



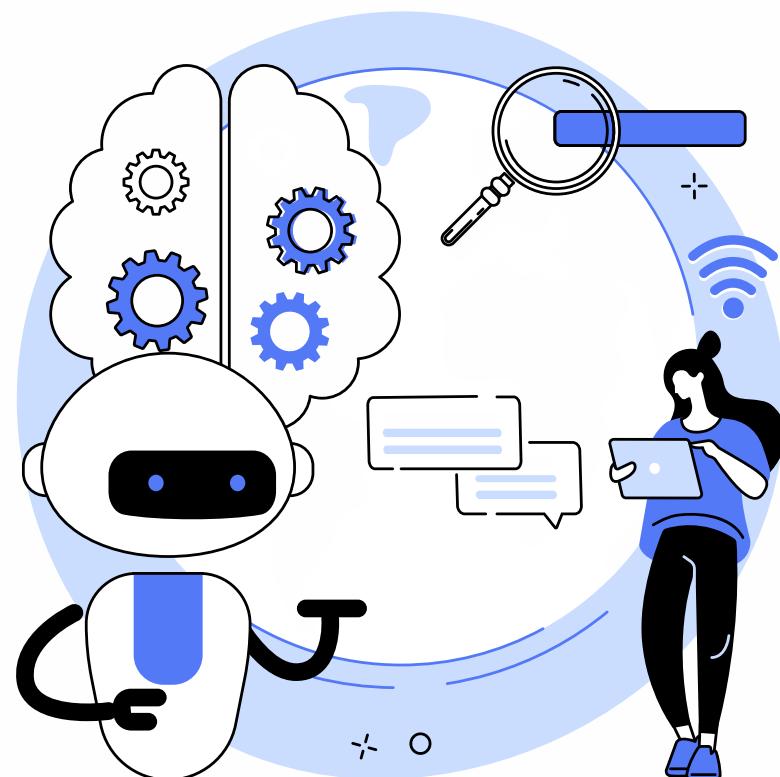
MULTI-MEDIA CUSTOMER SUPPORT CHATBOT FOR AI TESTING TOOL WITH BUILT-IN TESTABILITY

DATA 298A Project Presentation
Professor: Dr. Zeyu Jerry Gao

Group 3
Neha Shaikh, Supriya Kamble,
Praveen Kumar, Prasad Kalangi,
Quan Gu

Agenda

- 1** Project Introduction
- 2** Project Background
- 3** Project Requirements
- 4** Team Organization and Function Roles
- 5** Literature & Technical Review
- 6** Project Resources
- 7** Data Collection & Preparation
- 8** Model Selection
- 9** Model Implementation
- 10** Model Comparision
- 11** Future Work
- 12** Demos



Project Introduction

- Chatbots or virtual conversational agents are designed to mimic human conversation.
- Chatbot provides 24x7 enhanced customer support, faster responses to queries at any time, reduction in manual efforts and operational costs.
- AI chatbots are developed to interact in natural languages, information retrieval, problem-solving capabilities and virtual assistance.

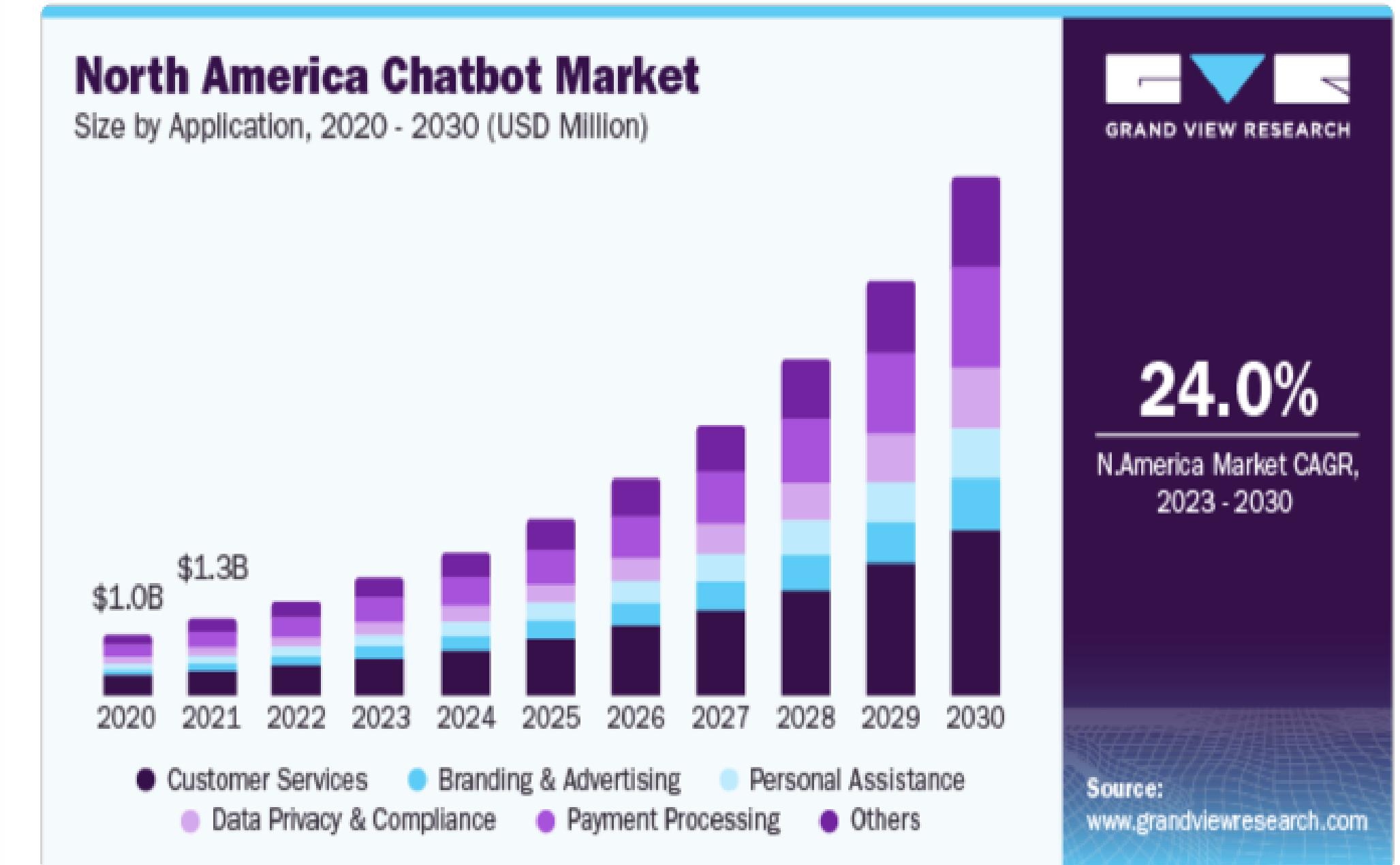


Project Background

Features	Traditional Chatbot	AI chatbot
Technology	Rule-based or Program-based	Powered by AI, ML/DL algorithms
Models Used	No models	NLP, ML and DL models
Chat Pattern and dialog flow	Responses based on pre-defined text and Q&A	Support diverse chat patterns and can generate human-like text
Self-improvement	No learning from user interactions	Continuously learns from conversations and improve their responses over time
Input Support	Only text	Yes (text, images, audio etc.)
Natural language understanding	Rely on pre-programmed scripts	Understands context, intent and sentiment based on the conversations
Cost	Lower costs	Build cost is high but offer long-term cost savings due to their capabilities

Market Analysis

- The global chatbot market was estimated at \$5,132.8 million in 2022 and is expected to increase at a compound annual growth rate (CAGR) of 23.3% from 2023 to 2030.
- In 2022, the market's largest sector by revenue was customer services, with a share of 31.2%.
- With AI chatbots, businesses are expected to save 2.5 billion customer service hours and \$11 billion annually on customer service by the end of 2023.



Problem Statement

Users of the AI Testing Tool face challenges while using the tool due to lack of knowledge or available resources.

Motivation and Goals

- To build a chatbot for AI Testing Tool capable of communicating with the users using diverse inputs such as natural language, image, audio etc.
- Make the platform user-friendly and accessible to everyone.
- **Functions** - Responding to questions, interaction, image and speech recognition etc.



Project Requirements

- **Functional Requirements**

Users should be able to communicate with the chatbot using text, image and audio inputs.

- **AI -Powered Requirements**

Be able to recall previous conversations and modify replies in response to changes in the user's intent.

- **Data Requirements**

Support diverse data inputs such as -

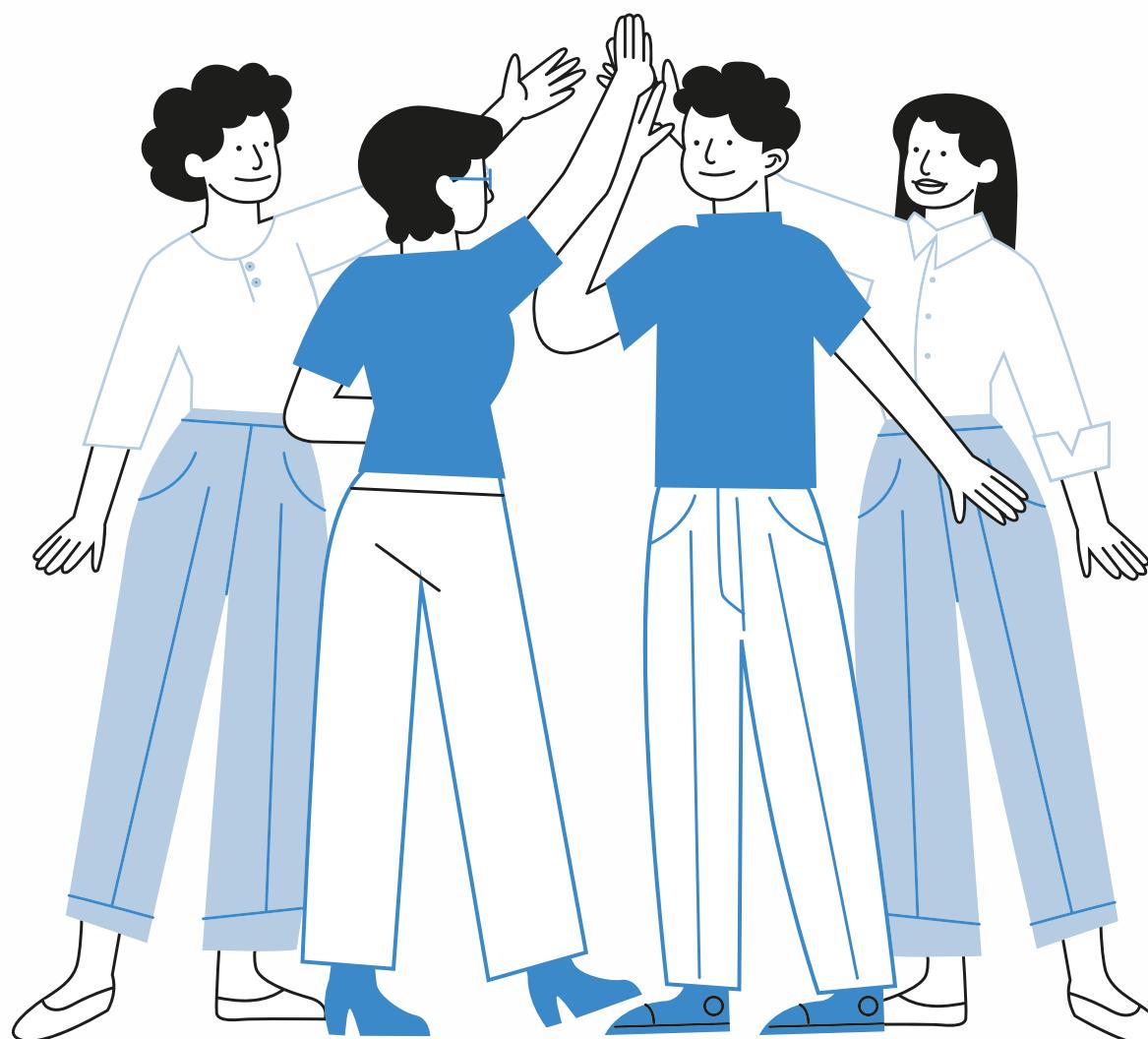
1. Text Data - User queries given in natural language.
2. Image Data - Different objects and texts in images.
3. Audio Data - Voice input

- **Non-Functional Requirements**

Ensure a reliable and seamless user experience.

Team Organization and Function Roles

Task Name	Start Date	End Date	Member	Status
Project Understanding	27-Aug	16-Sep		COMPLETE
Project Background	27-Aug	8-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Defining Problem statement	27-Aug	8-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Literature Review	27-Aug	16-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Technical Review	27-Aug	16-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Defining Project Scope	27-Aug	8-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Project Requirements	3-Sep	16-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Data Understanding	17-Sep	21-Oct		COMPLETE
Data Research and Collection	17-Sep	29-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Data Generation	17-Sep	29-Sep	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Data Preprocessing	8-Oct	13-Oct	Neha, Supriya	COMPLETE
Data Transformation	8-Oct	13-Oct	Neha, Supriya	COMPLETE
Data Preparation	8-Oct	13-Oct	Praveen, Prasad	COMPLETE
Data Storage Management	17-Sep	21-Oct	Praveen, Quan	COMPLETE
Model Development	22-Oct	16-Dec		COMPLETE
Model Selection	22-Oct	29-Oct	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Model Architecture	22-Oct	29-Oct	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Model Implementation	30-Oct	15-Nov	Neha, Supriya, Praveen, Prasad, Quan	COMPLETE
Text-based Model	30-Oct	15-Nov	Neha, Prasad, Quan	COMPLETE
Image-based Model	30-Oct	15-Nov	Praveen, Supriya	COMPLETE
Audio-based Model	30-Oct	15-Nov	Neha	COMPLETE
Knowledge Graph	30-Oct	18-Nov	Neha, Supriya, Praveen, Prasad, Quan	IN PROGRESS
Model Integration	18-Nov	16-Dec	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Model Evaluation	30-Jan	20-Feb		TO DO
Define Evaluation Metric	30-Jan	20-Feb	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Model Testing	30-Jan	20-Feb	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Fine Tuning	30-Jan	20-Feb	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Deployment	21-Feb	30-Apr		TO DO
Application Development	21-Feb	10-Mar	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Model Deployment	11-Mar	5-Apr	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Performance Analysis	6-Apr	15-Apr	Neha, Supriya, Praveen, Prasad, Quan	TO DO
Final Report and Presentation	16-Apr	30-Apr	Neha, Supriya, Praveen, Prasad, Quan	TO DO



Literature Review: Advanced Models in Chatbot

Classification	Models	Description
Cognitive Model	DQN, QR-DQN, BiRNN, NMT, Q-Learning	To overcome the challenges of context understanding with chatbots, three deep reinforcement approaches - Q-Learning, DQN (Deep Q Neural Network) and QR-DQN (Distributed Reinforcement Learning with Quantile Regression) are used along with BiRNN and NMT on CoQA and Cornell Movie datasets. The deep reinforcement models help the chatbots to learn from its errors thereby reducing the chances of providing irrelevant responses. BiRNN is utilized to provide the chatbots with previous and next states so that they can get the required information for answering the questions.
	LSTM, BERT, GAN	The research focuses on generating meaningful and relevant responses from chatbots while also making them sound natural by developing a chatbot using deep reinforcement learning and adversarial learning with a pre-trained BERT model. The models were evaluated using BERT, ROGUE and BLUERT scores which showed significant performance over the baseline models.
Knowledge Model	LSTM, CNN, GRU	The paper focuses on studying deep learning models, natural language generation, and evaluation of generated results. For implementation, LSTM and CNN are initially designed and trained on Customer Support on Twitter dataset. A new GRU model is also developed for comparing the results from CNN and LSTM. Seq2Seq learning is applied to map the correct responses to customer queries. BLEU score and cosine similarity are used for evaluation of the models. The paper showcases two results - sentence organization and generating relevant responses.
	BiRNN	This paper presents a study to improve the perplexity and learning rate of the model. The architecture of the chatbot includes a BiRNN with attention model which works for remembering long tokens at a time to understand the context of sentences. The model is trained on Reddit data to generate responses. The model achieved a BLUE score of 21.67 and perplexity of 56.10 for learning rate value 0.0001.

Technical Review: State-of-art Chatbot Technologies

Chatbot	Platform	Base Model	Type	Application
Speech-to-text chatbot	Google API	RNN, HMM, Bayesian network	Rule based + ML based + Knowledge Based	Customer service, automatic speech recognition, transcribe multimedia content
IBM Watson	IBM	NLP, REST API call, ML, Integrated dialog tools	Retrieval based	Healthcare, Finance, Legal, Retail, Insurance, Sports and Entertainment
GPT-2	Open AI	Open AI GPT (Transformer and NLP)	Generative pretraining and discriminative fine-tuning	Knowledge based customer Q&A platform
DialoGPT	Microsoft	GPT-2	Generative	Response Generation
ALEXA	Amazon	ASR based, NLU, NLG, TTS	Rule Based + Retrieval based + Generative	Voice assistance, automatic speech recognition
GPT-3	Open AI	GPT-2	Generative pretraining and discriminative fine-tuning	Knowledge based customer Q&A platform
SIRI	Apple Inc.	DNN, HMM	Retrieval based + Generative	Voice assistance, automatic speech recognition

Literature Review: Innovative Methods in Chatbot Development

- **Entertainment Chatbot (Sperlì, G. 2020):**
 - Approach: Seq2seq model using an NLP pipeline.
 - Justification: Encoder-Decoder framework with RNN.
 - Limitations: Limited user interactions, lacks online learning strategy.
- **UIED Toolkit (Xie et. al. 2022):**
 - Approach: Hybrid method for GUI object detection.
 - Justification: Coarse-to-fine strategy using computer vision.
 - Limitations: Higher object detection time, no automated GUI testing implementation.
- **Multi-modal Chatbot (Huang et al. 2021):**
 - Approach: Combines NLP and computer vision for intelligent manufacturing.
 - Justification: Enhances worker-machine interaction.
 - Limitations: Privacy and ethical concerns not addressed.
- **Emotion Detection (Kumar et al. 2022):**
 - Approach: Fine-tuning BERT for emotion detection in psychological texts.
 - Justification: Emotion-cause pair extraction for mental health insights.
 - Limitations: Dependent on the pre-trained BERT model's limitations.

Project Resources, Technology, Platform

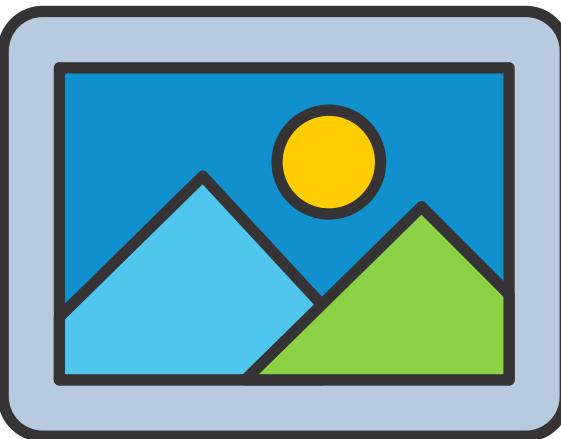
Requirement	Type of Resource	Resource	Duration	Estimated Cost
Model Development	Hardware	Google Cloud Platform Deep Learning VM with GPU(NVIDIA® Tesla® V100)	09/04/2023 - 12/15/2023	\$500
		MacBook Pro with Apple M1 Pro (8 core GPU)	09/04/2023 - 12/15/2023	Personal Laptop
AWS for Data Storage and Web Hosting	Hardware	AWS EC2 instance with 8GB memory	09/04/2023 - 12/15/2023	\$500 (\$1.46/per instance/hour)
Code Developmet	Software	Python 3.7 (Jupyter Notebook), NodeJS	09/04/2023 - 2/15/2024	Free
		HTML, JavaScript	03/01/2024 - 04/15/2024	Free
Database Management	Software	MongoDB	10/21/2023 - 4/15/2024	Free with Student ID
ML and DL Frameworks	Software	Pytorch, TensorFlow, Librosa, NLTK	10/21/2023 - 4/15/2024	Free

DATA PREPARATION PROCESS



Data Requirements

- Text-based Data - To understand and customer queries given in natural language.
- Image-based Data - To identify objects and texts in images to provide relevant responses.
- Audio-based Data - To understand audio inputs and respond



The ground or stays
iverse is vast, and you
; also beautiful. You a
nthing bigger than yo
t of something that ma
most of your time. Tal
e a blog post. Make a
...

