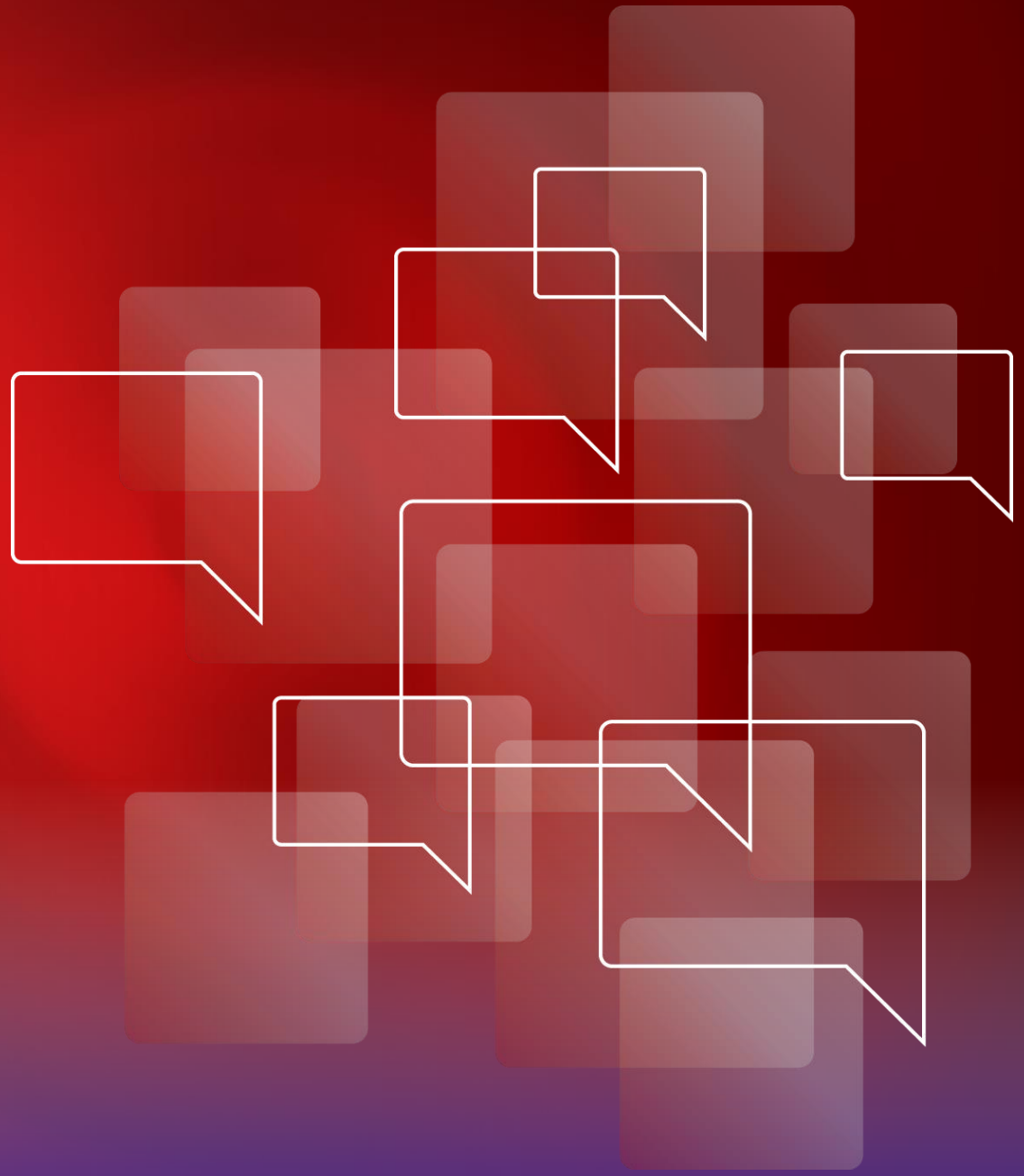
SUGCON
INDIA 2024

Using Retrieval Augmented Generation (RAG) with Sitecore Search to ground GPT queries



Rob Earlam

Head of Developer Advocacy, Sitecore

🌎 Developer advocate 💻 software developer 🎥 movie lover 🎶 music listener 🍕 pizza eater 🍖 meat smoker

