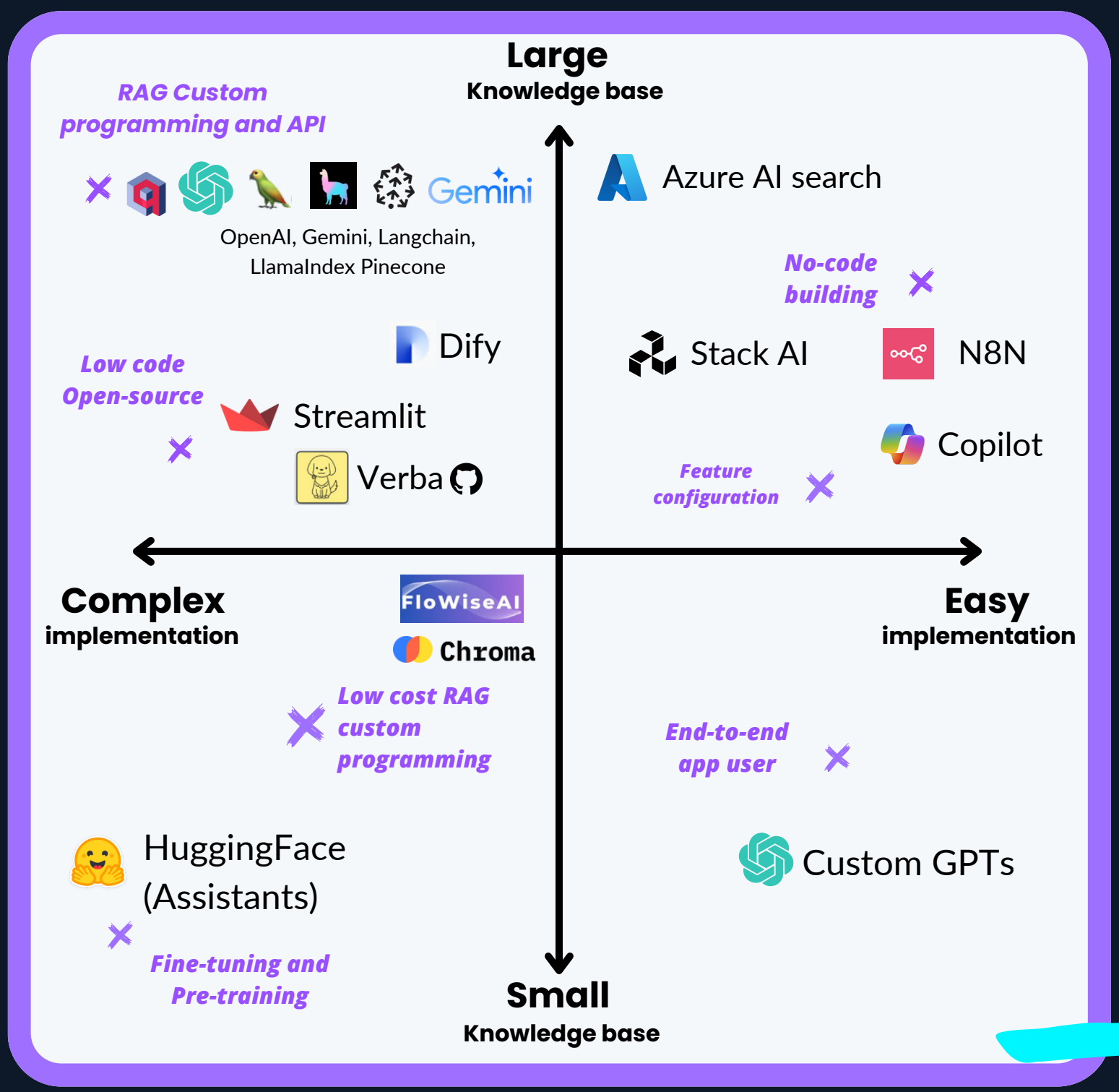
Labsbit.ai

Joanna Stoffregen
@JoannaStoffregen

Get started with RAG

BEGINNER TO ADVANCED - COMPLETE GUIDE



Labsbit.ai

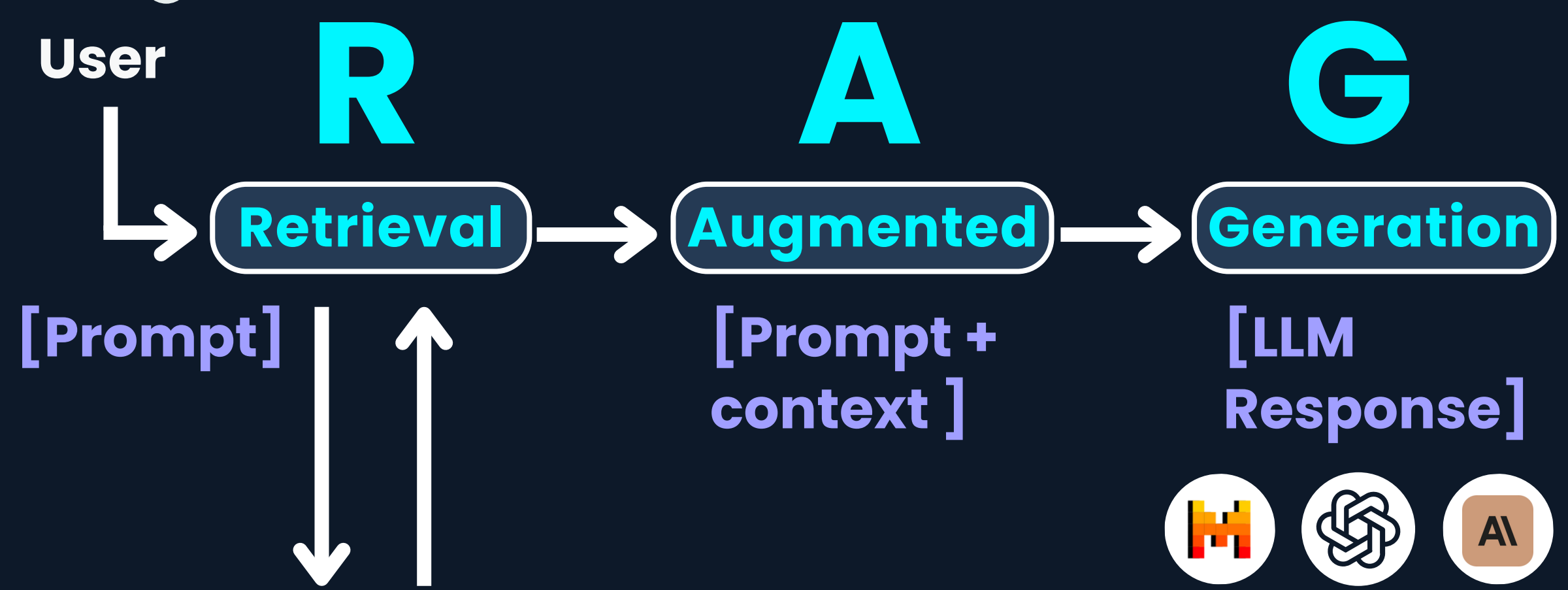
This guide will help you choose the best RAG solution !

Joanna Stoffregen
@JoannaStoffregen

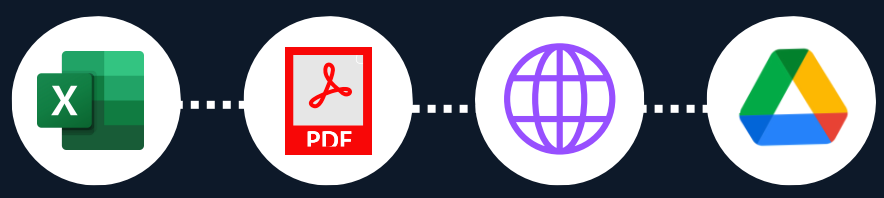
Labsbit.ai

Quick recap : What is RAG ?

RAG consists of augmenting an LLM (ie: ChatGPT) with your own data



[Context]



Websites, docs, databases... (proprietary and non proprietary)

Joanna Stoffregen
@JoannaStoffregen

Labsbit.ai

Today you will discover

Difficulty

- RAG for beginners : User mindset
- RAG for semi-pros : no code
- RAG for pros : low code
- RAG for developers : code
- RAG for experts : production-ready

Joanna Stoffregen
@JoannaStoffregen

Labsbit.ai

Very Easy

RAG for beginners: User mindset - plug & play

Tools

Custom GPT, AskYourPDF, OpenAI ADA, Dante AI

Upload data

1

Tool

Start chatting

2

Pros	Easy, no-code, cost effective
Cons	Limited size of docs and formats, no personnalized UI, scaling limitations
WHO	Individual use and occasional needs
Price	$20/month

