**Problem Statement** - Getting information from the file that contains research data of user/item interactions, star ratings, timestamps, product reviews, social networks, item-to-item relationships (e.g. co purchases, compatibility),product images, price, brand, and category information, GPS data, heart-rate sequences, other metadata.

**Solution Strategy** - Build a POC which should solve the following requirements:

- Users would responses from the researched data

- If they want to refer to the original data from which the bot is responding, the bot should provide a citation as well.

**Goal** - Solving the above two requirements well in the POC would ensure that the accuracy of the overall model is good and therefore further improvisations and customizations make sense.

**Data Used** - Recommendation datasets stored in one pdf file

**Tools used** - LlamaIndex (only for now) has been used due to its powerful query engine, fast data processing using data loaders and directory readers as well as easier and faster implementation using fewer lines of code.

The following are the steps in flowchart format:

Relevant Nodes

**Query**

**Engine**

**Retriever**

**Index**

**Document**s

**Final Response**

View over data and metadata