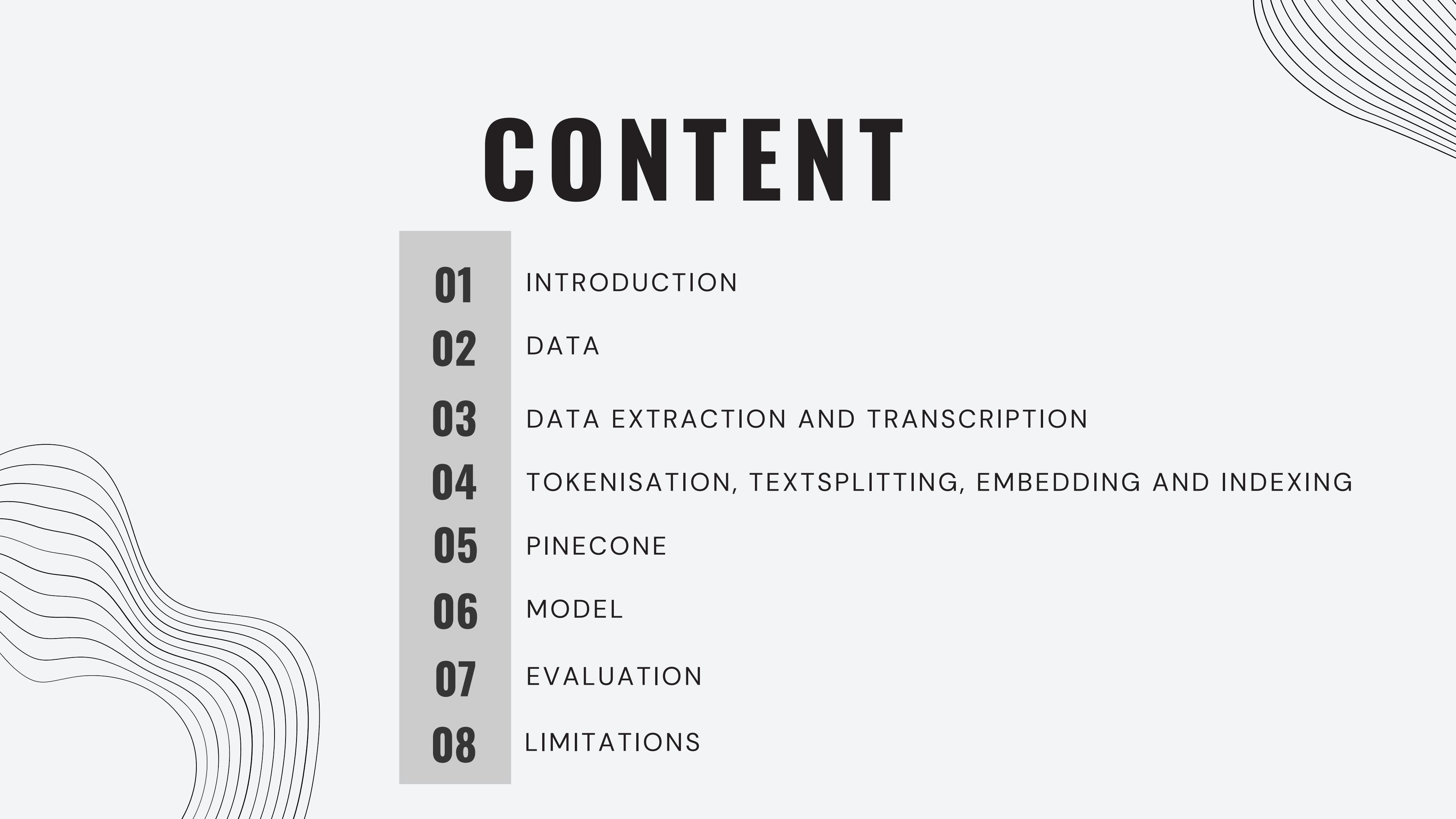




LOZ RIGBY

PROJECT III CHATBOT

CONTENT

- 
- 01** INTRODUCTION
 - 02** DATA
 - 03** DATA EXTRACTION AND TRANSCRIPTION
 - 04** TOKENISATION, TEXTSPLITTING, EMBEDDING AND INDEXING
 - 05** PINECONE
 - 06** MODEL
 - 07** EVALUATION
 - 08** LIMITATIONS

INTRODUCTION

The goal of this project is to develop an AI chatbot that combines the power of text and audio processing to answer questions about specific videos.

