

Introduction to RAGs & AI Agents

A Beginner's Guide to Modern AI Systems



AGENDA

What We'll Cover Today

01

Understanding RAGs

Retrieval-Augmented Generation explained simply

02

Understanding AI Agents

What they are and what they can do

03

Real-world Applications

How these technologies are used today

Large Language Models (LLMs)

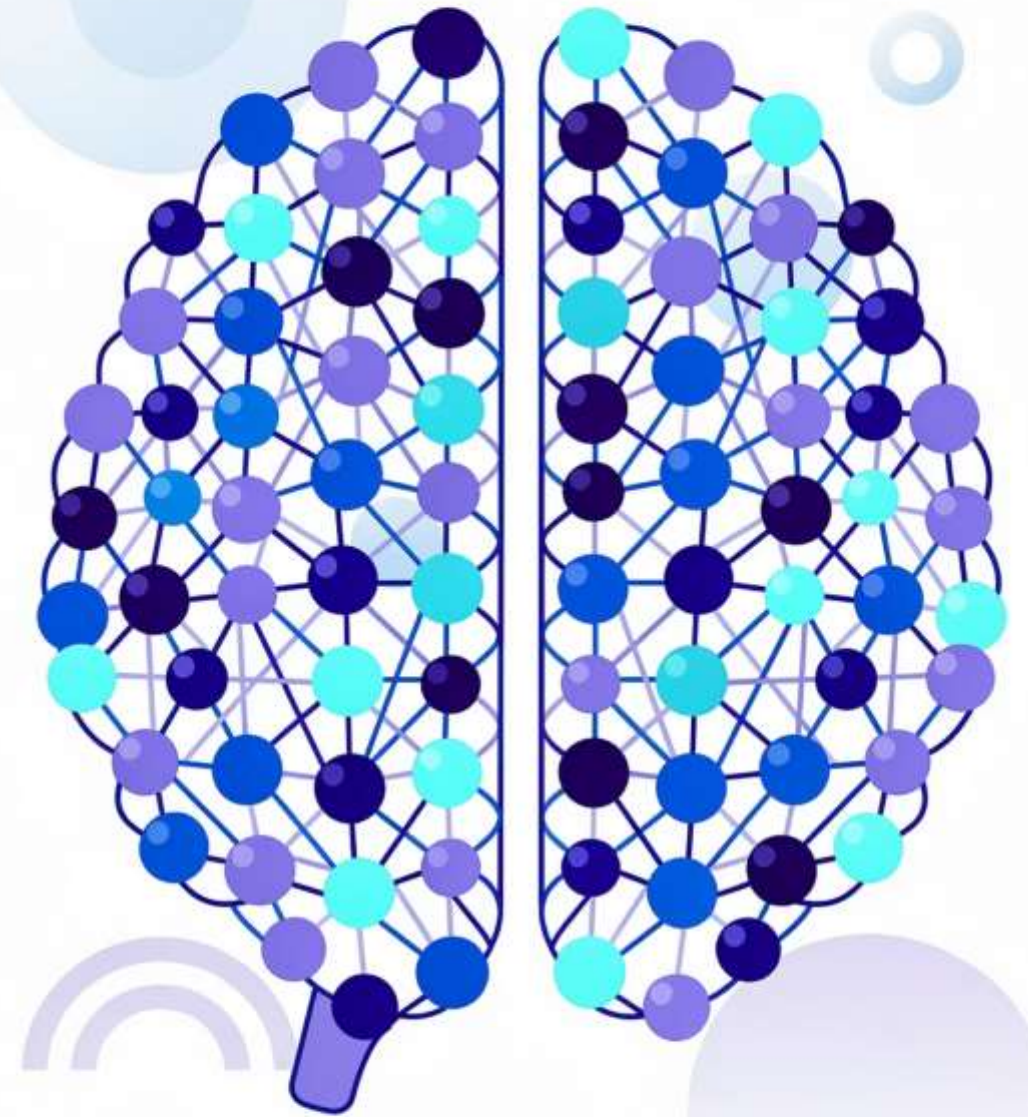
The foundation of modern AI assistants

What is an LLM?

An AI model trained on massive amounts of text to understand and generate human language

Think of it like this:

An LLM has read billions of pages and learned patterns in how language works. It can predict what words should come next, answer questions, write content, and more.



But LLMs Have Limitations



Knowledge Cutoff

Only knows information up to when it was trained



No Access to Private Data

Can't access your company documents or databases



No Memory

Doesn't remember previous conversations (unless designed to)



Can Make Mistakes

Sometimes generates plausible-sounding but incorrect information

PART 1

RAG

Retrieval-Augmented Generation

What is RAG?

