

AI Research Agent: Real-Time Knowledge Synthesis

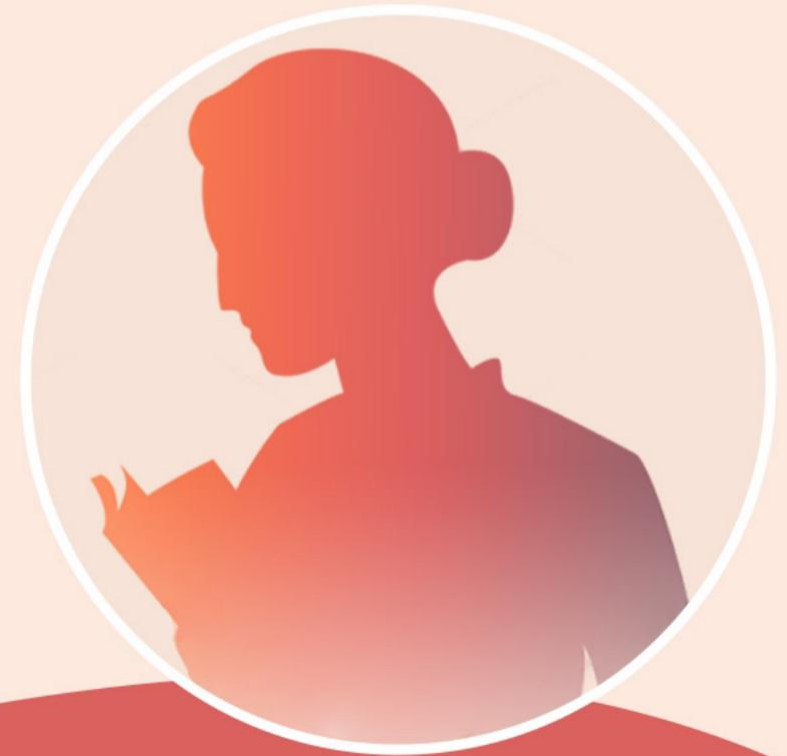


Teammates:

Mohd Anas Hussain
Sukant C



Date:2025-09-20



C O N T E N T S

01

PART

Introduction: The Research Revolution

02

PART

Solution: AI Research Agent

03

PART

Key Features and Capabilities

04

PART

Overcoming Challenges

05

PART

Measurable Outcomes and Benefits

06

PART

Revenue Model

07

PART

Future Development and Scalability

08

PART

Conclusion

PART 01

Introduction: The Research Revolution



Problem: The Modern Research Challenge

Manual Research Bottlenecks

Research processes are currently manual, which leads to slower insights and fragmented information gathering.

Information Overload

Information is scattered across numerous sources, making verification a tedious and reliability a major concern.

Outdated Conclusions

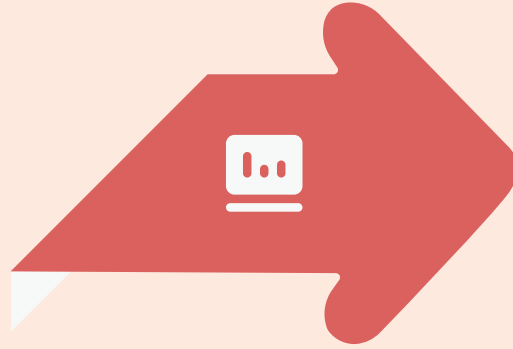
Results often become outdated quickly, impacting decision-making processes and strategic planning.

The Motivation Behind Automation



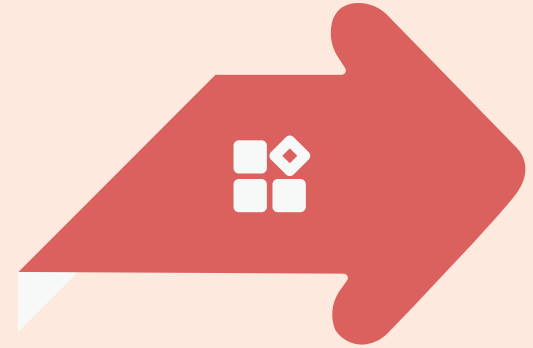
Accurate Decisions

Effective research is essential for informed decisions, strategic development, and driving innovation across various sectors.



LLM Limitations

LLMs are robust, but their reliance on static knowledge and susceptibility to hallucinations pose significant challenges.



Need for Real-Time Insight

A solution is needed that provides fast, dependable, and current insights for real-time research needs.