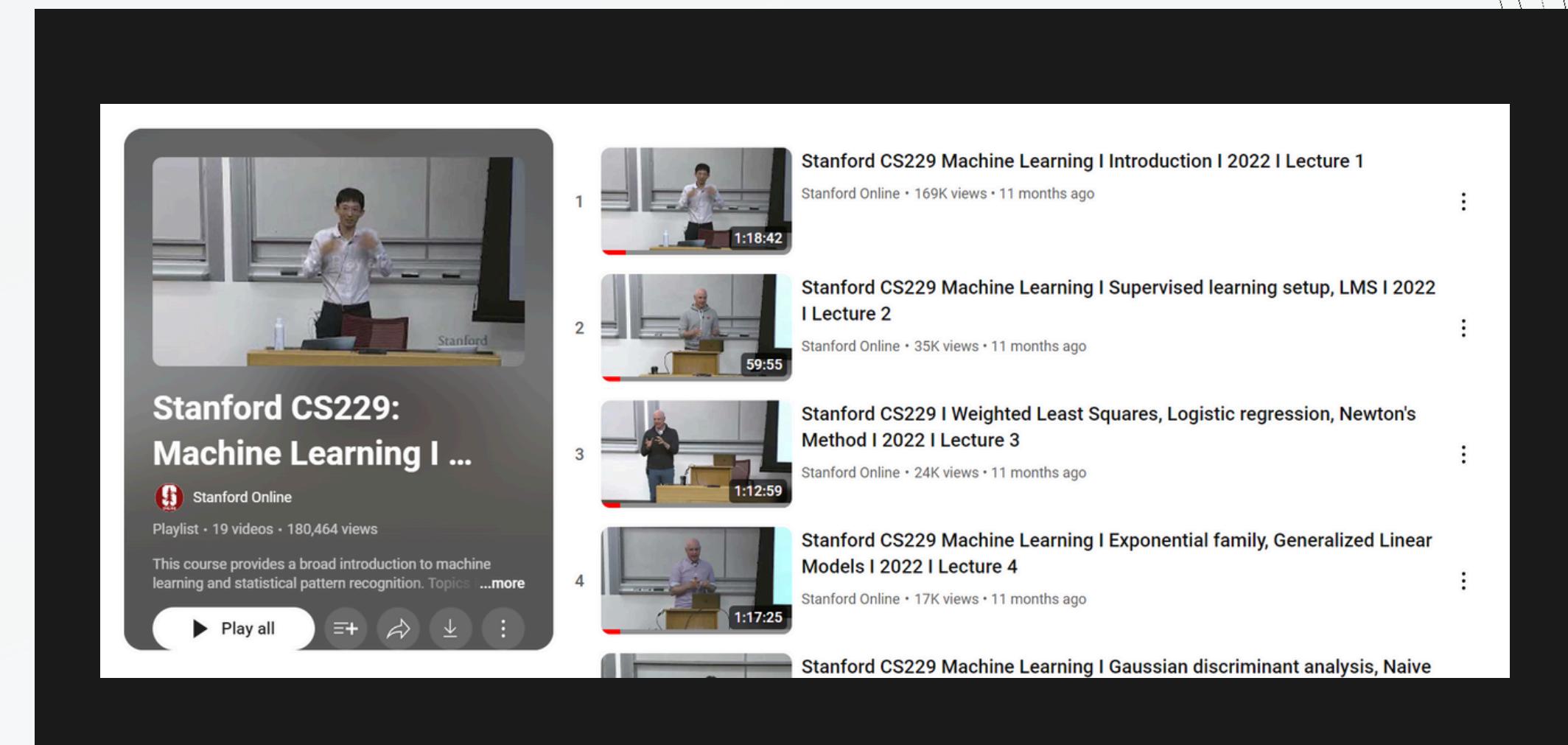
The OBJECTIVE was for students or anyone seeking to learn from the Machine Learning module at Stanford University to have a chat bot that can be used for summarisation and question answering



DATA

The data was selected and collected from the Stanford Online youtube channel where they have uploaded their lectures from their Machine Learning module.

There are nineteen videos, cumulating over 26 hours of footage.



DATA EXTRACTION AND TRANSCRIPTION

Downloading

Audio files were downloaded from individually from youtube

Audio chunking

In order for efficient and manageable downloads so they can be processed in smaller segments

Whisper

The audio chunks are then transcribed using the Whisper Model to create full transcripts for each video

