1. **Tên Đề Tài**  
   \_ Thiết kế và triển khai AI platform multi tools
2. **Mục tiêu**  
   \_ Xây dựng một nền tảng AI cho phép tích hợp các công cụ AI và non-AI khác nhau. Nền tảng nhắm đến mục tiêu là cung cấp hỗ trợ toàn diện cho nhiều loại công cụ AI và non-AI khác nhau, bao gồm nhưng không giới hạn ở các tác vụ như: Xử lí ngôn ngữ tự nhiên cho tuỳ các tác vụ khác nhau, ocr, recommend, tìm kiếm thông tin và tổng hợp,…
3. **Mục đích**  
   \_ Với nền tảng xây dựng trên, em nhắm đến các mục đích sau
   * **Một hệ sinh thái các công cụ AI hợp nhất:** Tạo dựng một môi trường mà các công cụ AI có thể kết hợp hoặc tương tác với nhau liền mạch. Cung cấp các giải pháp linh hoạt cho các vấn đề đa dạng.
   * **Khả năng mở rộng và tính linh hoạt**: Đảm bảo nền tảng có thể mở rộng hiệu quả để xử lí lượng lớn dữ liệu, cung cấp phương án giải quyết các yêu cầu bài toán khác nhau, tích hợp dễ dàng các công cụ và công nghệ mới trong tương và loại bỏ các công cụ và công nghệ lỗi thời.
   * **Xử lý dữ liệu tự động:** Phát triển các đường ống dữ liệu tự động để thu thập, xử lí và lưu trữ dữ liệu, hạn chế các thao tác thủ công, tăng độ tin cậy và hiệu suất chung của hệ thống.
   * **Xử lý dữ liệu nâng cao:** Tận dụng các khả năng NLP tiên tiến để chuyển đổi các dạng dữ liệu lộn xộn thành thông tin chi tiết có giá trị cao, nhằm cải thiện các quyết định dữ liệu tốt hơn và chính xác hơn.
   * **Lấy người dùng làm trung tâm:** Tập trung vào việc tạo giao diện người dùng và trải nghiệm trực quan cho phép người dùng dễ dàng tương tác với nền tảng, bất kể trình độ chuyên môn kỹ thuật của họ.
4. **I/O:**  
   \_ Input cho hệ thống hoặc một tiến trình sẽ là một dạng dữ liệu, tín hiệu hoặc một chỉ đạo từ người dùng được truyền vào để đạt được đầu ra mong muốn. Các đầu vào dữ liệu tuỳ vào các bài toán khác nhau mà sẽ có các kiểu loại khác nhau.   
     
   * **Bài toán truy vấn:**  
     \_ Đầu vào: Chỉ đạo của người dùng (Prompt), file, chat bot, kết hợp với dữ liệu được truy vấn từ các cơ sở dữ liệu dựa trên chỉ đạo của người dùng.  
     \_ Đầu ra: Một text report trả ra kết quả theo yêu cầu của người dùng hoặc một API, SDK, chứa các thông tin liên quan.  
     \_ Ví dụ: Một truy vấn tới LLM như "Doanh thu từ dịch vụ di động trong quý 2 năm 2023 là bao nhiêu?" với dữ liệu được lấy từ cơ sở dữ liệu chứa hồ sơ tài chính.
   * **Bài toán thu thập thông tin:**  
     \_ Đầu vào: Chỉ đạo của người dùng (Prompt) kết hợp với dữ liệu được cào nguồn dữ liệu mục tiêu dựa trên chỉ đạo của người dùng.  
     \_ Đầu ra: Một text report trả ra kết quả thông tin hoặc một API chứa các thông tin liên quan đến thông tin cần tìm.  
     \_ Ví dụ: “Cho tôi biết lịch bay của hãng Vietnam Airline trong ngày 10/8/2024”
   * **Bài toán giám sát log và api**  
     \_ Đầu vào: Log được tạo bởi các ứng dụng và yêu cầu/phản hồi của API, file, db, message queue.  
     \_ Đầu ra: Phân tích văn bản hoặc tóm tắt dữ liệu log hoặc API end point cung cấp thông tin đã xử lý.  
     \_ Ví dụ: Log hệ thống từ máy chủ cho biết hoạt động hoặc lỗi của người dùng và các lệnh gọi API theo dõi tương tác giữa các microservice.
   * **Bài toán xử lí File**  
     \_ Đầu vào: Nhiều loại tệp khác nhau như PDF, tài liệu Word, tệp CSV, v.v.  
     \_ Đầu ra: Văn bản được trích xuất và xử lý từ các tệp hoặc SDK cung cấp khả năng xử lý tệp.  
     \_ Ví dụ: Một tài liệu PDF cần được dịch từ Tiếng Anh sang Tiếng Việt.
5. **Keywords:**
   * **Công nghệ Web-Server:**  
     \_ Front-end: ReactJS  
     \_ Back-end: SpringBoot
   * **Công nghệ lưu trữ dữ liệu:**  
     \_ RDBMS: MySQL, PostgreSQL, MSSSQL, Oracle  
     \_ NoSQL:

+ Key-value: redis

+ Document: Elasticserach

+ Graph: Neo4j, Redis-graph

+ Wide column: Hbase, cassandra

* + **Công nghệ Big Data:**  
    \_ Intergration: Apache Nifi, Apache Kafka, Logstash  
    \_ Computing: Apache Spark, Nvidia
  + **Công nghệ Machine Learning/Deep Learning:**  
    \_ LLM model: LLAMA3, GPT, Gemini  
    \_ LLM technique: Retrieval-Augmented Generation (RAG)
  + **Công nghệ ảo hoá và điều phối tài nguyên:**  
    \_ Container: Docker  
    \_ Orchestration: Kubernetes
  + **Công nghệ CI/CD**  
    \_ Jenkins

1. **Kế hoạch:**  
   Để nhằm đảm tiến độ, em xin được lên kế hoạch triển khai dự án như sau:  
     
   \_ Giai đoạn 1: Thu thập yêu cầu và thiết kế hệ thống (2 tuần)
   * Xác định yêu cầu cụ thể bài toán từ người hướng dẫn
   * Mô tả mục đích sử dụng, các tính năng, các yêu cầu chi tiết
   * Thiết kế kiến trúc tổng thể của hệ thống
   * Lên kế hoạch triển khai chi tiết cho từng giai đoạn

\_ Giai đoạn 2: Xây dựng và Triển khai Cơ sở Hạ tầng (2 tuần)

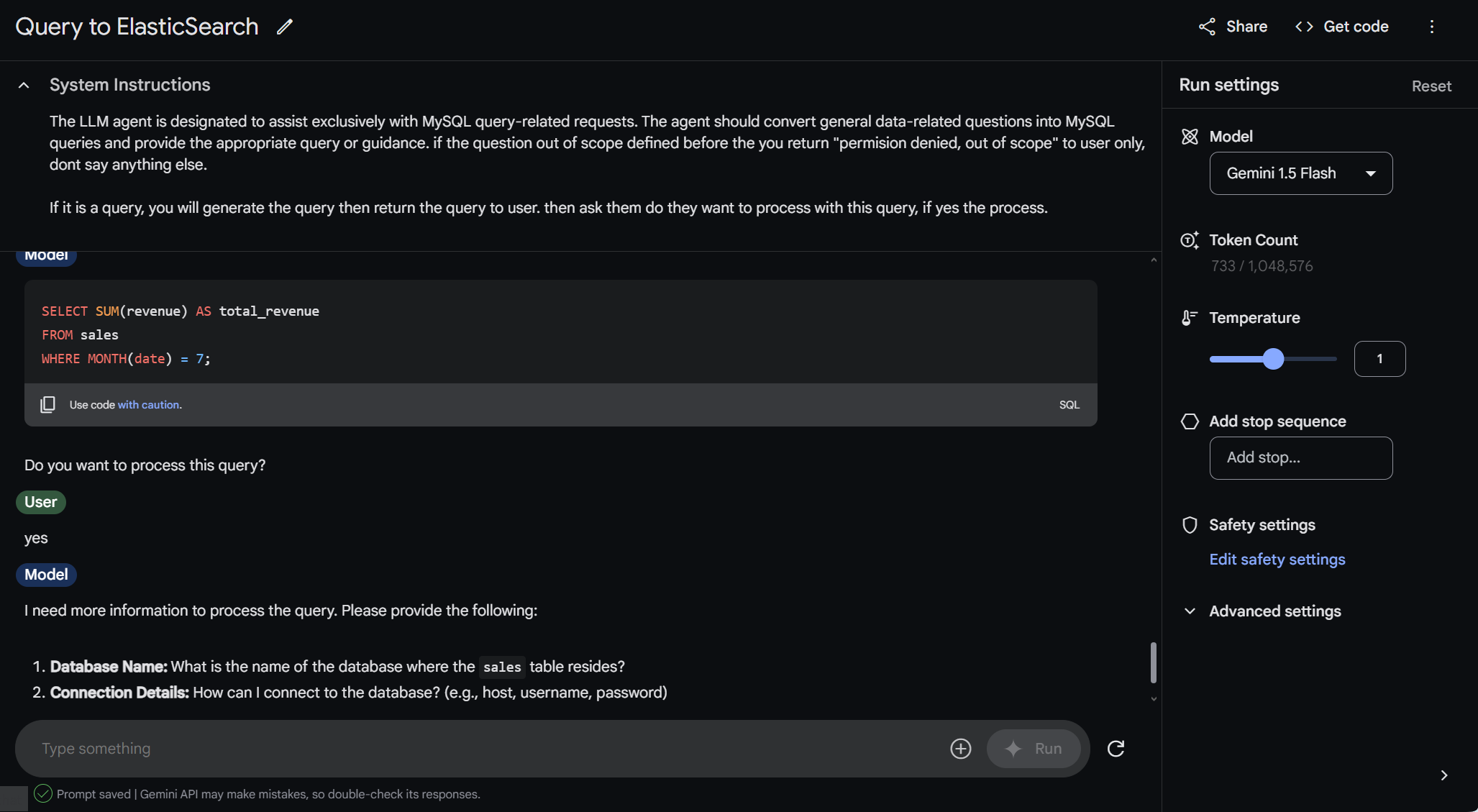
* Thiết lập và cấu hình các công cụ đã được nêu
* Đảm bảo các công cụ hoạt động ổn định, tương thích và tương tác được với nhau