PART 02

Solution: AI Research Agent



Overview of the AI Agent



Technology Stack

The AI Research Agent is built using React, FastAPI, and Gemini, to deliver a robust research solution.



Modular Pipeline

The process is broken into distinct modules, including Planner, Searcher, Reader, Verifier, Reflector, and Briefer.



Real-Time Data

Pathway-powered Real-Time RAG ensures the ingestion of fresh data and continuous updates.

Building Blocks of the System

Frontend Tech Stack

The frontend is built with React, Vite, and Tailwind CSS to provide a user-friendly interface.

AI Model Utilization

The system integrates Gemini API for planning, synthesis, and verification tasks.

Backend Architecture

FastAPI is used as the backend framework to efficiently manage the asynchronous pipeline.

Live Data Integration

Pathway integration enables live data ingestion and vector index updating for enhanced retrieval.

API Endpoint Summary

The API offers `/research`` for long tasks and `/api/endpoint`` for single queries.



React, Vite, Tailwind: Streamlined Frontend Solution



User-Friendly Interface

React + Vite + Tailwind offers an intuitive interface for query input and progress tracking.

FastAPI Backend Orchestration



Backend: FastAPI. FastAPI efficiently orchestrates the asynchronous pipeline to ensure seamless operation and quick response times.



Enhanced Planning with Gemini API Integration

01

Advanced Planning Capabilities

Leveraging Gemini API for enhanced planning processes.

02

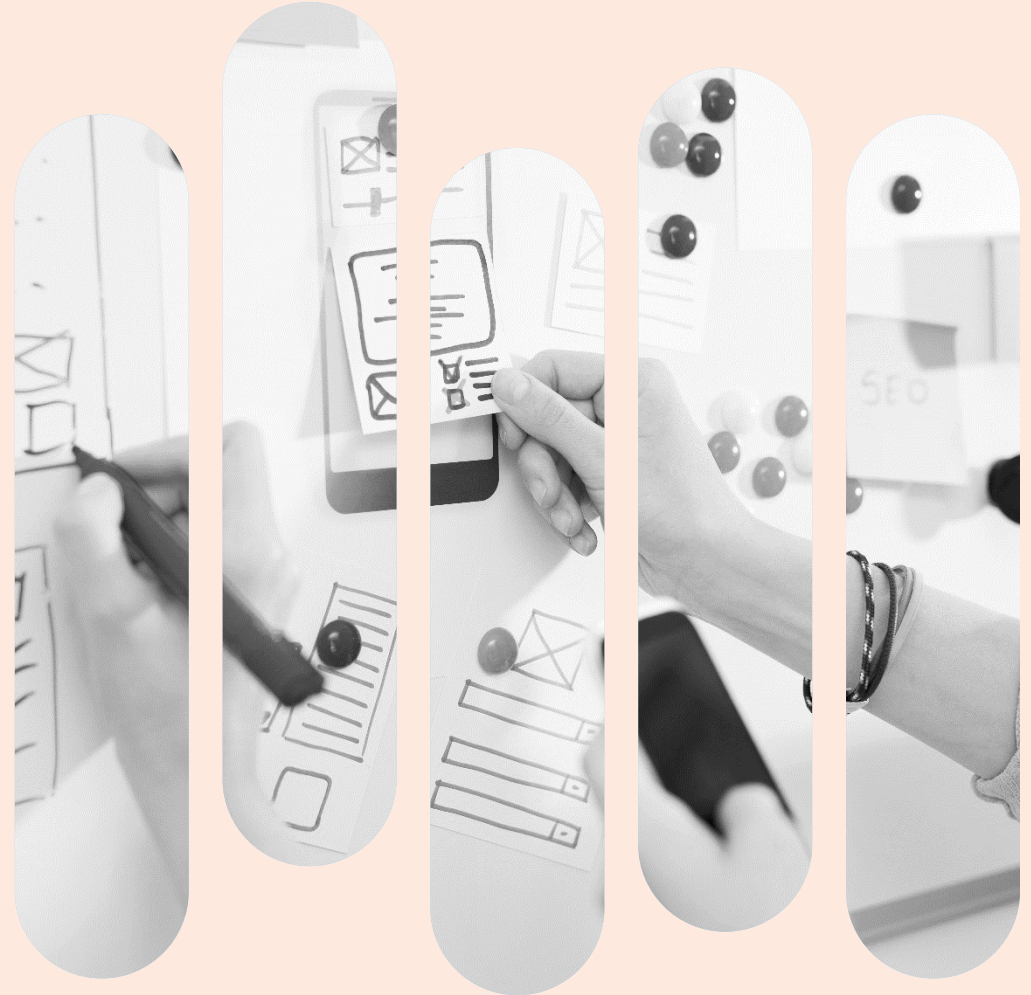
Precise Synthesis

Gemini API enables precise synthesis of information.

