research prompt.
2. Planning Agent generates a research plan.
3. Human Review: User approves or revises plan.
4. Research Agent executes plan, using research tools (via MCP server).
5. Senior Research Agent reviews research completeness. Loops up to 3 times as needed.
6. Write Agent drafts the report.
7. Review Agent checks report quality, using document tools as required. Up to 3 revision loops.
8. Revision Agent implements feedback.
9. Formatting Agent applies document styling, templates, charts/tables, and format conversion.
10. Summary Agent produces an executive summary.
11. Final Document is output as a styled PDF.

**4. Tooling & Technology Stack**

* LLM: OpenAI GPT-4.1-mini (for all agents' natural language reasoning tasks).
* Frameworks: Gradio (for MCP servers and UI), LlamaIndex (for agent orchestration), Python (primary language).
* APIs: See Section 2.1 (Tool lists).
* Templating: Jinja2, Pandoc, or docxtpl (for document styling).
* Charts/Tables: Matplotlib, Plotly, Pandas.
* PDF Conversion: WeasyPrint, pdfkit, or equivalent.

**5. Agent-to-Tool Mapping**

|  |  |
| --- | --- |
| Agent | Tool Access (via MCP) |
| Research Agents | All research tools (search, scraping, scholarly APIs, Wikipedia) |
| Write Agent | Citation Formatter |
| Review Agent | Checklist Validator |
| Revision Agent | (No dedicated tools; acts on feedback) |
| Formatting Agent | Template Engine, Chart/Table Generator, Format Converter, Citation Style Converter |
| Summary Agent | (LLM-based summarization) |

**6. Human-in-the-Loop Controls**

* Pre-research: Users review and approve research plan.
* Escalation: If max research/revision iterations are reached, system notifies the user for manual intervention or decision.

**7. Error Handling & Iteration**

* Each research/revision decision point allows up to 3 automated iterations.
* On exceeding iteration limits, the process escalates to the user or logs for admin attention.
* All failures, retries, and tool errors are logged for audit and debugging.

**8. Security & Compliance**

* All API keys and credentials managed via secure environment variables.
* Data privacy and usage in accordance with relevant regulations (e.g., GDPR).

**9. Extensibility**

* Additional agents or tools can be added to either MCP server with minimal code changes.
* Modular agent design allows for new workflow steps or alternative toolsets as needed.