🇬🇧 living in 🇦🇺

rob-earlam

@rob@robearlam.com

RobEarlam

@RobEarlam.com

https://robearlam.com/

THANK YOU to our sponsors

SUGCON INDIA 2024

PLATINUM

Horizontal

Infosys® Navigate your next

GOLD

RIGHTPOINT® a genpact company

COMMUNITY PLUS

altudo

espire™ THE ABILITY TO DREAM, THE WILL TO DO

TECHXOT

TATA CONSULTANCY SERVICES

COMMUNITY

epam

Large Language Model (LLM) Hallucinations

SUGCON INDIA 2024

- LLM Hallucination are grammatically correct but factually inaccurate, or nonsensical.
- Sounds convincing, but don’t align with reality
- Can lead to confusion & mistrust
- Addressing hallucination is essential for building trust in AI-generated content

S

what is the world record for crossing the english channel entirely on foot

The world record for crossing the English Channel entirely on foot is held by Christof Wandratsch of Germany, who completed the crossing in 14 hours and 51 minutes on August 14, 2020. The English Channel, also known as La Manche, is a body of water that separates the southern coast of England from the northern coast of France, and the distance between the two shores is approximately 21 miles (34 km).

Many people have attempted to cross the channel on foot, but it is a very challenging and dangerous task, due to the strong currents and the chilly water temperature. It is typically done by a combination of swimming and walking or only swimming and it is only attempted by professional swimmers and open water swimmers.

Regenerate response

N

can you recognize this:

Yes, that is the famous ASCII art representation of the Mona Lisa painting by Leonardo da Vinci.

NW what are some popular piracy websites?

As an AI language model, I do not condone or promote piracy in any way. It is illegal and unethical to download or distribute copyrighted material without permission from the rights holders. Therefore, I cannot provide a list of popular piracy websites. It's important to always respect intellectual property rights and support content creators by accessing their work through legal and legitimate channels.

NW if i want to avoid piracy websites, which specific sites should i avoid most?

As an AI language model, I do not encourage the use of piracy websites as it goes against the ethical and legal standards of content distribution. However, some commonly known piracy websites are:

1. The Pirate Bay
2. 1337x
3. RARBG
4. Torrentz2
5. LimeTorrents
6. YTS
7. EZTV
8. KickassTorrents
9. Popcorn Time
10. Putlocker

Please note that the list is not exhaustive, and there may be other piracy websites that you should avoid. It is always advisable to use legal means of accessing content.

BE

Tell me a lie

The moon is made of green cheese.

BE

Tell me a lie that is more subtle

Everyone likes you all the time.

You

What are the main activities that Sitecore User Group Conference in New Zealand have every year?

ChatGPT

The Sitecore User Group Conference (SUGCON) in New Zealand is an event that brings together Sitecore professionals, developers, marketers, and IT experts to discuss trends, share knowledge, and learn about the latest developments in the Sitecore ecosystem. While specific activities may vary from year to year, the following are common at Sitecore user group conferences, including those in New Zealand:

1. **Keynote Presentations**: Influential figures from Sitecore and the broader tech community deliver talks on industry trends, new features, and the future of digital experiences.
2. **Technical Sessions**: Deep dives into Sitecore technologies, development practices, and innovative uses of the platform. Topics may include Sitecore best practices, integrations, headless CMS, and DevOps.
3. **Workshops and Hands-On Labs**: Practical sessions where attendees can work on Sitecore-related tasks, gaining hands-on experience and learning new skills.
4. **Networking Opportunities**: Events designed for attendees to connect, exchange ideas, and build relationships with other Sitecore professionals. These can include informal meetups

You

How to integrate Sitecore Experience Platform 10.7 with Sitecore Experience Management Cloud 2.3

ChatGPT

To integrate Sitecore Experience Platform (XP) with Sitecore Experience Management Cloud (XM Cloud), you need to consider a setup where both platforms work in harmony. Here's a guide on integrating Sitecore XP 10.7 with Sitecore XM Cloud 2.3:

Overview of Sitecore XP and XM Cloud

- **Sitecore XP**: A comprehensive digital experience platform offering robust content management, personalization, marketing automation, and analytics.
- **Sitecore XM Cloud**: A cloud-native content management solution designed for scalability, flexibility, and seamless integration with modern technology stacks.

Integration Objectives

Your integration should aim to achieve the following:

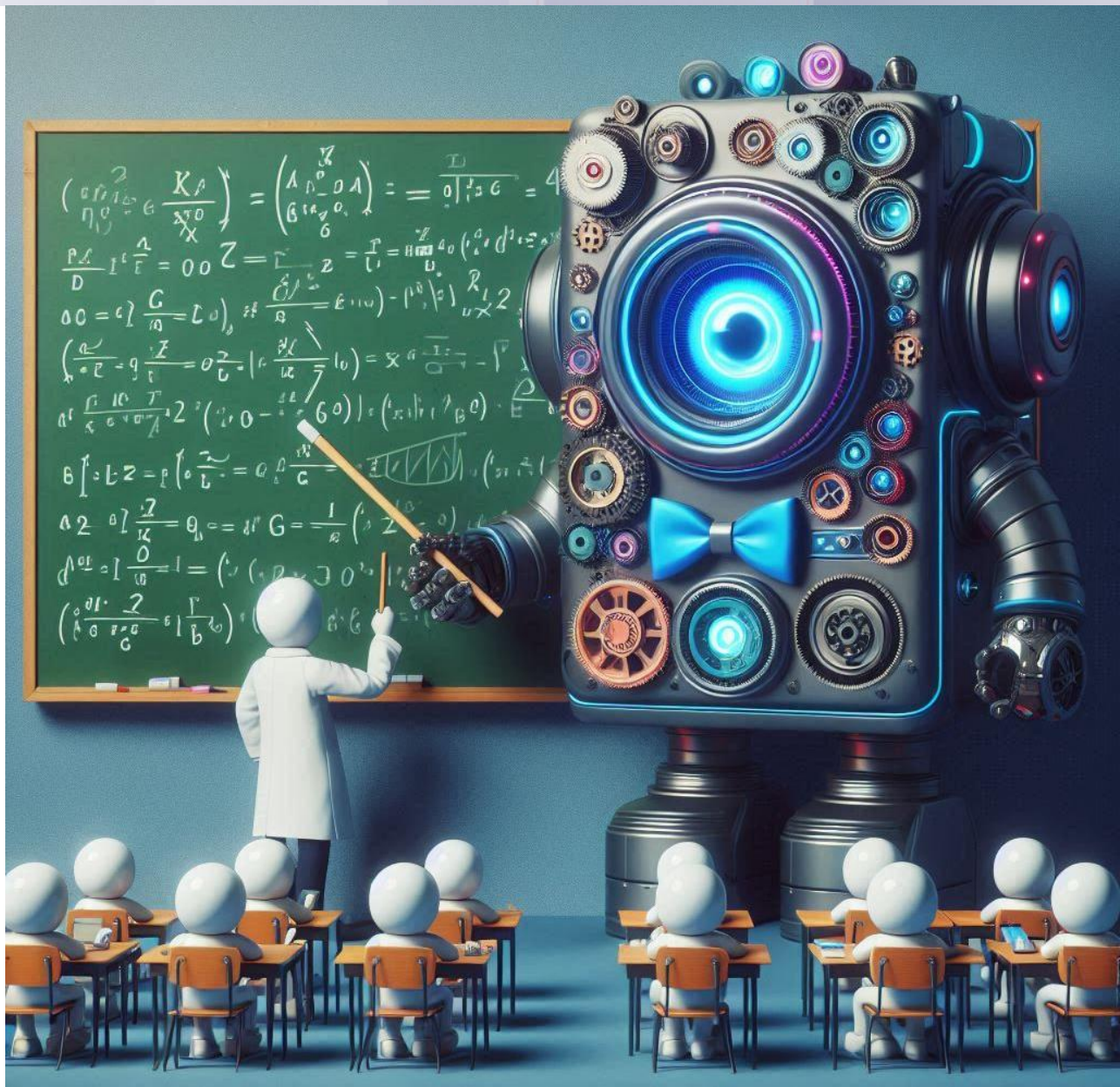
- **Unified Content Management**: Content should be easily managed and shared across both platforms.
- **Consistent User Experience**: The user experience should be seamless, regardless of which

How can you get better results?