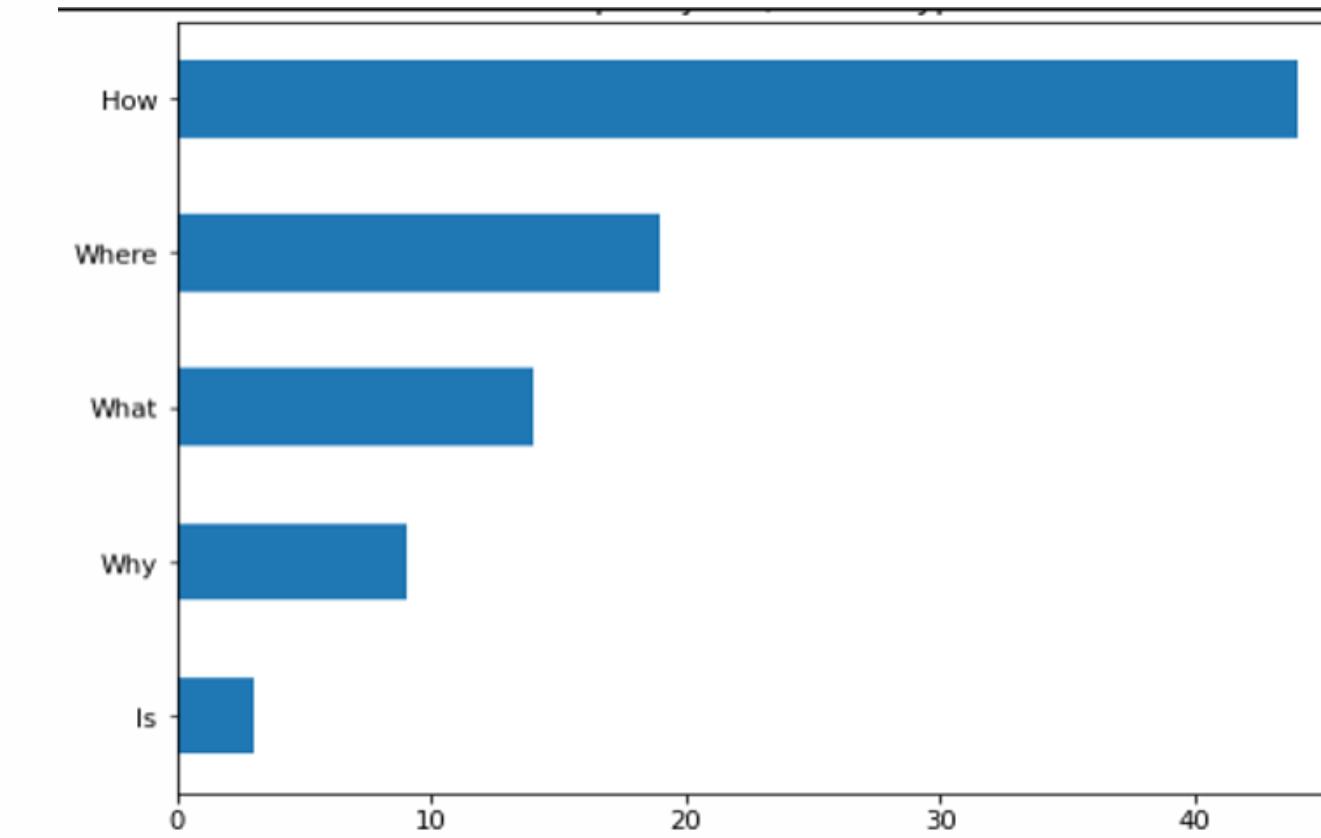
Text Q&A Data Sample

Self Collected Dataset for AI Testing Tool	
Source	Primary Data
Size	21KB
Format	CSV
Number of Instances	354
Number of Features	2
Sample 1	<p>Ques: How to signup for the AI testing tool</p> <p>Ans: Click on the sign -up button</p>

Self collected Dataset Sample

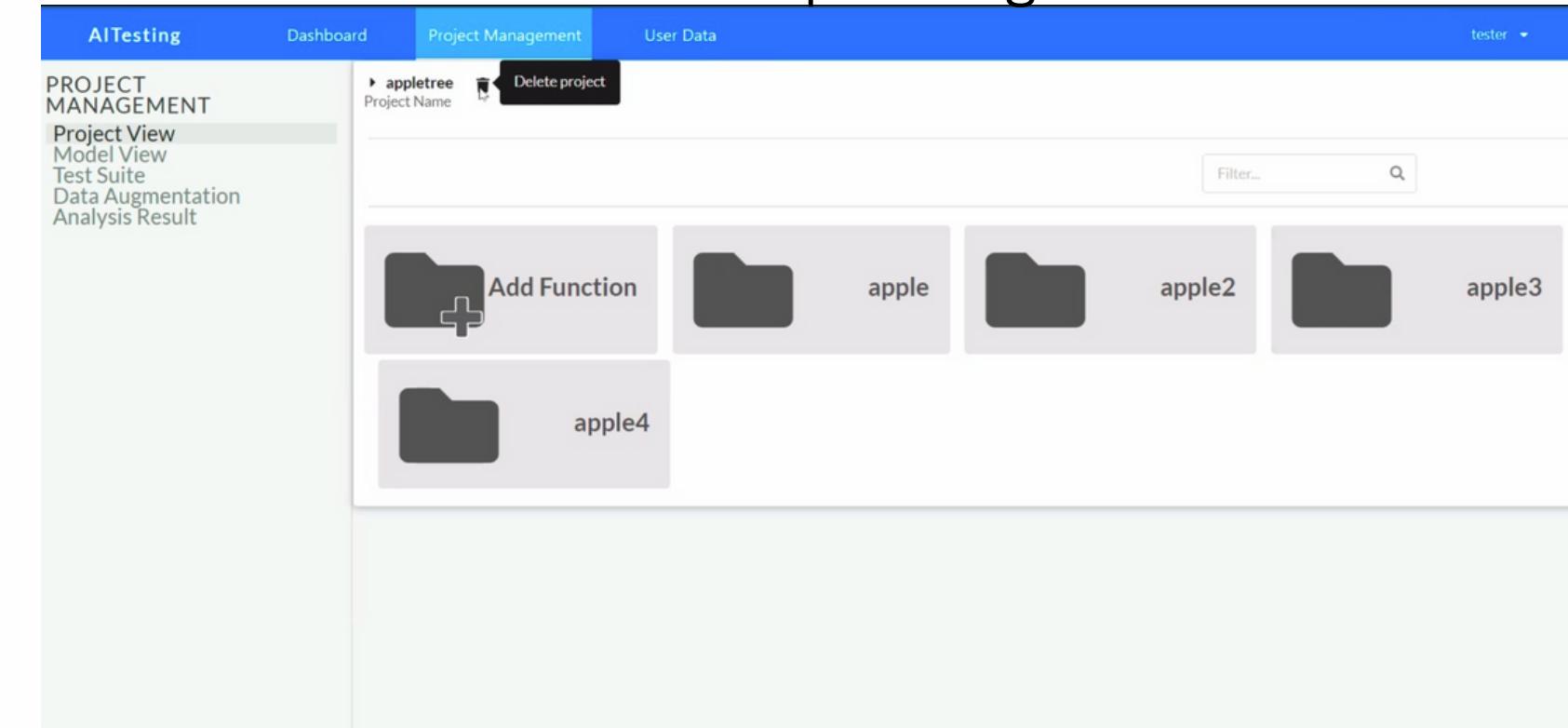
Questions	Answers
How to signup in the AI testing tool?	Click on the sign up button
How to sign up for AI testing tool	Click on the sign up button as shown in the image
Why I am getting an error while login?	Either the password or username is incorrect. Enter the correct details.
Why I am not able to login to AI testing tool?	Username and Password is incorrect. Please enter the correct details.
How to create a new project?	Go on to Dashboard and click on add project. You can start creating a new project.
How to look at previous created function?	Click on project management module to view the previous created functions.
How to add a project?	Go on to Dashboard and add project
Where can you start creating a new project?	Go on to the Dashboard to create project
How to look at previous project?	Go to the Dashboard and click on the summary section
How to check all uploaded projects?	specific project name in the filter

Frequency of Question Type



AI Testing Tool Image-Based Data Description

AI testing Tool Data	
Project Management	318
Dashboard	326
Common	544
User	114
Number object per image	0 - 20
Number words per image	10 - 200



Text2Audio Converted dataset samples

Text to Audio using GGTS	
Source	Primary Data
Size	57.9 MB
Format	Mp3
Number of instances	348
Number of Features	1

 audio351.mp3	Nov 14, 2023	:
 audio350.mp3	Nov 14, 2023	:
 audio349.mp3	Nov 14, 2023	:
 audio348.mp3	Nov 14, 2023	:
 audio347.mp3	Nov 14, 2023	:



Open Source Data Collection

Text-based Open Source Dataset Description

Twitter Customer Service	
Source	Open Source
Size	516.5 MB
Number of Rows	2.8 M
Number of Features	7
Format	.CSV

Raw Sample of Twitter Data

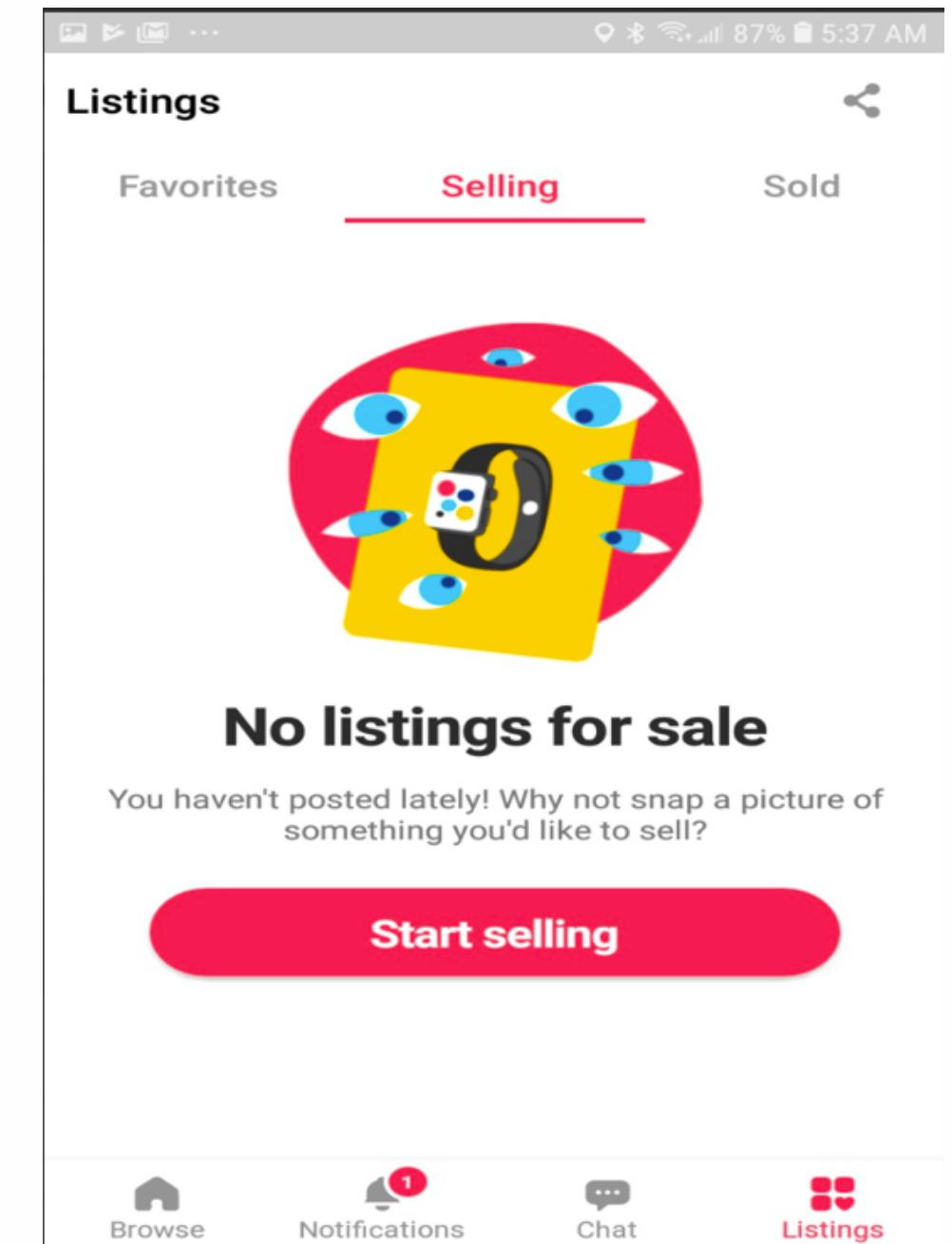
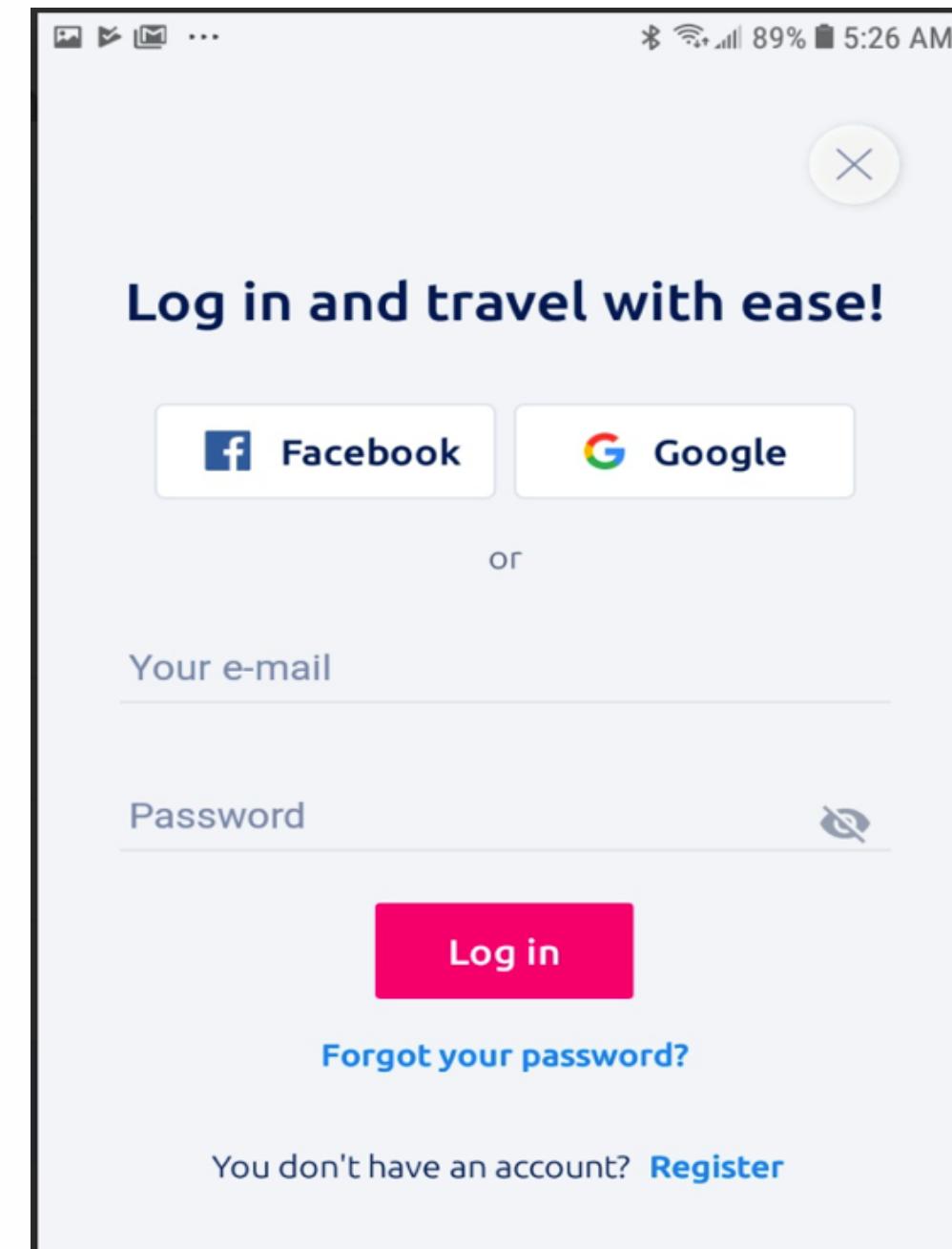
tweet_id	author_id	inbound	created_at	text
1	sprintcare	FALSE	Tue Oct 31 22:10:47 +0000 2017	@115712 I understand. I would like to assist you. We would need to get you into a private secured link to further assist.
2	115712	TRUE	Tue Oct 31 22:11:45 +0000 2017	@sprintcare and how do you propose we do that
3	115712	TRUE	Tue Oct 31 22:08:27 +0000 2017	@sprintcare I have sent several private messages and no one is responding as usual
4	sprintcare	FALSE	Tue Oct 31 21:54:49 +0000 2017	@115712 Please send us a Private Message so that we can further assist you. Just click 'Message' at the top of your profile.
5	115712	TRUE	Tue Oct 31 21:49:35 +0000 2017	@sprintcare I did.
6	sprintcare	FALSE	Tue Oct 31 21:46:24 +0000 2017	@115712 Can you please send us a private message, so that I can gain further details about your account?
8	115712	TRUE	Tue Oct 31 21:45:10 +0000 2017	@sprintcare is the worst customer service
11	sprintcare	FALSE	Tue Oct 31 22:10:35 +0000 2017	@115713 This is saddening to hear. Please shoot us a DM, so that we can look into this for you. -KC
12	115713	TRUE	Tue Oct 31 22:04:47 +0000 2017	@sprintcare You gonna magically change your connectivity for me and my whole family ? 😞 100

Image-based Open Source Data Description

Raw Sample of VINS image dataset

VINS Data Description

VINS Open Source Dataset	
Number of Images	4800
Source	(Bunian et.al., 2021)
Format	.jpg



Audio-based Open Source Data Description

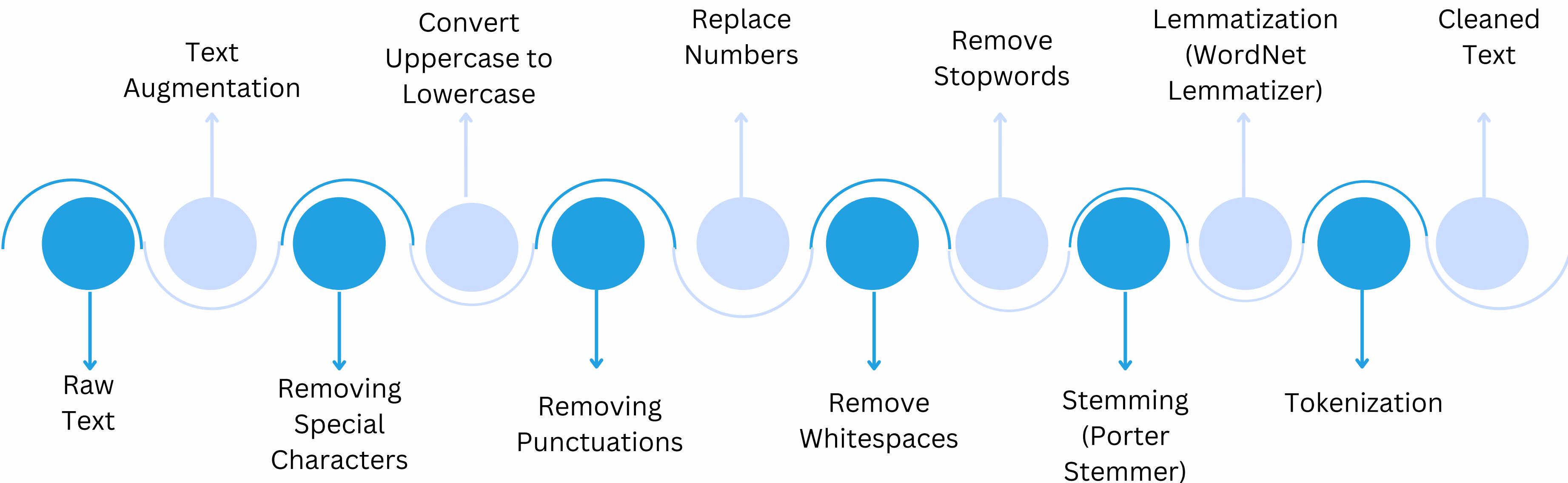
LibriSpeech Dataset Description

LibriSpeech Corpus	
Source	LibriVox audiobooks
Size	304.47GB
Number of Parameter	71,825,408
Recorded Hours	1000
Format	FLAC

