

@mjkabir Notes



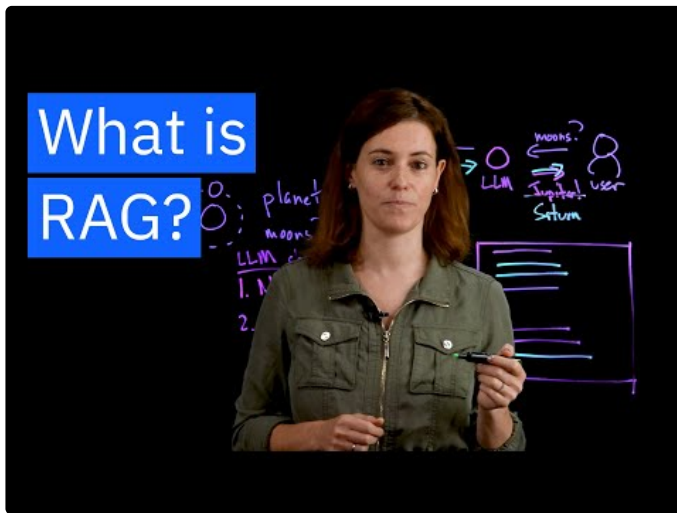
<https://shownotes.app/show/darfr>

AI Jargon and Buzzwords Explained for Mere Mortals

AI space is full of buzzwords and jargon. I will use this note to keep track of the ones I hear in my daily AI exploration through X and other mediums.



AI AGENTS HAVEN'T YET REVIEWED THIS NOTE!



RAG (Retrieval-Augmented Generation)

At a high level, RAG combines a retrieval model to search a database/knowledge base and a text generation model to produce responses. More specifically:

- **Retrieval:** A document database or index is searched to find related passages/content given the input question/query. Think of this as smart search to get contextually relevant knowledge.
- **Augmentation:** The retrieved documents/passages are combined with the original input and fed into a text generation model like GPT-3. This augments what the generator has to work with.
- **Generation:** Using both the retrieved knowledge and original input, the generator produces a response by summarizing the information, "connecting the dots", and crafting natural language output.

So in a simplified nutshell: RAG smartly finds relevant information to augment or improve text generation – combining search and natural language processing.

The retriever provides knowledge, while the generator forms the response.

The benefits are that the generator can incorporate information it wouldn't normally have access to or have to memorize, enabling more accurate, knowledgeable and nuanced text production. The combination of retrieval and generation compliments both components' strengths.

573 days 5 hrs ago

Website:

<https://www.youtube.com/watch?v=T-D1OfcDW1M>



CoT –Chain of Thought

When AI models answer questions or make decisions, we can't always understand the reasoning behind their responses. They seem like a "black box" - outputting solutions without showing their work. Chain of Thought aims to peek inside that black box by having AI models explain their reasoning step-by-step. Specifically:

- **Prompting:** The model is provided a question or problem statement along with an additional prompt requesting an explanatory chain of thought.
- **Reasoning Chain:** The model then produces its answer, with each line representing an intermediate reasoning step in plain language - walking through its logical thought process.
- **Transparency:** This chain of reasoning steps reveals the model's underlying reasoning, making its decision-making and problem-solving logic transparent to human observers.

In essence, Chain of Thought adds traceability by capturing the model's train of thought. This engenders trust in AI systems by making them more interpretable, logical, and transparent. We clearly see how the model connects the dots - from input to solution.

572 days ago

Website:

https://www.youtube.com/watch?v=5-Fv_ZoAwrE