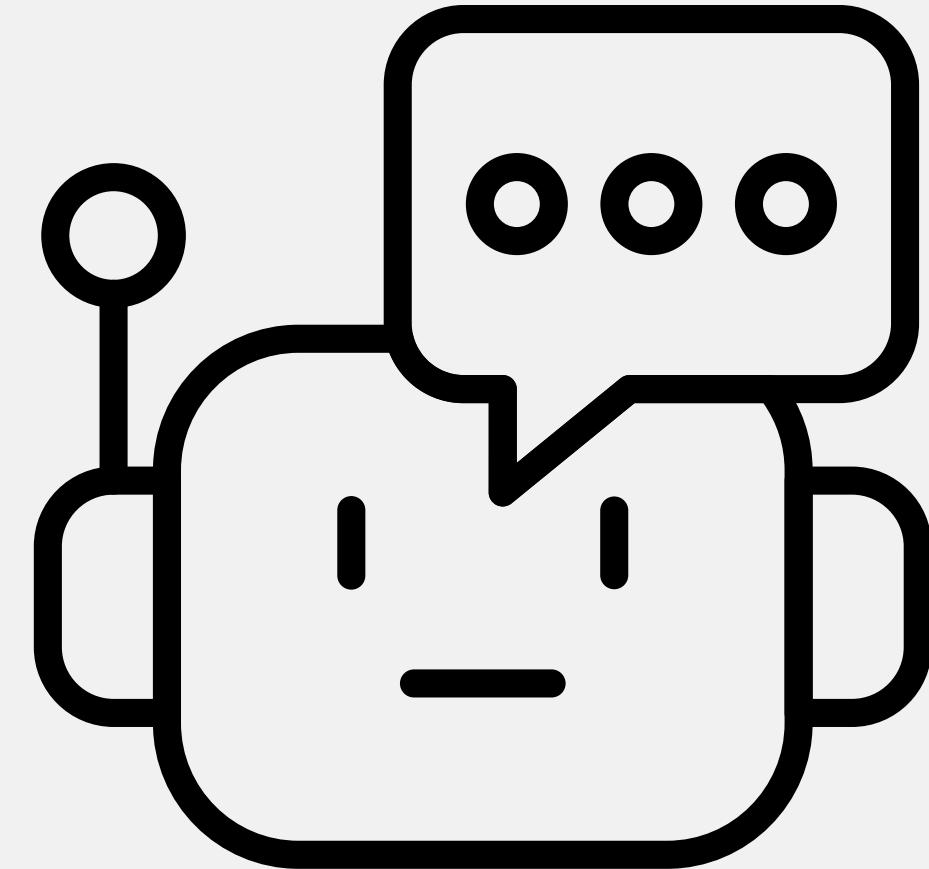


ZIM AI Assistant: Enhancing User Experience with Machine Learning



Machine Learning Capstone Project

BY: SUHA ISLAIH

suha@mcode.club

Sept 4, 2024

Dan's Problem & the ZIMJS Website

- **Dan's Problem:** Users struggle to find quick, accurate answers to their code-related questions on the ZIMJS website.
- **ZIMJS Website Overview:** A platform using the ZIMJS library to create animations, drawings, games, and apps with JavaScript. It includes documentation, tutorials, and a community forum.