Raw Sample of LibriSpeech dataset

 84-121123-0000.flac	Jul 8, 2014 at 1:32 AM	33 KB	FLAC Audio File
 84-121123-0001.flac	Jul 8, 2014 at 1:32 AM	82 KB	FLAC Audio File
 84-121123-0002.flac	Jul 8, 2014 at 1:32 AM	265 KB	FLAC Audio File
 84-121123-0003.flac	Jul 8, 2014 at 1:32 AM	125 KB	FLAC Audio File
 84-121123-0004.flac	Jul 8, 2014 at 1:32 AM	86 KB	FLAC Audio File
 84-121123-0005.flac	Jul 8, 2014 at 1:32 AM	298 KB	FLAC Audio File
 84-121123-0006.flac	Jul 8, 2014 at 1:32 AM	105 KB	FLAC Audio File
 84-121123-0007.flac	Jul 8, 2014 at 1:32 AM	39 KB	FLAC Audio File
 84-121123-0008.flac	Jul 8, 2014 at 1:32 AM	138 KB	FLAC Audio File
 84-121123-0009.flac	Jul 8, 2014 at 1:32 AM	54 KB	FLAC Audio File
 84-121123-0010.flac	Jul 8, 2014 at 1:32 AM	139 KB	FLAC Audio File
 84-121123-0011.flac	Jul 8, 2014 at 1:32 AM	63 KB	FLAC Audio File
 84-121123-0012.flac	Jul 8, 2014 at 1:32 AM	52 KB	FLAC Audio File
 84-121123-0013.flac	Jul 8, 2014 at 1:32 AM	44 KB	FLAC Audio File
 84-121123-0014.flac	Jul 8, 2014 at 1:32 AM	53 KB	FLAC Audio File
 84-121123-0015.flac	Jul 8, 2014 at 1:32 AM	57 KB	FLAC Audio File
 84-121123-0016.flac	Jul 8, 2014 at 1:32 AM	163 KB	FLAC Audio File
 84-121123-0017.flac	Jul 8, 2014 at 1:32 AM	184 KB	FLAC Audio File
 84-121123-0018.flac	Jul 8, 2014 at 1:32 AM	73 KB	FLAC Audio File
 84-121123-0019.flac	Jul 8, 2014 at 1:32 AM	52 KB	FLAC Audio File
 84-121123-0020.flac	Jul 8, 2014 at 1:32 AM	138 KB	FLAC Audio File
 84-121123-0021.flac	Jul 8, 2014 at 1:32 AM	91 KB	FLAC Audio File
 84-121123-0022.flac	Jul 8, 2014 at 1:32 AM	53 KB	FLAC Audio File
 84-121123-0023.flac	Jul 8, 2014 at 1:32 AM	89 KB	FLAC Audio File
 84-121123-0024.flac	Jul 8, 2014 at 1:32 AM	202 KB	FLAC Audio File
 84-121123-0025.flac	Jul 8, 2014 at 1:32 AM	125 KB	FLAC Audio File
 84-121123-0026.flac	Jul 8, 2014 at 1:32 AM	262 KB	FLAC Audio File
 84-121123-0027.flac	Jul 8, 2014 at 1:32 AM	62 KB	FLAC Audio File
 84-121123-0028.flac	Jul 8, 2014 at 1:32 AM	99 KB	FLAC Audio File
 84-121123.trans.txt	Aug 15, 2014 at 10:13 PM	3 KB	Plain Text

Data Preprocessing: Text-Based Data



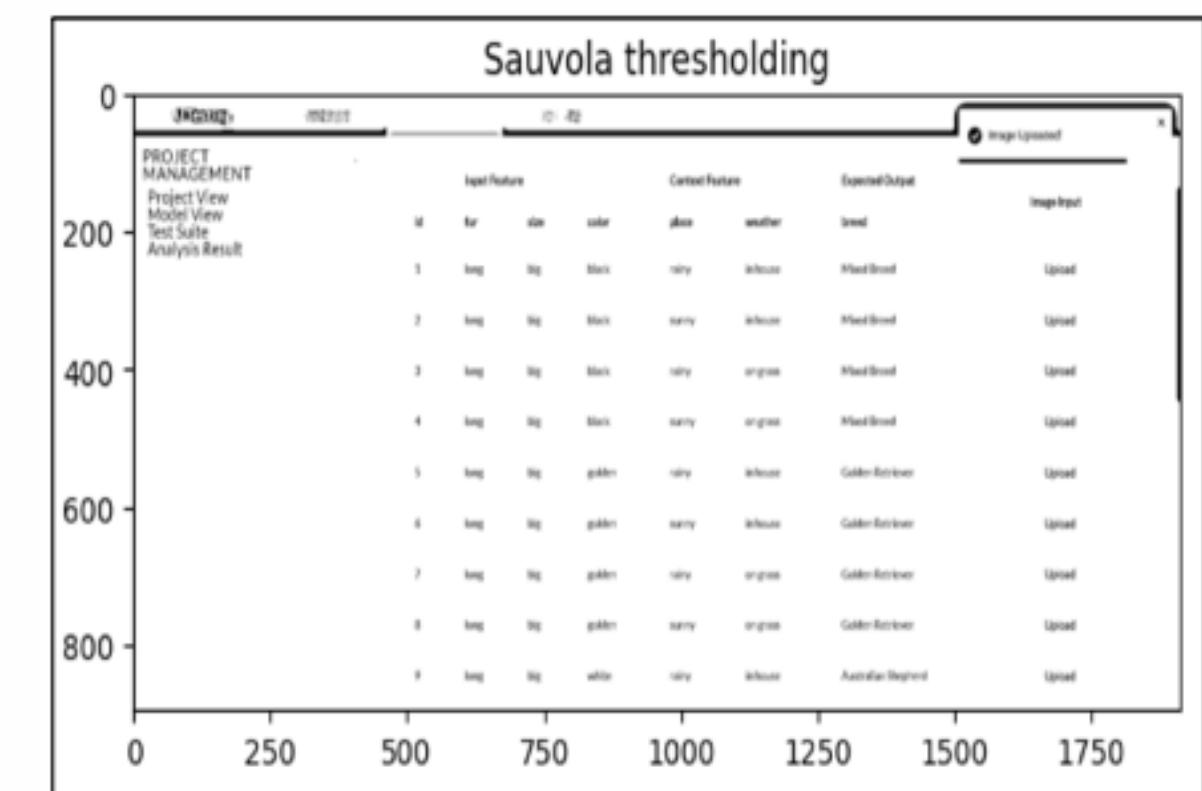
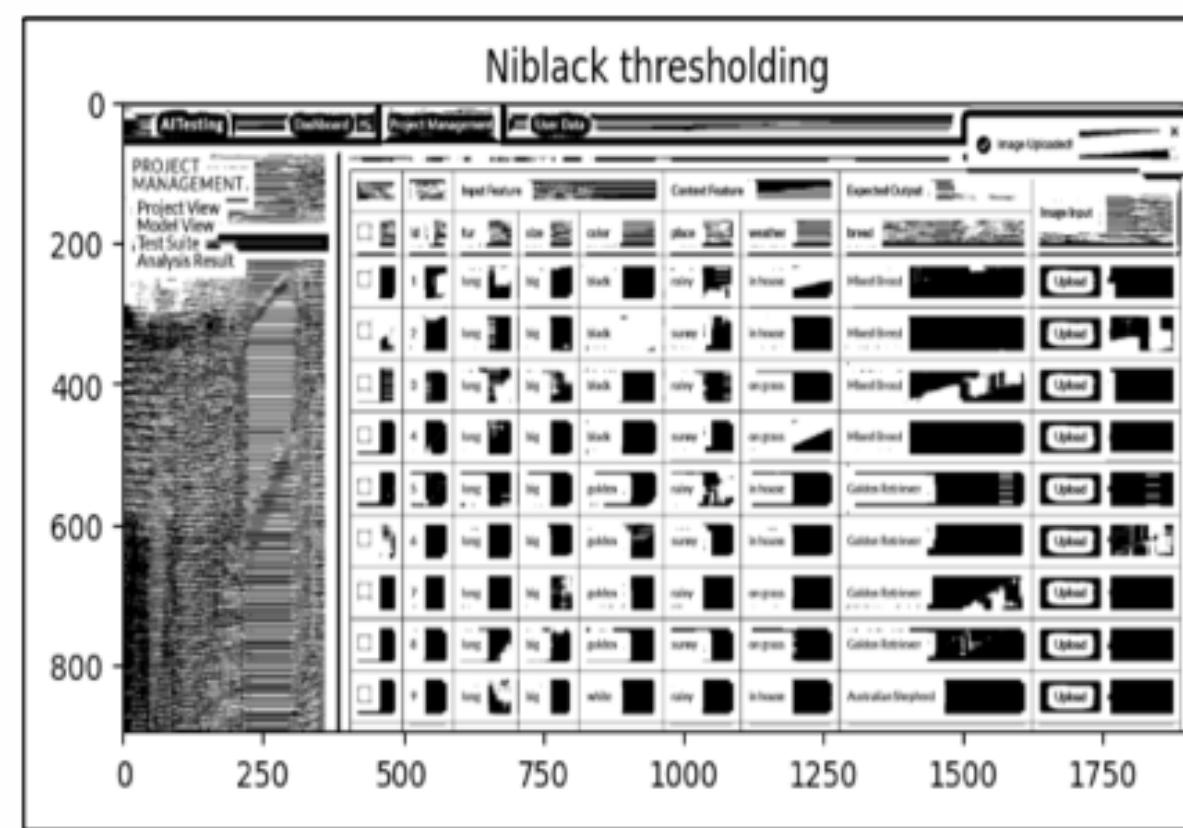
Data Preprocessing: Image-Based Data

Removing blurred images

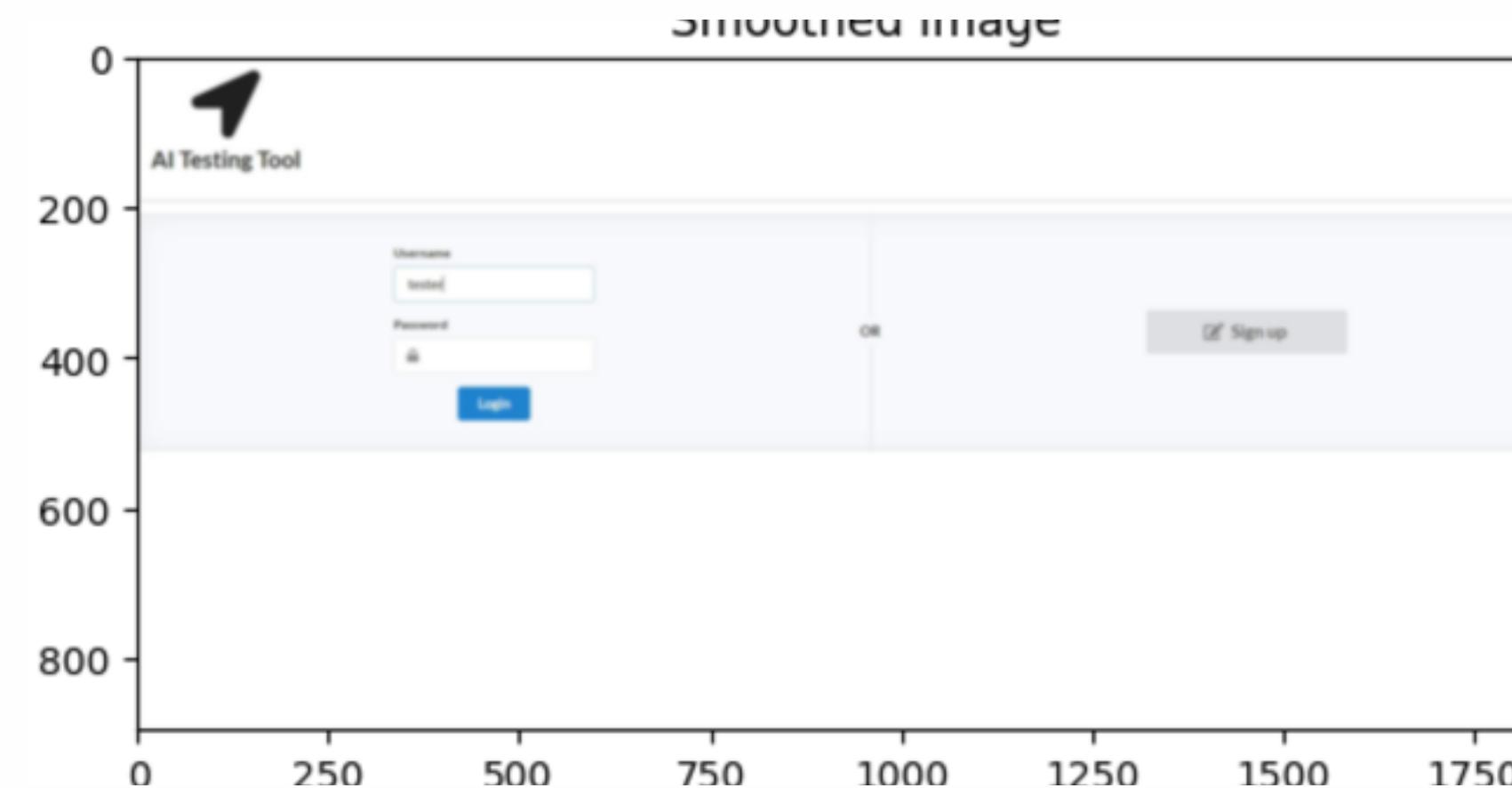
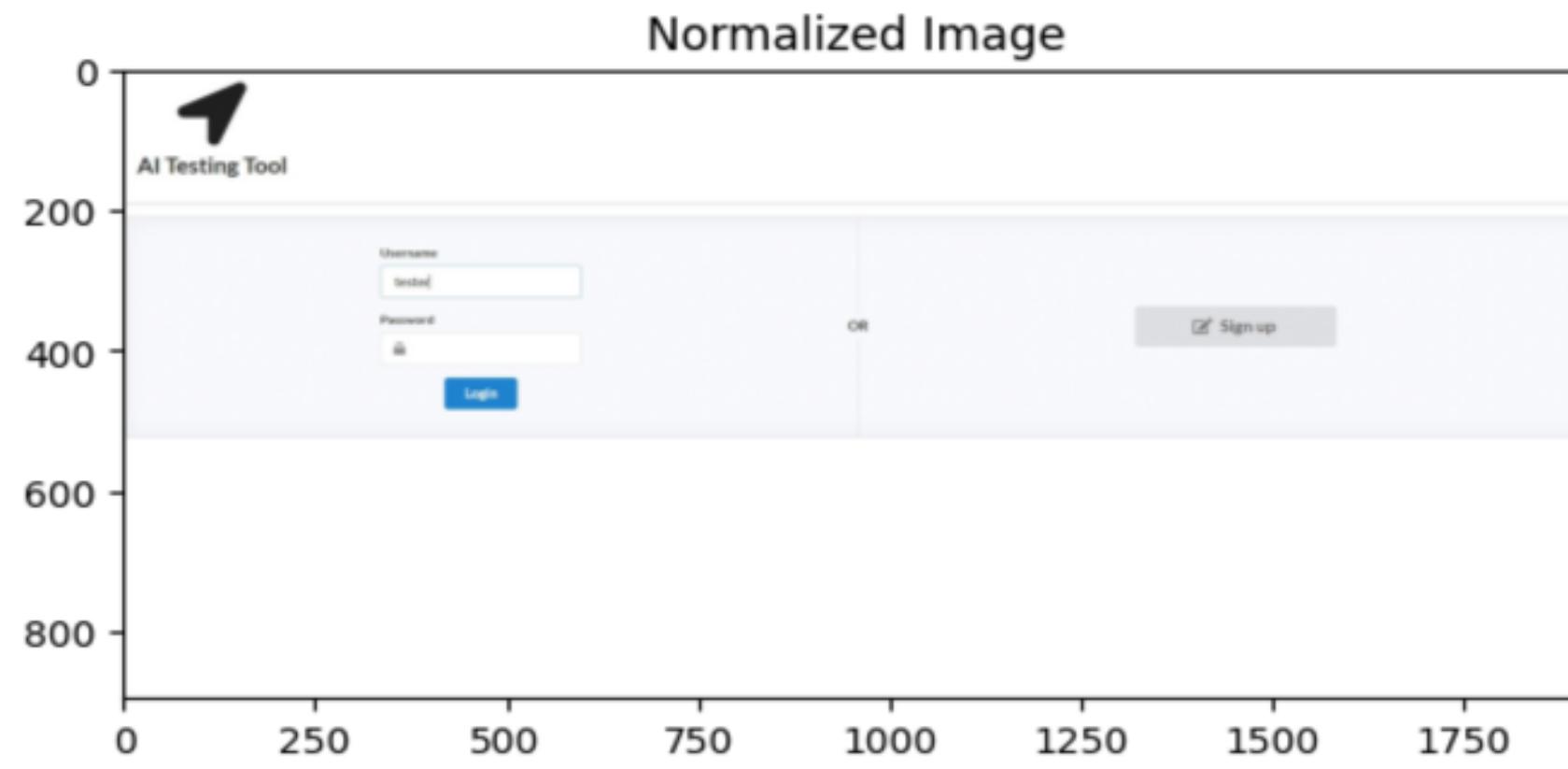
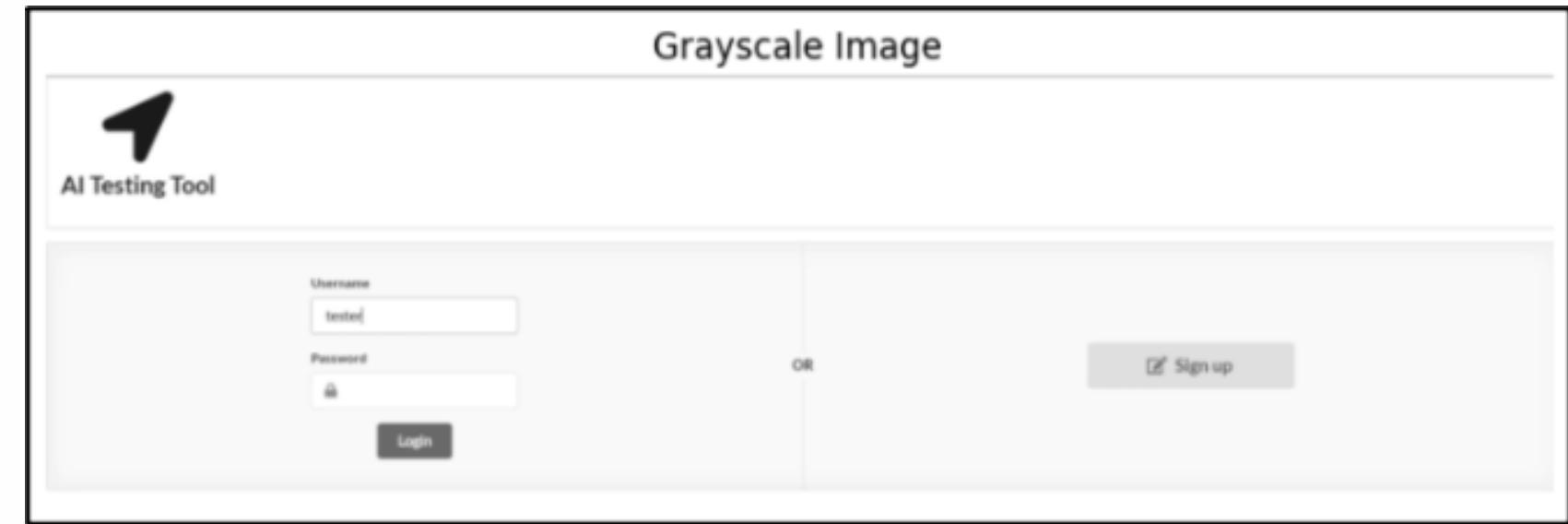
The blurry images are identified and eliminated by comparing the Laplacian metric with a threshold value of 75.

