Leveraging Large Language Models for Conceptualizing Health Economic Models: A Feasibility Study in Oncology

EE504

Tushar Srivastava, Shilpi Swami, Thaison Tong

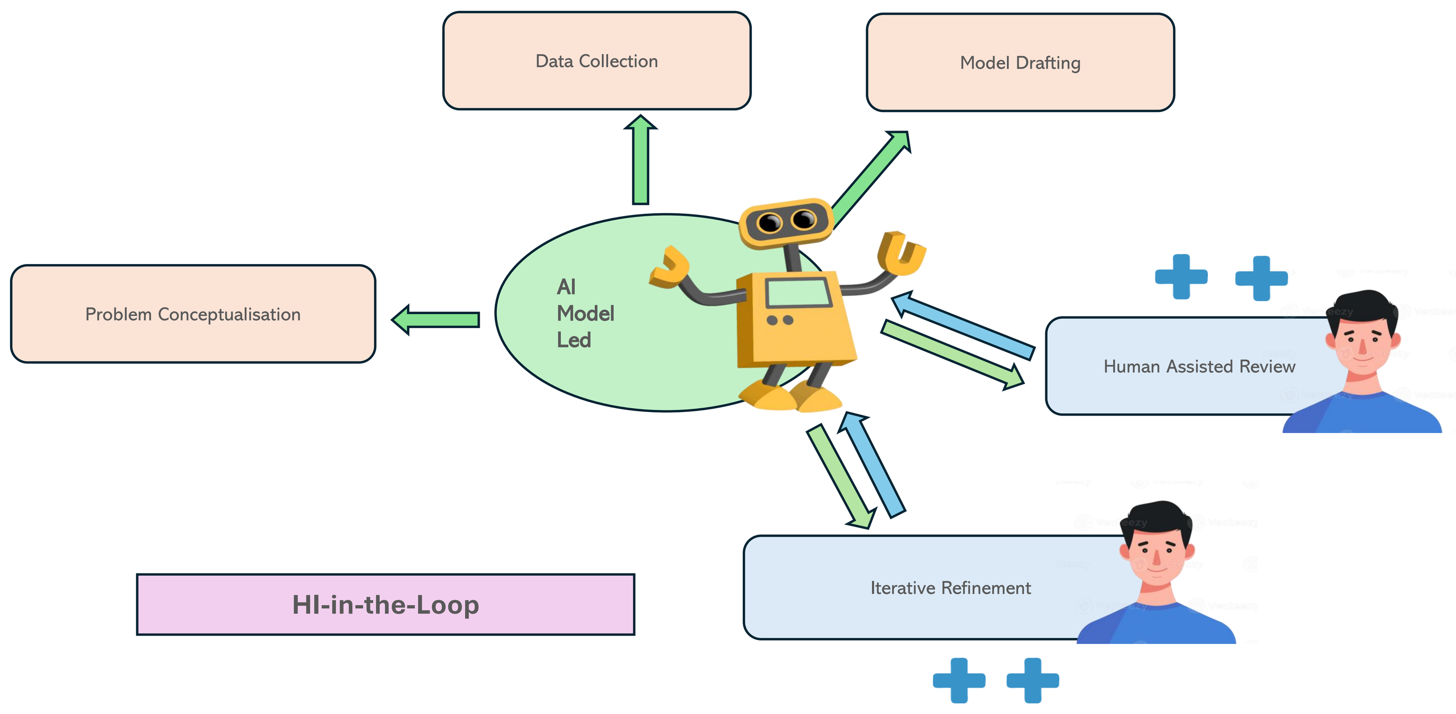
ConnectHEOR Ltd., London, UK. Email: Tushar.srivastava@connectheor.com

