QuantumBasel

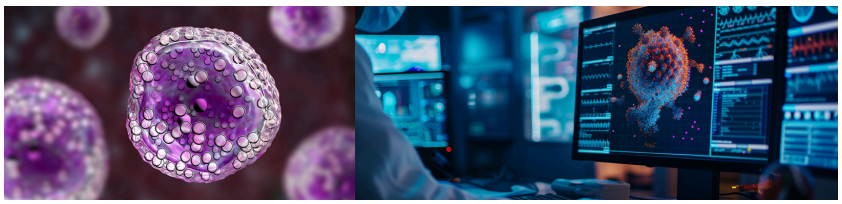
Center of Competence
for Quantum and Artificial Intelligence

LLM GenAI

Precise Diagnostics Improve Treatment Options for Cancer Patients

2024 | Use Case | Model Optimization Solution

Cancer Genomics Consults (CGC) is dedicated to advancing cancer care through precision genomic analysis. With years of experience interpreting complex cancer genomics cases worldwide, the company provides timely, expert analysis and consultation to healthcare providers, ensuring that world-class expertise is available when needed.



THE CHALLENGE

Cancer genomics is a critical element of precision oncology, which uses broad genomic profiling of tumor tissue or blood to identify genomic biomarkers of sensitivity or resistance to various cancer therapies. The lab reports from these tests are routinely 20-50 pages long and impractical for integration into the workflow of a practicing oncologist. Even for oncologists who are experts in the field, these machine-generated reports can be too complicated, resulting in incorrect treatment recommendations.

40-60
page long oncology reports to read

2-3h
laboratory staff time required to condense complex report

PROJECT GOAL

To improve this process, the goal is to develop an AI workflow that integrates retrieval (searching for relevant information) with generation (creating text based on retrieved information) to improve the quality of output and produce more informed and accurate results.

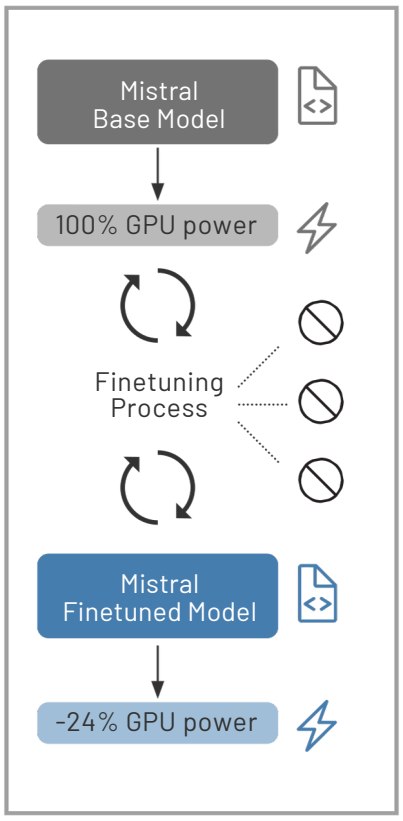
9x
reduction in review time needed

SOLUTION APPROACH

For this case, several state-of-the-art open source components and a customized large language model (LLM) with local data storage and processing were used to improve the accuracy, explainability, and transparency of the data flow. An open source LLM was fine-tuned through supervised and reinforcement learning based on input from a genomics expert. CGC experts were able to provide the AI with the correct answers to typical genomic questions, which served as ground truth. The model was then tested in multiple iterations on new datasets to evaluate its ability to perform the task. Through a Retrieval-Augmented-Generation (RAG), the user is able to upload a document (the cancer report) and chat with the document.

„The AI efficiently scans long cancer genomics reports (40- 60 pages) to help analysts capture important details and flag missing clinically relevant information for follow-up. It's a great example of explainable AI in routine use today."

James Creeden, GCG Cancer Genomics



The new solution resulted in a 90% reduction in text reading time and analysis work and a 24% increase in hardware efficiency. This hardware optimization was achieved by quantizing the model, making it leaner and more efficient. Finally, the AI was made highly trustworthy by significantly reducing hallucinations. In order to improve this even further, the model provides all the sources for the given answer in a visually appealing way within the uploaded report and has a self-check function. This allows the user to review only the relevant text in the report as needed.