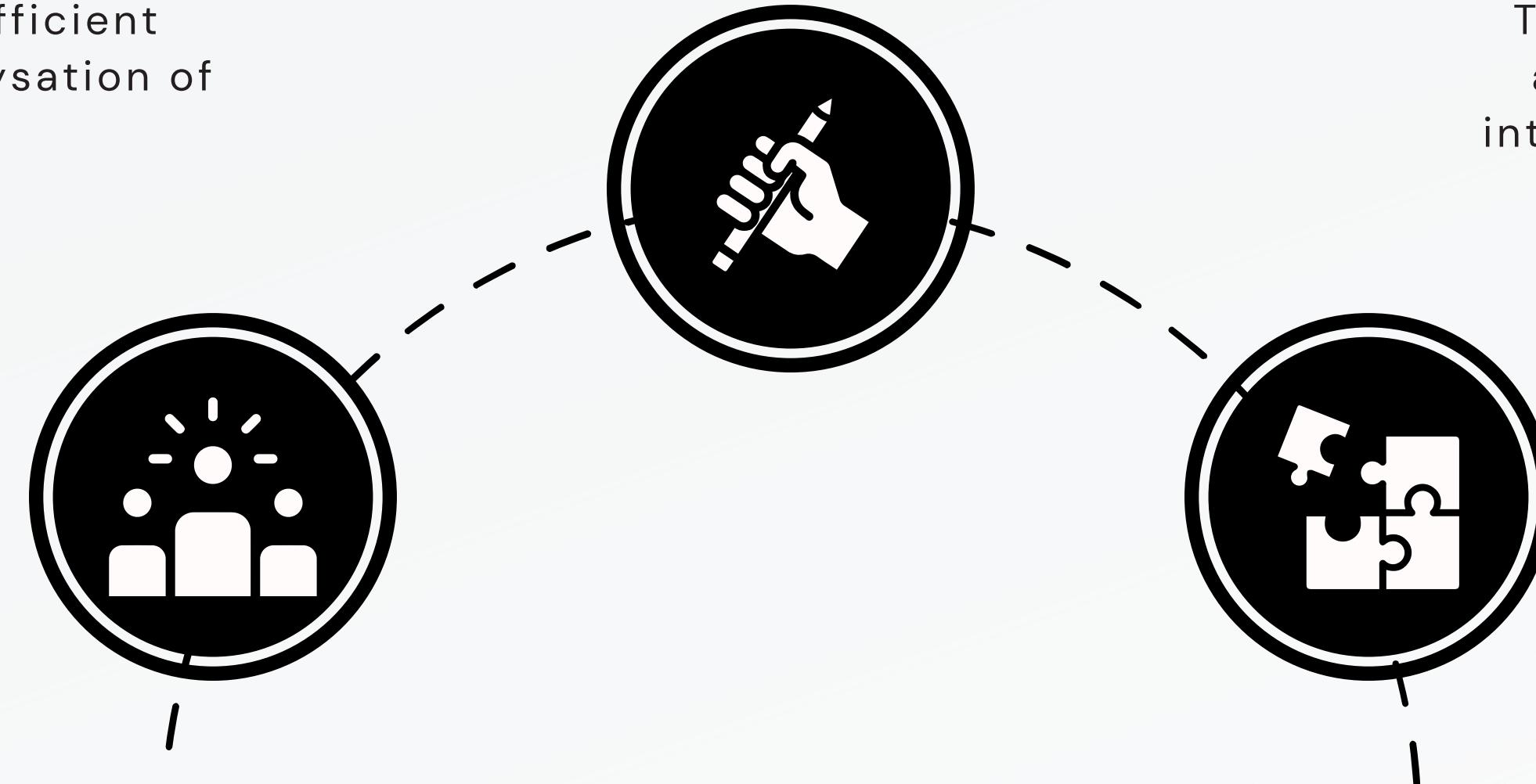
```
data[1]

{'id': 1,
'url': 'https://www.youtube.com/watch?v=ZMxfDWPXmjc',
'title': 'Stanford Cs229 Machine Learning I Neural Networks 1 I 2022 I Lecture 8.Mp3.Txt',
'transcript': "Hello everybody. Hi. My name is Masha. Some of you may have met me already as part of office hours and seen me post on Ed and things like that. I'm really excited to be giving lectures today. It's going to be in a slightly different format than ten use or criss-os. So feel free to give me"}
```

TOKENISATION, TEXTSPLITTING, EMBEDDING AND INDEXING

Objectives^{° 1}

The text is tokenised used tailored for GPT-3.5turbo model.
It is then split into manageable chunks for more efficient processing and analysation of long texts.

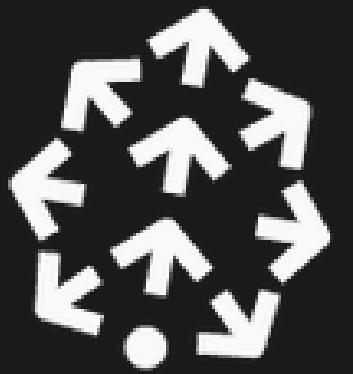


Objectives^{° 2}

The model is embedded, which are numerical representations of text to capture meaning. Essential for language processing tasks.