Image Segmentation

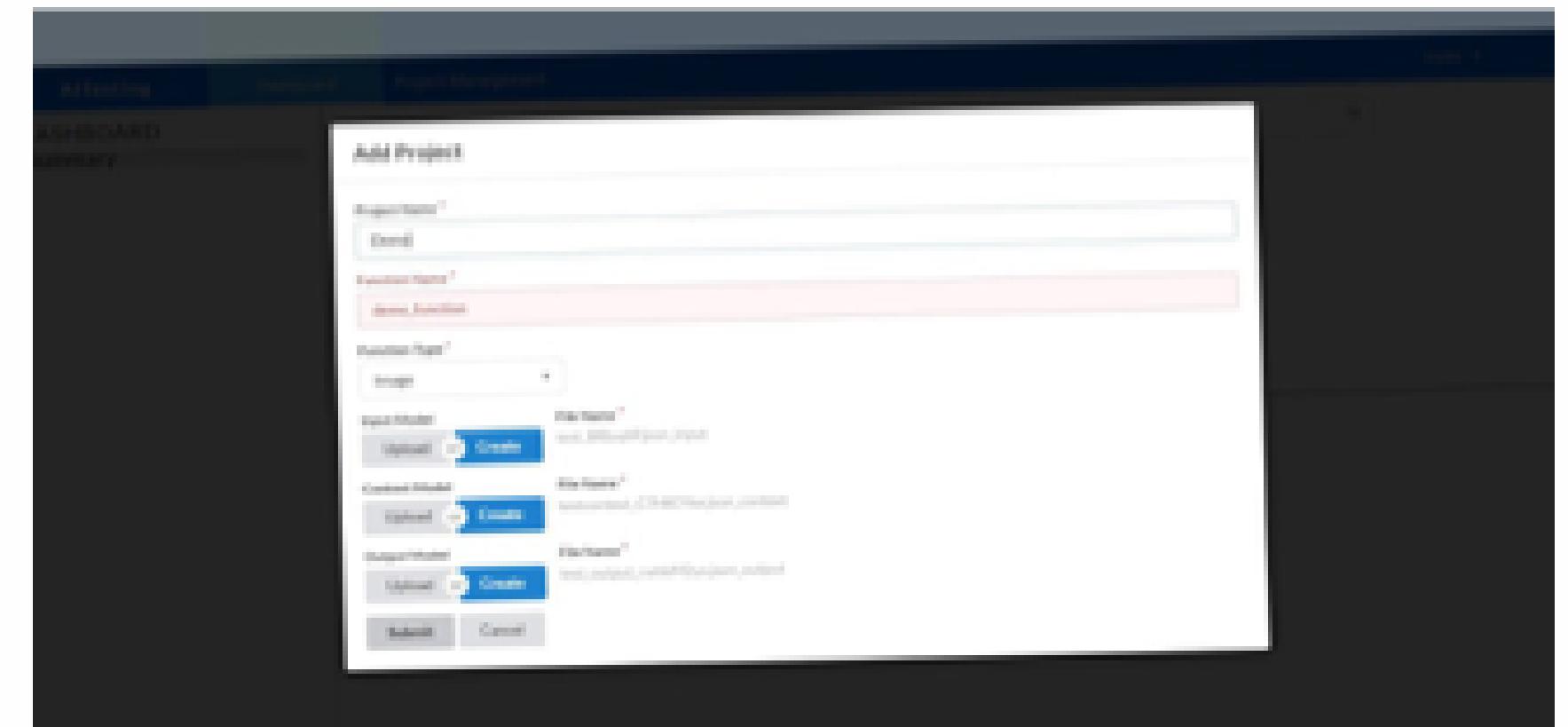
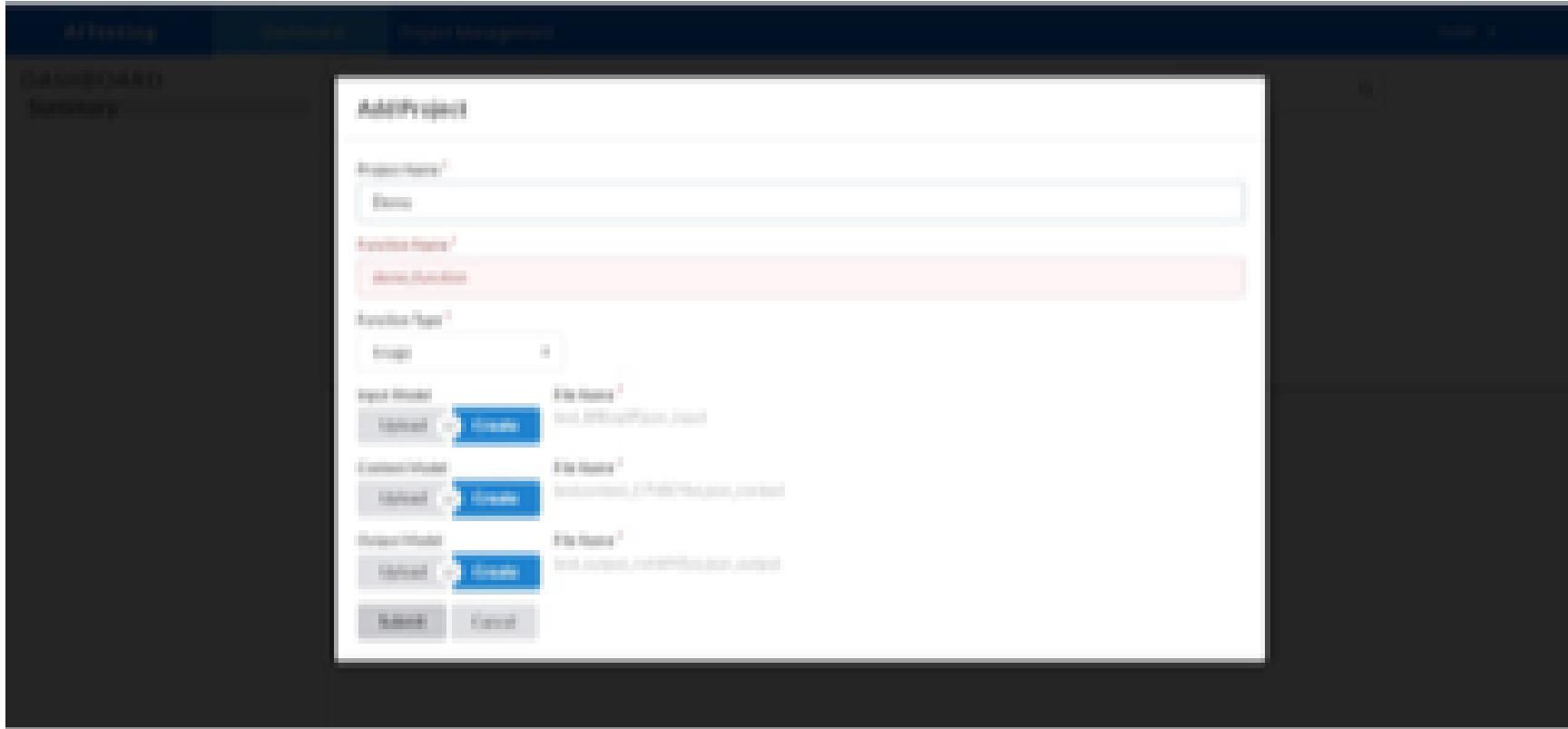
Segmented images to extract the objects of interest.



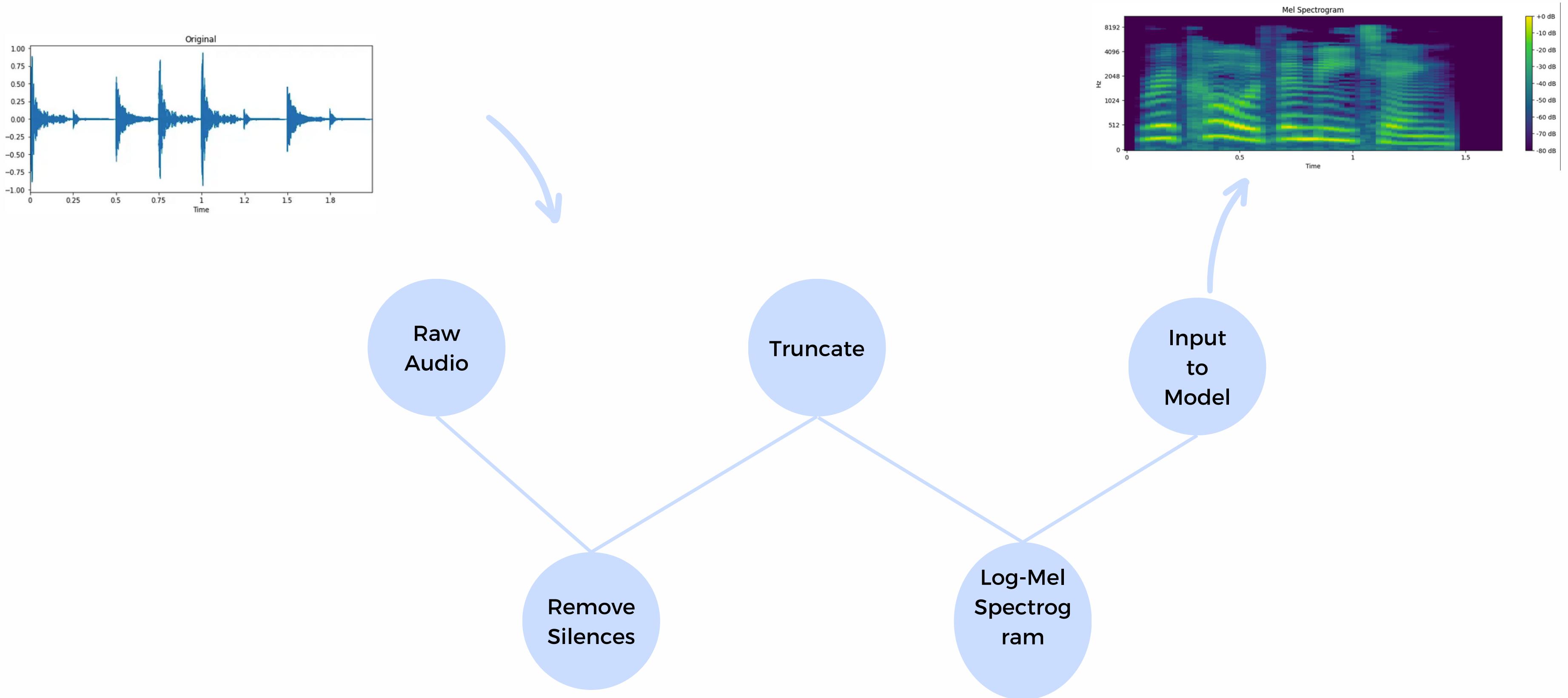
Data Transformation: Image-Based Data



Data Augmentation: Image-Based Data

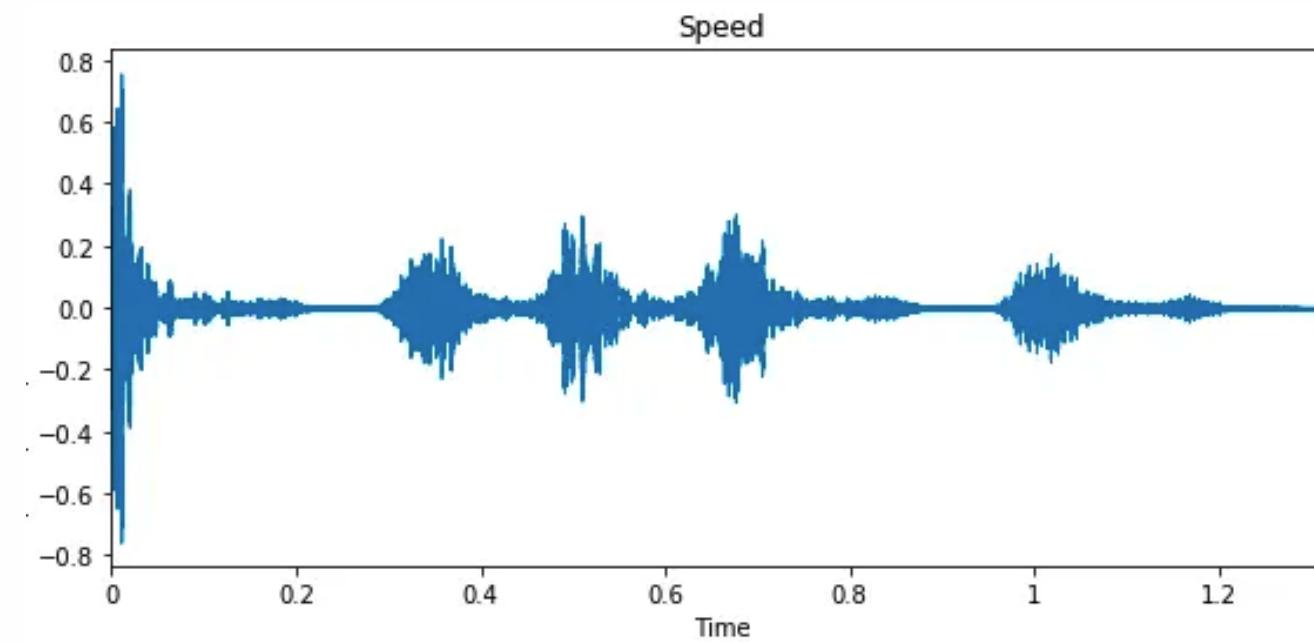
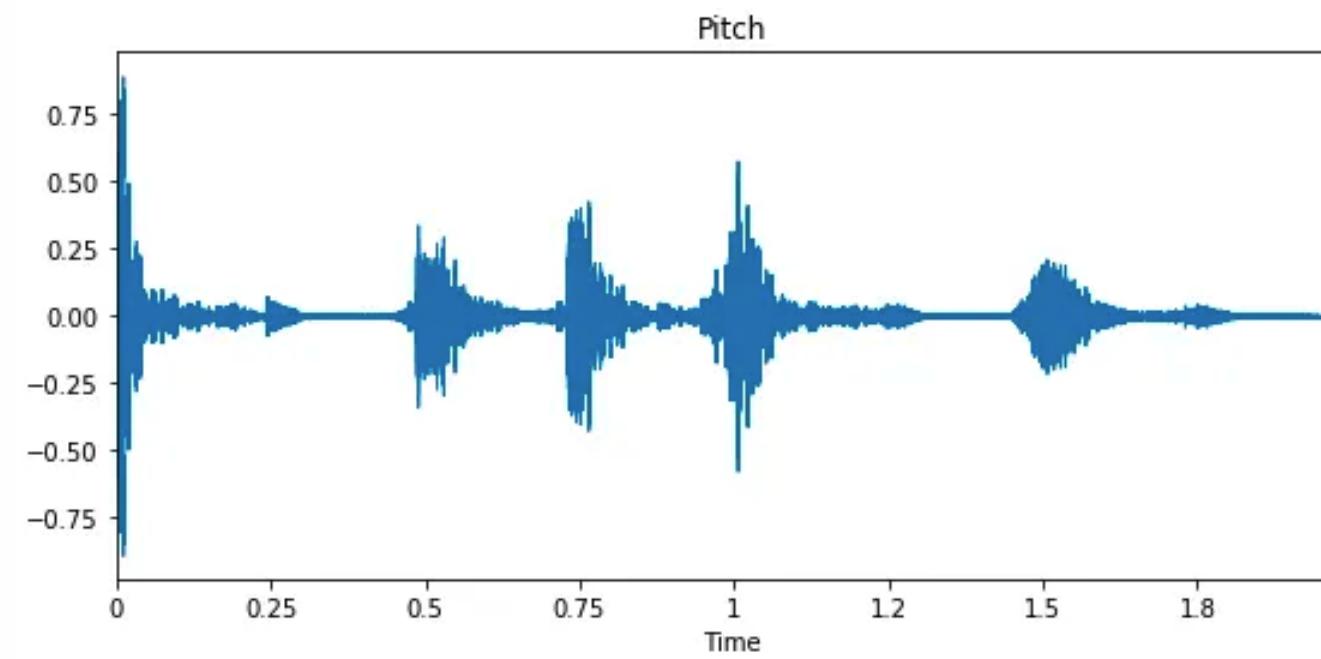
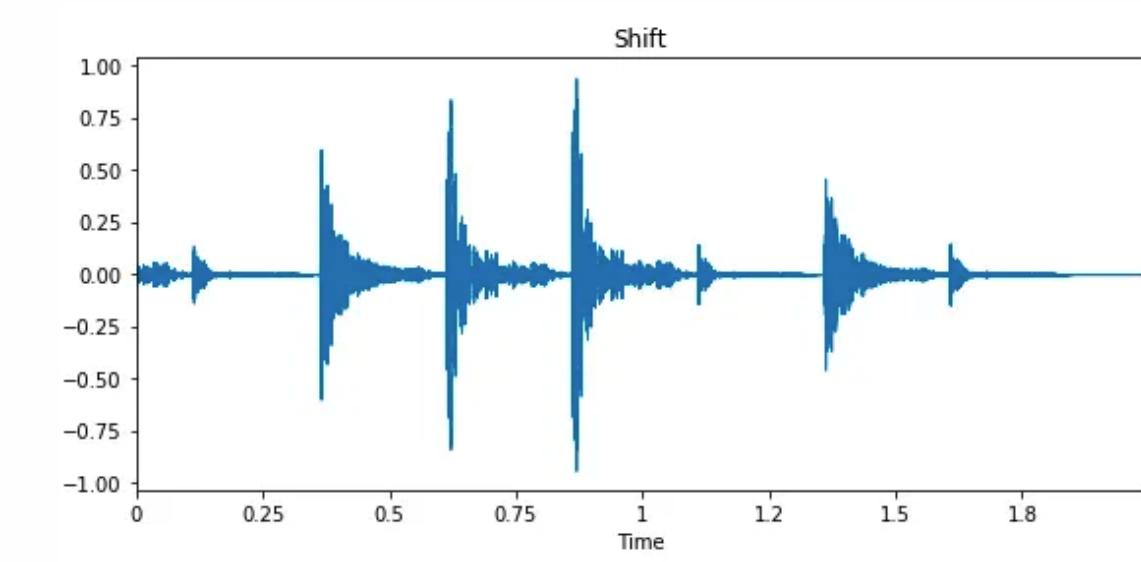
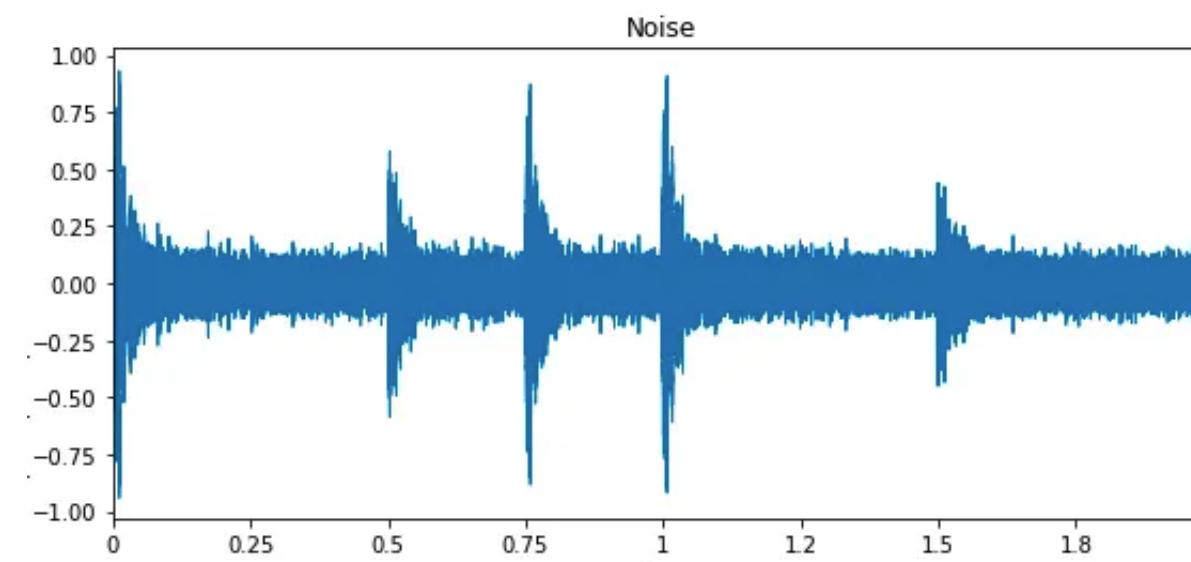
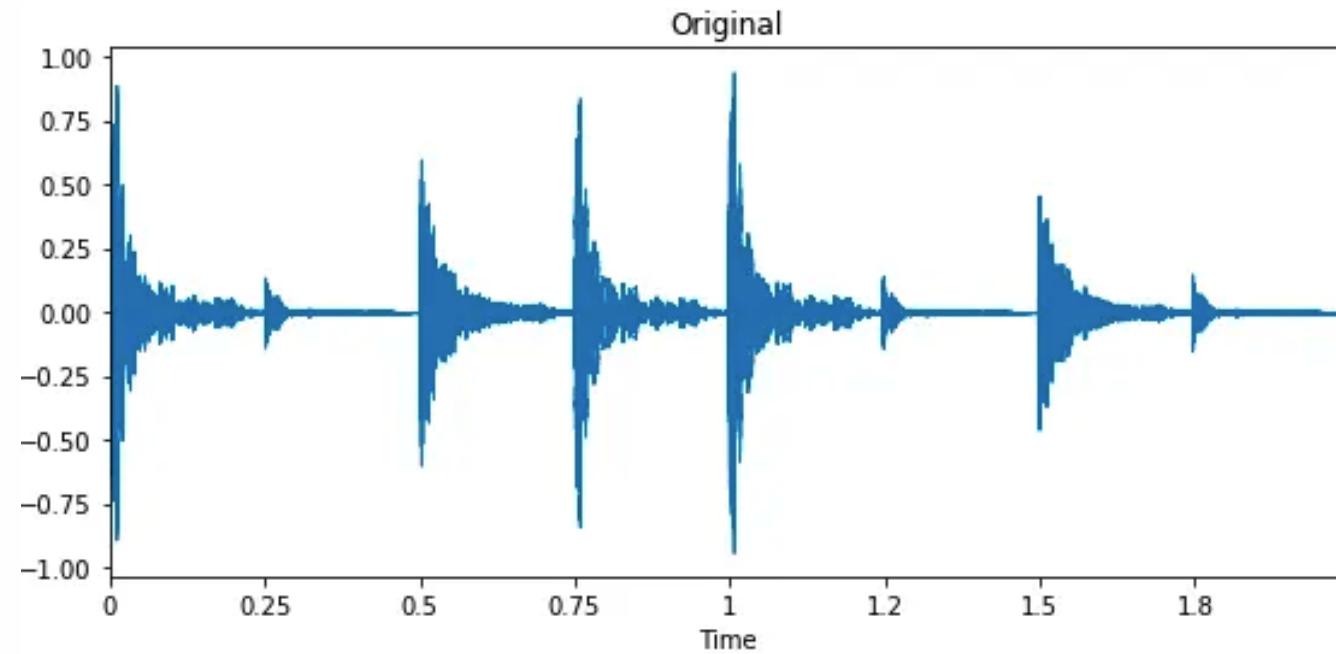