Joanna Stoffregen
@JoannaStoffregen

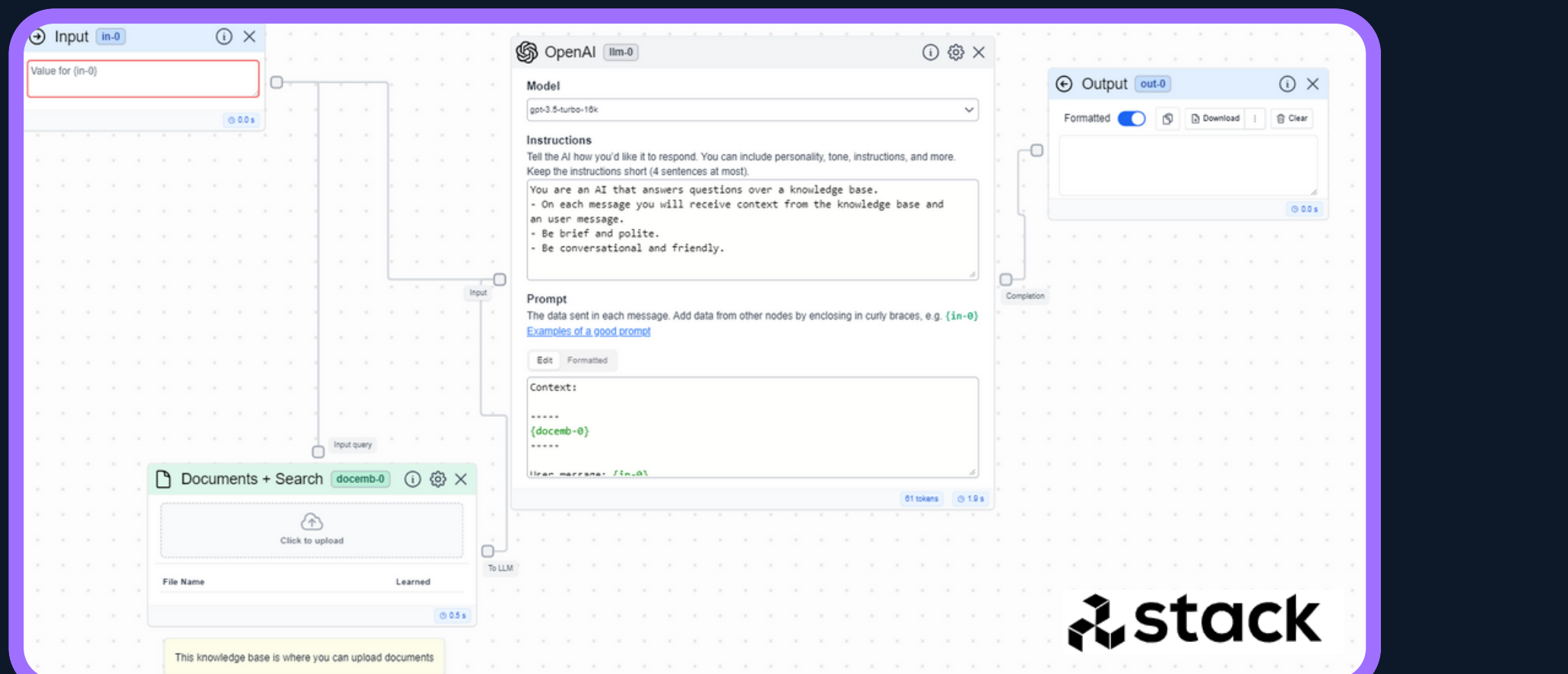
Labsbit.ai

Easy

RAG for semi-pros: No-code closed source

Tools

Stack AI, AirOps, N8N



Pros Easy, no-code, end-to-end solution (chat interface, embeddings, vector database, deployment)

Cons High costs for individuals and high costs for scaling, UI customization limitations

WHO Ideal for startups, SMEs, solopreneurs who want an end-to-end solution with a **medium budget**