connectHEOR

BACKGROUND

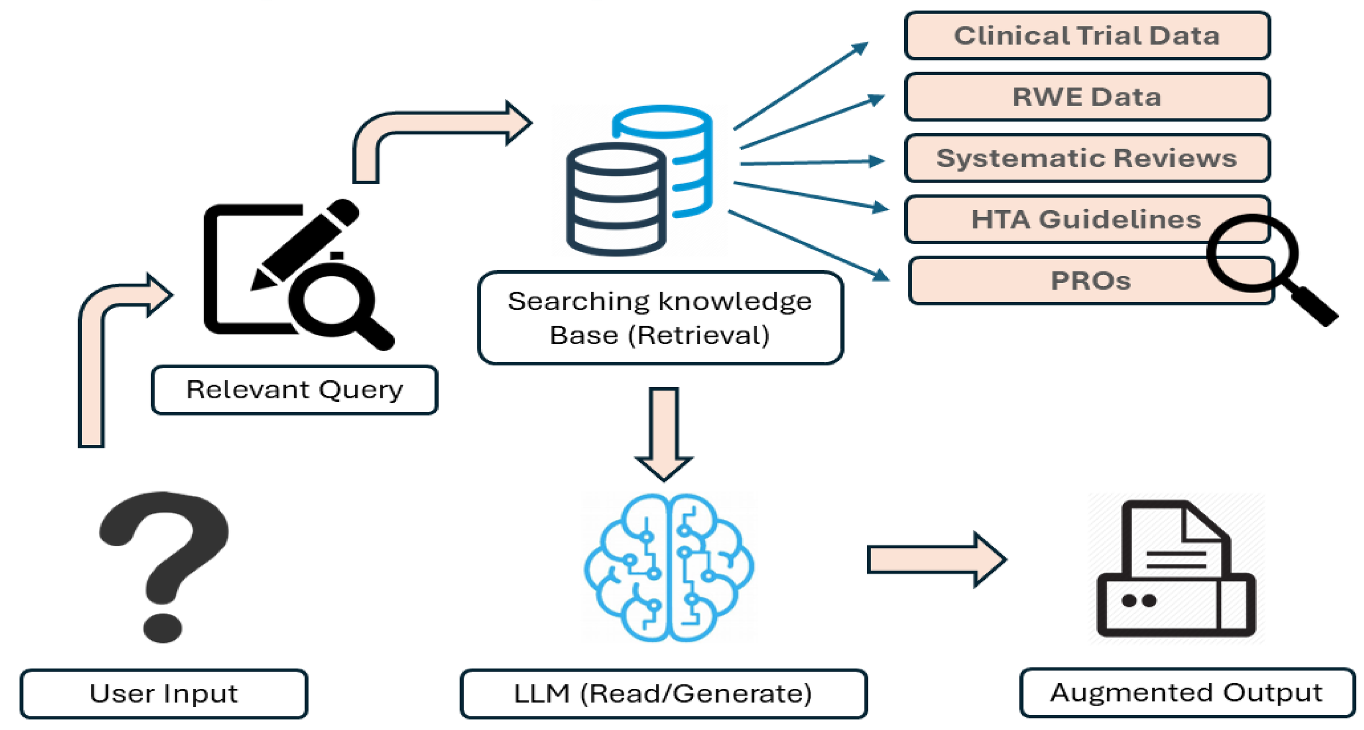
- **Context: Health economic modelling (HEM)**, crucial for assessing the cost-effectiveness of healthcare interventions, is a labour-intensive process requiring extensive expertise and time. However, advancements in **artificial intelligence (AI),** particularly with **large language models (LLMs) such as GPT-4**, offer new opportunities to streamline this process.
- **Aim:** We explore the feasibility of using LLMs for conceptualizing HEMs by leveraging advanced reasoning algorithms and prompt engineering techniques. A **proof-of-concept** exercise was undertaken and a **cost-effectiveness model for an anti-cancer therapy** in advanced breast cancer was developed using a **human intelligence (HI) in-the-loop approach (Fig. 1)**.

Figure 1 : Human in the loop approach to incorporating artificial intelligence



- Various reasoning algorithms using **prompt engineering**, such as Chain of Thought (CoT)[1], Tree of Thought (ToT), and CoT-Self-Consistency, were explored.
- To augment the knowledge base of the LLM with domain-specific data**, retrieval augmented generation (RAG)**[2] was employed. RAG database was populated with HEOR-related guidelines as well as disease specific documents to provide background and context. The framework was developed in Python along with **PostgreSQL** for database management (**Fig. 2**). The user input in the form of prompting techniques coupled with augmented knowledge base enabled the LLM to produce highly specific and **human-expert-like responses**.