Data Preprocessing: Audio-Based Data



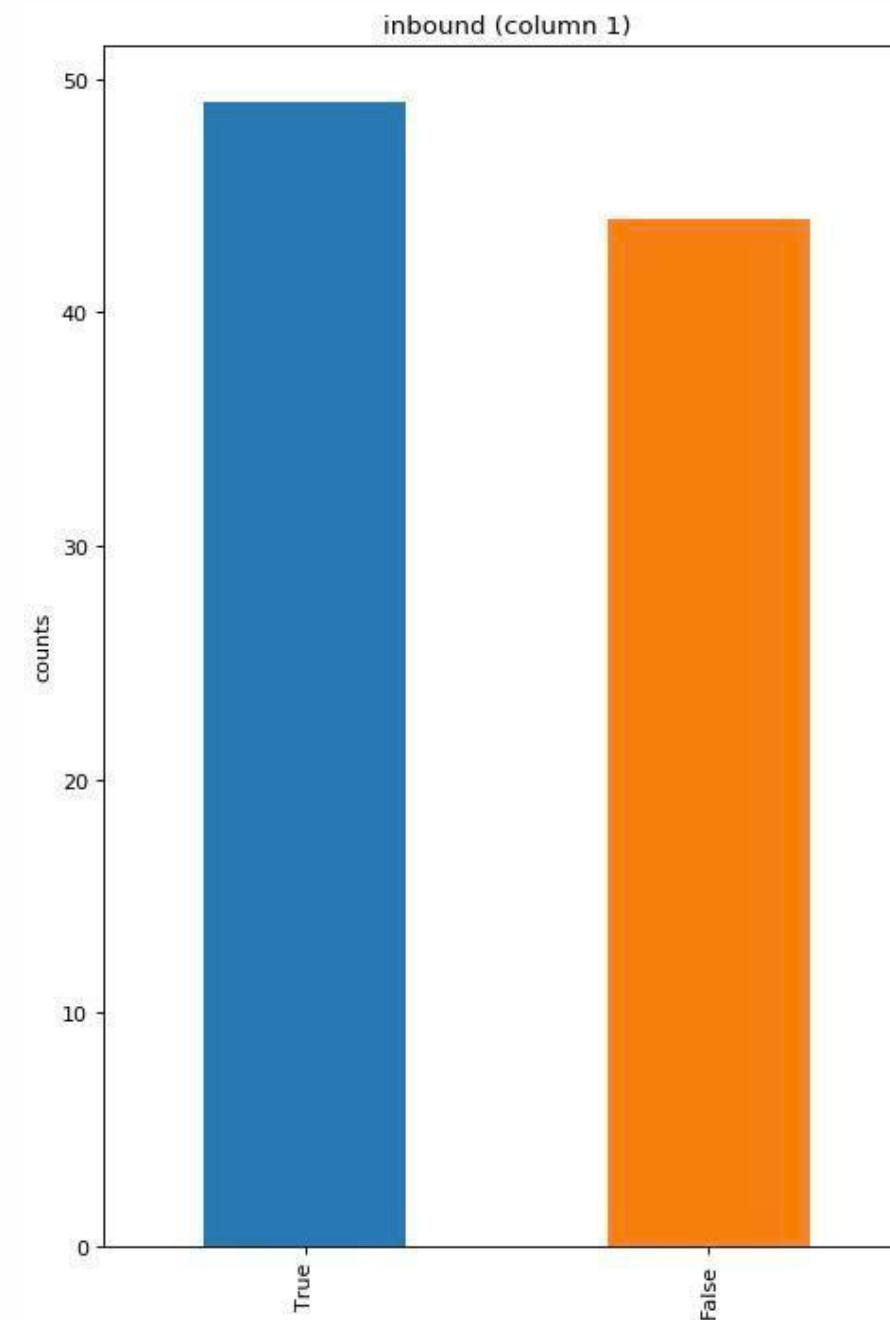
Data Preprocessing: Audio-Based Data

Augmented text Q&A data is converted to audio mp3 files

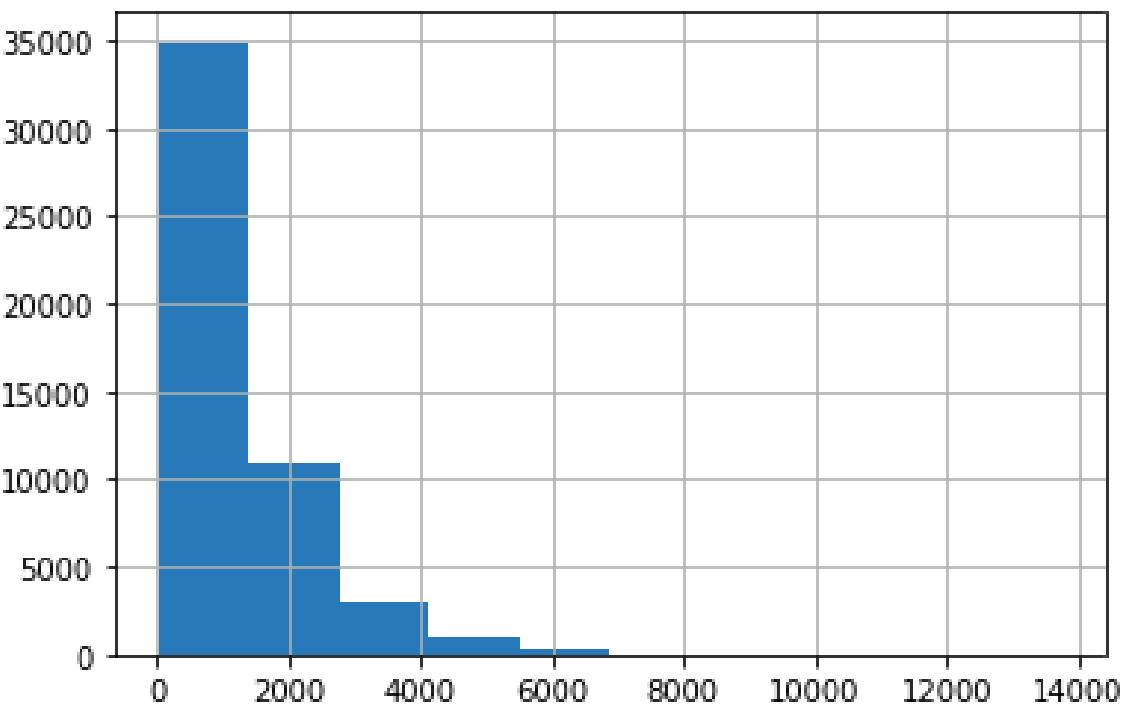


Data Analysis - Text Data

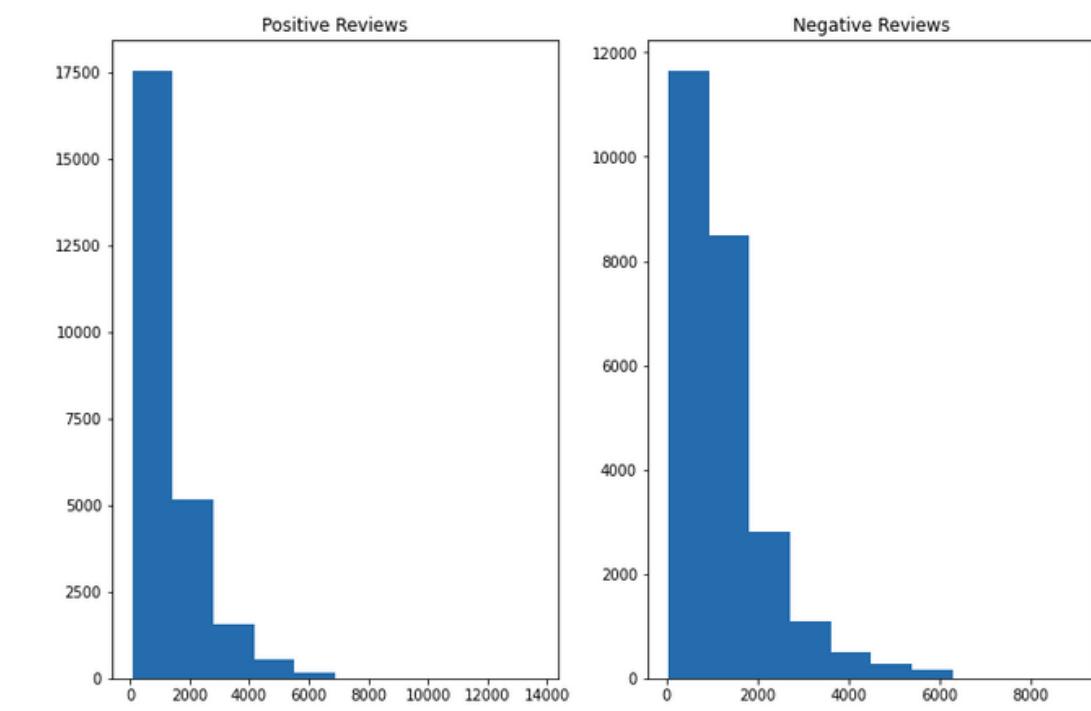
Bar Graph b/w Inbound & Outbound
Tweets



Histogram of Avg Text Length

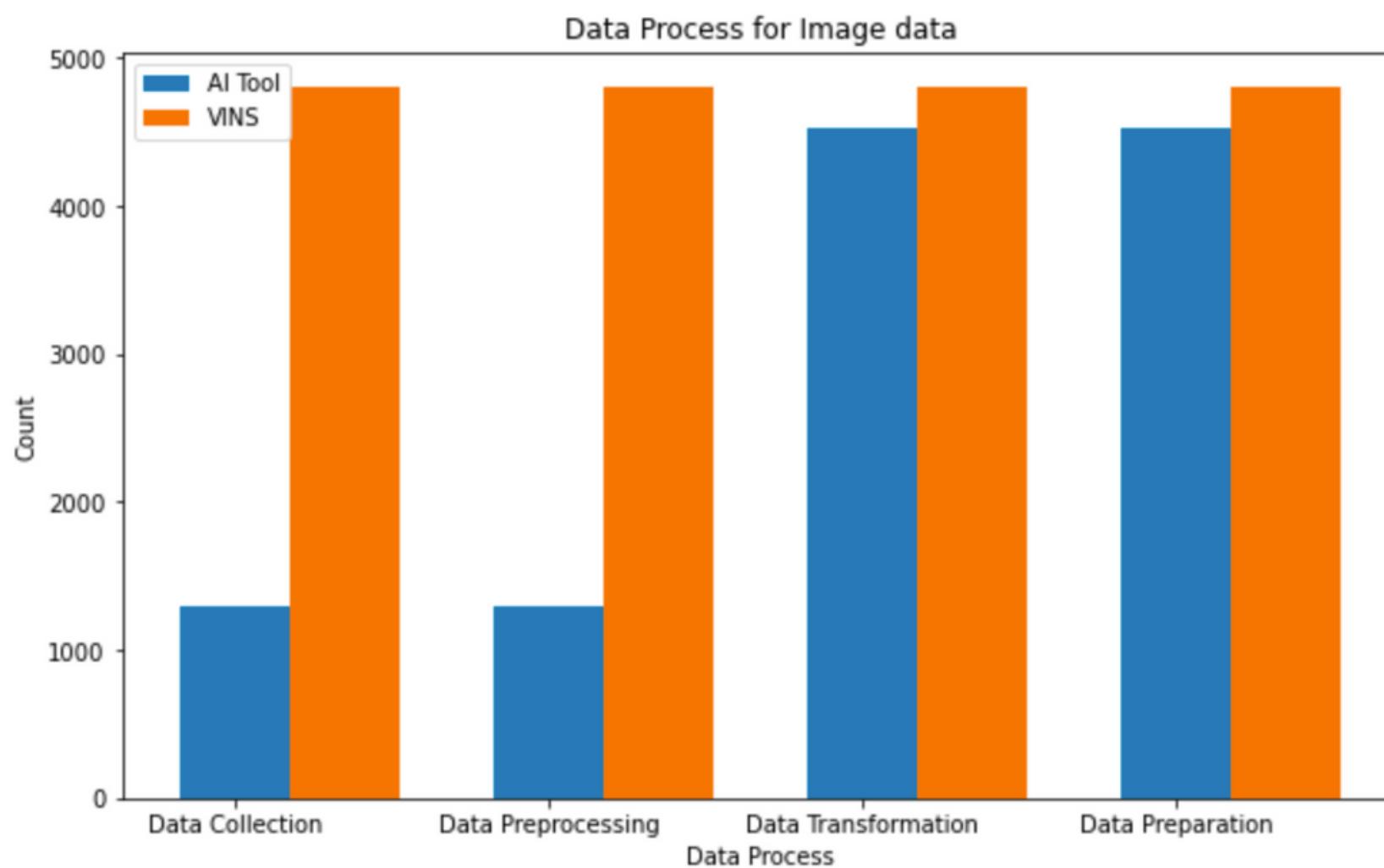


Sentiment analysis of the Tweets

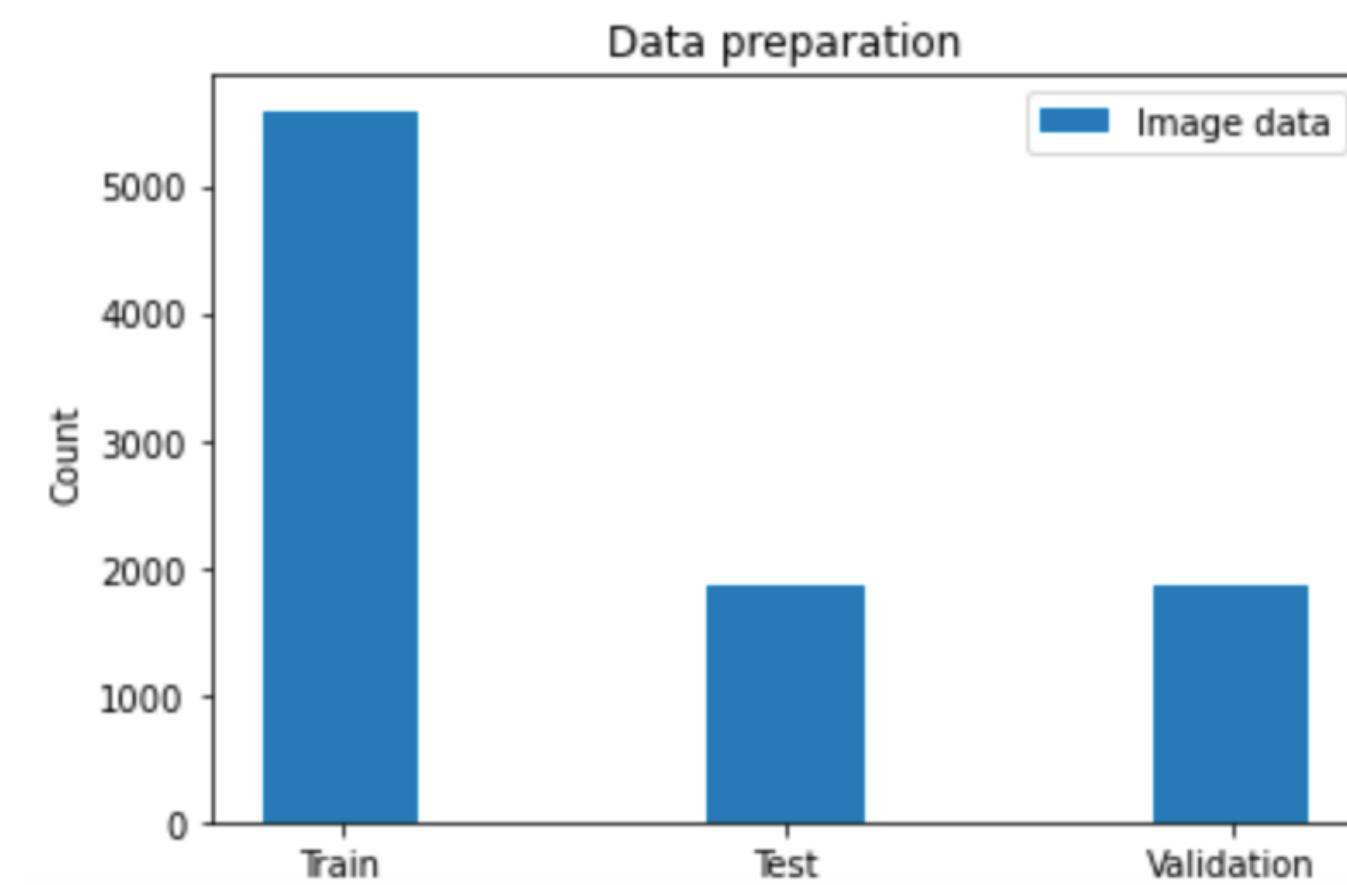


Data Analysis - Image Data

Data process statistics



Data splitting statistics



Data Analysis - Image Data

Data Process	Method	AI Tool dataset	VINS dataset
Data Collection	Raw data	1302	4800
Data Preprocessing	Removing blurred images	1300	4800
Data Preprocessing	Image segmentation	1300	4800
Data Transformation	Grayscale conversion	1300	4800
Data Transformation	Normalization	1300	4800
Data Transformation	Image smoothing	1300	4800
Data Transformation	Dimensionality reduction	1300	4800
Data Transformation	Image augmentation	4518	4800
Data Preparation	Data splitting	4518	4800