Price Pro plan $200/m ; Team plan 900$/month

Joanna Stoffregen
@JoannaStoffregen

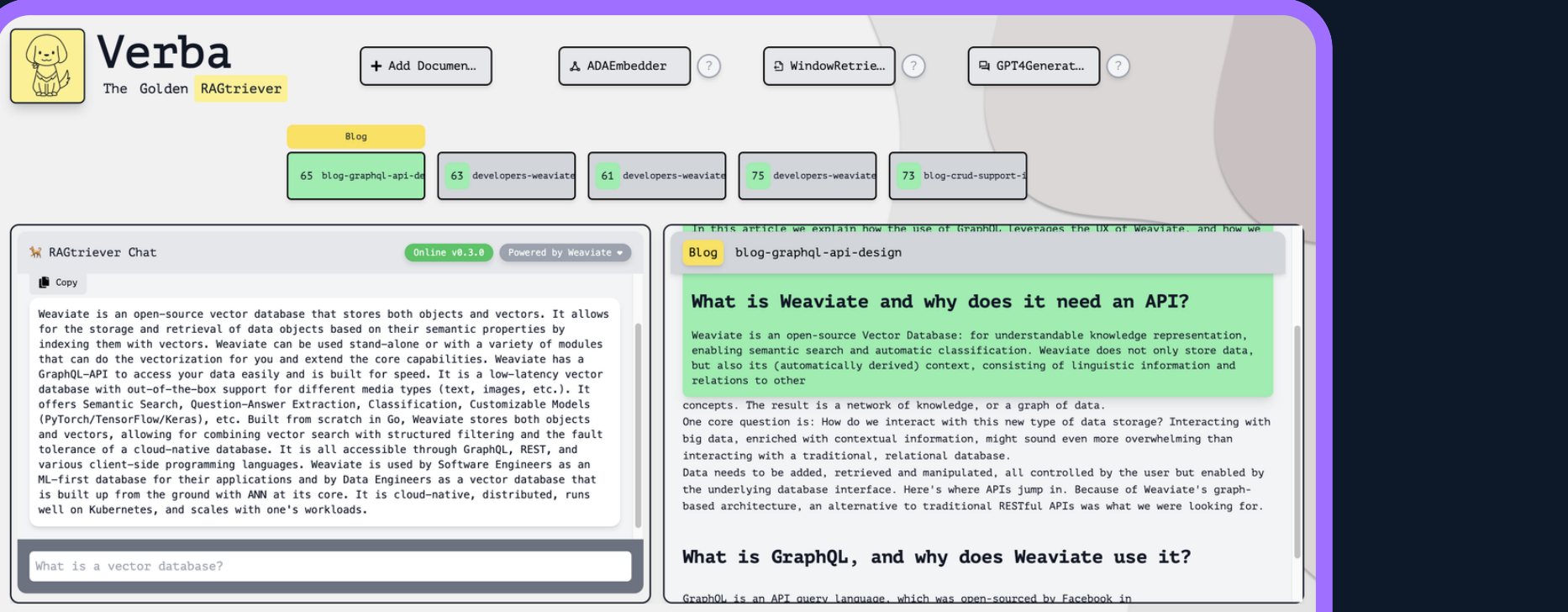
Labsbit.ai

Medium

RAG for pros: Low-code open source

Tools

Verba AI, Dify, Flowise



Pros Low code, free, perfect for an MVP, possibility to run locally (data privacy)

Cons UI customization and cloud deployment requires technical knowledge, not ready for production

WHO Ideal for startups, SMEs, solopreneurs who want an MVP solution.