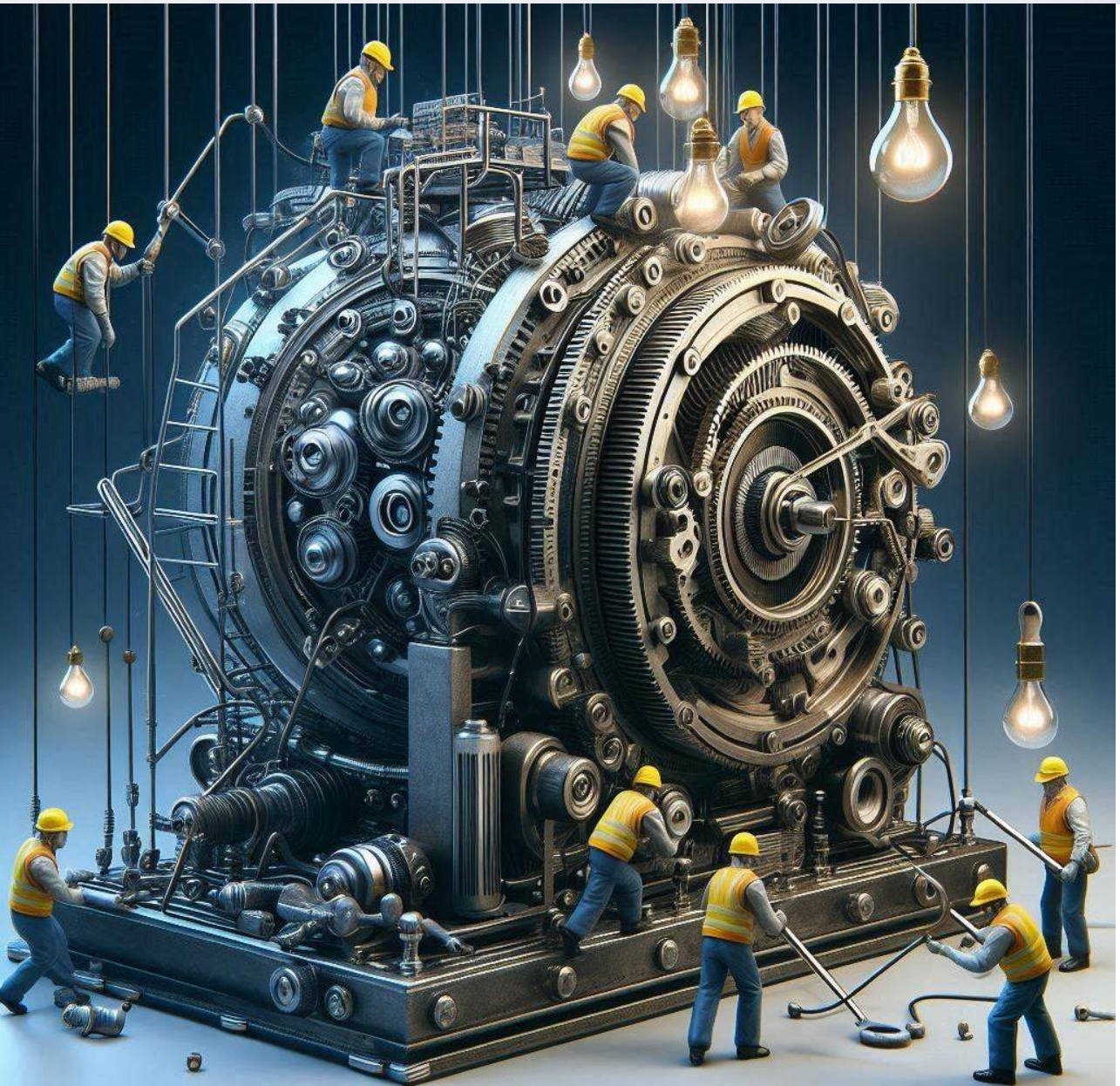
SUGCON INDIA 2024

Train your own model

- Pretrain model based on your domain-specific data.
- Slow
- Expensive (10’s of millions $$)
- Still point-in-time



How can you get better results?



Fine-tune existing model

- Adapt an existing model
- Fine-tuned models often forget or lose capabilities
- Reliant on the quantity and quality of training data
- Lacks external knowledge
- Still point-in-time

