1. **Project Background:**

We will be building a **Semantic Spotter AI** project related to the uber 10-K financial and insurance industry forms. An AI-powered project to extract information from these documents can revolutionize how policies are analyzed and understood. Using natural language processing (NLP) / LlamaIndex framework, this system can efficiently scan, interpret, and extract key details such as coverage limits, exclusions, claim procedures, and premium details from the policy documents

1. **Solution Strategy:**

Build a project which should solve the following requirements:

- Users would get responses from the financial statement & the Life Insurance policy statements

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

1. **Project Goal:**

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

1. **Problem Statement:**

Provided with Uber's 2021 financial statement Form 10-K & HDFC Bank Life Insurance policy statements. Our goal here is to build a RAG application on the annual financial statement pdf document.

1. **Tools Used:**

*LlamaIndex* 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. Also, if we want to process heavy documents for RAG application then *LlamaIndex* provides various custom settings such as override chunk sizes, metadata filters for exact search, prompt engineering & overriding embedding model

1. **Design choice:**

Please refer to the document attached, ‘**SystemDesign\_Sahil\_Avasthi.pptx’**

1. **Challenges faced:** Segmentation of the document and creation of a coherent response across different documents