ZIM 016



► Discord ► Forum



ABOUT

EXAMPLES

LEARN

EDITOR

CODE

DOCS

DEVS

GOLD

ABOUT

INTRO

VISION

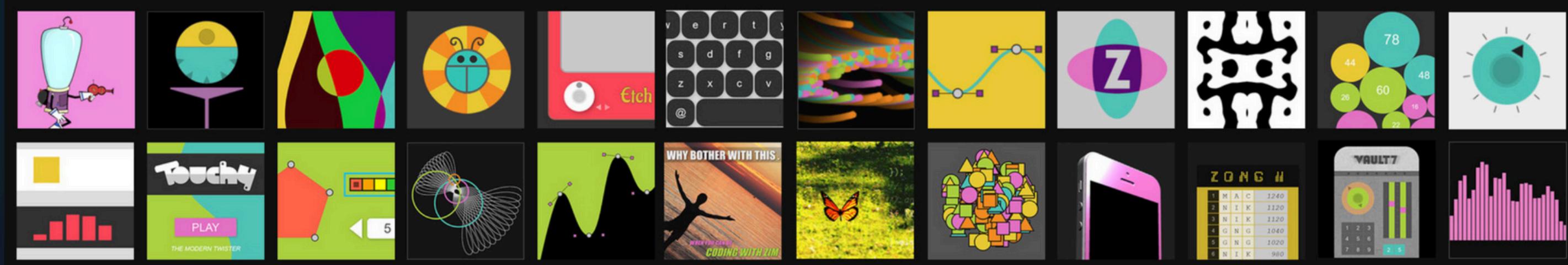
FEATURES

VERSIONS

REVIEWS

FOUNDER

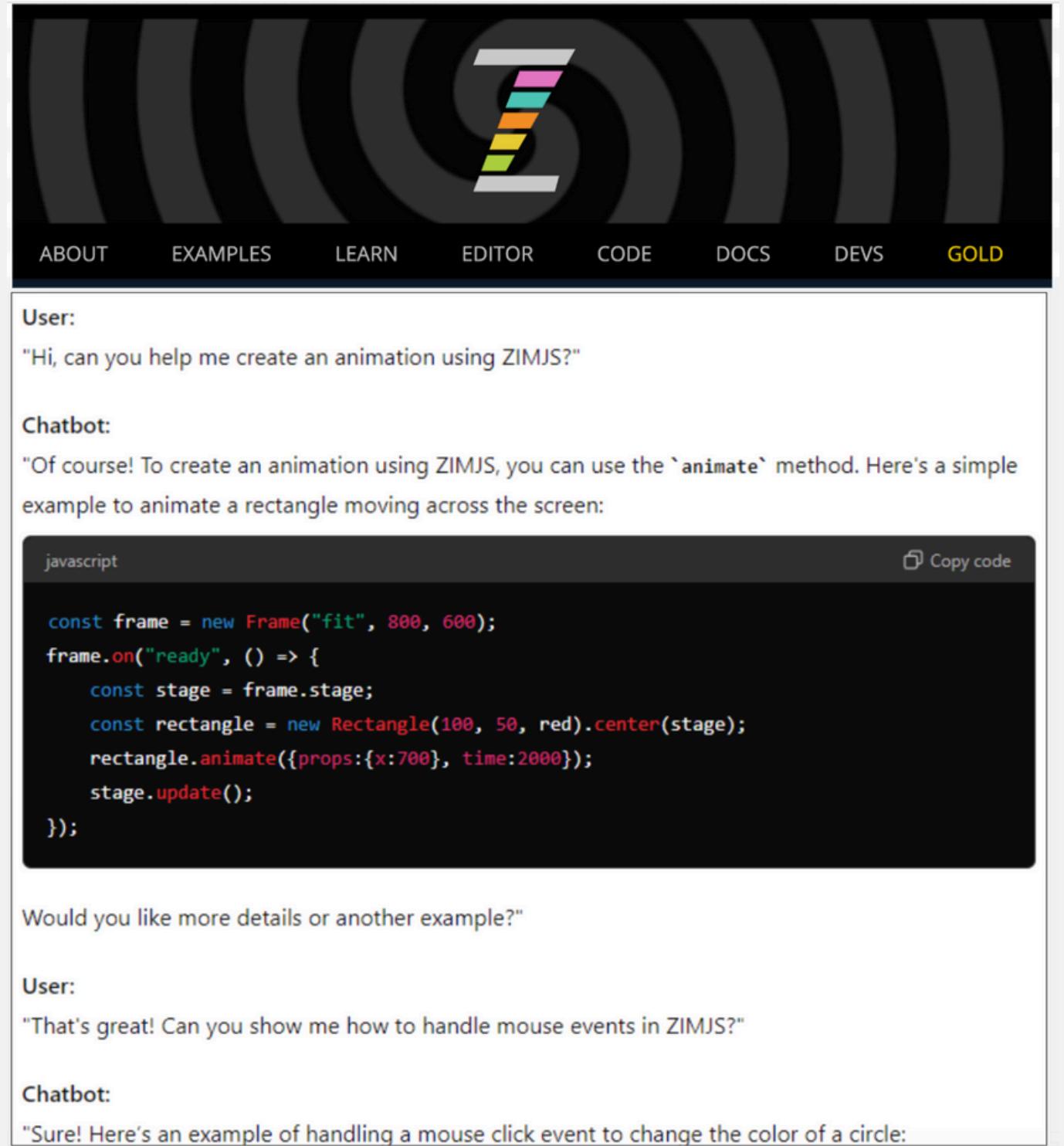
INTRO



Introduction

Project Goal

- **Objective:** To develop a chatbot to answer user questions related to the ZIMJS website.
- **Key Contribution:** Enhancing the user experience by providing quick, accurate responses to code-related queries, saving time and improving engagement.



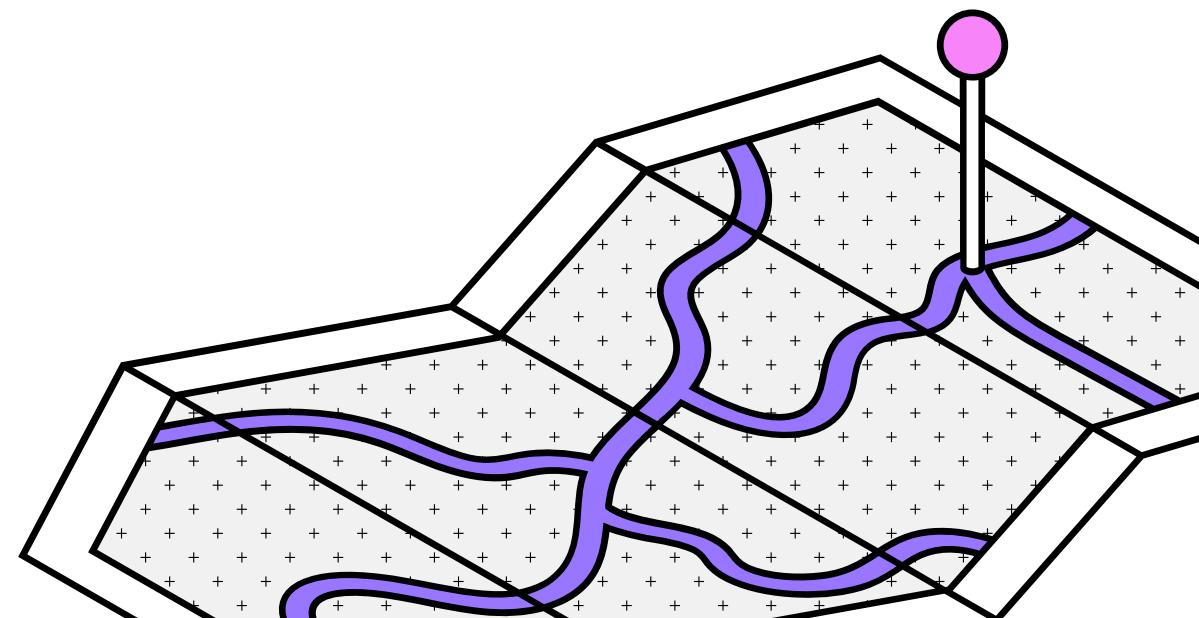
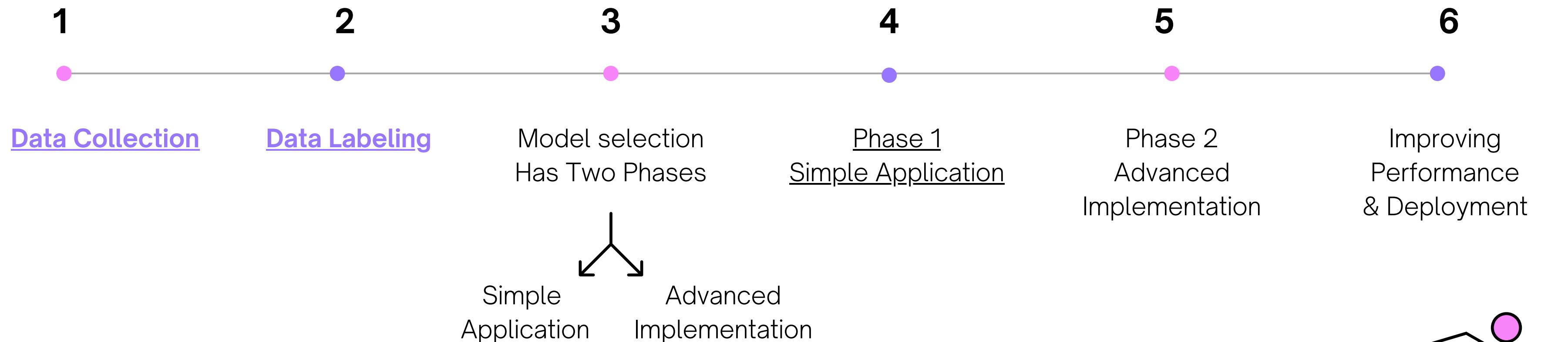
The screenshot shows the ZIMJS website's interface. At the top is a dark header bar with the ZIMJS logo and navigation links: ABOUT, EXAMPLES, LEARN, EDITOR, CODE, DOCS, DEVS, and GOLD. Below the header is a light-colored main area. A user message "Hi, can you help me create an animation using ZIMJS?" is followed by a chatbot response: "Of course! To create an animation using ZIMJS, you can use the `animate` method. Here's a simple example to animate a rectangle moving across the screen:". Below this is a code block in JavaScript:

```
javascript
const frame = new Frame("fit", 800, 600);
frame.on("ready", () => {
    const stage = frame.stage;
    const rectangle = new Rectangle(100, 50, red).center(stage);
    rectangle.animate({props:{x:700}, time:2000});
    stage.update();
});
```

At the bottom, the user asks if there's more information, and the chatbot responds with another question about handling mouse events.

Approach

Data Collection, Labeling and Model Selection



Data Collection

Comprehensive data from the ZIMJS website

Documentation

ZIM FRAME

- EXPAND ▾ Frame(scaling, width, height)
- EXPAND ▾ Pic(file, width, height)
- EXPAND ▾ Aud(file, volume, loop)
- EXPAND ▾ Vid(file, width, height)
- EXPAND ▾ SVG(svg, width, height)
- EXPAND ▾ Speech()
- EXPAND ▾ Fonts – loaded
- EXPAND ▾ Keys – keyboard
- EXPAND ▾ Cursors – custom
- EXPAND ▾ Tilt – device motion
- EXPAND ▾ PermissionAsk(permission, message)
- EXPAND ▾ Colors – orange
- EXPAND ▾ Constants – FIT, FILL, CENTER
- EXPAND ▾ Globals – F, S, L, P, S, Pizza

Lessons & Tutorials

REFERENCE

LITERALS

You will recall from Lesson 01 that objects are made from classes called a **literal**:

01

```
7; // a Number literal for new Number(7);
"hi"; // a String literal for new String("hi");
true; // a Boolean literal for new Boolean(true);
[1,2,3]; // an Array literal for new Array(1,2,3);
{property:"value"}; // an Object literal for new Object
```

OBJECT LITERALS ► [What is Object Literal](#)

We are interested in the Object Literal. An **Object Literal** holds pairs of names and values. A colon (:) separates the name from the value. A comma (,) separates the pairs.

