Proposed Flow: AI Researcher

⇒) Objective : When given a complex research query by the user, Perform Web search and RAG and provide a well researched Response

2) Guidance : • Use CrewAI + Qdrant + Any LLM

⇒) Scope & Delimiters :
* Should Answer Questions about:
   → Technology / Innovations
   ⇒ Science / History / Philosophy
   ⇒) Mathematics / Engineering
   2) Business / Economics
   2) Current Affairs / Politics
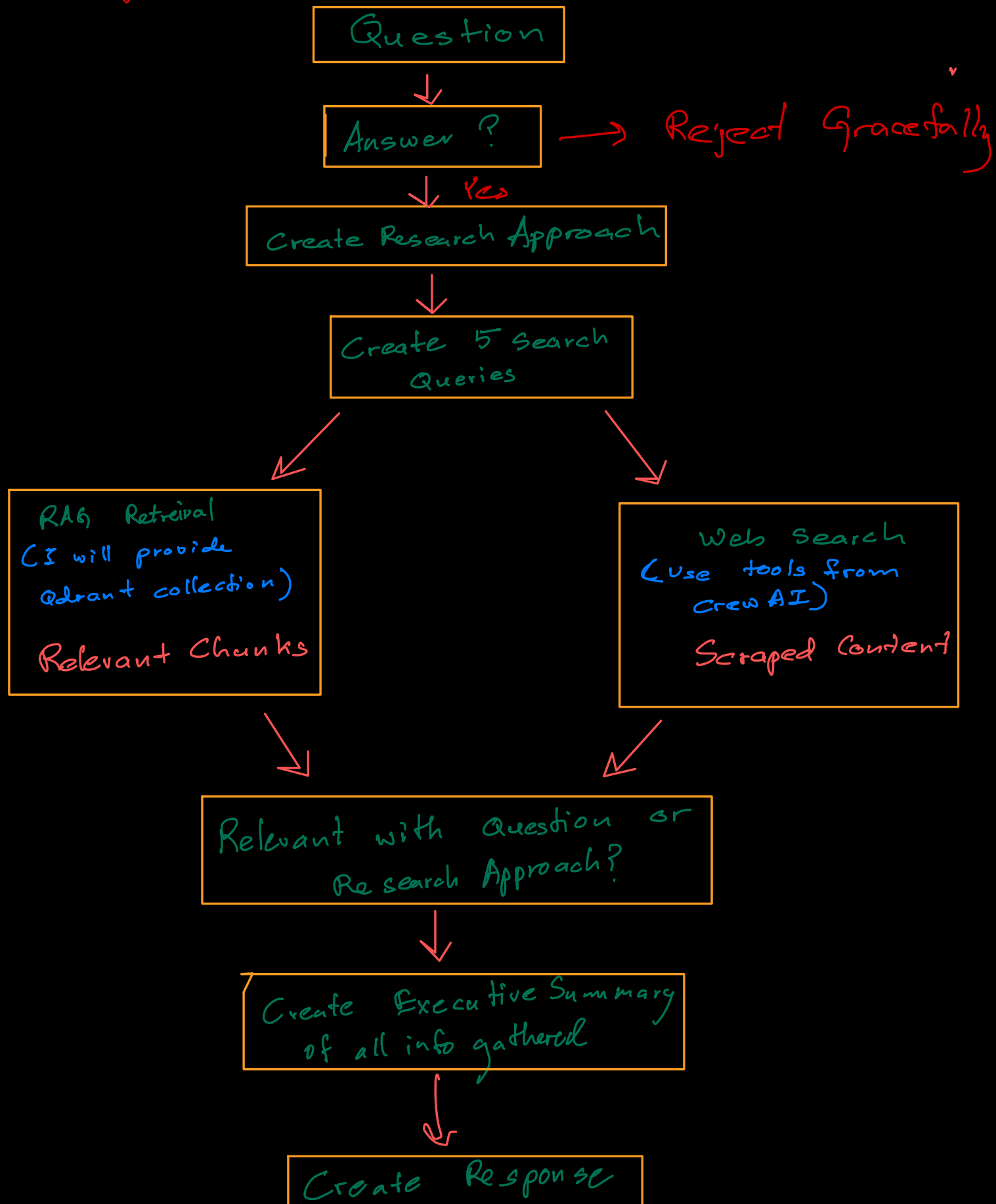
* Should Reject Questions About
   ⇒) Sports   → Non Questions (statements)
   ⇒) Harmful Content
   ⇒) Entertainment

Flow diagram (Proposal Only) (Tasks)

```
┌─────────────────┐
│    Question     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│    Answer ?     │  ──────→   Reject Gracefully
└─────────────────┘
         │ Yes
         ▼
┌──────────────────────────┐
│  Create Research Approach │
└──────────────────────────┘
         │
         ▼
┌──────────────────────────┐
│     Create 5 Search      │
│        Queries           │
└──────────────────────────┘
       │              │
       ▼              ▼
```

┌────────────────────────┐   ┌────────────────────────┐
│ RAG Retreival   │   │   Web Search   │
│ (I will provide   │   │ (use tools from   │
│ Qdrant collection)   │   │   crew AI)   │
│   │   │   │
│ Relevant Chunks   │   │ Scraped Content   │
└────────────────────────┘   └────────────────────────┘
            │   │
            ▼   ▼

┌────────────────────────────────────┐
│ Relevant with Question or   │
│ Research Approach?   │
└────────────────────────────────────┘
            │
            ▼
┌────────────────────────────────────┐
│ Create Executive Summary   │
│ of all info gathered   │
└────────────────────────────────────┘
            │
            ▼
┌────────────────────────────────────┐
│   Create Response   │
└────────────────────────────────────┘

* This is just a proposal
* Be mindful of dates ⚠

Tips :

* Optimize for latency:
    - => It takes 10 mins = Bad
    - => It takes 5 mins = Bad
    - => It takes < 3 mins = Perfect!
    - => Answer is Bullshit = Bad

* Use built-in tools in Crew AI

* This is Sequential! 🙆

* Given task Breakdown is just an example Optimize and Customize

* Don't use Reasoning LLMs

* LLM Suggestions:
    - => gpt-4o-mini
    - => mistral-small-latest
    - => ministral-8b
    - => GPT-4o / Claude 3.7

Must <u>Be</u> FastAPI Service!

- Must use uv (NO pip)
- Use a sensible folder structure.
- Use Pydantic V2 for endpoint Validation

- Use:
  - * type hints
  - * doc-strings/comments
  - * functional programming whenever possible
  - * RORO [Google it!]
  - ⟹ * No API keys on VC ⚠️

- Create a public git repo (must!)
- Use .env for all API keys and other secrets
- Use clear commit messages and sensible commits
  - ↳ don't dump code in single commits