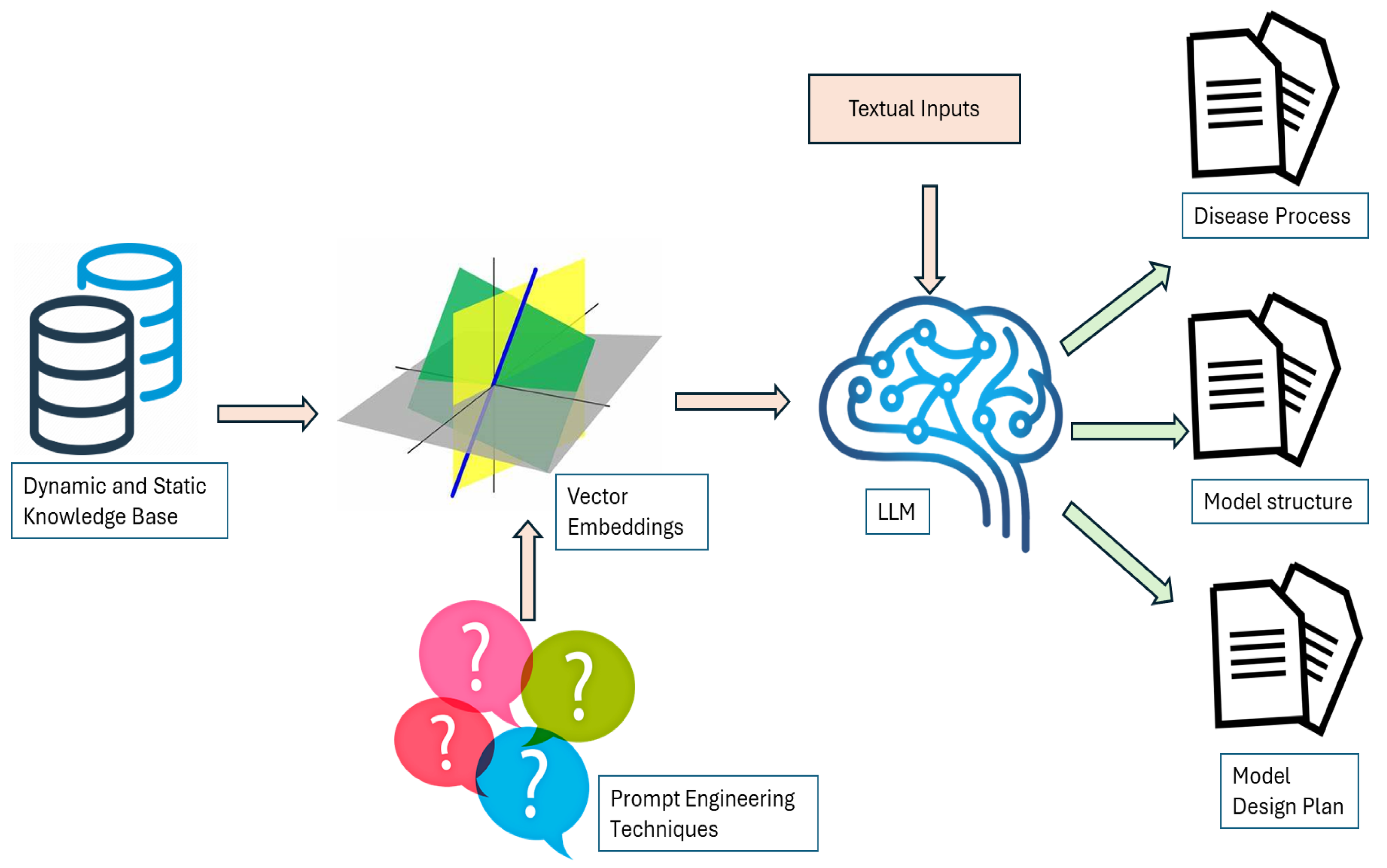
Figure 2: The working of RAG architecture with an LLM