03

Rigorous Verification

Ensuring thorough verification throughout the process with Gemini API.



Pathway Integration for Live Data

Step1

Live Data Ingestion

Pathway enables the ingestion of live data streams.

Step2

Vector Index Updates

It updates vector indexes to improve data retrieval.

Research vs. API Endpoints: Key Differences



/research Endpoint

Task Complexity: High

Handles comprehensive and complex tasks that require significant computational resources.

Execution Time: Long

Intended for long-running operations, potentially spanning minutes or hours.

Update Frequency: Ongoing

Provides ongoing updates and status reports throughout the task's execution.



/api/endpoint Endpoint



Task Complexity: Low

Suited for simple, single queries that demand minimal processing.

Execution Time: Short

Designed for quick responses, typically within seconds.

Update Frequency: Single

Delivers a single, final response upon completion of the query.

PART 03

Key Features and Capabilities



Live Research Pipeline

Stepwise Updates

Provide live research pipeline with stepwise updates to facilitate instant feedback and monitoring.

Transparency

Enables users to observe the research process and enhance trust in the AI agent's insights.



Source Summarization & Verification

Automated Summary

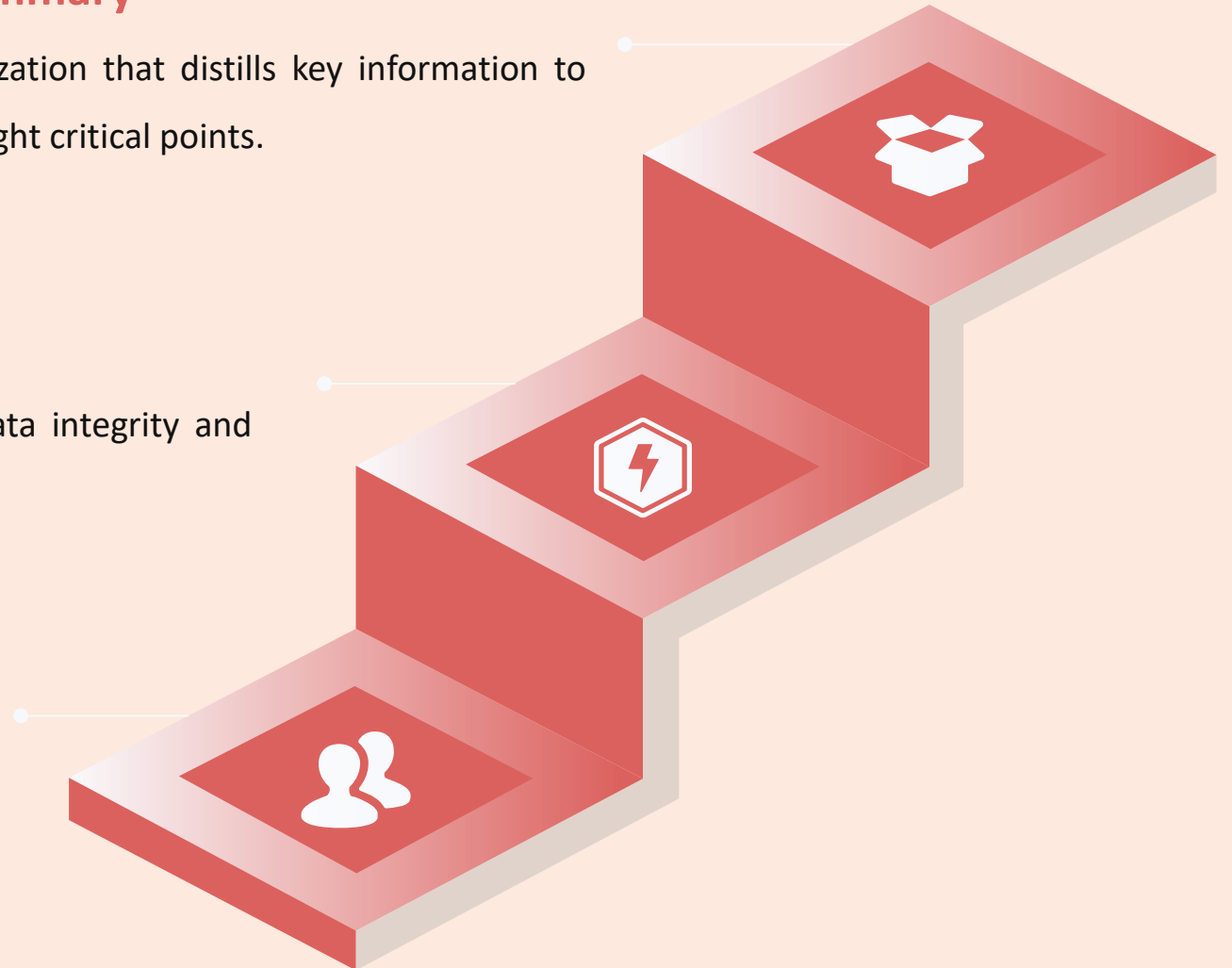
Automatic summarization that distills key information to save time and highlight critical points.