Price Free

Joanna Stoffregen
@JoannaStoffregen

Labsbit.ai

medium hard

RAG for developers: Code - Do it yourself

Tools

OpenAI API, Gemini API, Ollama

OpenAI Assistant API

Knowledge retrieval function

+

Code interpreter

Pros Embedding and vector store included

Cons Low UI customization

WHO Startups, SMEs, solopreneurs

Price $0.20 /GB/assistant

Google Gemini API

Gemini

LangChain

Chroma

Streamlit

Pros Free, Google drive easy integration

Cons Coding skills required

WHO Startups, SMEs, solopreneurs

Price Free

Ollama + Mistral

MISTRAL AI_

LangChain

Chroma

Ollama

Pros Free, open source, full data privacy

Cons Coding skills required

WHO Businesses with high confidential data

Price Free

Joanna Stoffregen
@JoannaStoffregen

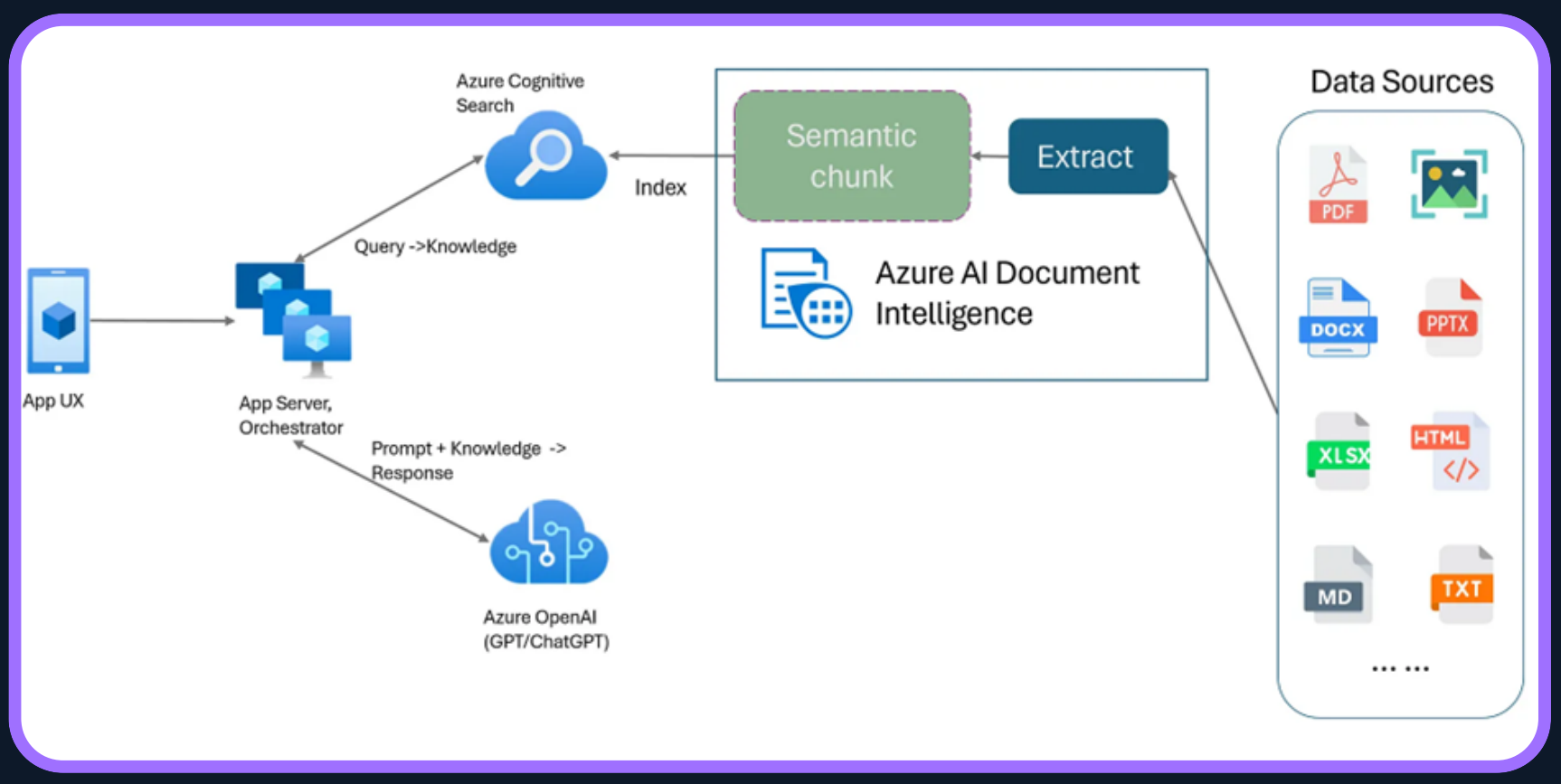
Labsbit.ai

Hard

RAG for experts: Azure end-to-end solution

Tools

Azure AI search



Pros End-to-end solution, low code, can be combined with OCR, variety of formats available

Cons Microsoft enterprise account required, high scaling costs for small businesses and individuals

WHO Ideal for companies who are already using Azure and Microsoft suite

Price $245/month/user (25 GB) $5600/month/user (2 TB)