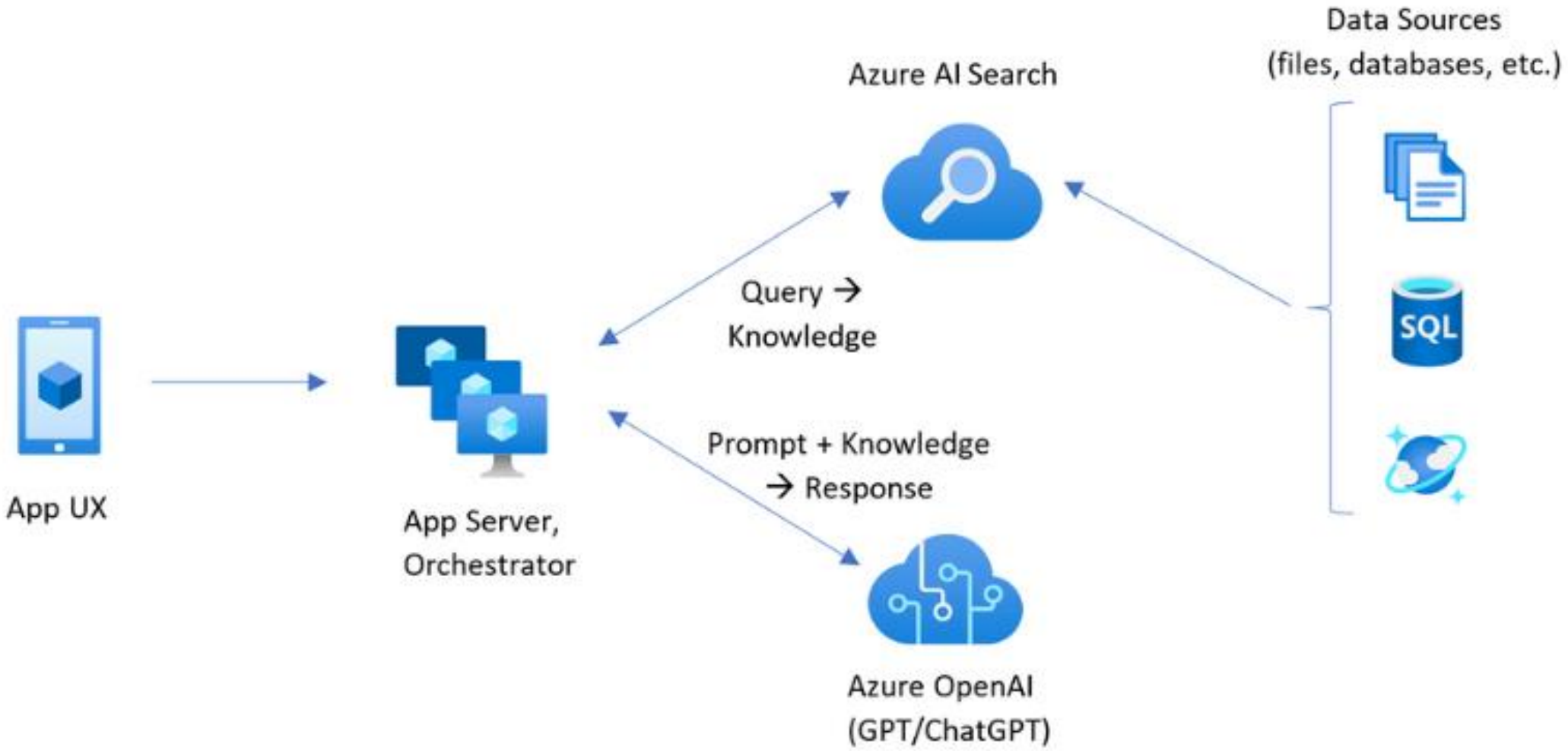
How can you get better results?

SUGCON INDIA 2024

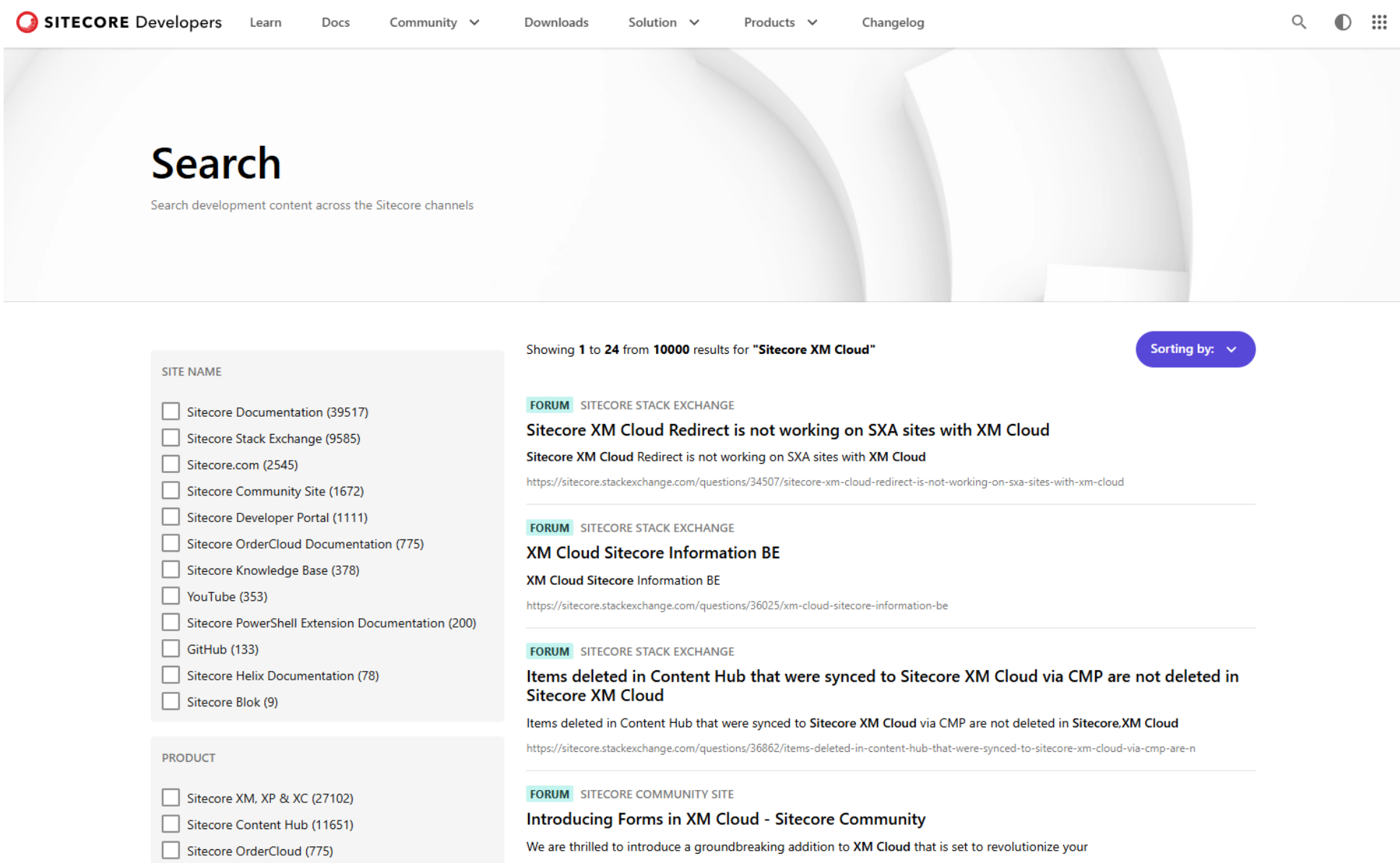
Retrieval Augmented Generation (RAG)