Retrieval-Augmented Generation

A technique that enhances LLMs by giving them access to external information

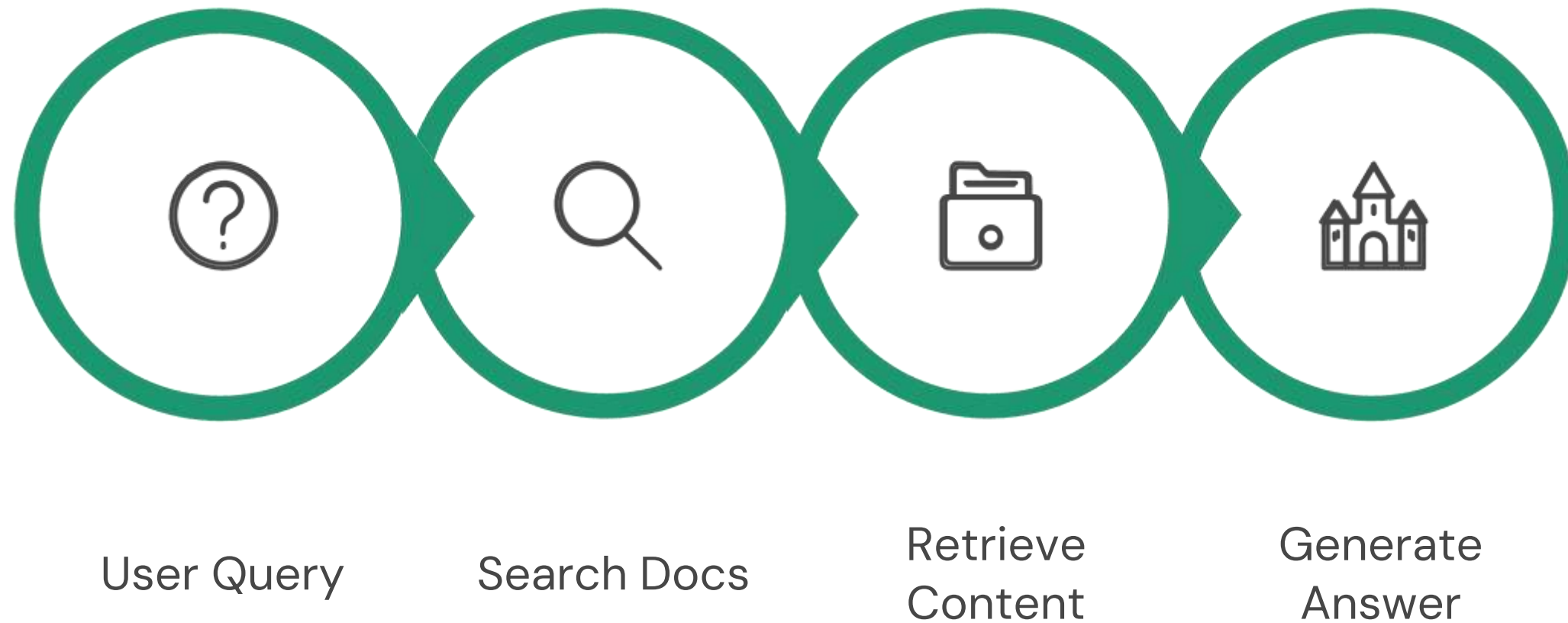
📄 Simple Analogy:

Imagine you're taking a test. Instead of answering from memory alone, you're allowed to search through your textbook first, find relevant information, and then use that to answer the question.

That's basically what RAG does for AI!



How RAG Works



The system acts as an intelligent intermediary, finding the right information and helping the AI provide accurate, contextual responses based on real data rather than just training memory.

Why Use RAG?



Up-to-Date Information

Access current data beyond the model's training cutoff



Domain-Specific Knowledge

Use your company's documents, policies, or specialized content



Improved Accuracy

Answers based on actual documents, not just model memory



Transparency

Can show sources and cite where information came from

RAG in Action: Example

Question: "What is our company's vacation policy?"

Without RAG

"I don't have access to your company's specific policies."

With RAG

"According to your employee handbook, full-time employees receive 15 days of paid vacation per year..."

RAG searched your company documents and found the answer!



PART 2

AI Agents

AI that takes action

What is an AI Agent?

An AI system that can take actions to achieve goals

Regular AI vs AI Agent

Regular AI

- Responds to questions
- Provides information
- Waits for instructions

AI Agent

- Takes actions
- Uses tools
- Works toward goals

Key difference: Agents can perform tasks autonomously, not just answer questions

How AI Agents Work

Perceive
Understand the goal and current situation

Repeat
Continue until goal is achieved



Reason

Decide what actions to take

Act

Execute actions using available tools

Learn

Observe results and adjust approach

Tools AI Agents Can Use



Search the web
Find current
information



Send emails
Communicate on
your behalf



Manage
calendars
Schedule meetings



Access
databases
Read and update
data



Run calculations
Perform complex
math



Create reports
Generate
documents

AI Agent in Action: Example

Goal: "Research competitors and create a summary report"



Search the web for competitor information



Visit competitor websites and gather data



Organize findings into categories



Create a formatted report document



Email you the completed report



All from a single instruction!



RAG + AI Agents

The power of combining both technologies

Example: Customer Support Agent

- Uses RAG to access your knowledge base and documentation
- Acts as an agent to perform tasks like creating support tickets
- Can search for solutions, apply them, and follow up with customers
- Combines accurate information with autonomous action

Real-World Applications



Customer Service

Chatbots that access company info and resolve issues



Research Assistants

Gather data from multiple sources and create reports



Code Assistants

Help developers by searching docs and writing code



Personal Assistants

Manage schedules, emails, and tasks autonomously



Content Creation

Research topics and generate articles or presentations



Data Analysis

Query databases and generate insights

Key Takeaways



1

RAG enhances AI with access to external knowledge and current information



2

AI Agents can take autonomous actions to achieve goals using various tools



3

Combining RAG and Agents creates powerful systems that are both informed and capable



4

These technologies are already being used across many industries to automate and enhance work



Thank You!

Questions?

Ready for the exercises!