

Knowledge Management for a synthetic diamond manufacturer

In manufacturing, a fractional increase in efficiency or quality can scale into a decisive competitive edge. Data and scientific research hold the key to these gains; the challenge lies in turning that knowledge into something accessible and actionable.

Client context

The client is a global leader in synthetic diamond and super materials manufacturing, providing abrasives, tools and parts in industries from optics to aerospace. Decades of innovation have produced thousands of research documents dating back to the 1940s, but this knowledge was effectively inaccessible due to the time and effort required to comb it for specific insights. They needed a solution to get value out of this rich mine of research while minimizing the time required to do so.

The solution

The solution, developed in late 2024, was a Knowledge Management tool, consisting of a web chat application using a Retrieval Augmented Generation (RAG) model which users could chat with to answer questions using the Research and Development (R&D) repository as its knowledge base. Users could ask follow-up questions to enrich their understanding of research topics while giving them confidence that the information is grounded in verified research documents.

The back end of the solution would run entirely on the company's data platform in Databricks, using

Mosaic AI tools to build and serve the model to the front-end application.

Solution architecture

The architecture of the Knowledge Management tool includes an automated ingestion pipeline. This regularly scans the R&D SharePoint repository for new and updated documents, ingesting them into the Databricks Lakehouse before extracting and formatting the text and tables and storing them in a Databricks Vector Search Index for semantic retrieval by the model.

The RAG model itself was built using LangChain and Llama 3.3 70b and hosted on Mosaic AI Foundation Model Serving. A user's query and chat history are passed to the model as an input, with the chat history used to contextualize the query so that it can be understood as a standalone question. The rephrased question and any hard filters, such as the date of publication, are passed to the document retriever. This queries the Vector Search Index to find relevant documents for the RAG model to use as the basis of its response to the user. The context provided by the retriever, and the original user question and chat history, are passed to the model, which generates the final response to the user's question.



A single data platform

One of Thorogood's operating principles is to use the right technology for every business case. The team used Databricks and Mosaic AI to build, test, and deploy the solution within one secure, unified environment. Data storage, vector indexing, model development, custom model serving, and evaluation remained inside the Databricks platform, avoiding the complexity of stitching together external services for hosting, security, or retrieval infrastructure. This centralization significantly reduced development time and made governance, privacy, and scalability far easier to manage.

Mosaic AI added semantic vector search, as well as foundation and custom model serving, all running natively within the platform. This enabled fast experimentation through tools like MLflow, easy swapping between supported models, and strict control over sensitive R&D documents without sending data to third-party providers. The result was a streamlined development experience that accelerated delivery of the project while ensuring enterprise-grade security and flexibility.

Adoption, the right way

Any tool is only effective if people find that it allows them to do something they couldn't do before, or if it increases the efficiency of an existing process. Crucially, the tool must provide useful and correct results that users trust.

The Thorogood team developed the tool in an iterative manner to address these objectives. By getting to a working prototype of the Knowledge Management tool as early as possible and involving users in feedback sessions, the team was able to identify problems with outputs early in the development process to improve responses and build user trust.



Find out more:

Contact: ethan.hinton@thorogood.com
Ethan Hinton, Data & AI Consultant

The results

"The tool dramatically reduced the time required to find and connect relevant insights, while increasing confidence through traceable, source-backed answers drawn directly from our research corpus. Overall, the project transformed how scientific knowledge is accessed and reused, turning decades of archived expertise into a living, decision-ready asset for the organization."

Head of Data & AI

The project was widely adopted by the R&D team, and users appreciated the natural language interface and the veracity of the generated responses. It proved useful not only for this department, but potentially for employees in other arms of the business who needed to ask questions and leverage the repository quickly. The Knowledge Management tool unlocked decades of research for this manufacturer, allowing them to make use of a store of knowledge and insight that was previously inaccessible.