Objectives^{° 3}

A pinecone index is initialised, which is where we will store our data.
The text embeddings are batch uploaded into the Pinecone index



Pinecone

The VectorStore
Pinecone was initialised
to facilitate efficient
storage, retrieval, and
querying of text
embeddings.

transcription-index ●

METRIC DIMENSIONS
dotproduct 1536

HOST
<https://transcription-index-viq75a9.svc.aped-4627-b74a.pinecone.io> ⓘ

CLOUD REGION TYPE
aws AWS us-east-1 ⓘ Serverless

VECTOR COUNT
754

Note:

no external knowledge agent added as queries want to be specifically answered using the vectorstore data

MODEL



The LLM uses Chat-GPT-3.5-turbo with a temperature of zero, to ensure deterministic responses. This model will be used to generate the queries.
A retriever is used to access our created vectorstore

CHATGPT-3.5-TURBO



Conversational Buffer Memory is created to manage memory, allowing context to remain on track.

MEMORY



ReAct agent is utilised to handle conversation queries and provide descriptive responses as well as combine the tools and memory into the LLM

REACT AGNET

EVALUTAION

Query: What is an example of using AI systems in healthcare for heart conditions?

Model answer: One example of using AI systems in healthcare for heart conditions is the development of a computer vision system that analyzes ultrasound videos of the human heart to assess different cardiac conditions in patients. This system was developed at Stanford and has been deployed in healthcare settings to help healthcare professionals in diagnosing and monitoring heart conditions.

Correctness

The submission accurately describes the use of AI systems in healthcare for heart conditions, specifically mentioning the development of a computer vision system that analyzes ultrasound videos of the human heart to assess different cardiac conditions in patients. The submission also correctly states that this system was developed at Stanford and has been deployed in healthcare settings. The information provided in the submission aligns with the reference provided. Therefore, the submission meets the criteria.', 'value': 'Y', 'score': 1

Conciseness

The submission provides a clear and concise example of using AI systems in healthcare for heart conditions. It directly mentions the development of a computer vision system that analyzes ultrasound videos of the human heart to assess different cardiac conditions in patients. The information is presented in a straightforward manner without unnecessary details. Therefore, the submission meets the criteria for conciseness. Y', 'value': 'Y', 'score': 1

Coherence

The submission is coherent as it clearly explains the example of using AI systems in healthcare for heart conditions. The submission is well-structured, starting with the example and providing details about the development and deployment of the system. The submission is organized, with a clear flow of information from the introduction of the example to its application in healthcare settings. Therefore, the submission meets the criteria.', 'value': 'Y', 'score': 1

EVALUTAION (NOT SO GOOD)

Query: What is the type of generative learning algorithms discussed?

Model answer: The type of generative learning algorithms discussed are Gaussian discriminative analysis.

Correctness

Correctness: The submission states that the type of generative learning algorithms discussed are Gaussian discriminative analysis. However, the correct type should be Gaussian discriminant analysis, not Gaussian discriminative analysis. This is an incorrect statement. Therefore, the submission does not meet the criteria of correctness.', 'value': 'N', 'score': 0

Conciseness

The submission directly answers the question without any unnecessary information. The submission is clear and to the point, mentioning Gaussian discriminative analysis as the type of generative learning algorithms discussed. The submission is concise and does not include any extra details that are not relevant to the question. Therefore, based on the criteria of conciseness, the submission meets the criteria. Y', 'value': 'Y', 'score': 1

Coherence

The submission states that the type of generative learning algorithms discussed are Gaussian discriminative analysis. However, the correct term should be Gaussian Discriminant Analysis (GDA) as per the reference provided. The submission is clear and concise in its response, directly answering the question asked. The submission is not well-structured as it contains a minor error in the term used. Therefore, based on the coherence criterion, the submission does not fully meet the criteria. 'value': 'N', 'score': 0

LIMITATIONS

Time

Time limits played a big factor on how much the outcome of the project

Personal doubts

Personal doubts at the beginning made progress slow, especially when combined with time constraints

Complexity

The project complexity means that there many separate areas which can be improved



THANK YOU!

Any questions?

Example of how youtube URL was used in deployment

The screenshot shows a user interface for transcribing a video. At the top, there is a text input field labeled "Enter YouTube URL" containing the URL <https://www.youtube.com/watch?v=CT0aB-EORMU>. Below the URL, a message says "Downloading and transcribing video...". Once the transcription is complete, a message indicates "Transcript saved to files/transcripts\How your General Practice team can help you | NHS.". A text input field for asking questions is present with the placeholder "Ask a question about the transcribed video!". A specific question is typed: "what is a general practice team ?". The "Answer:" section displays a JSON response:

```
▼ { ↴
  "query" : "what is a general practice team ?"
  "result" :
  "A general practice team typically refers to a group of healthcare professionals working together in a primary care setting, such as a doctor's office or clinic. This team may include doctors, nurses, nurse practitioners, physician assistants, medical assistants, and administrative staff who collaborate to provide comprehensive and coordinated care to patients."
  ↴
}
```

At the bottom, a message states "Temporary Pinecone index 'youtube' deleted."