02

```
const obj = {name:"Dr. Abstract", age:327};
```

Code Examples

EDITOR

Phone L P S Pizza

```
1 // Given F, S, W, H or frame, stage, stageW, stageH
2
3 F.color = red;
4
5 // FEATURE - ZIM DUO
6
7 // In many cases, ZIM lets you add parameters in two ways:
8 // 1. regular with parameter values in order
9 // 2. a single configuration object with properties
   matching parameters
10 // You pick whichever is more convenient!
11
12 new Rectangle(600, 300, yellow, null, null, 70).center();
13 new Rectangle({width:500, height:200, corner:50}).center();
14
15 // The docs will have a SUPPORTS ZIM DUO at the top
16 // of the parameters section if ZIM DUO is available
```

ZAPPS VIEW PROMO DOCS

Forums

categories ▾ tags ▾ Categories Latest Top

An Announcements ZIM release information with emoji level. Also see versions in ZIM About.

Qu Questions Please ask any ZIM questions here. Also see ZIM Tips.

Re Requests Request a feature to be added to ZIM. Also see ZIM Docs for existing...

Sh Showcase Post your works, all are welcome, from prototype to professional!

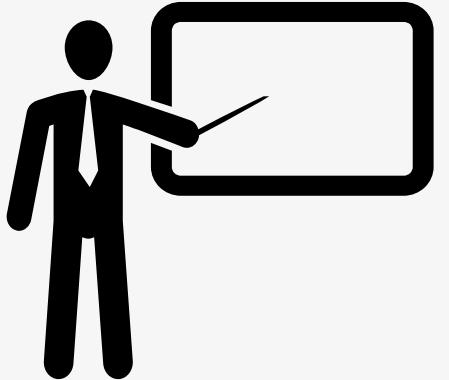
Data Labeling

Categorizing the data into labels

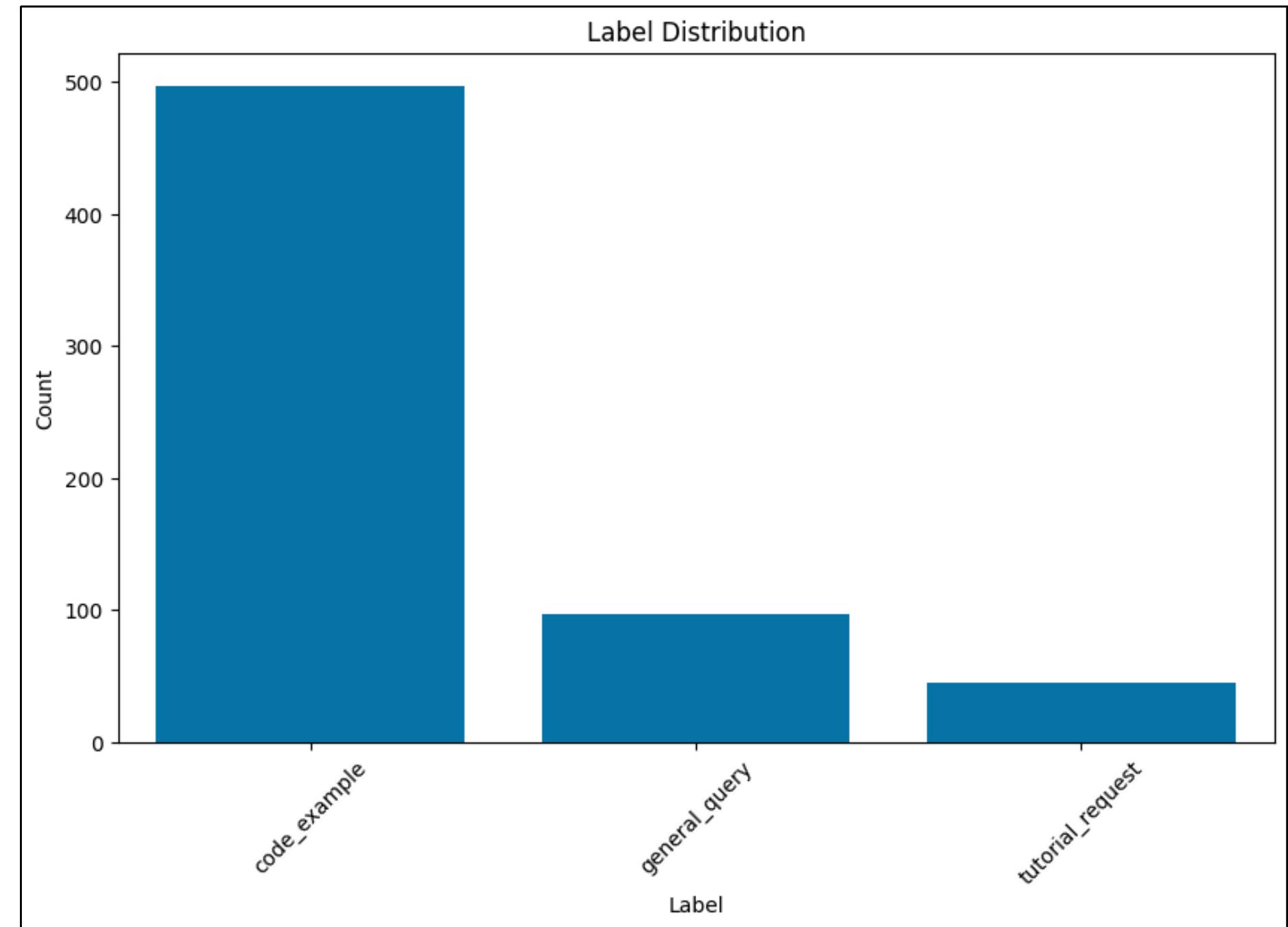
general_query

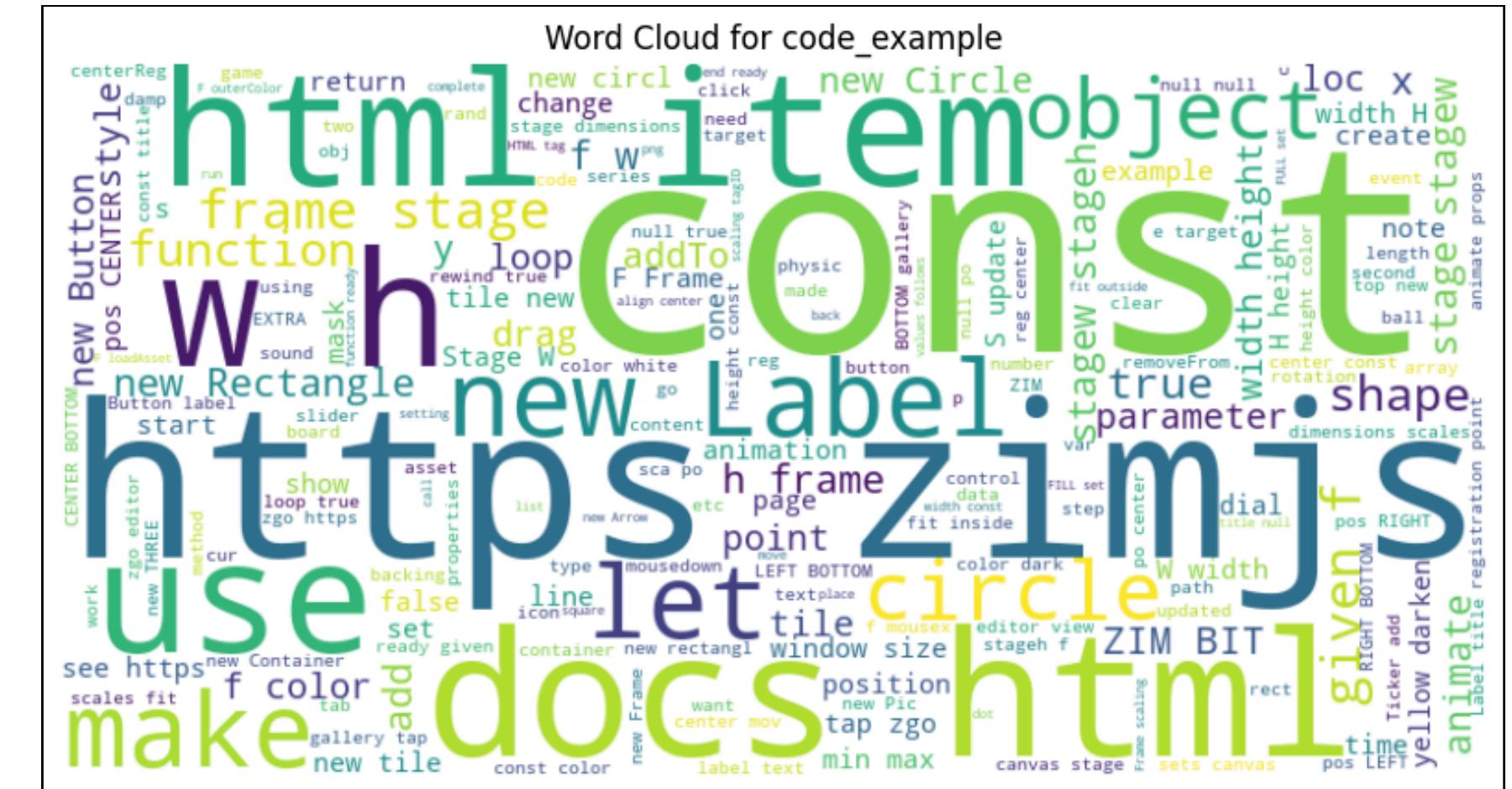
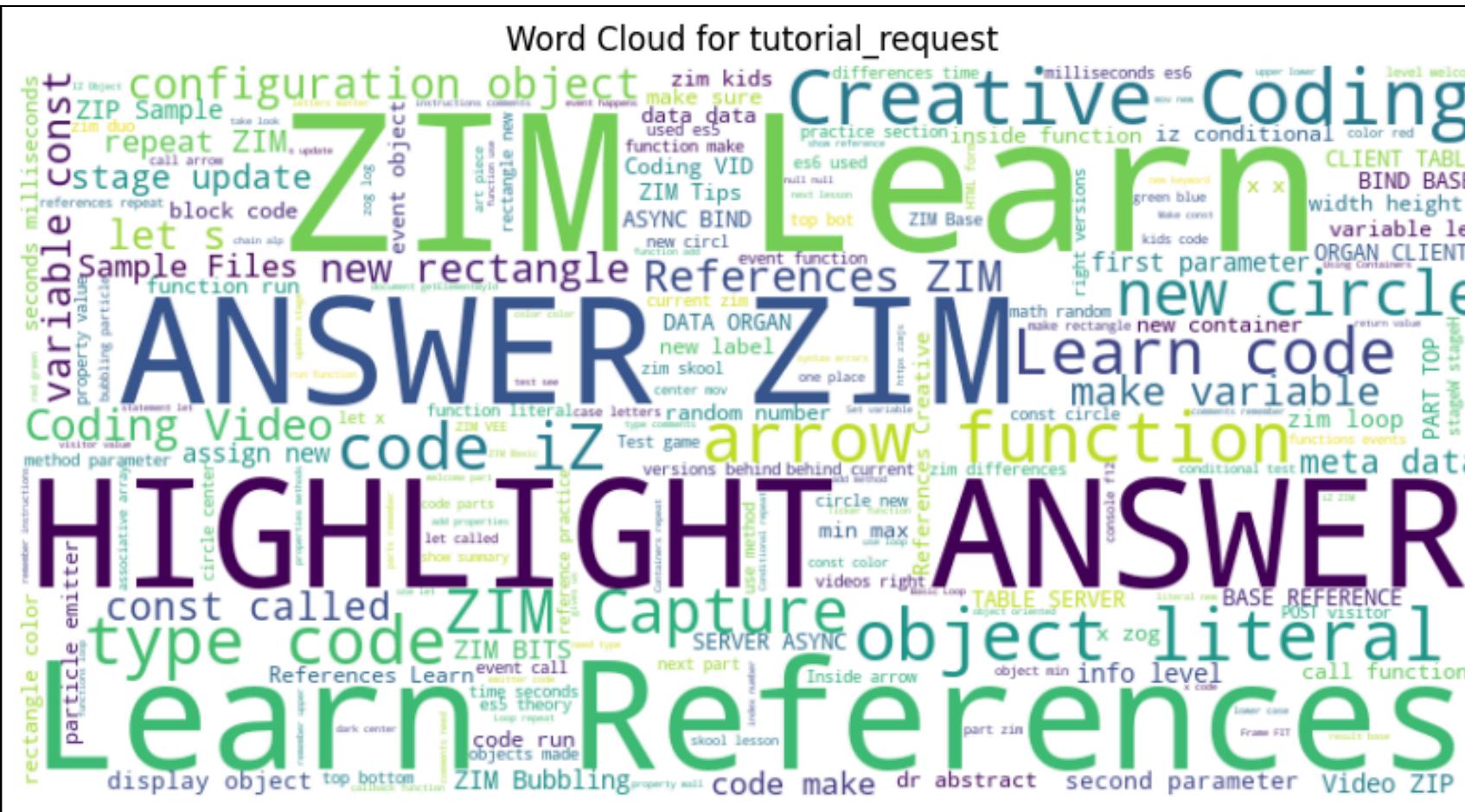
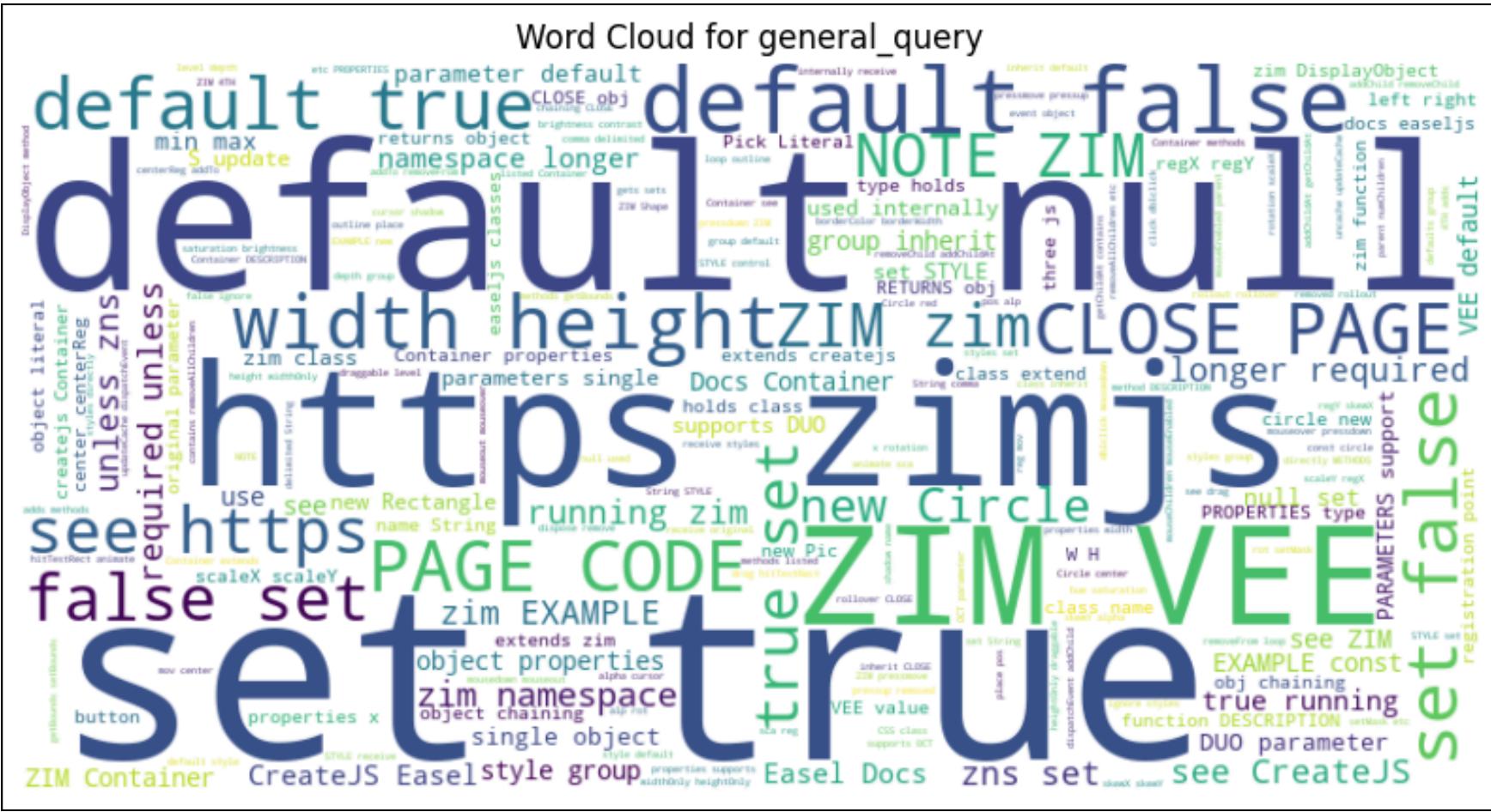