Verification of Data Sources

Verification of data sources to ensure data integrity and reliability.

Reduce Hallucinations

Reduces instances of misinformation and enhances the trustworthiness of results.



Flexibility and Adaptability

Flexible Solutions

Operates both with and without a Gemini API key, offering a fallback mode.

Extensible Design

Highly extensible, designed to meet needs of teams and larger enterprises by accommodating custom features.

Demo Ready

Demo-friendly to showcasing features and capabilities for demonstrations and presentations.

PART 04

Overcoming Challenges



Asynchronous Orchestration



Task Queues

Solved asynchronicity by implementing effective task queues coupled with result polling, minimizing delays.

Handling Noisy Webpages

Readability

Utilized readability-lxml and BeautifulSoup, to extract clean and relevant content for improved analysis of the content.



Addressing CORS & Development Friction

Permissive Middleware

Implemented permissive middleware to mitigate cross-origin issues, ensuring smoother development and compatibility.



Hallucination Mitigation

Verification Stages

Verification and reflection phases integrated to proactively address and diminish inaccuracies and biases.



PART 05

Measurable Outcomes and Benefits



Structured Research Briefs

Single Query Advantage

Transforms simple queries into structured, comprehensive research briefs, covering multiple aspects of the topic.



Efficiency and Transparency

Time Savings

Substantial time and effort savings through automation, allowing researchers to focus on strategic tasks.

Transparent Processes

Transparency in research process provides visibility, enhancing trust and credibility in outcomes and insights.



Wide Application

Cross-Disciplinary Use

Extensible for academia, supporting in-depth studies, business optimizing strategic decisions, and more.



PART 06

Revenue Model



Freemium Subscription



01 Accessible Entry

Free tier provides limited access, allowing initial use and exploration of core features.



02 Incentive to Upgrade

Encourages users to upgrade for more comprehensive and advanced functionality.



Premium Subscriptions

01

Enhanced Access

Unlimited queries, advanced features, and integrations for premium users to maximize research capabilities.

02

Institutional Packages

Customized solutions cater to the specific needs and demands of different enterprises and users.



Future Expansion

01

Enterprise Licensing

Enterprise licensing enables large organizations to access research capabilities.

02

API Monetization

API monetization allows integrating research agent capabilities, broadening market penetration.

PART 07

Future Development and Scalability





Deep RAG Implementation

Enhanced Search

Richer vector search techniques and integrations with academic databases for comprehensive data retrieval.

Comprehensive Insight

Deeper insights and more accurate information synthesis from diverse sources.

Automation of Citations



Auto Generation

Automated citation and bibliography generation to maintain research integrity.

Streamlined Process

Improving efficiency across research workflows by reducing manual tasks and errors.



Team Collaboration Features

Collaborative Environment

Team collaboration features enable shared projects, facilitating cooperation among researchers and across teams.

Enhanced Productivity

Improving overall team productivity with collaborative efforts and insights capabilities.



Enterprise Scalability

01

Enterprise-Ready Solutions

Designed for enterprise environments with job persistence, retries, and observability for robustness.

02

Increased Reliability

Assured reliability, scalability, and manageability with comprehensive support and maintenance features.

PART 08

Conclusion





Transformation Overview

01

From Manual to Automated

Transitioning from manual to automated research processes for increased speed and accuracy.

02

Orchestration

Orchestration fragmented information into cohesive strategies, to enhance clarity and understanding.

Technology Integration



Core Components

Built using a robust combination of key technologies: Gemini, Pathway, FastAPI, and React.



Integrated Platform

Offers fully integrated platform for scalable research management and strategic initiatives.

Future Implications

Next-Gen Assistants

A blueprint for next-generation research assistants to redefine future of information discovery.

Product Scalability

Fully demo-ready, product-scalable, to demonstrate its potential across diverse and challenging research environments.



Thank you

● Date:2025-09-20