\_ Giai đoạn 3: Phát triển và Tích hợp (5 tuần)

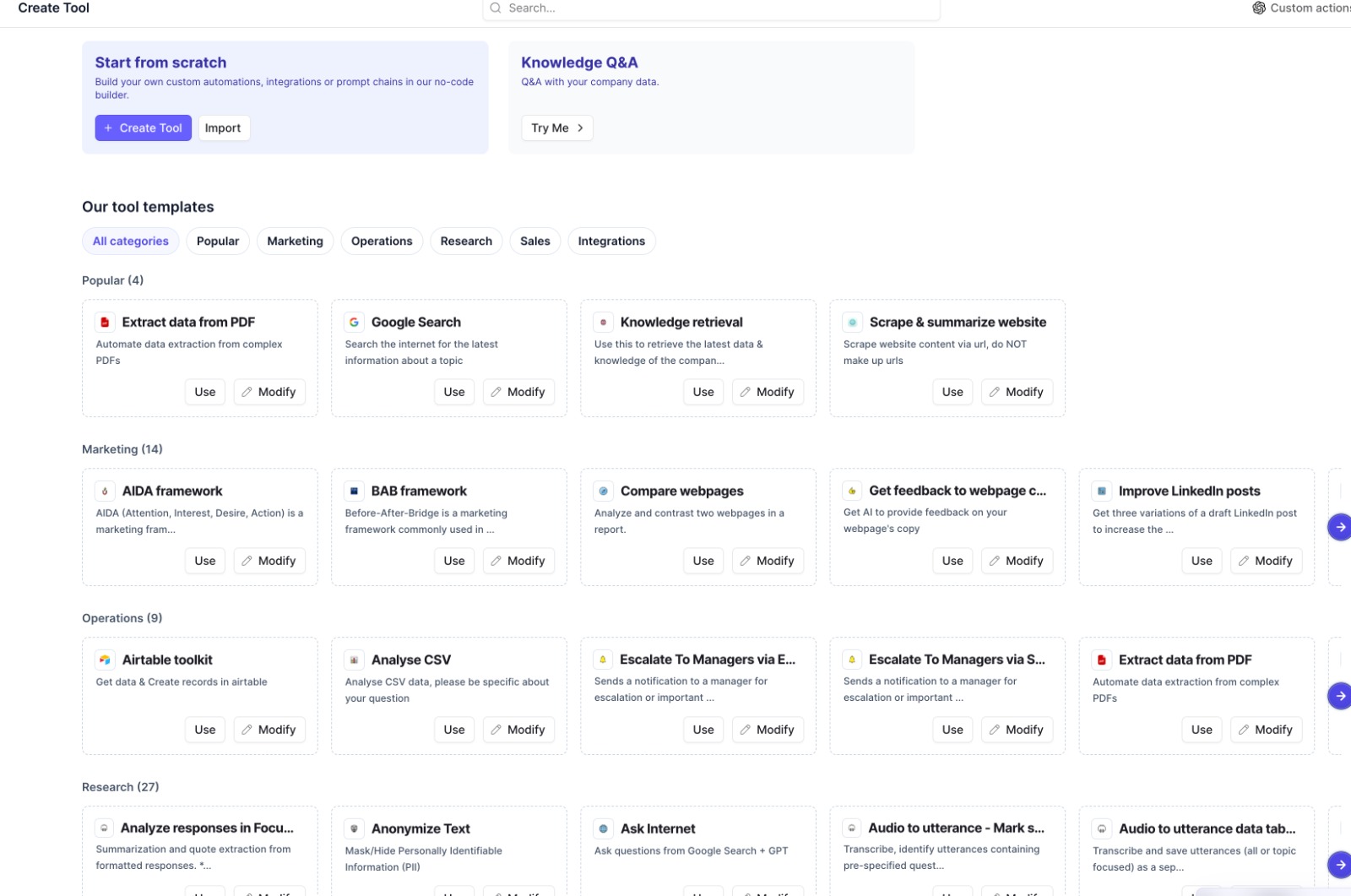
* Phát triển module Back-end và Front-end (1 tuần)
* Phát triển các module được liệt kê như sau: (4 tuần)  
  + Bài toán truy vấn  
  + Bài toán thu thập thông tin  
  + Bài toán giám sát log & Api  
  + Bài toán xử lí file