tutorial_request



code_example



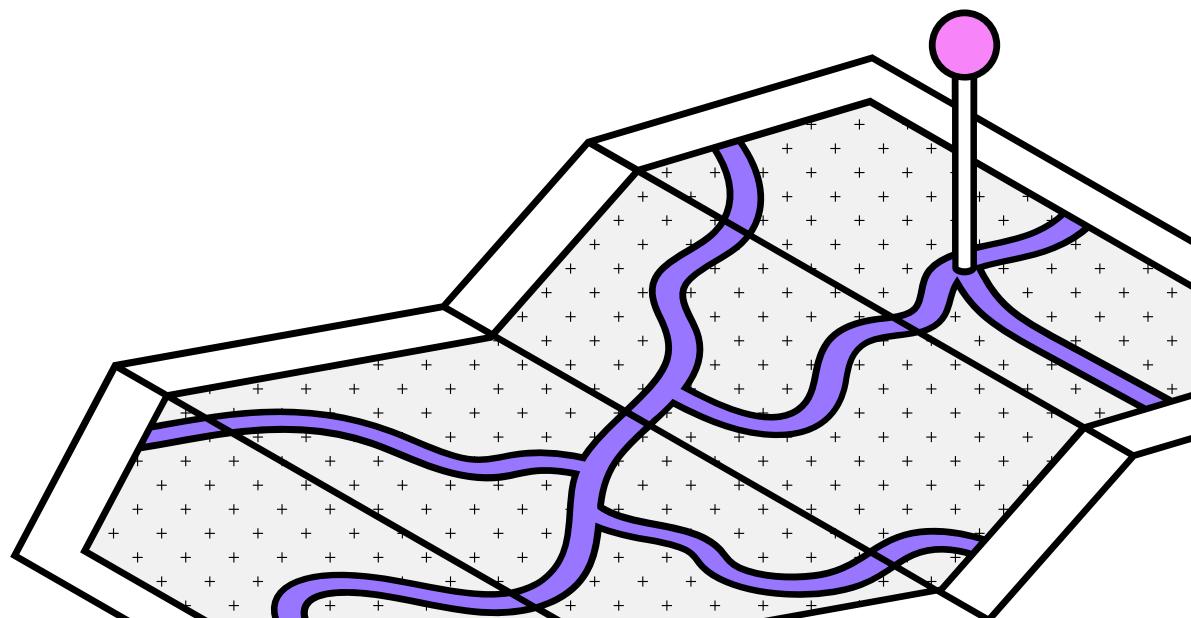
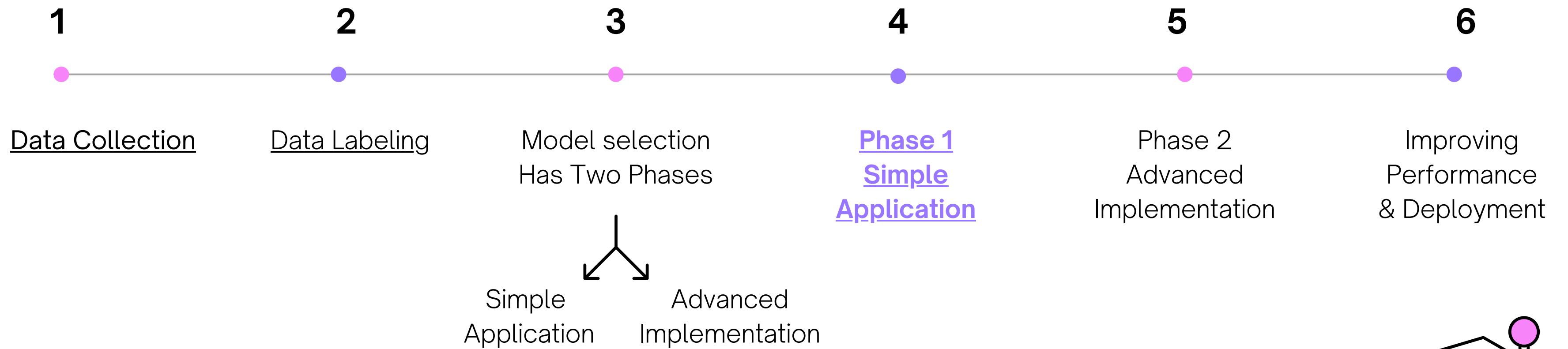


Dataset for ZIM website

label	title	url	content
code_example	accent test	https://zimjs.com/editor/code/Z_97DYP	// Given F , S , W , H or frame , stage , stag...
code_example	Slice a Blob (Egg) to split into two Blobs	https://zimjs.com/editor/code/Z_XTREU	// Given F , S , W , H or frame , stage , stag...
code_example	Pages with Arrows	https://zimjs.com/editor/code/E_A5DXC	// Given F , S , W , H or frame , stage , stag...
code_example	Fireworks!	https://zimjs.com/editor/code/Z_6TMBS	// Given F , S , W , H or frame , stage , stag...
code_example	DoDoDots 1	https://zimjs.com/editor/code/Z_Q9SR8	// Given F , S , W , H or frame , stage , stag...
code_example	Parallax	https://zimjs.com/editor/code/Z_5W9MF	// Editor template built in see https : //zimj...

