

How to Implement a RAG Use Case With The Generative AI

Nora von Thenen, Developer Advocate, SAP

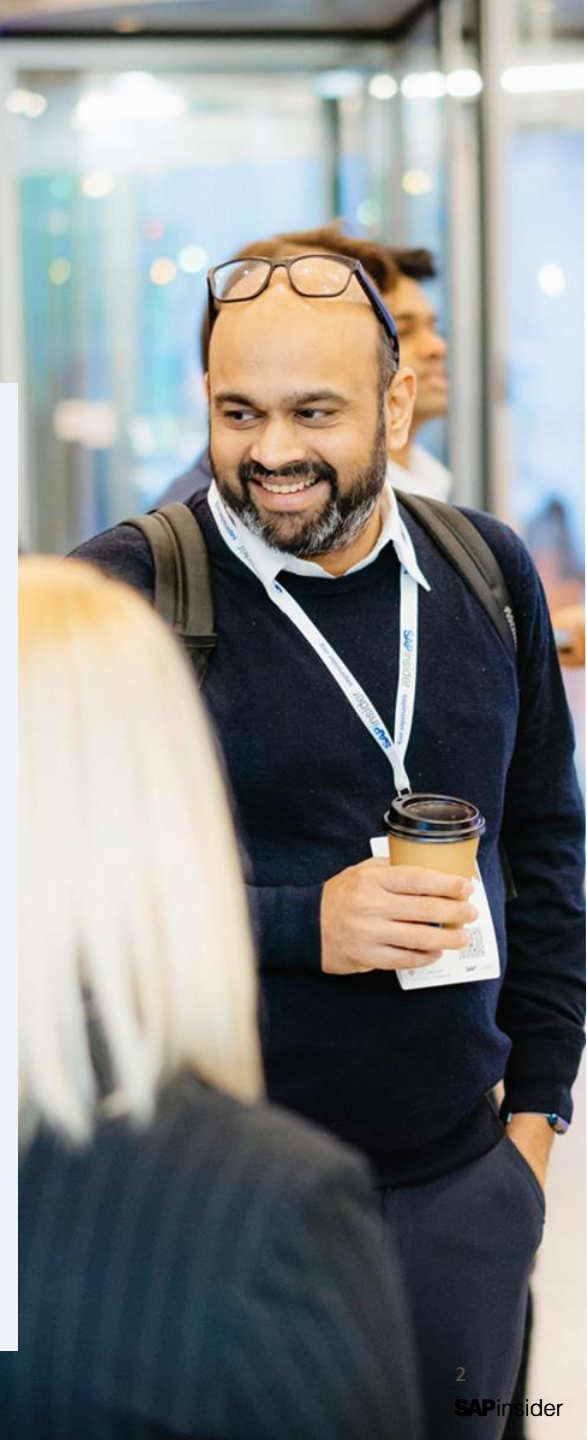
Barcelona

2024

SAPinsider

What We'll Cover

- AI Foundation
- Use the generative AI hub on SAP AI Core
- Learn how to leverage and customize language models with custom data
- Use the Python SDK of the generative AI hub
- Wrap-Up



AI buzzwords

ChatGPT

langchain

hallucination

Foundation Model

Generative Pretrained Transformer (GPT)

Agents

embeddings

fine-tuning

vector database

retraining

Large Language Model (LLM)

Transformer

Retrieval Augmented Generation (RAG)

Prompt Engineering

- Make prompt as specific as possible
- Be creative and try different prompts/wording
- Chain-of-Thought Prompts: clear instructions: “first do this then do that”, “if this do that”
- Zero-shot prompting
 - “Translate this sentence from German to English:
Wir sind in Las Vegas!”
- Few-shot prompting (give examples/shots)
 - “Ich heiße Nora. – My name is Nora.
Wir sind in Las Vegas! – ?”
- More complex prompts you can find here:
<https://www.promptingguide.ai/techniques>
- Feed output again and ask for feedback
- Roles in GPT interface to feed history: system, user and assistant

Demo

Key Points to Take Home

- Compare performance of LLMs – use cheapest and fastest
- Prevent model hallucinations
- Python SDK for generative AI hub for complex use cases
- Biases in LLM output
- Ethical considerations of use case

Where to Find More Information

<https://www.youtube.com/watch?v=MEJJnzNjGKE>

- SAP Developers YouTube Channel tutorial video on prompting the generative AI hub

Nora von Thenen, “SAP Tech Bytes: Prompting Large Language Models in the generative AI hub on SAP AI Core”

<https://developers.sap.com/tutorials/ai-core-generative-ai.html>

- Tutorial on how to use the generative AI hub on SAP AI Core

Dhrubajyoti Paul, “Prompt LLMs in the generative AI hub in SAP AI Core & Launchpad”

<https://pypi.org/project/generative-ai-hub-sdk/>

- SDK documentation for the Python SDK of the generative AI hub

<https://github.com/SAP-samples/sap-tech-bytes/tree/2024-29-01-generative-ai-hub>

- Example code for a retrieval augmented generation use case for question answering using the generative AI hub and the Python SDK

https://api.sap.com/api/AI_CORE_API/overview

- Postman collection for SAP AI Core API

<https://www.youtube.com/watch?v=PoZxanevWa4>

- Devtoberfest Session on Large Language Models by Karim Mohraz

Thank you! Any Questions?

Nora von Thenen

[Twitter.com/nora_von_thenen](https://twitter.com/nora_von_thenen)

[Linkedin.com/in/nora-von-thenen](https://linkedin.com/in/nora-von-thenen)

Please remember to
complete your session
evaluation.

SAPinsider



SAPinsider.org

PO Box 982Hampstead, NH 03841
Copyright © 2024 Wellesley Information Services.
All rights reserved.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies. Wellesley Information Services is neither owned nor controlled by SAP SE.

**SAPinsider
comprises the
largest and fastest
growing SAP
membership group
with more than
800,000 members
worldwide.**
