EDM Webinar

Scaling Responsible GenAI: 4-Step Guide

Live Date: October 11, 2023

Featuring:

Lakshmi Krishna, Associate Director of Data & AI services, Kyndryl

Wiem Sabbagh, Director, Global Data & AI Consult Leader, Kyndryl

Moderator: **Jim Halcomb,** Head of Product Management, EDM Council

Recording: View webinar **Presentation: View slide deck** **Relevant links:** Learn more about EDM Council	Learn more about Kyndryl Connect with Kyndryl experts Follow Lakshmi Krishna on LinkedIn Follow Wiem Sabbagh on LinkedIn

Thank you to Kyndryl and the panelists for providing the below answers to all questions posed during the live webinar. For more information or additional questions, contact us here.

WEBINAR Q&A:

Quality of the input to the GenAI is important. What about the output of the GenAI? Should you QA/QC the output? If so, why? Will it cause overheads in a project and drive up operational cost?

Human in the loop feedback (internal) and user feedback (external) should be implemented as part of your QA mechanism. The reason being, GenAI can be prone to hallucination - where it can generate false or incorrect responses. Some content, such as marketing / legal content needs to be considered as a first draft from Large Language Models (LLMs) as well.

Apart from that output, QC for content moderation, content filtering and sensitive data needs to be implemented as well to determine accuracy. If there are no tools to support this in your current ecosystem, it will add to project costs depending on the criticality of your business case.

Is Prompt for LLMOps, similar to Agent (for AI)?

No. Prompts are text inputs which you enter into your LLMs for which you get the appropriate response. Agent is a program that integrates with your LLM. It can activate various tools such as calculators, search, etc. Agents/tools can be triggered through prompts as part of your prompt flow.

What are the anticipated next steps in the evolution of GenAI as it relates to self-learning?

Enterprises are still prototyping and releasing beta version applications with GenAI. AGI (Artificial General Intelligence) surely will be an open research area for a considerable time. As hardware processing, model architecture evolves to support more efficient ways of handling information, training models with massive datasets, advancement in AI regulations will also influence the way GenAI evolves into self-learning or AGI.

How can data management be effectively handled in real-time?

Data Management can be handled in real-time by having a strong foundation for DataOps / LLMOps which brings together security, governance and observability features. Also, with tools that help provide a single pane of visibility into your data fabric, provide unified access, role management, identification of business relationships within your data, rule-based policy packs that can apply confidentiality rules, intelligent metadata discovery, data quality rules, tagging and cataloging your data will be crucial for maintaining data integrity and reliability. Combining this with observability and alerting will be critical to ensuring guardrails against failure. Kyndryl provides these services through the LLMOps services; do get in touch with us for more information.

What are the challenges and considerations for managing data on the fly?

The challenges with real-time data are mainly to preserve the data integrity and reliability. Ensuring content generated by LLM does not breach privacy policies and ensuring veracity of the generated content are key to successful GenAI application.

If organizations have large data shifts into SharePoint online, and other integrated Data stores, with policy management, will this be defined as data ready for LLM access, and operations for GenAI?

Vector databases such as Cognitive Search would be the best/highly efficient way to manage data. Cognitive search can connect to other data sources such as SharePoint and index data. *(Note: SharePoint is still under preview as the data source supported by Cognitive Search*). As long as the data sources are supported by Cognitive Search - it should be GenAI ready.
(https://learn.microsoft.com/en-us/azure/search/search-data-sources-gallery)

Otherwise, special integration connectors will have to be developed. In addition, depending on your use cases, the right solution can be determined.

What are the key areas where we can apply GenAI?

GenAI - is highly applicable in the following areas:

1. Customer experience - Conversational chatbot, sales, personalized marketing, etc.
2. Employee Experience - Conversational chatbots for IT/ HR functions
3. Developer Productivity - Application development, legacy code transformation, etc.
4. IT services - Agent Assist, Faster RCA, Contact Center Analytics, etc.
5. Others - Document summarization, contract compliance, business intelligence Q&A, Information retrieval Q&A