Approach

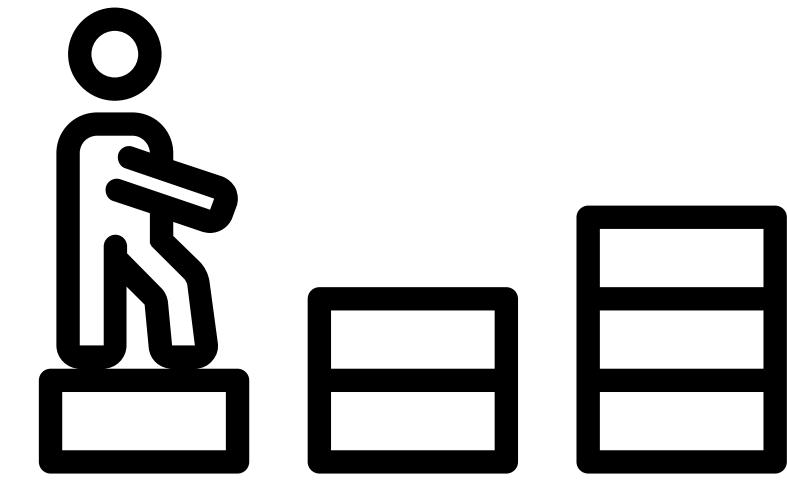
Data Collection, Labeling and Model Selection



Simple Application

Phase 1

Initial implementation of pre-trained 'bert-base-uncased' model from the Hugging Face library for intent classification, information retrieval, and basic conversational AI.



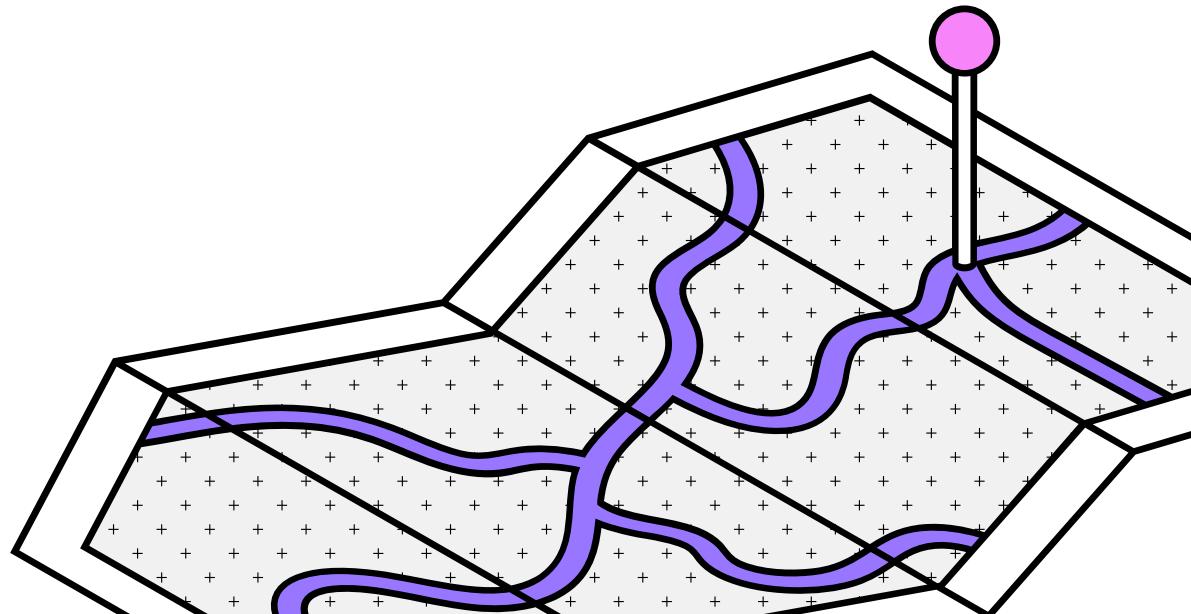
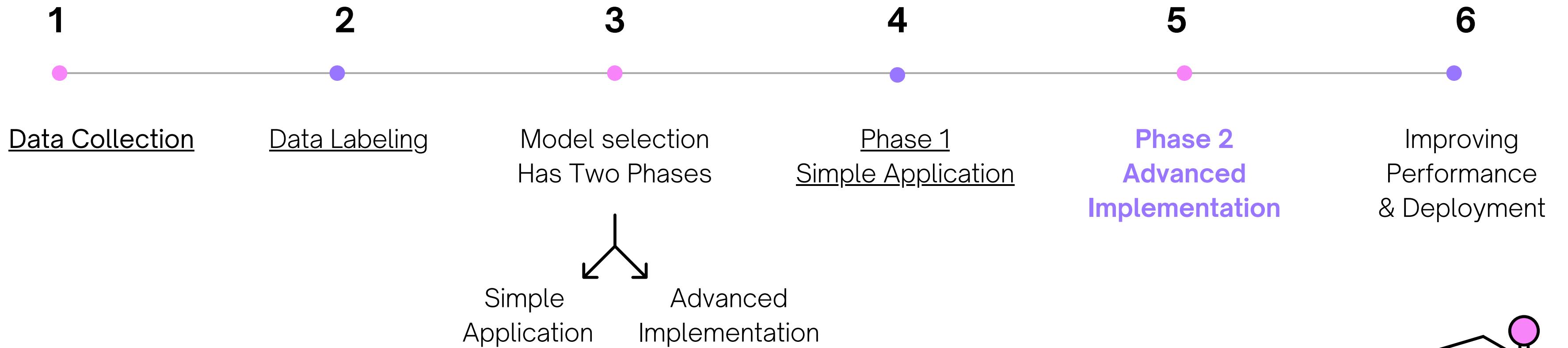
```
Question: "Parameters must be passed in order and cannot be skipped to get to a distant parameter."  
Predicted label: general_query
```

```
Question: "ZIM is a general Canvas Framework with simple, powerful JavaScript that lets everyone"  
Predicted label: general_query
```

```
Question: "const circle = new Circle(50, green, pink, 5) // radius, color, borderColor, borderWidth.loc(200, 200);"  
Predicted label: code_example
```