Data Preparation

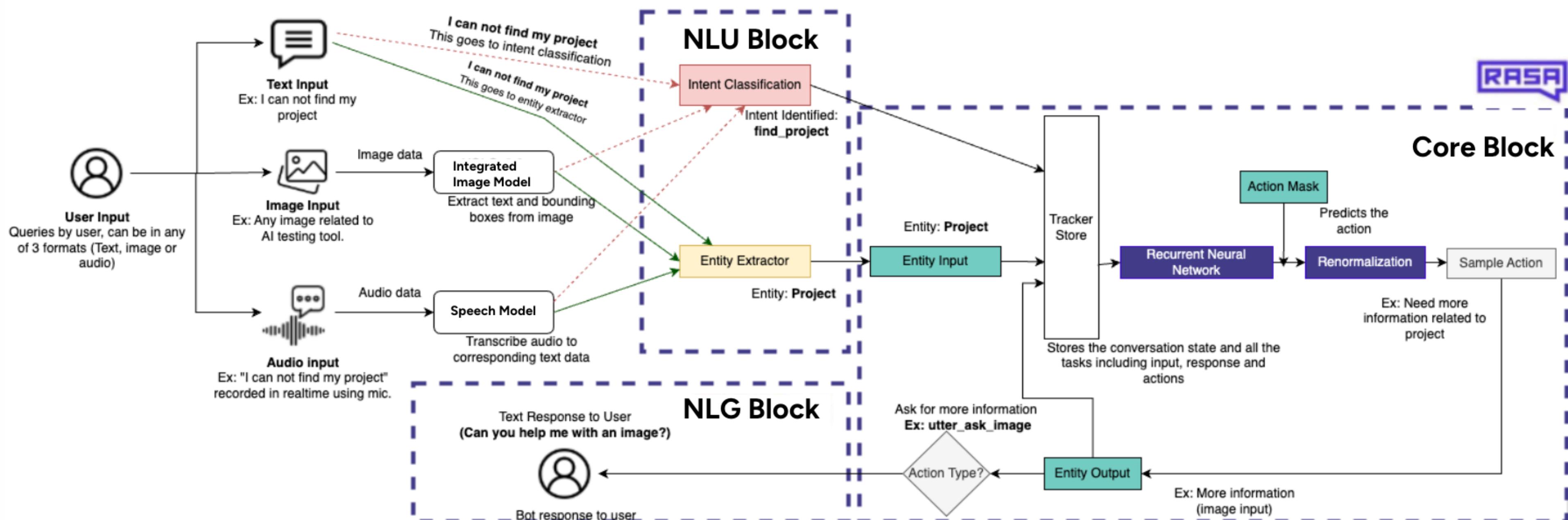
Split Statistics for Open-Source Dataset

Dataset	Total Dataset	Train	validation	Test
Text Data (Twitter)	2.9M	1.74M	580,000	580,000
Audio Data (LibriSpeech)	54741	27539	2620	1000
Image Data (AI Tool + VINS)	9318	5591	1864	1864

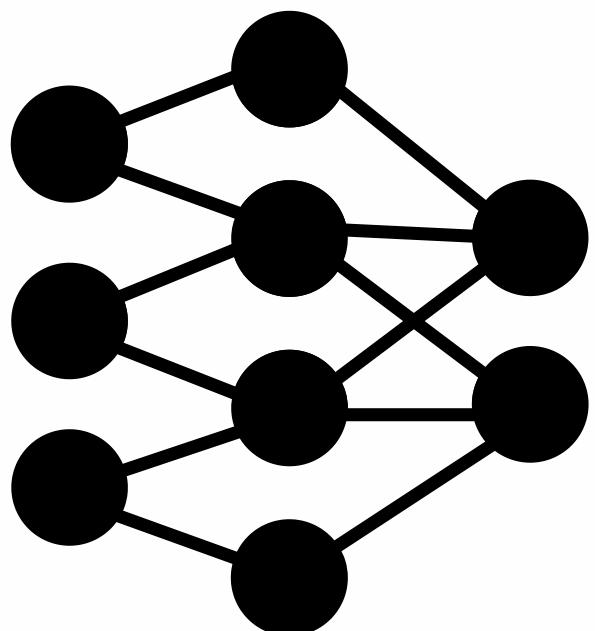
Split statistics for AI Tool Text and Audio

Dataset	Total Dataset	Train	validation	Test
Text Q&A Data	355	213	71	71
Text-2-Audio Data	355	213	71	71

PROPOSED ARCHITECTURE



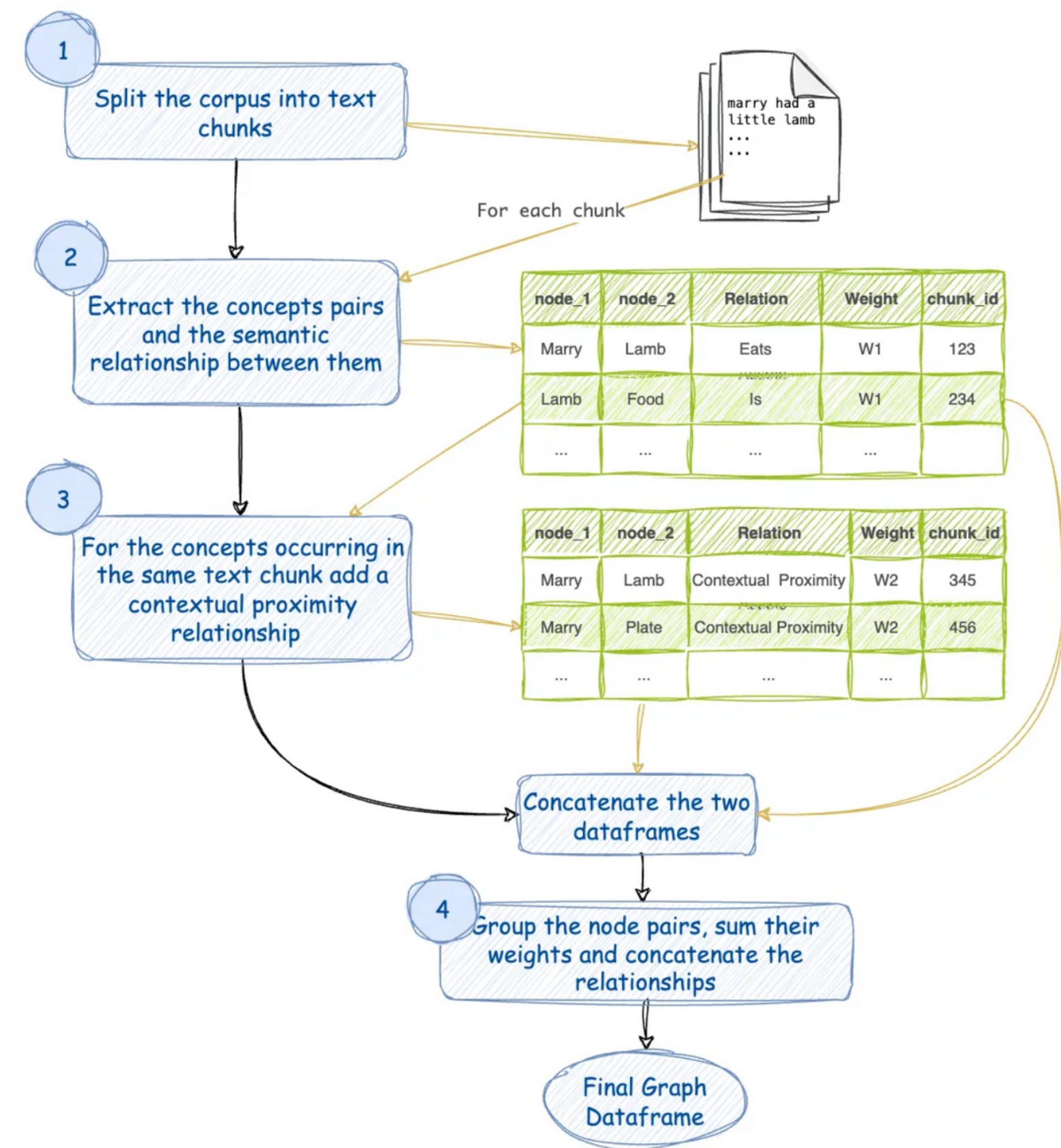
MODEL SELECTION



Text-Based Models 1:2

LLMs for converting any text to Knowledge Graph

- Knowledge Graphs are useful in a variety of ways. We can run graph algorithms and calculate the centralities for any node, to understand how important a concept (node) is to the body of work
- We can also use knowledge graphs to implement Graph Retrieval Augmented Generation (GRAG or GAG) and chat with our documents.
- **To extract the concepts and their relationships, we are using the Mistral 7B model.**



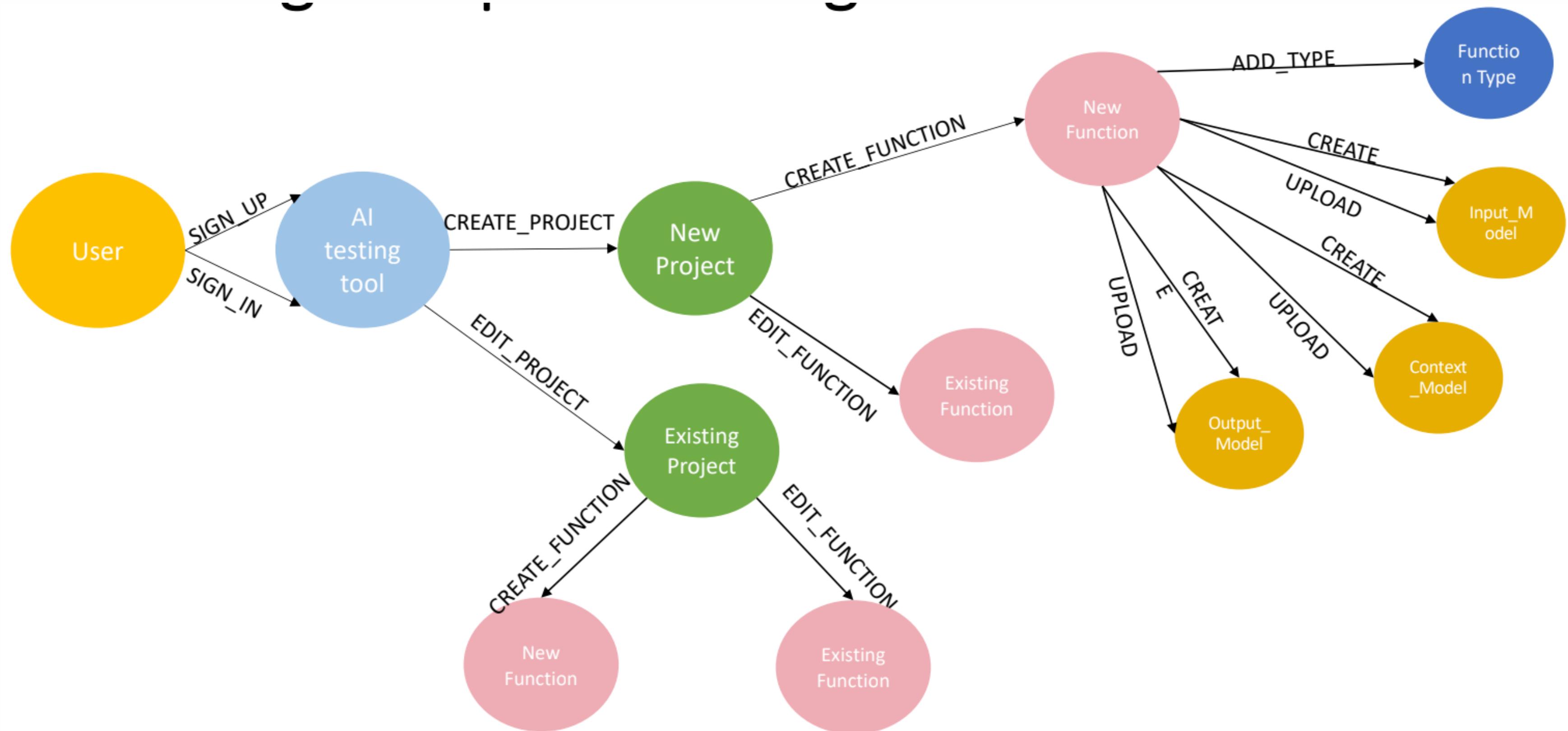
Data Preparation

```
def extractConcepts(prompt: str, metadata={}, model="mistral-openorca:latest"):
    SYS_PROMPT = (
        "Your task is extract the key concepts (and non personal entities) mentioned in the given context."
        "Extract only the most important and atomistic concepts, if needed break the concepts down to the simpler concepts."
        "Categorize the concepts in one of the following categories:"
        "[event, concept, place, object, document, organisation, condition, misc]\n"
        "Format your output as a list of json with the following format:\n"
        "[\n"
        "  {\n"
        "    \"entity\": The Concept,\n"
        "    \"importance\": The concontextual importance of the concept on a scale of 1 to 5 (5 being the highest),\n"
        "    \"category\": The Type of Concept,\n"
        "  },\n"
        "{ },\n"
    ]\n"
)
----- - client.generate(model_name=model, system_SYS_PROMPT=prompt=prompt)

SYS_PROMPT = (
    "You are a network graph maker who extracts terms and their relations from a given context."
    "You are provided with a context chunk (delimited by ```) Your task is to extract the ontology "
    "of terms mentioned in the given context. These terms should represent the key concepts as per the context. \n"
    "Thought 1: While traversing through each sentence, Think about the key terms mentioned in it.\n"
    "\tTerms may include object, entity, location, organization, person, \n"
    "\tcondition, acronym, documents, service, concept, etc.\n"
    "\tTerms should be as atomistic as possible\n\n"
    "Thought 2: Think about how these terms can have one on one relation with other terms.\n"
    "\tTerms that are mentioned in the same sentence or the same paragraph are typically related to each other.\n"
    "\tTerms can be related to many other terms\n\n"
    "Thought 3: Find out the relation between each such related pair of terms. \n\n"
    "Format your output as a list of json. Each element of the list contains a pair of terms"
    "and the relation between them, like the follwing: \n"
    "[\n"
    "  {\n"
    "    \"node_1\": \"A concept from extracted ontology\", \n"
    "    \"node_2\": \"A related concept from extracted ontology\", \n"
    "    \"edge\": \"relationship between the two concepts, node_1 and node_2 in one or two sentences\"\n"
    "  }, ...\n"
]
)
```

Design KG

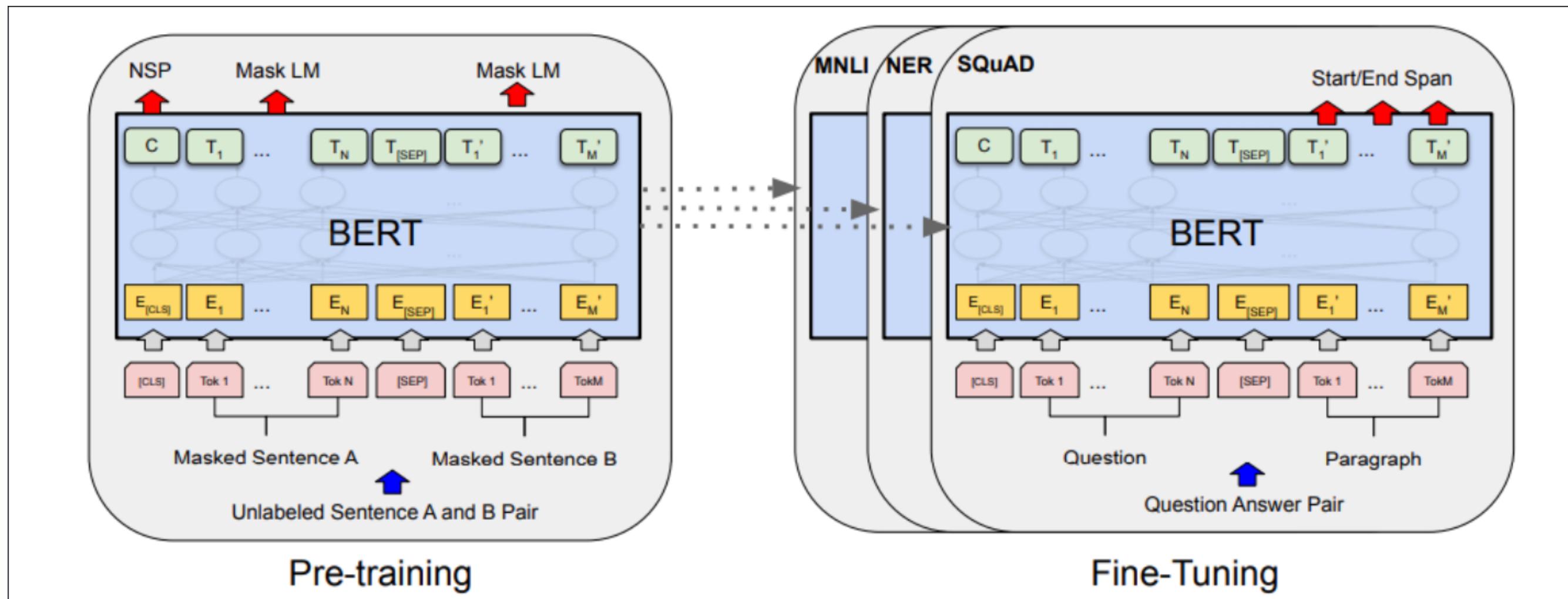
AI testing Tool KG -ER



Text-Based Models

BERT for QnA:

- BERT stands for Bidirectional Encoder Representations from Transformers developed by Google in 2018.
- Because of its bidirectionality, it has a deeper sense of language context and flow and hence, is used in a lot of NLP tasks nowadays.



