Title: Building and Leading a Team for Generative AI Product

Abstract:

The advent of generative artificial intelligence (AI) has opened new possibilities for businesses to enhance customer interactions through conversational agents. Building a successful generative AI product requires assembling a team with diverse expertise and effective leadership to navigate the complex challenges associated with this cutting-edge technology.

This lecture focuses on the process of building and leading a team for developing a generative AI conversation bot specifically tailored for Booking.com, a leading online travel platform. It explores the key considerations involved in assembling a proficient team to deliver a high-quality conversational experience to customers.

By combining technical expertise and domain knowledge, the team created an AI system that understands user queries and provides accurate and relevant responses, that maintain a consistent conversational style that aligns with Booking.com's brand voice.

In the product, we considered aspects such as data privacy and security, implementation of a moderation process, adherence to UX constraints, and an integration of the bot with actual Booking.com recommendation cards. We have faced challenges with prompt engineering and how to split the prompt instructions to best fit our needs and flow. We constructed a new infrastructure that yielded several noteworthy insights and lessons learned, which we would like to discuss and share.

The team incorporates data scientists, engineers, and user experience specialists to overcome these challenges, enabling the bot to deliver accurate information, assist in bookings, and provide exceptional and personalized recommendations while preserving latency and scale of users.

Potential discussion points:

- **Expertise needed to develop GenAI based products.** Simply forming some prompts and hoping for good results will not provide the needed outcome. The only way to develop this expertise is by building GenAI based products hands-on.
- **Prompt engineering tricks-** Different prompts techniques need to be used to improve the accuracy of LLMs - like breaking prompts to smaller tasks, Chain-of-thought (CoT), providing examples in the prompt itself (Few-shot).
- **Understanding the team roles** - discuss the importance of teamwork in developing generative AI products. Explain how diverse skill sets and perspectives can contribute to a successful team.
- **Leadership in Generative AI Projects** - Discuss the unique challenges faced by leaders in the field of generative AI. Explore leadership styles and strategies that are effective in motivating and guiding the team towards successful outcomes.
- **How to deliver fast?**- how to break the GenAI product to MVP, what should it include?
- **Future Trends and Opportunities-** discuss emerging trends, research directions, and potential future opportunities of generative AI. How what we built can be still relevant if new models are coming out.

An explanation about relevance of this talk/panel discussion to the workshop

The lecture on "Building and Leading a Team for Generative AI Product" is extremely pertinent to the workshop as it equips the audience with valuable practical knowledge and strategies for effectively constructing and guiding teams in the realm of generative AI projects. Given the increasing significance and applicability of generative AI in numerous machine learning companies today, the insights shared during the lecture hold immense value and relevance. By incorporating the knowledge imparted in the talk, workshop participants can directly apply it to their roles as managers and leaders of applied machine learning teams, empowering them to steer their teams towards accomplishing successful and impactful generative AI projects.

Presenter information (LinkedIn)

Moran is a machine learning manager at booking.com, researching and developing computer vision and NLP models for the tourism domain. Recently launched a generative AI chatbot product designed for the travel domain.

Moran is a Ph.D candidate in information systems engineering at Ben Gurion University, researching NLP aspects in temporal graphs.

Previously worked as a Data Science Team Leader at Diagnostic Robotics, building ML solutions for the medical domain and NLP algorithms to extract clinical entities from medical visit summaries.