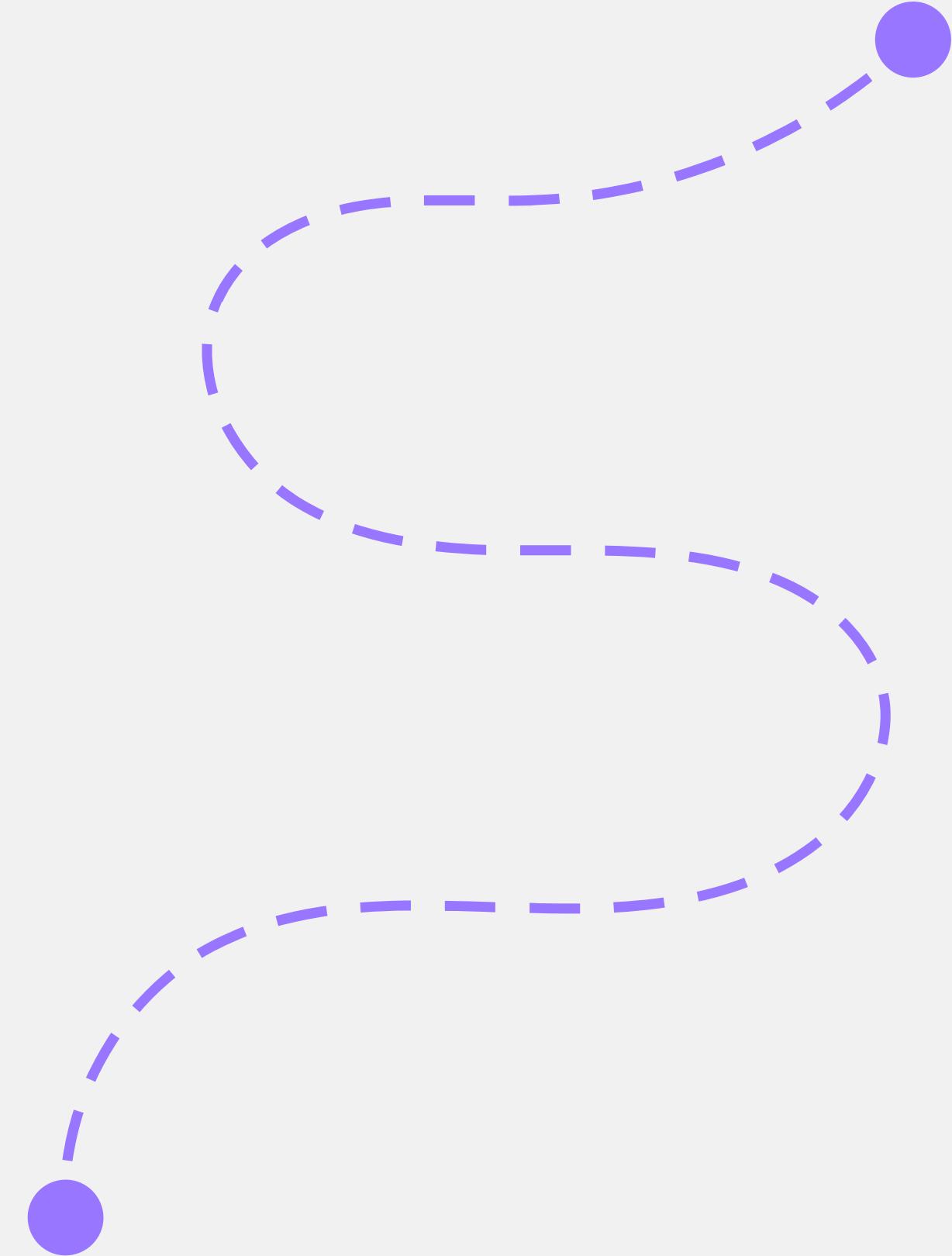
Approach

Data Collection, Labeling and Model Selection



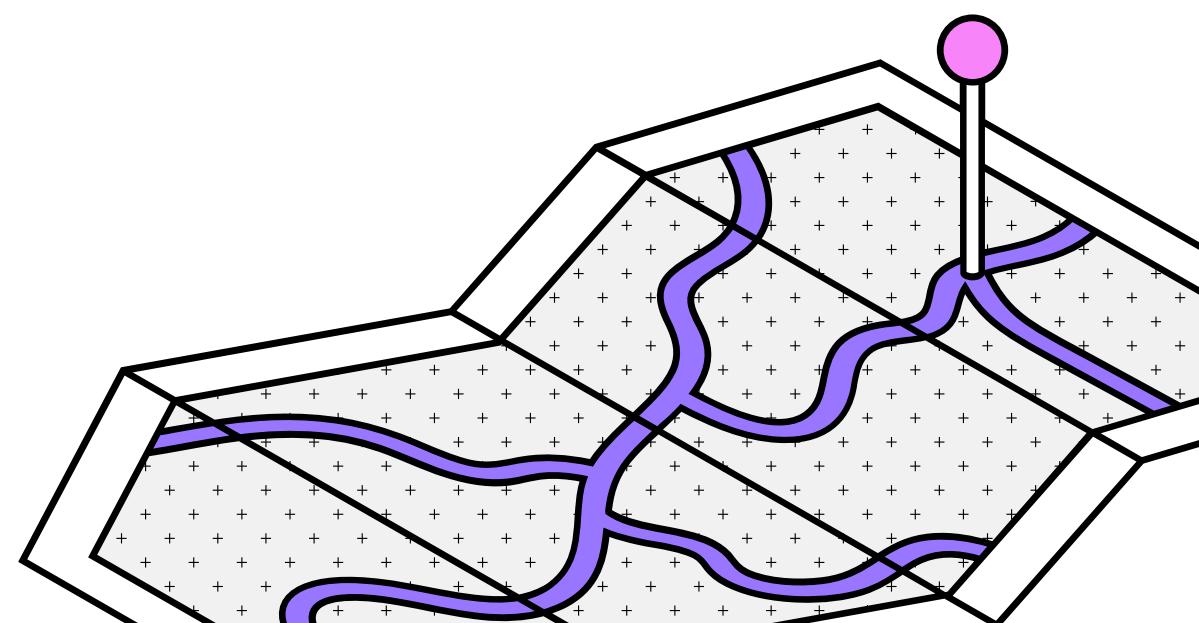
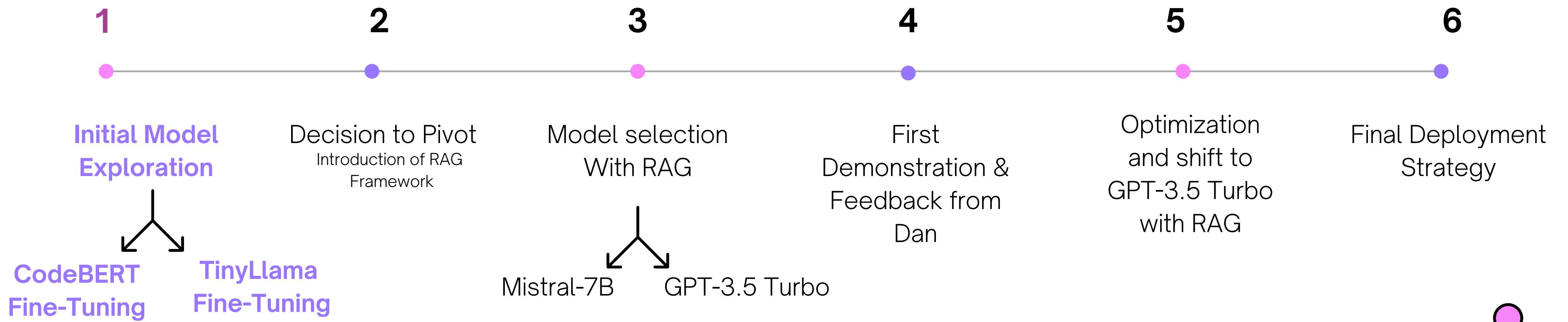
Advanced Modeling Pathway