IMPLEMENTATION

- **Static documents** pertaining to HEOR guidelines and **dynamic documents** related to disease-specific data was **uploaded to the database.**
- The said data was further **embedded as high-dimensional vectors** and stored in a vector database to be queried efficiently using **cosine similarity matching**.
- **CoT prompts** were designed emulating human-like thought process. The prompts were further supplemented by user inputs in text format to provide context to the LLM.
- The prompts, input and query were passed to the LLM which further queried the vector database to find the documents with the **highest similarity index.**
- The retrieved data was used by the LLM to conceptualize the problem question.
- The output was given in the form of **1) a disease process diagram, 2) model structure diagram,** and **3) model design plan.**

Figure 3: The workflow of the AI HEM conceptualisation tool



RESULTS

Figure 4: Disease process diagram produced by the AI

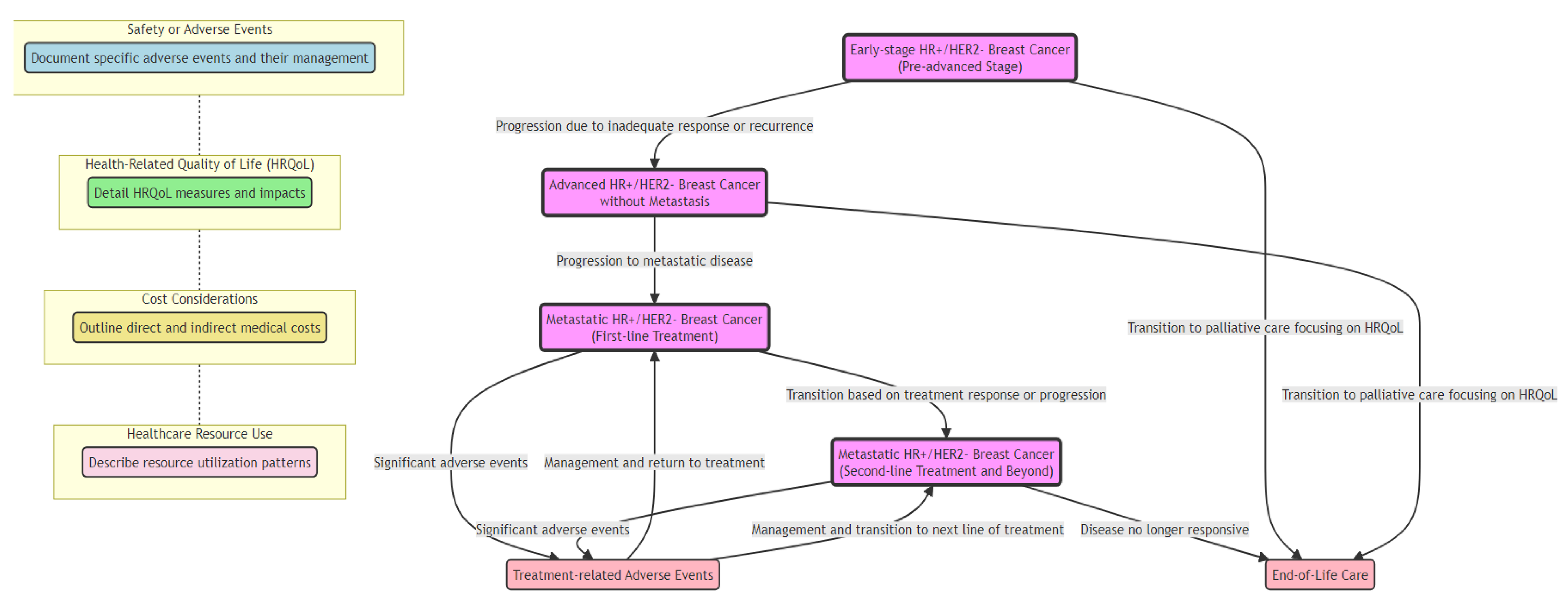
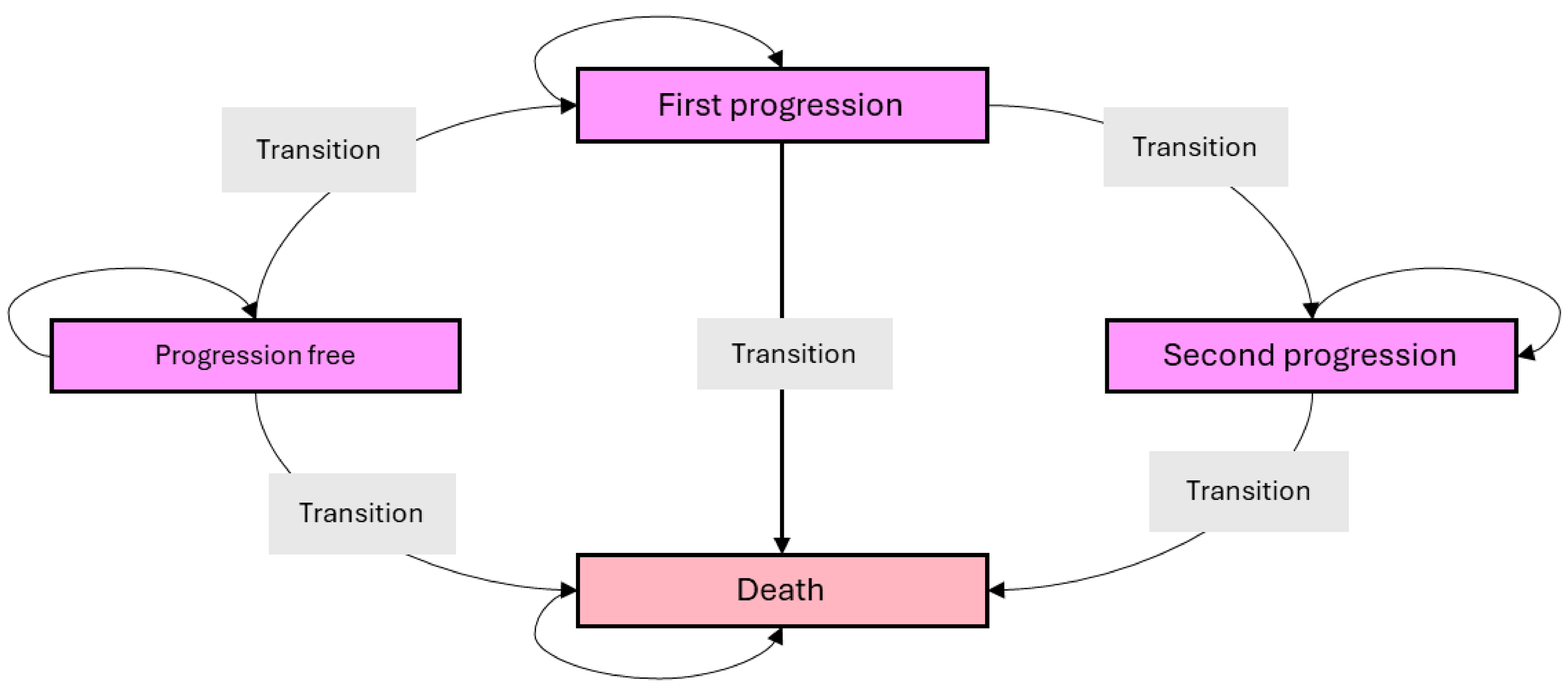


Figure 5: Model schematic produced by the AI



For the model structure, LLM suggested a Markov model with four health states: "Progression-Free Survival" (PFS), "First Progression," "Second Progression," and "Death." Key parameters and gaps were highlighted. The LLM recommended a natural history which was further refined using the HI-in-loop approach. The initial recommendation was promising and closely aligned with what human experts might have generated. However, a prime facie limitation would be the need for regular human review to ensure adherence to best practices.

References.

1. Sun, B. W. (2023). Towards Understanding Chain of Thought Prompting: An Empirical Study of What Matters.
2. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Lewis, M. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.

Acknowledgments

Thanks to Hanan Irfan and Yash Kumar (ConnectHEOR, Delhi, India) for their support on poster content and design development.

Financial Disclosure

The authors are employees of ConnectHEOR Limited and no external funding was received to conduct this research. The authors have no conflict of interest to declare.

Poster presented at the ISPOR EU 2024, November 17-20, Barcelona, Spain