Text-Based Models 2:2

Llama 2 Index

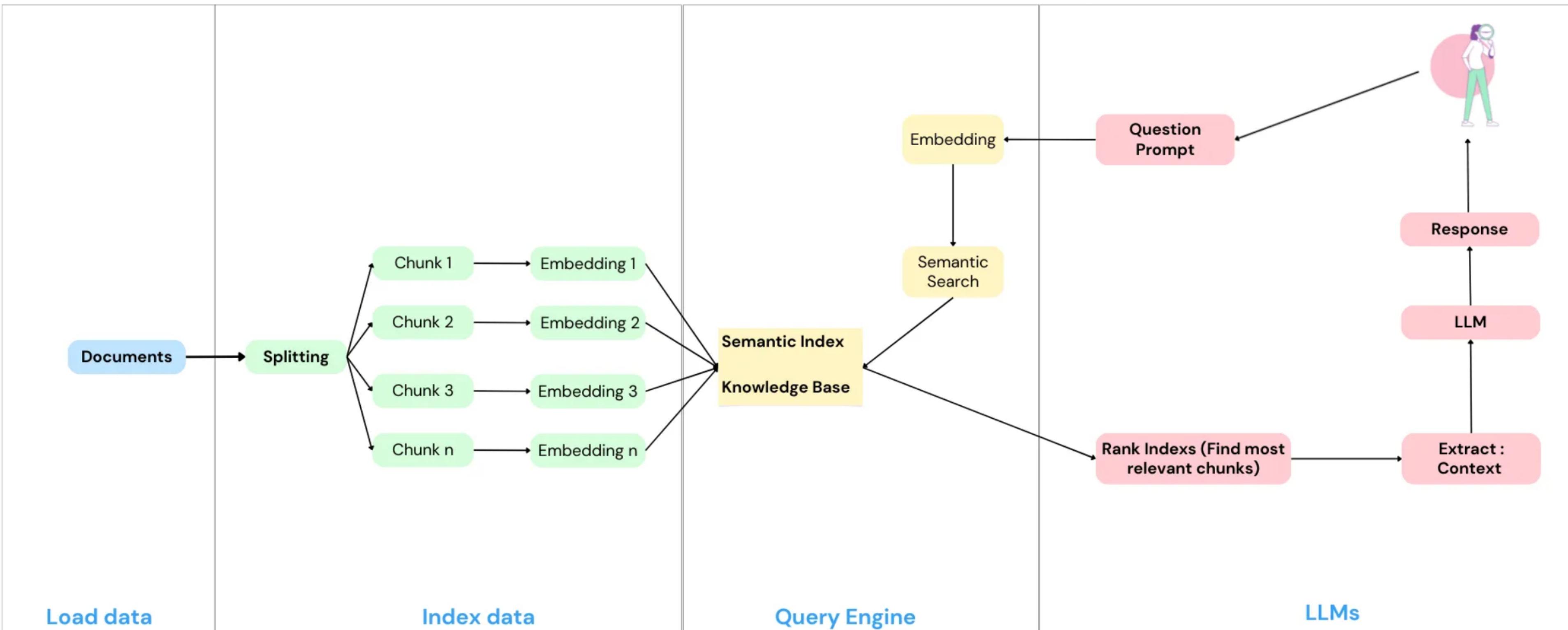
- (Llama 2-Chat), which have been optimized for dialogue applications using Reinforcement Learning from Human Feedback (RLHF). An **open alternative to closed-source chatbots**
- It breaks down huge documents into chunks and assigns an index, during retrieval, these indexes will be used to search for relevant context-related information.

MODEL SIZE (PARAMETERS)	PRETRAINED	FINE-TUNED FOR CHAT USE CASES
7B	Model architecture:	Data collection for helpfulness and safety:
13B	Pretraining Tokens: 2 Trillion	Supervised fine-tuning: Over 100,000
70B	Context Length: 4096	Human Preferences: Over 1,000,000

Model Justification

Model	License	Commercial use?	Pretraining length [tokens]	Leaderboard score
Falcon-7B	Apache 2.0	✓	1,500B	47.01
MPT-7B	Apache 2.0	✓	1,000B	48.7
Llama-7B	Llama license	✗	1,000B	49.71
Llama-2-7B	Llama 2 license	✓	2,000B	54.32
Llama-33B	Llama license	✗	1,500B	*
Llama-2-13B	Llama 2 license	✓	2,000B	58.67
mpt-30B	Apache 2.0	✓	1,000B	55.7
Falcon-40B	Apache 2.0	✓	1,000B	61.5
Llama-65B	Llama license	✗	1,500B	62.1
Llama-2-70B	Llama 2 license	✓	2,000B	*
Llama-2-70B-chat*	Llama 2 license	✓	2,000B	66.8

Chatbot Architecture with Llama and LLM

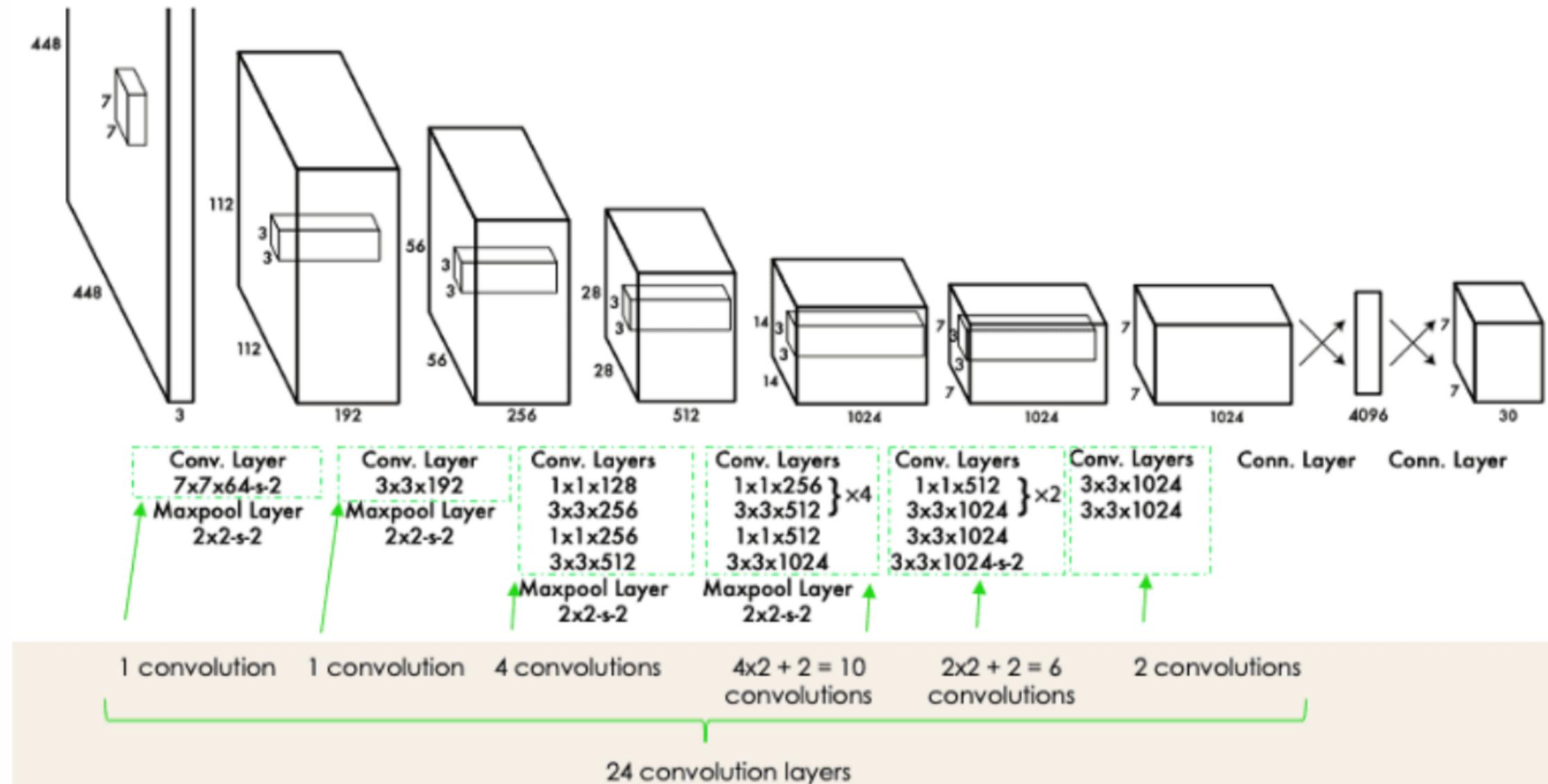


Comparision based on performance metrics

Metrics	Model 1	Model 2
	BERT	llama 2
Perplexity	115.23	113.85
BLUE	71.12	92.50
Precision	87.43	98.61
Recall	76.21	98.68
F1 Score	81.43	98.65
Rouge	80.14	93.46

Model Selection - Object Detection

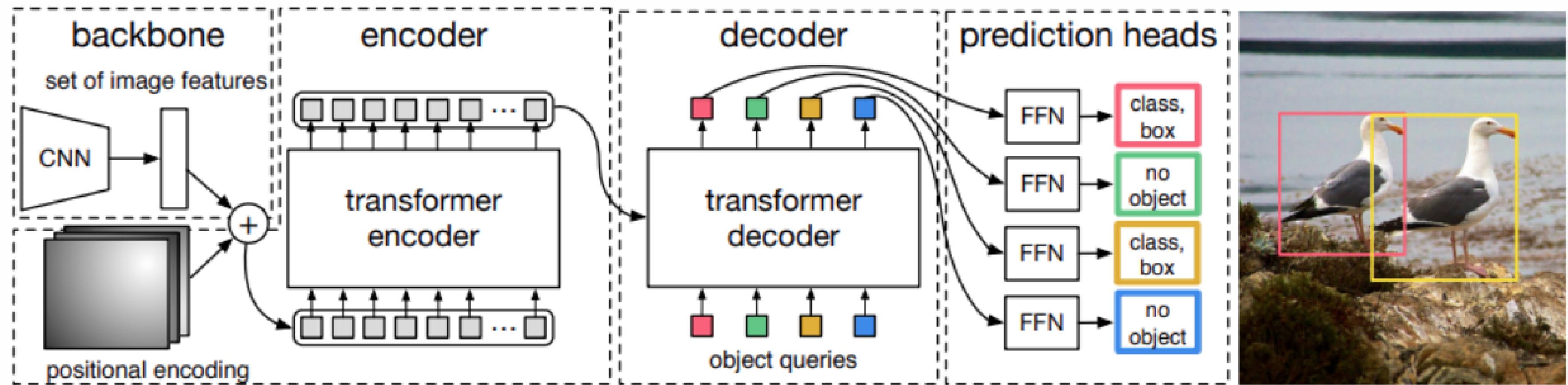
YOLO Object Detection Model



- Single-shot detection: Object detection in a single pass.
- Real-time capabilities: Suitable for applications requiring speed.
- Efficient for capturing text in images.
- used to detect custom objects like dropdown, buttons, icons and other UI webpage elements

Model Selection - Object Detection

DETR Object Detection Model



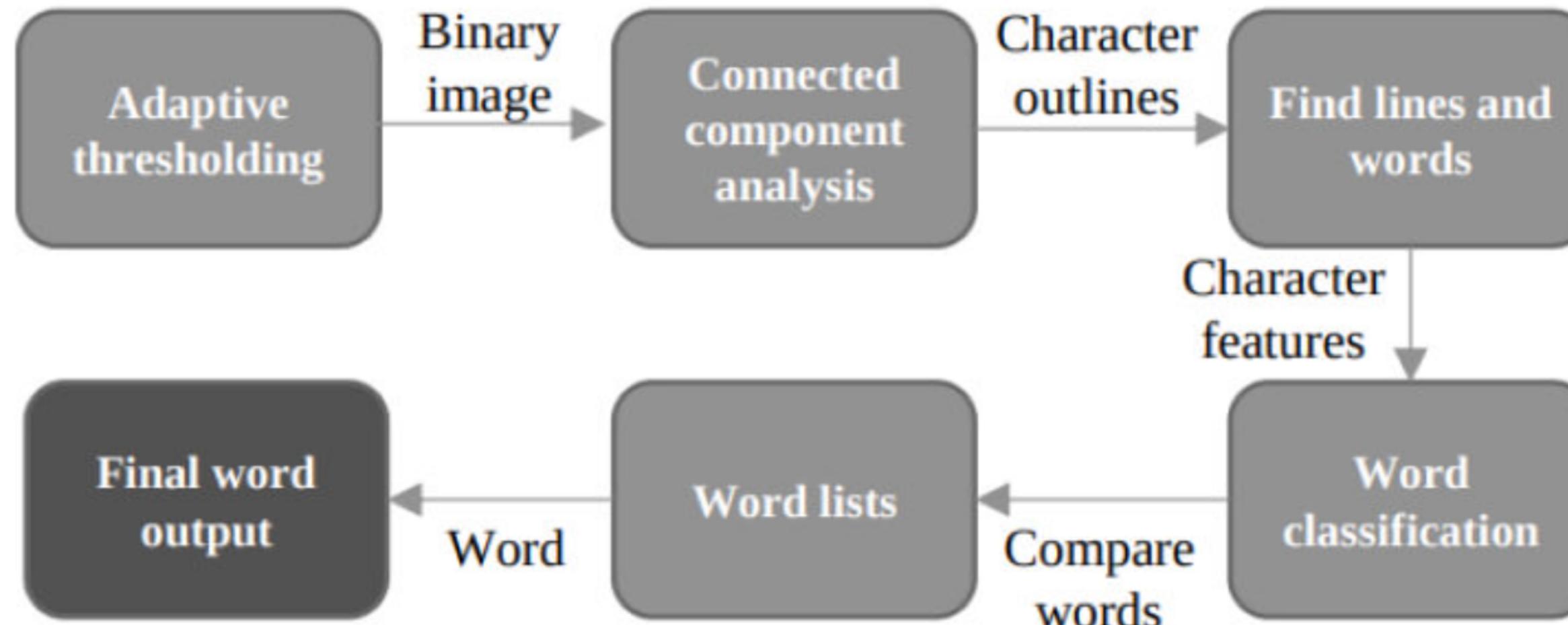
- DETR introduces transformers to object detection.
- Transformer architecture: Leveraging attention mechanisms.
- DETR can be applied to detect and extract text regions.

Comparison & Justification - Object Detection

	YOLO	DETR
Features	Single-shot object detection.	Transformer architecture.
Hardware Requirements	Efficient on GPU for real-time inference.	Requires significant computational resources.
Accuracy	High accuracy in object detection.	Competitive accuracy in various tasks.
Approach	Grid-based object detection.	Transformer encoder-decoder architecture.
Strengths	Real-time processing for video streams. Efficient for scenarios with multiple objects.	Contextual understanding with attention mechanisms. Handles object detection and segmentation jointly.
Limitations	Prone to false positives in complex scenes. Limited accuracy in detecting very small objects.	Training can be slower compared to other methods. Limited interpretability due to transformer complexity.

Model Selection - Optical Character Recognition

Tesseract OCR



- Uses image processing techniques and deep learning approach (CNN + RNN) for text recognition
- Used for extracting the text from the GUI elements detected from YOLO/DETR

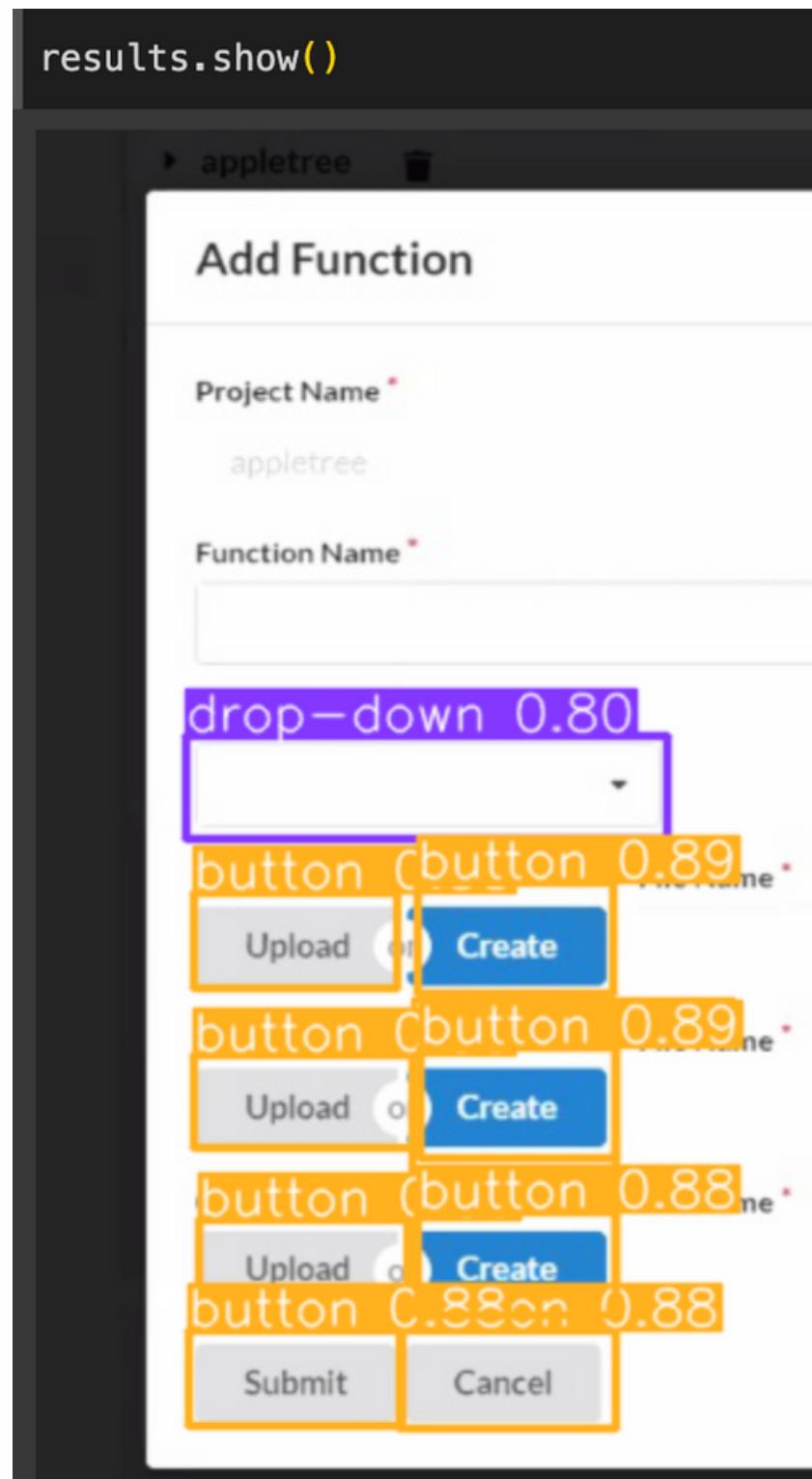
Sample Model Results - Text Q&A

Model Results from Transformers

True Answers	Predicted Answers
Click on the project name for which you want to download the data. After clicking on project, click on either context, input or output. ... Click on the sign up button	Click on the project name for which you want to download the model. After clicking on project, click on either context, input or output.... Click on the sign up button
Click on the Test Suite module from left side. You can see the existing Test Cases. You can regenerate test cases by clicking on an 'reg... Firstly, make sure that you have an expected output for the test-cases for which you will perform data augmentation. Select the test cas...	Click on the Test Suite module from left side. You can see the existing Test Cases. You might have selected an existing test case. Anoth... After selecting a project, click on Data Augmentation category from the left panel. In this screen, you need to add test case(s). After ...
After selecting a project, click on Data Augmentation category from the left panel. In this screen, you need to add test case(s). After ... To download the input json file, click on project name then click on output button, you can download the file by clicking on the arrow b...	Firstly, make sure that you have an expected output for the test-cases for which you will perform data augmentation. Select the test cas... To download the input json file, click on project name then click on output button, you can download the file by clicking on the arrow b...
On the 'Test Suite' page, click on the 'Export Testcases' button. You have an option to export 'Base Test Cases' , 'Augment Test Cases' ... Click on the 'Project Management' tab , select the required function and click on the 'Test Suite' tab. Upload the APK file on the 'Test...	On the 'Test Suite' page, click on the 'Export Testcases' button. You have an option to export 'Base Test Cases', 'Augment Test Cases'or... Click on the 'Project Management' tab, select the required function and click on the dropdown at the top right corner. After clicking on...
Yes absolutely, the tree can be downloaded in png format!	Yes absolutely, the tree can be downloaded in pn format!