-90%
reading time required

8x
increased capacity for cancer laboratories

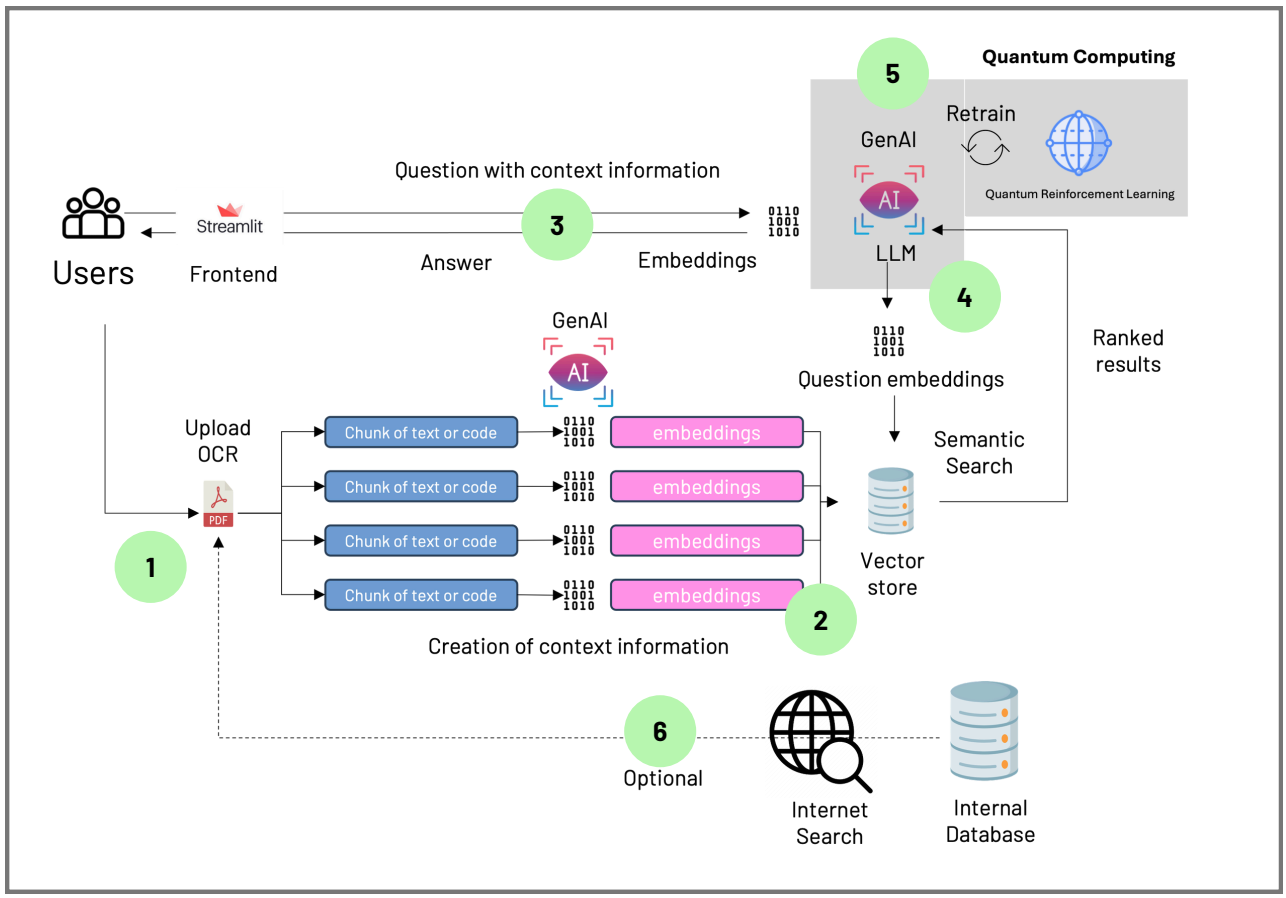
325%
improvement in user scores

NEXT STEPS

The investigation of how Quantum Reinforcement Learning and other quantum machine learning techniques could enhance the computational pipeline for interpreting genomic reports will be explored in the next steps.

DETAIL INFORMATION

Explanable Generative AI-based Cancer Genomics Report Interpretation infused with Quantum Computing reduces 90% of the work involved in reading and analyzing text.



1. User can upload a cancer report as a PDF in 80+ languages.

2. The provided PDF is split into text chunks and then converted into embeddings. These embeddings are represented by vectors and stored in a vector memory.

3. The user can ask questions with the information provided in the report.

4. When context information is provided by the user, the AI performs a semantic search on the vector store.

5. A key step in retraining our GenAI model is Reinforcement Learning (RFL). Quantum computing has shown great advantages for speeding up the RFL training process.

6. In addition, we are providing the AI with the ability to retrieve information from other sources such as the web and databases.