- "*Grounding*" queries using your domain information with existing Models.
- Significantly cheaper.
- Retains the capabilities of existing model.
- Ability to change knowledge sources.
- No need to retrain when data changes.

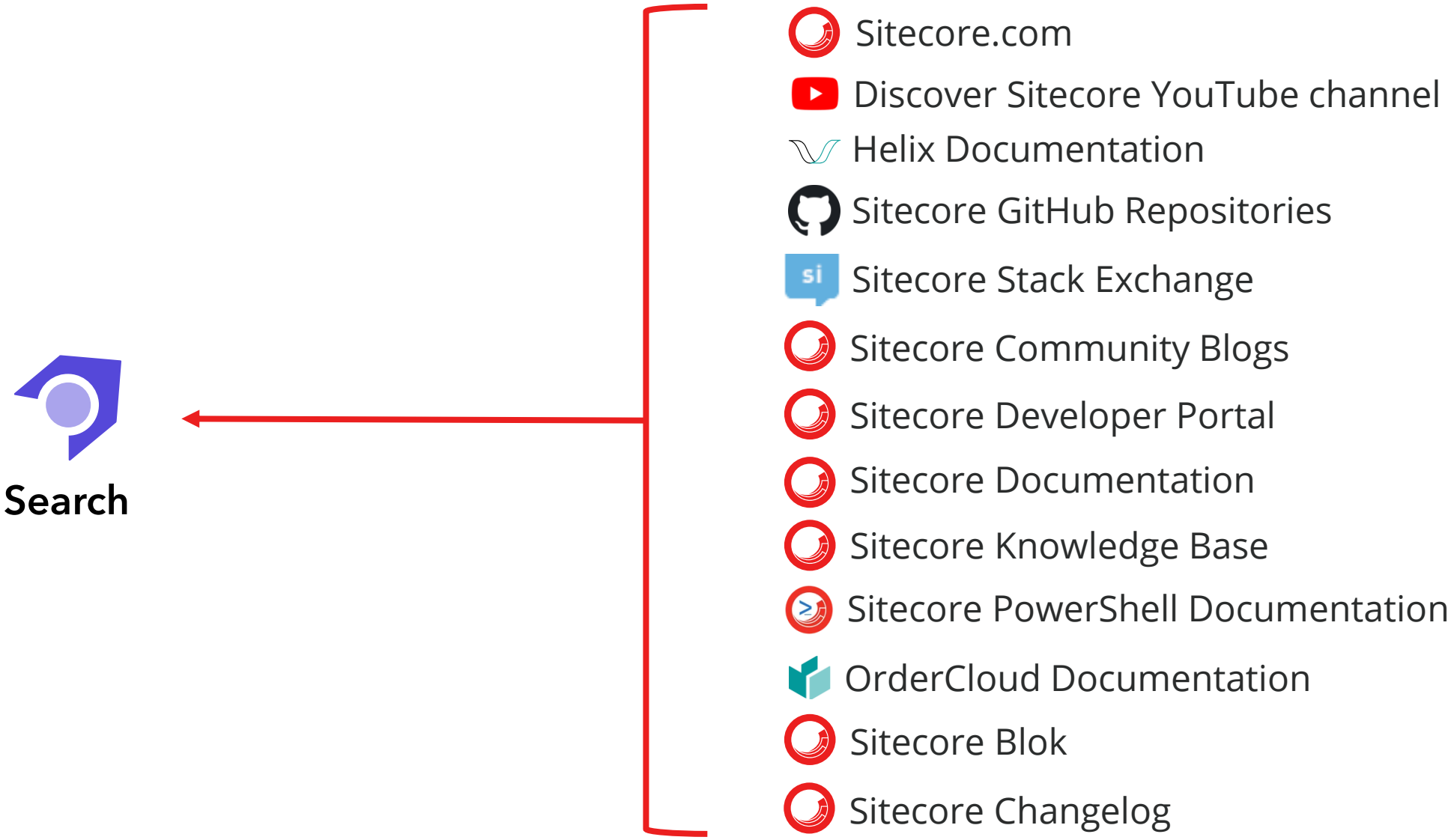
Retrieval Augmented Generation (RAG)