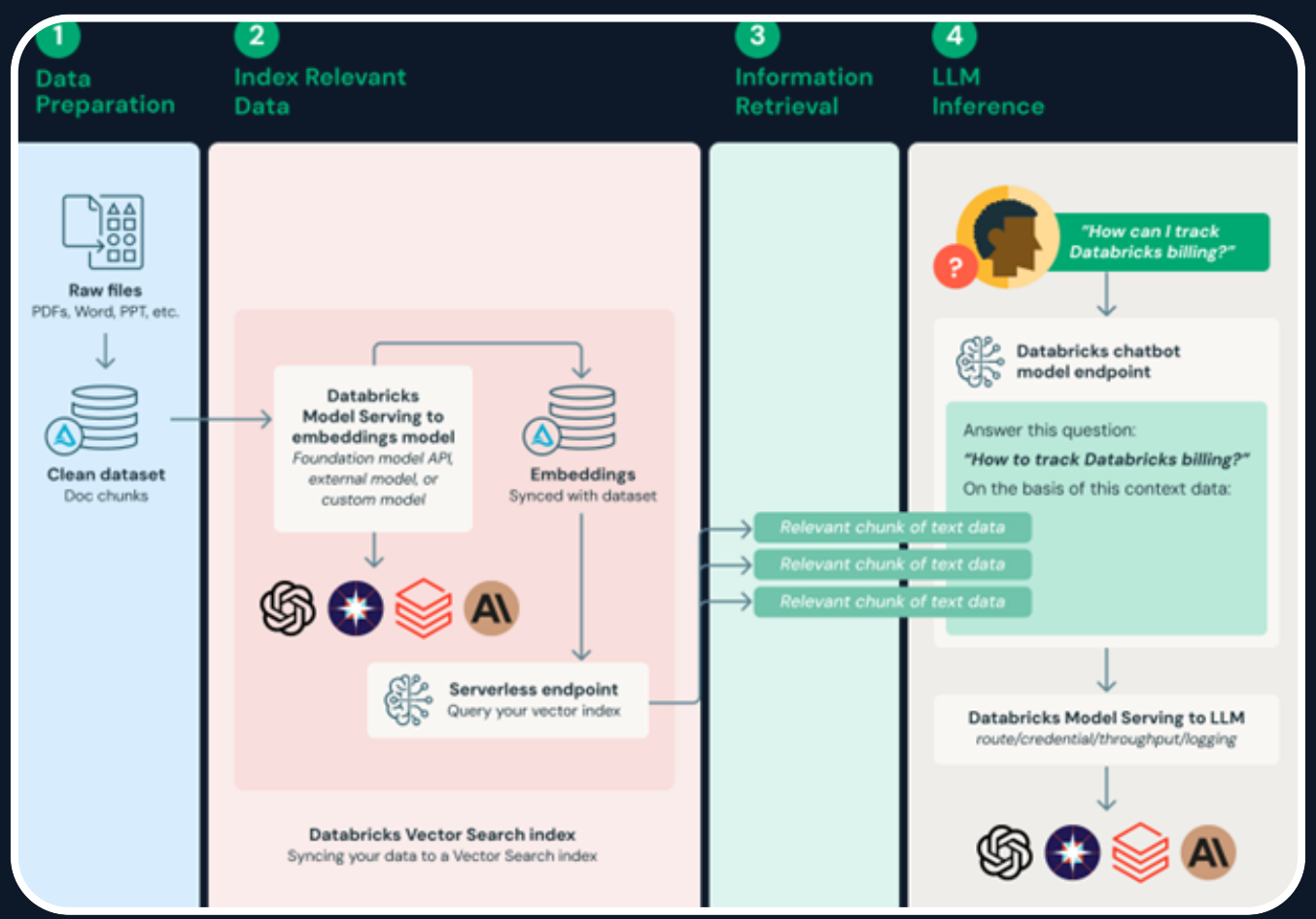
Joanna Stoffregen
@JoannaStoffregen

Labsbit.ai

Very hard

RAG for experts: Databricks – for production

Tools | Databricks



Pros Wide choice of cloud providers, End-to-end solution, serverless endpoint , variety of formats available

Cons Requires deep technical skills, high scaling costs and complexity for small and medium businesses

WHO Ideal for very large companies (Fortune 500)

Price Pay-as-you go (complex)

Joanna Stoffregen
@JoannaStoffregen

Labsbit.ai

Hey, I'm Joanna from Labsbit.ai

An AI/ML Product Development
& Automation Company

Have trouble with RAG implementation ?

Reach out to learn more about how we can help.

Joanna Stoffregen
@JoannaStoffregen