



Empowering Individuals with Accurate and Up-to-Date Information

Snowflake- The RAG powered diabetes chatbot



Overview

Diabetes can be tough to live with. Individuals face numerous challenges in managing their condition effectively, including staying informed about the latest research, treatment options, and lifestyle recommendations.

Just the overwhelming amount of information online can stump anyone. Not to mention all the false product claims, misinformation and outdated advice prevalent.

Meet Snowflake, the diabetes chatbot grounded in facts and state-of-the-art research on diabetes, here to help you navigate diabetes.



Meet the Team



SpaceCats

Darrius

Jensen

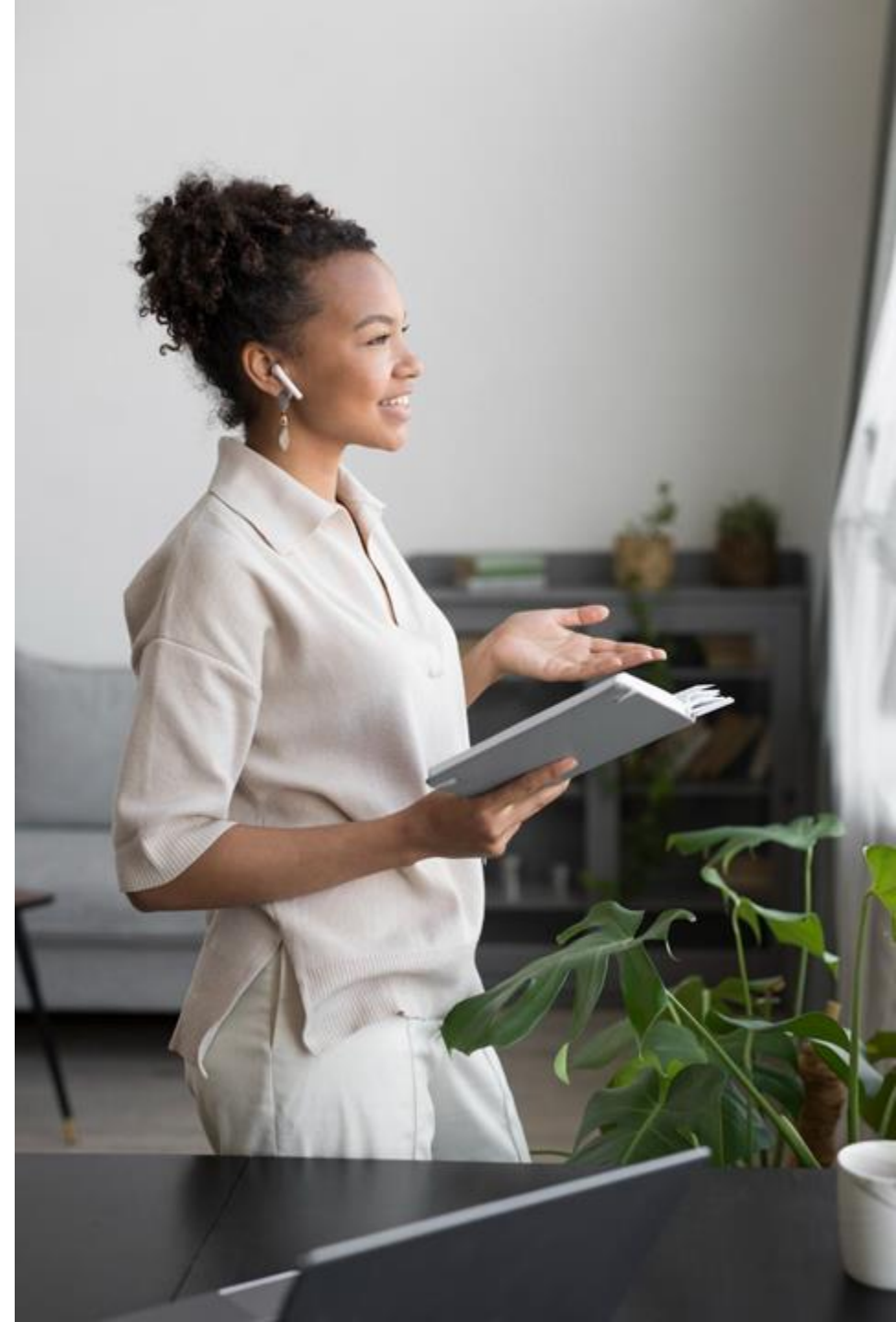
Chloe

Limitations of current RAG

Retrieval Augmented Generation (RAG) are resistant to hallucination. However, majority of current RAG products are based on keyword search, and are thus unable to capture intricacies in semantics within sentences.

