



OpenTech is an information technology company headquartered in Riyadh, Saudi Arabia. They focus on delivering a wide range of technology solutions and services, including business consulting, web portal development, data analytics, process management, custom software development, and cloud computing. The company emphasizes innovation and the adoption of the latest technologies to help clients stay competitive. OpenTech often collaborates with both private and public sector organizations, providing tailored solutions that address the specific needs of businesses in the region. Additionally, they offer training and support services to ensure that clients can effectively utilize the technologies and solutions they implement.

Businesses in all sectors around the world are looking at use cases for generative AI in their organization to extract new value from data and increase productivity for employees. However, much of this data is confidential and sensitive, and cannot be stored in a public cloud environment to train a Large Language Model (LLM). This presents a unique challenge for businesses on how to take advantage of the possibilities of generative AI while at the same time ensuring data security and compliance with regulations such as GDPR, HIPAA, etc. OpenTech is a company that embraces the challenges of "it can't be done". When the opportunity was presented to build an on-prem chatbot solution that talks to private documentation for clients, OpenTech turned to [IBM® Power®](#) for the infrastructure on which to build on.

Solution

The OpenXAI platform from OpenTech is a powerful, LLM-based, private document sourced knowledge-based generative AI chat application that generates content for users based on their own private data. It is meticulously trained on localized user data, and what you feed the platform is what you retrieve. RAG (Retrieval-Augmented Generation) is used to further restrict the information sources to just the desired and targeted documentation and not include any pre-trained data. This training approach enables OpenXAI to respond with heightened efficiency and accuracy to a wide array of service requests and inquiries. OpenXAI delivers a consistently superior user experience, significantly boosting client satisfaction.

The choice of hardware on which to run OpenXAI was extremely critical for OpenTech. Since the chat application was being trained on sensitive and confidential documentation, it was mandatory that the solution stay on-prem within the country to comply with local regulations regarding data privacy and sovereignty. Cloud options that other vendors located outside of the country were pushing were not an option for their clients. Because of this and the performance and security requirements that OpenXAI needed, OpenTech knew IBM Power10 processor-based servers were the only direction to go.

Benefits

With the ease of scalability of Power, OpenTech can dynamically spin up down additional CPU cores for dynamic loads and power boosts based on demand and usage of the OpenXAI solution. The ability to configure a system exactly the way a client required was also a key benefit of Power. The Power10 processor excels at multi-threading and vectorization, critical for enterprise AI workloads and applications such as OpenXAI. This is a major advantage for not just the OpenXAI solution, but the entire OpenTech application portfolio as well because it allows for better processing of background tasks and resource management of the systems. The reliability, uptime, and security capabilities of Power are also all critical for OpenTech serving client needs.

OpenTech is deploying the OpenXAI solution running on IBM Power10 processor-based servers for a variety of industries and use cases, such as personalized education, real-time translation and transcription, smart home technologies, and more. With OpenXAI running on-prem on Power, OpenTech can ensure security for sensitive client information, allowing users to take advantage of the possibilities and productivity improvements of generative AI while keeping the confidential and proprietary data that the chatbot is trained on private. With Power, OpenXAI can scale according to client needs and more easily be customized for a specific client or use case.

© Copyright IBM Corporation
2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
August 2024

IBM, the IBM logo, and Power are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on ibm.com/trademark.

Red Hat and OpenShift are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Statement of Good Security Practices: No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