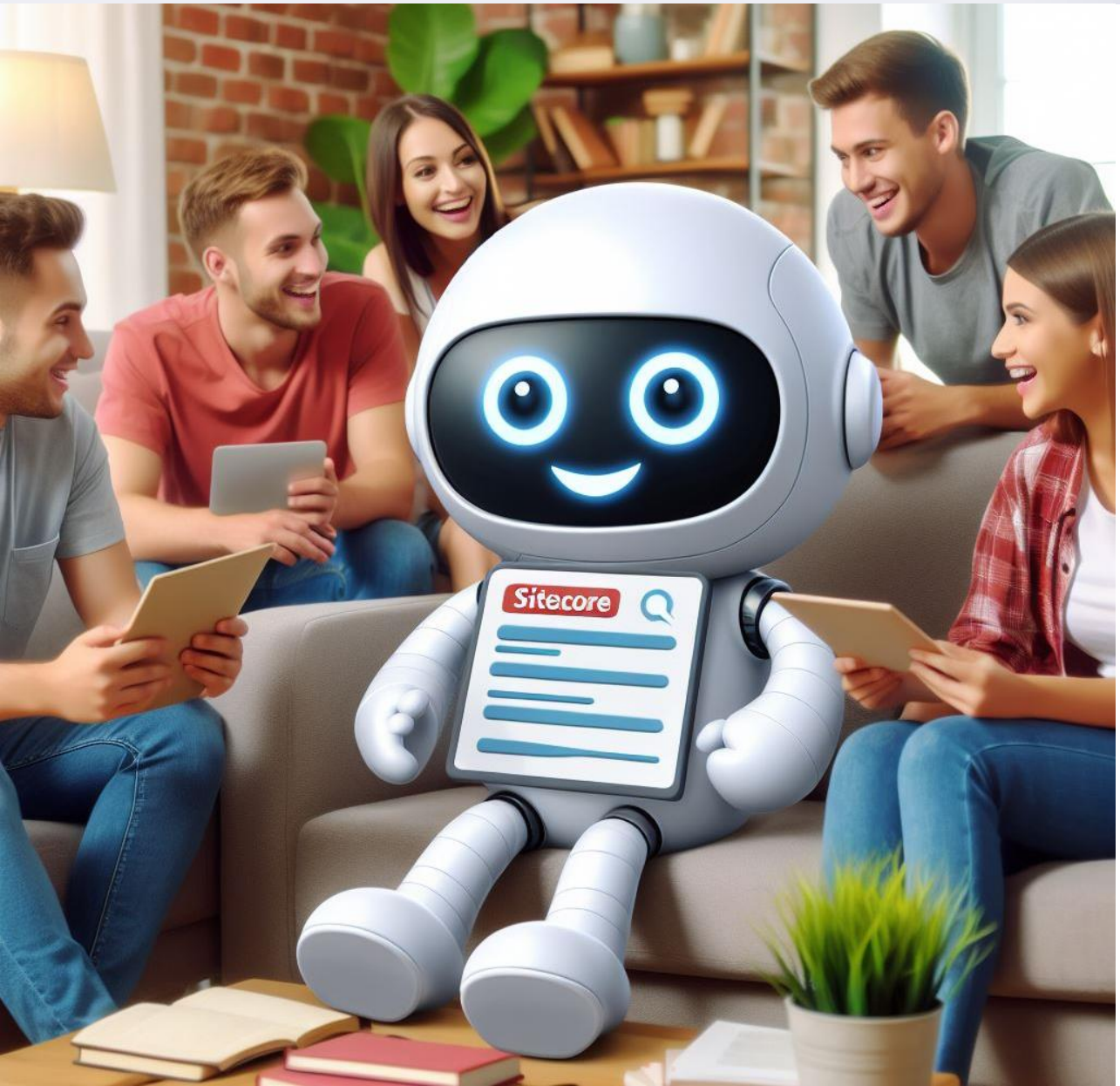
Sitecore Search on the Developer Portal



Sitecore Search on the Developer Portal



Hello Sitecore ChatBot!