Results - Object Detection & Text Extraction

UI elements



Coordinates

```
tensor([149.61563, 356.34805, 225.73930, 399.19644, 0.88746, 3.00000], device='cuda:0')
Create
tensor([148.13445, 292.52332, 224.45210, 335.79282, 0.88678, 3.00000], device='cuda:0')
Create
tensor([148.33340, 420.55496, 225.63832, 460.18738, 0.88468, 3.00000], device='cuda:0')
Create
tensor([ 61.91154, 296.69974, 140.44658, 333.08649, 0.88407, 3.00000], device='cuda:0')
Upload
tensor([ 59.36609, 466.30600, 142.46973, 503.95398, 0.88120, 3.00000], device='cuda:0')
Submit
```

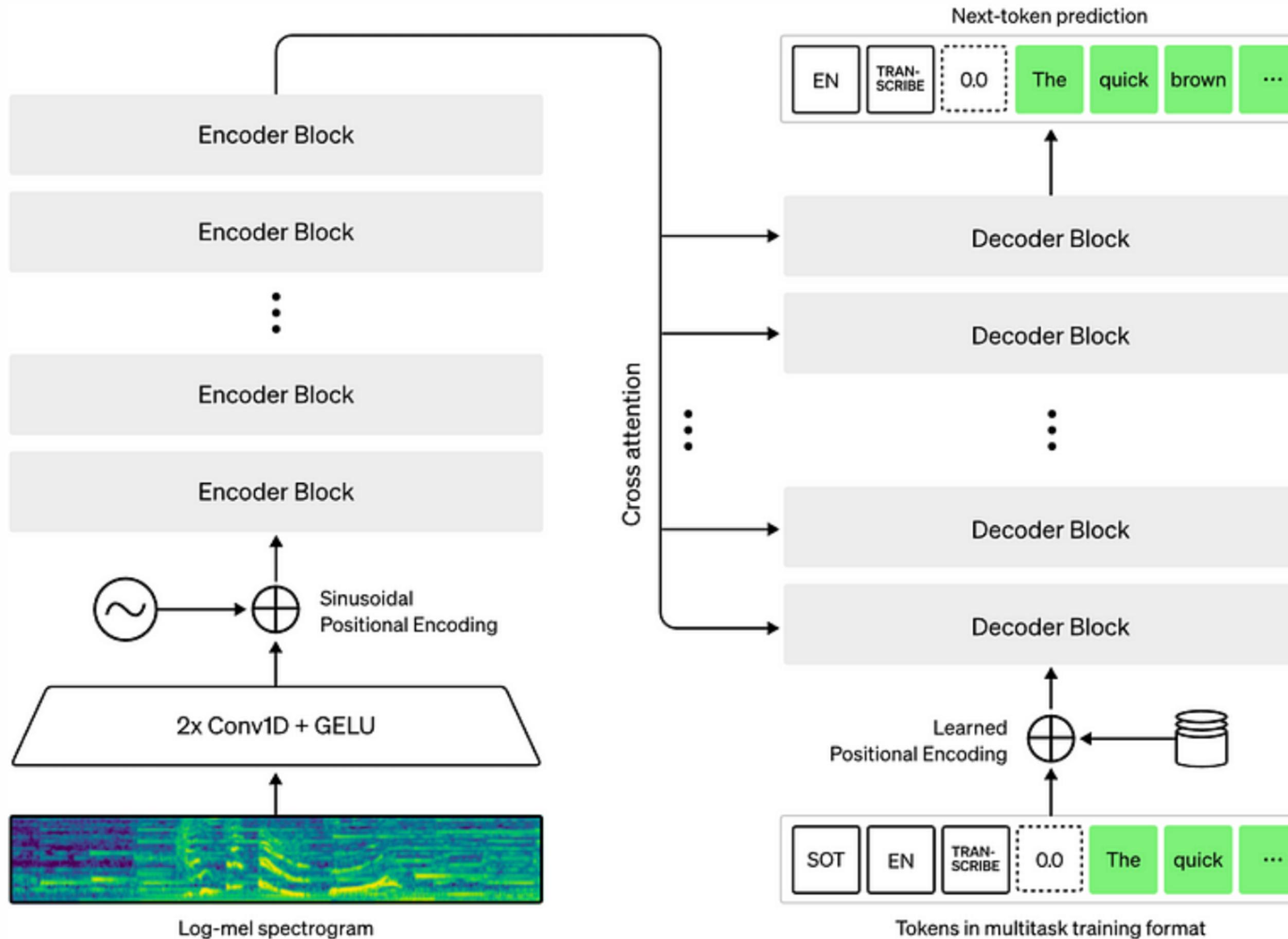
Text extraction

```
▶ for i in temp:
    print(i[0].replace('\n', '').replace('\x0c', ''))
```

👤 Create
👤 Upload
👤 'Submit'
👤 Cancel
👤 Upload
👤 Upload

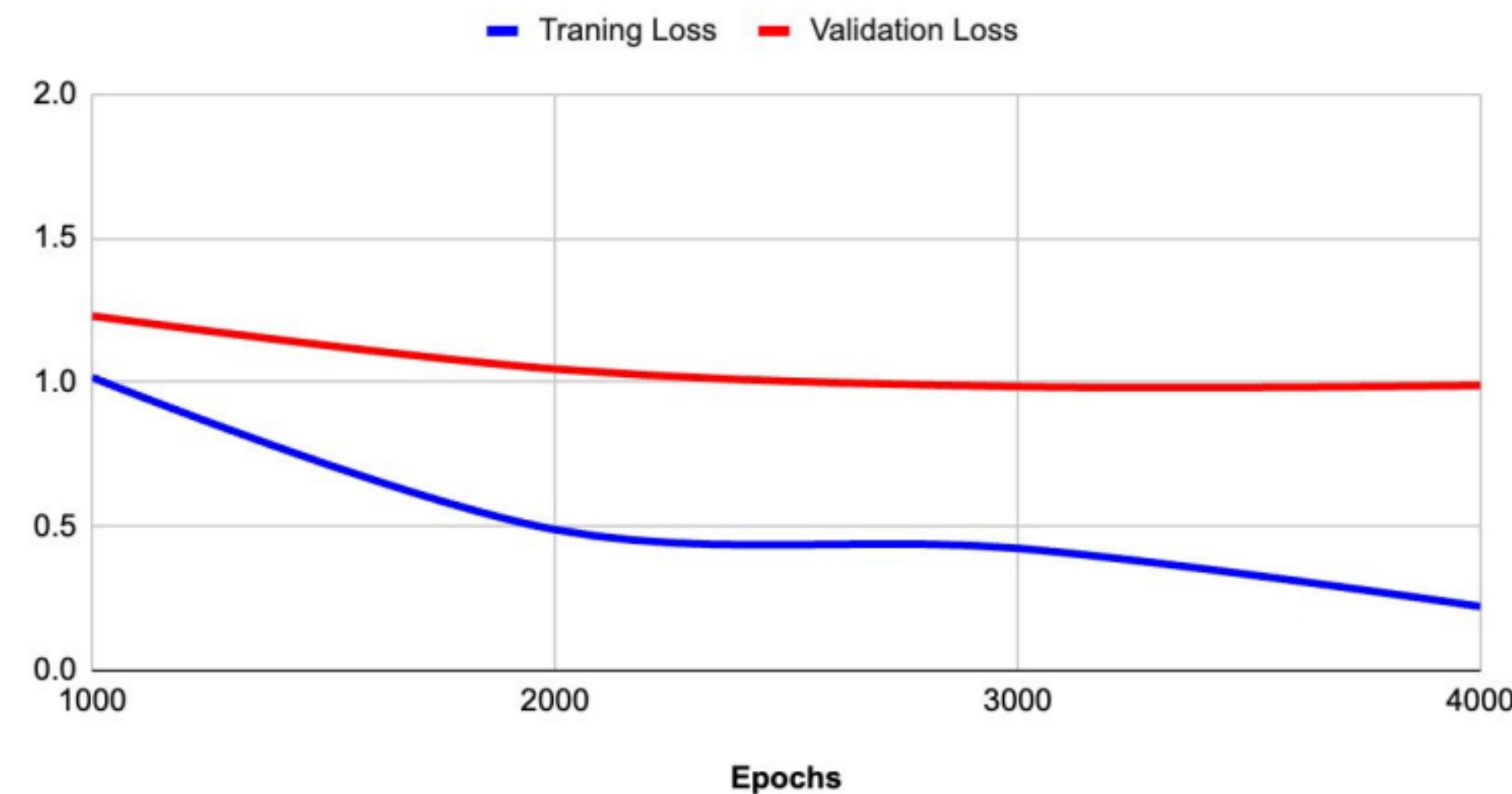
ferredCreate

Whisper ASR Model

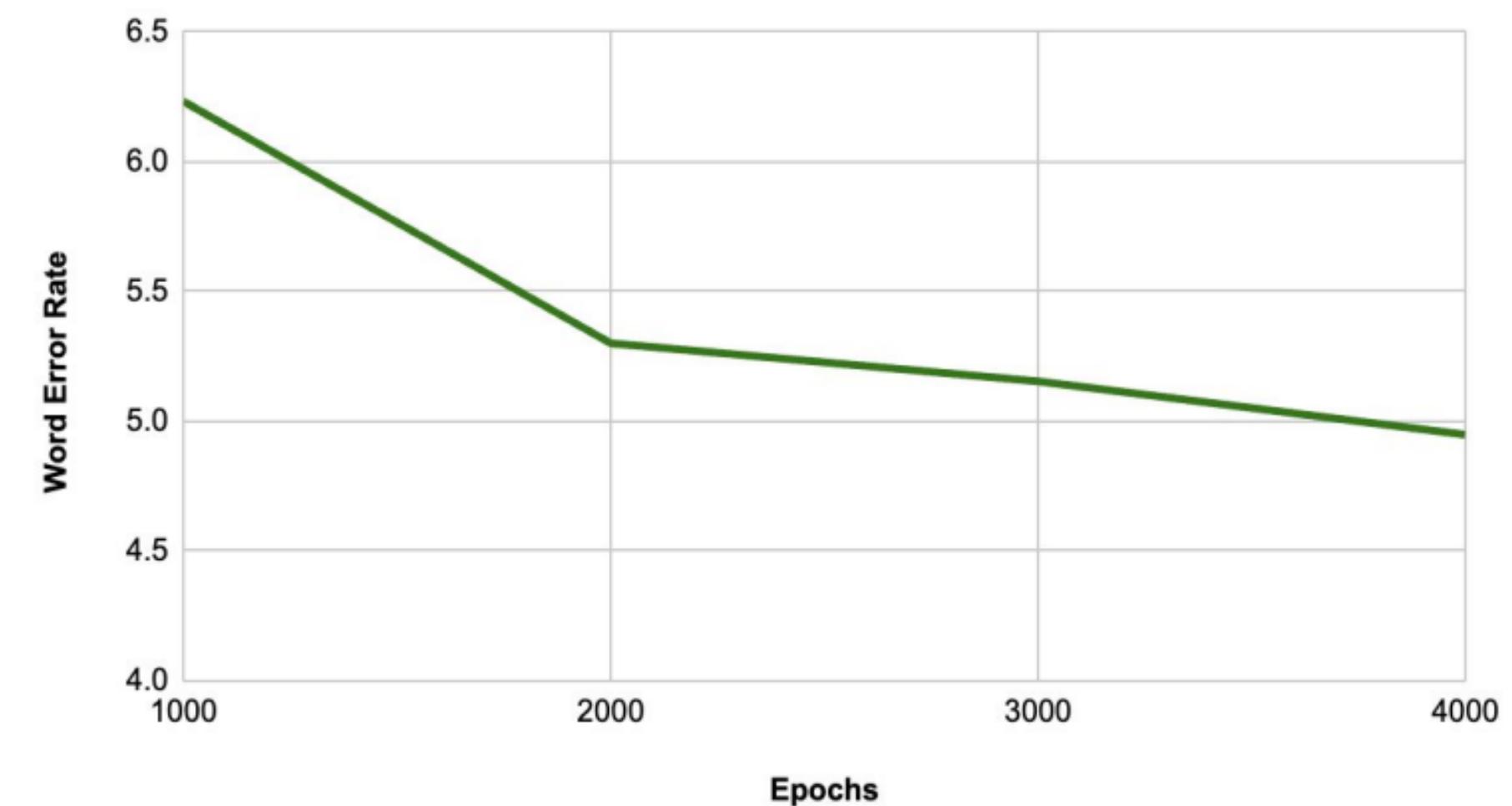


Whisper model results

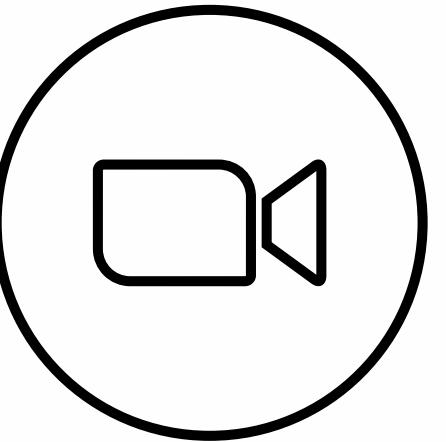
Training v/s Validation Loss



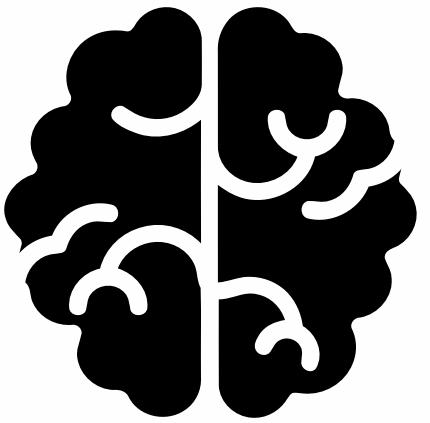
Word Error Rate



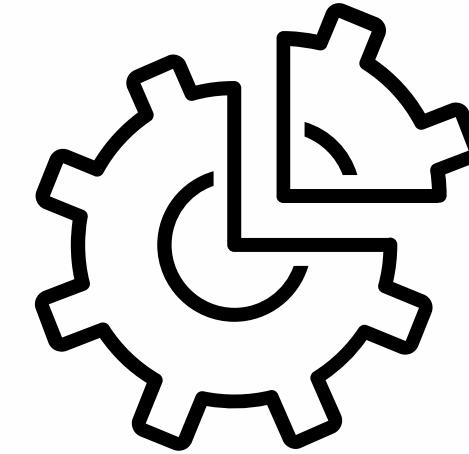
Future Implementation



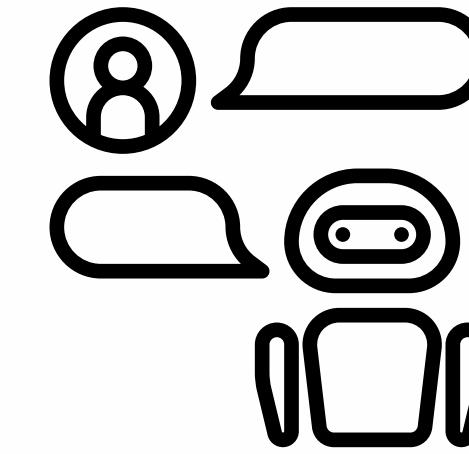
Zoom meeting
API integration
in the tool



Automatically
convert any
text corpus into
knowledge
Graph

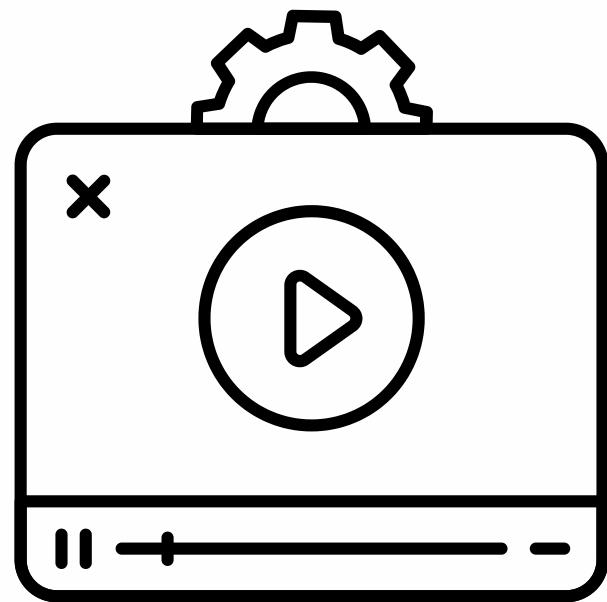


Integration Model



Chatbot GUI for users

CHATBOT DEMO



Llama_2 Text Chatbot

The screenshot shows a Google Colab notebook titled "Llama_2_Chatbot.ipynb". The notebook interface includes a toolbar at the top with various icons, a tab bar showing the current file, and a sidebar on the left with icons for code, text, copy to Drive, and a search bar.

The main area contains the following Python code:

```
[13]: #!pip install llama-index  
#!pip install openai  
  
[5]: from llama_index import VectorStoreIndex, SimpleDirectoryReader  
import os  
  
os.environ['OPENAI_API_KEY'] = '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'  
  
[6]: from pathlib import Path  
from llama_index import download_loader  
  
JSONReader = download_loader("JSONReader")  
  
loader = JSONReader()  
documents = loader.load_data(Path('/content/QnA.json'))  
  
[7]: # UPDATE: Since llama_index changed their library, the following code should replace the code above  
index = VectorStoreIndex.from_documents(documents)  
  
[8]: # Saving Index for future use. Run this cell if you need to save the index  
index.storage_context.persist()  
  
# Loading Index from local storage. Run this cell if you want to load an index to resume  
from llama_index import StorageContext, load_index_from_storage  
  
storage_context = StorageContext.from_defaults(persist_dir='./storage')  
index = load_index_from_storage(storage_context)
```

Below the code, there is a section titled "Querying the index and getting a response". At the bottom of the notebook, a status bar indicates "0s completed at 7:44PM".

YOLO_OCR Image Models

The screenshot shows a Mac desktop with a Jupyter Notebook titled "YOLO_OCR.ipynb" open in a browser window. The notebook is running on Google Colab. The code cell contains Python code for YOLOv5 inference:

```
[ ] # YOLOv5 PyTorch HUB Inference (DetectionModels only)
import torch
from google.colab.patches import cv2_imshow
import cv2

#model = torch.hub.load('ultralytics/yolov5', 'yolov5s', force_reload=True) # yolov5n - yolov5x6 or custom
model = torch.hub.load('.', 'custom', path='/content/best.pt', source='local')
im = '/content/yolov5/images/image_16.jpg' # file, Path, PIL.Image, OpenCV, nparray, list
results = model(im) # inference
results.print() # or .show(), .save(), .crop(), .pandas(), etc.
```

The output of the code cell shows the model's configuration and performance metrics:

```
YOLOv5 v7.0-168-gec2b853 Python-3.10.11 torch-2.0.0+cu118 CUDA:0 (Tesla T4, 15102MiB)

requirements: /usr/local/lib/python3.10/dist-packages/requirements.txt not found, check failed.
Fusing layers...
Model summary: 267 layers, 46194438 parameters, 0 gradients, 107.9 GFLOPs
Adding AutoShape...
image 1/1: 550x900 9 buttons, 1 drop-down
Speed: 40.7ms pre-process, 50.0ms inference, 52.8ms NMS per image at shape (1, 3, 416, 640)
```

The next code cell shows the results being converted to a Pandas DataFrame:

```
[ ] temp = results.pandas()
```

```
[ ] temp.xyxy[0]
```

	xmin	ymin	ymax	confidence	class	name
0	149.615631	356.348053	225.739304	0.887462	3	button
1	148.134445	292.523315	224.452103	0.886779	3	button
2	148.333405	420.554962	225.638321	0.884682	3	button

ASR_Whisper Model



Inmute (m)