GenAI Architecture Case Study

Student Guide

This is the student guide for the advanced ILT workshop: **Generative AI on Vertex AI Bootcamp** to be delivered in 24H1.

Customer profile

- We are a large component manufacturer for compute and network infrastructure products.
- We provide our business partners access to technical information and training at no cost via a private web portal we call the Partner University.
- Increasing the utilization of the Partner University resources is a key performance indicator for our internal partner-enablement staff.

Current state

- Content libraries
    - Our content is divided into two separately indexed repositories: one for technical documentation, the other for educational resources.
    - The technical documentation repository contains hundreds of PDFs, all indexed by the full text of the document.
    - The educational resources are a composition of videos, PowerPoint presentations, Word documents, Excel spreadsheets, and PDFs. These resources are indexed by the title, summary, and the keyword metadata stored in a database.
    - The indices are separate and must be searched independently.
- Content ingestion
    - Between 1-10 new resources are added to the university every month. These resources may include a set of videos and presentations, or they may be a bundle of PDFs and other documents. Occasionally existing content is removed due to age/relevance, or replaced due to errors/omissions.
    - All new educational resources are reviewed by a human, who assigns metadata tags in the form of keywords, category, and a summary or description of the content. The metadata and object URLs are stored in a database indexed by the search engine.
    - The content is then posted in a hierarchy of categories available for the users to browse.
- User experience
    - The users have several ways by which they can discover and view content:
        - Browse the catalog hierarchy by topic/category.
        - View a list of all content, filterable by selecting tags.

        - Search by keyword
        - The hosting engine provides recommendations based on the user's history and on the most popular videos. The recommendations are provided by our video hosting platform and are not connected to the keyword metadata.
    - On a monthly basis, the content curators send the users email newsletters containing announcements and links to new content..
- Access control
    - All users of the university are authenticated and all resource access is tracked to provide consumption metrics for the university and a history for the user.
    - User access to the university is contingent upon them signing a Non-Disclosure Agreement (NDA).

Current challenges

- Content ingestion
    - Labor-intensive and error prone Content ingestion process
        - The human labeling process is a bottleneck in the content publishing process
        - The labelers occasionally mislabel content, rendering it as difficult to find in the University as a misshelved book in a library. No safety net exists to catch labeling mistakes.
        - The product managers would like to publish more content, and at a faster pace, but the human labeling process has become a significant bottleneck.
- Content discovery
    - Content discovery is hampered by the reliance on keyword search and a loosely structured set of metadata tags.
        - Users report having a difficult time discovering new content that is relevant to their needs.
        - Users report that searching for content is often frustrating due to the seemingly random relevance of the search results.
        - The search engine does not index the contents of the videos. It is based on the title and keywords contained in the metadata. This causes many false negative search results.
    - Categories are not helpful
        - Despite multiple recategorization efforts, users still report that the hierarchical categories are not intuitive and limit their ability to find content more than improve it.
        - This is partly because the content is often relevant to multiple categories, yet the content can only be displayed in one location in the category hierarchy.
    - Lack of cross referencing between pieces of content
        - Users cannot simultaneously search for technical documentation and educational content.
        - The pages hosting technical documents have no links or references to relevant educational content, and vice versa.
- Lower than anticipated viewership/engagement
    - We have increased user enrollments, but overall content consumption has not increased
    - Based on the content usage history and user feedback, we believe users aren't finding the content they need.
    - We also track viewer engagement by video.

        - A percentage calculated per user for each video viewed.
        - Number of minutes watched divided by the video duration.
        - We believe that low percentages indicate that the content was not relevant to the user.
    - For select partners, we provide curated custom learning paths containing training videos and technical resources.
        - Static (out of date) Learner experience. New courses and docs may be left out of learning paths; Old courses may not be removed.
        - Content selection is subjective and prone to human error.

Desired outcomes

- Minimize human involvement in the content ingestion process.
- Increased user satisfaction with regard to the searchability and discoverability of content on our site.
- Increased content viewership and engagement (user retention).

Project success metrics

- Minimize human involvement in the content ingestion process.
    - Number of labor hours incurred indexing content or correcting resource metadata.
    - Number of labor hours incurred maintaining the proposed system.
- Increased user satisfaction with regard to the accessibility (discoverability) of content on our site.
    - Gather user feedback while conducting A/B testing of the UI and search capabilities
    - Compare metrics from before and after the project implementation.
- Increased content viewership and engagement (user retention).
    - Quantify dormant learner reengagement by separately reporting the activity of newly enrolled users versus existing users.
    - Collect statistics to report on the rolling 30-day average:
        - New user enrollments,
        - number of visits per user
        - number of unique user visits daily
        - Video engagement time (by video); minutes watched.