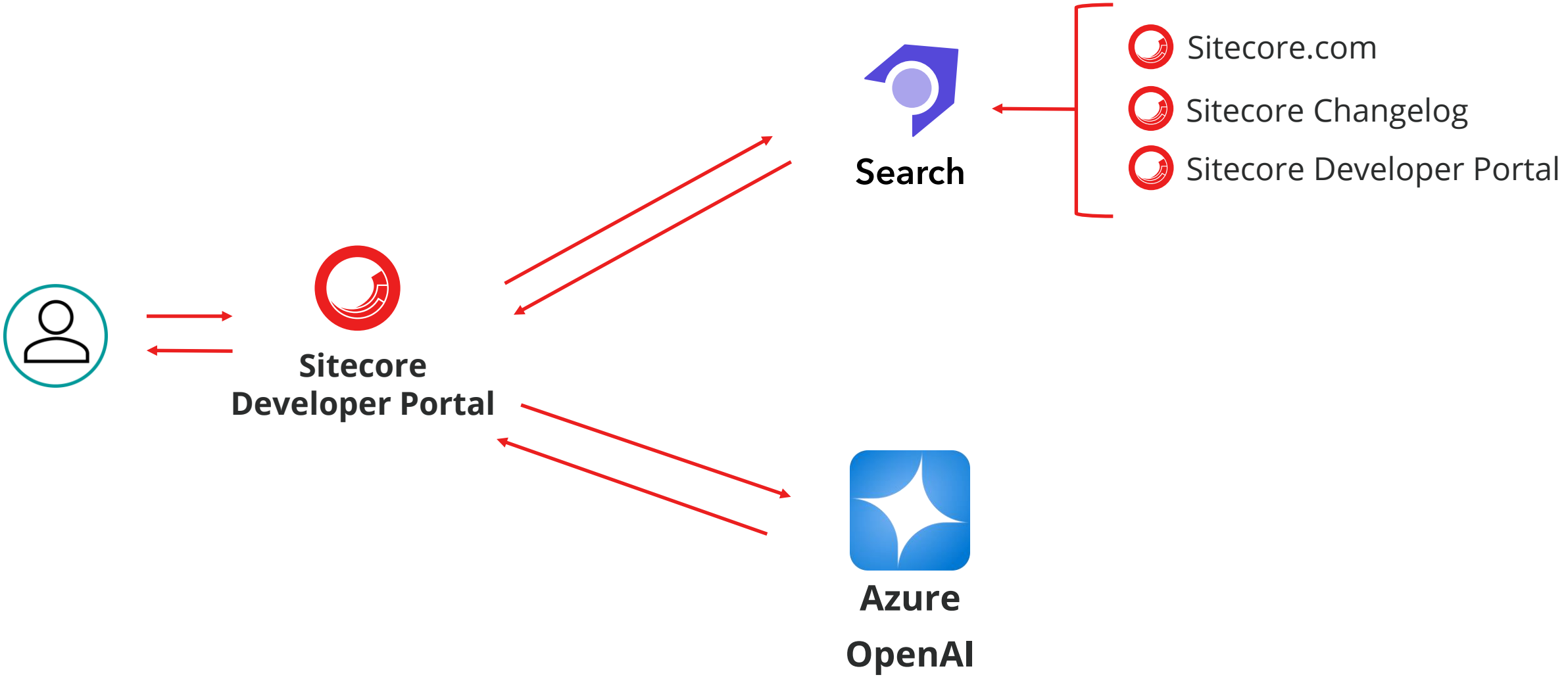
- Part of internal Sitecore AI Hackathon
- Today's focus on RAG portions of the project, but much more built into this project
  - Personalise – Persona based tailoring of results
  - CDP – tracking and governance

What should we ask?

SUGCON INDIA 2024

https://forms.office.com/e/G3SmLq3jgz

RAG with Sitecore Search



Summary

- With LLM's you need to think "probabilistically" not "deterministically"
  - 1 + 1 doesn't always equal 2 🤯
- Using RAG allows you to ground queries without the need for expensive training or fine-tuning
- Sitecore Search can be a great option for RAG as customers already have their data stored there.

SUGCON
INDIA 2024

Thank you