Our Innovation:

- Combining Semantic Search with data structures such as vector databases, keyword databases and summary indexes to augment the retrieval process in our RAG pipeline





01

Ingestion

1

Vectara API is used to integrate Semantic Search



- *Boomerang LLM* – Uses NLP to capture semantics during chunking
- *VectaraIndex* - Strong Semantic Search Capabilities

2

Additional Data Structures

U N S T
R U C T
U R E D

Unstructured.io

Document structure aware chunking is used to effectively parse scientific papers.



LangChain

Semantic Chunking

together.ai

Using Together.ai, the Mixtral 8x7B LLM is integrated into the ingestion process, generating summaries and keywords to enhance retrieval process.

The lightweight Mxbai Embed Model is used to generate embeddings.

Vector Database

Keyword Database

Summary Index

The ingestion process produces a comprehensive selection of data structures to augment semantic search.

02 Retrieval

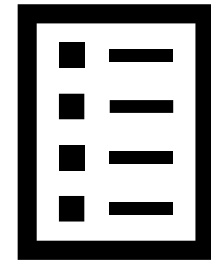


LlamaIndex's *FusionQueryRetriever* integrates the retrievers and indices.

together.ai

LLM generates extra queries based on the original query.

Creates a robust set of retrieved nodes.



Top k ranking to select the most relevant nodes

03 Generation



Retrieved nodes are reranked using the SBERT model, prioritizing relevant nodes.

together.ai

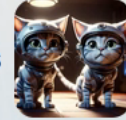
Selected nodes are used with Mixtral 8x7B LLM to conduct Q&A.



The query process is integrated into the Snowflake chatbot, inspired by Llamaindex's ChatBot

04 Demo

Diabetes Chatbot | Built by Space Cats



Based on the latest research, list the current diabetes risk factors that young adults are facing

Send message

05 Performance

Snowflake Chat

- Outperforms Vectara
- Robust, integrates multiple retrievers and databases
- Hallucination resistant

Vectara

The screenshot displays the Vectara user interface. On the left, a sidebar contains a 'Choose app type' dropdown and a list of features: Chat, Retrieval, Generation, Chat, Evaluation, Filters, Show API request, and Hybrid search. The main area shows a 'Query' input with the text 'What is diabetes?'. Below the query, a 'Summary' box contains the text 'I do not have enough information to answer this question.', which is highlighted with a red border. Underneath the summary, the 'Factual Consistency Score' is 0.12, and there is a 'What's this?' link. The 'Retrieved results' section lists two items. Item 1 is 'What is gestational diabetes?' with a score of 1.0838, followed by a snippet of text from a clinical acta and a 'lang: eng' tag. Item 2 is 'Lower Your Risk The good news is you can manage your risk for type 2 diabetes.' with a score of 1.0256, followed by a snippet of text from an adapted article and a 'lang: eng' tag. Both items have 'Filter on ID' and 'Details' buttons.

Choose app type

Query

What is diabetes?

Summary

I do not have enough information to answer this question.

Factual Consistency Score: 0.12 [What's this?](#)

Retrieved results

1 What is gestational diabetes? [Filter on ID](#) [Details](#)

Score: 1.0838

Clin Chim Acta 2015;445:107–114 200. Buchanan TA, Xiang A, Kjos SL, Watanabe R. What is gestational diabetes? Diabetes Care 2007;30(Suppl. Noctor E, Crowe C, Carmody LA, et al.; ATLANTIC-DIP Investigators.

lang: eng

Doc ID: Diagnosis and Classification of Diabetes Standards of Care in Diabetes—2024.pdf

2 Lower Your Risk The good news is you can manage your risk for type 2 diabetes. [Filter on ID](#) [Details](#)

Score: 1.0256

...h less than the amount in the left column: 0 points Adapted from Bang et al., Ann Intern Med 151:775–783, 2009 □ Original algorithm was validated without gestational diabetes as part of the model. Lower Your Risk The good news is you can manage your risk for type 2 diabetes. Small steps make a big difference in helping you live a longer, healthier life. If you are at high risk, your first step is to visit your doctor to see if additional testing is needed.

lang: eng

Doc ID: Diagnosis and Classification of Diabetes Standards of Care in Diabetes—2024.pdf

Thank you!