\_ Giai đoạn 4: Triển khai và Kiểm thử (1 tuần)

* Triển khai hệ thống trên môi trường thực tế
* Thực hiện các test case để đảm bảo hệ thống hoạt động chính xác

1. **Nguồn kham thảo:**  
   **Google AI Studio**  
   \_ Ở tài liệu này, em kham thảo về mô hình tạo các prompt để người dùng tương tác với hệ thống kèm theo sự hỗ trợ của các model LLM. Mo hình này giúp em hiểu rõ hơn về việc thiết kế giao diện công cụ thân thiện với người dùng cũng như là luồng xử lí của công cụ.  
   

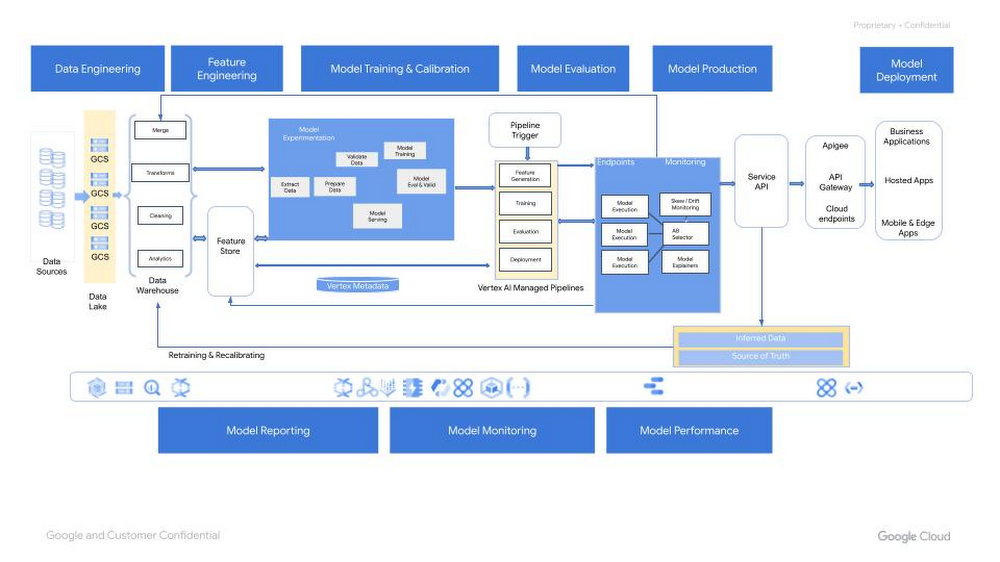
**Hình 1: Giao diện công cụ**

**Relevance AI:**  
\_ Ở tài liệu này, em kham thảo về mô hình multi-tool để người dùng tương tác với hệ thống hoặc thiết kế các công cụ phù hợp với bài toán của họ. Mo hình này giúp em hiểu rõ hơn về việc thiết kế giao diện platform thân thiện với người dùng cũng như là luồng xử lí và tạo mới công cụ.  


**Hình 2: Giao diện tổng quan công cụ**

**Google Vertex AI**

\_ Ở tài liệu này, em kham thảo về mô hình AI platform cho phép huấn luyện và triển khai mô hình ML, ứng dụng AI và fine-tuning LLM. Mô hình này giúp em hiểu về các thống nhất các thành phần như: Data Engineering, Data Science và Machine Learning engineering để tạo thành một hệ sinh thái các công cụ AI hợp nhất.



**Hình 3: Kiến trúc tổng quan của Vertex AI**