**From Initial Attempts to Optimized
Deployment**



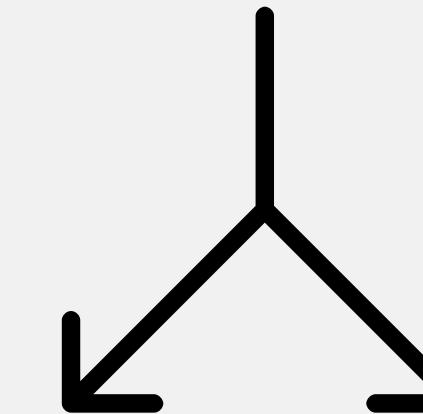
Advanced Modeling Pathway

From Initial Attempts to Optimized Deployment





Initial Model Exploration With Fine-Tuning

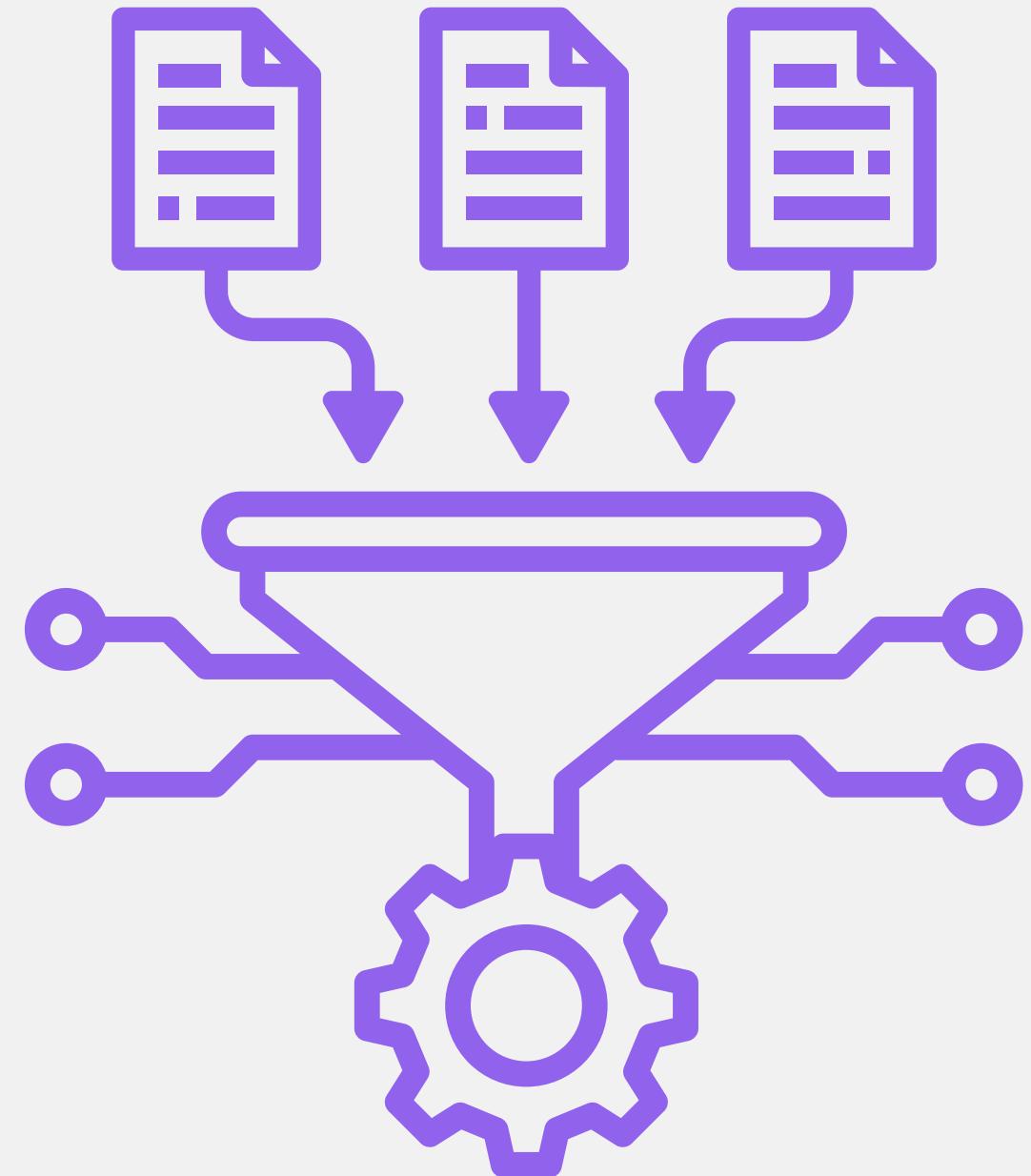


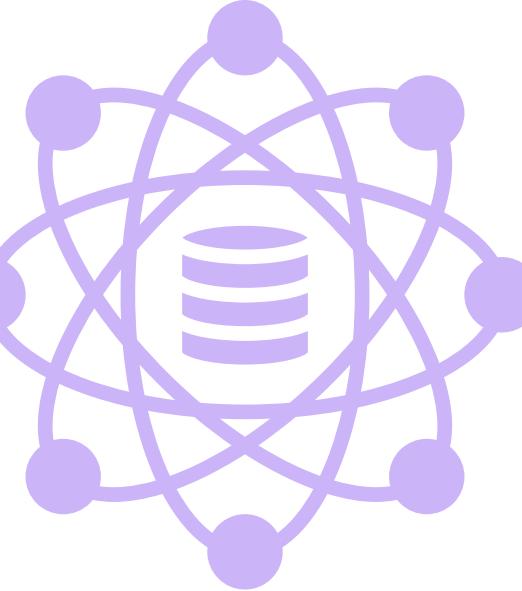
CodeBERT
Fine-Tuning TinyLlama
 Fine-Tuning



Data Preparation

Prepare the dataset with the same ZIMJS question-answer pairs and converted it into a format suitable for training with selected models

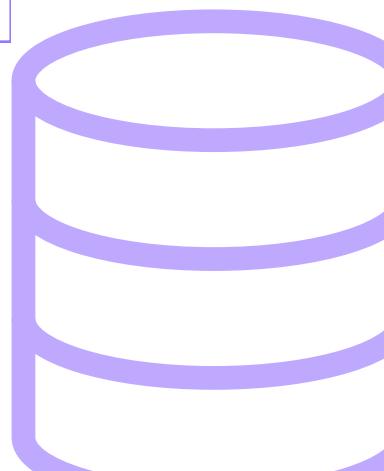




Original Dataset

```
df.head()
```

	file_name	title	url	content	label
0	Hit Tests.txt	Hit Tests	Code https://zimjs.com/editor/code/Z_FJQWU	// Editor template built see https://zimjs.com/editor/code/Z_FJQWU	code_example
1	Snapping objects to align.txt	Snapping objects to align	Code https://zimjs.com/editor/code/Z_MGRQS	// given f , , w , h frame , stage , stagew , ...	code_example
2	Blob & Squiggle.txt	Blob & Squiggle	Code https://zimjs.com/editor/code/Z_9FHGJ	// Editor template built see https://zimjs.com/editor/code/Z_9FHGJ	code_example
3	Timeline Test.txt	Timeline Test	Code https://zimjs.com/editor/code/Z_AMGMT	const head = new layer (400 , 400 , head) .c...	code_example
4	DoDoDots 1.txt	DoDoDots 1	Code https://zimjs.com/editor/code/Z_Q9SR8	// given f , , w , h frame , stage , stagew , ...	code_example



Initial Model Exploration With Fine-Tuning

Startin by preparing and reshaping the ZIMJS dataset, which contained question-answer pairs. This dataset was then converted into a format suitable for training

	file_name	title	url	content	messages
✓ 0.0s					Python
0	part_5_level_level_2.txt	PART 5	https://zimjs.org/kids/part_variables.html	// BASIC CONDITIONALS CODE - LEVEL 2 // 1 . TY...	[{"content": "What purpose of `getMood ()` ..."}]
1	part_5_level_level_3.txt	PART 5	https://zimjs.org/kids/part_variables.html	// BASIC CONDITIONALS CODE - LEVEL 3 // 1 . TY...	[{"content": "What purpose of `const ` keyword..."}]
2	part_5_level_level_1.txt	PART 5	https://zimjs.org/kids/part_variables.html	// BASIC CONDITIONALS CODE - LEVEL 1 // Here c...	[{"content": "What purpose of `hitTestCircleR..."}]
3	part_4_level_level_2.txt	PART 4	https://zimjs.org/kids/part_variables.html	// BASIC EVENTS CODE - LEVEL 2 // 1 . TYPE mak...	[{"content": "How create circle on stage using..."}]
4	part_1_level_level_1.txt	PART 1	https://zimjs.org/kids/part_variables.html	// BASIC VARIABLES CODE - LEVEL 1 // Here code...	[{"content": "What purpose of using variables ..."}]
...
991	File 56	Using ZIM with InertiaJs & Vue3 - One Way	https://forum.zimjs.com/t/using-zim-with-inert...	[{ 'author' : 'pettis' , 'content' : 'Forg...	[{"content": "nan", "role": "user"}, {"content": "..."}]
992	File 57	Shaders make cool visual experiences!	https://forum.zimjs.com/t/shaders-make-cool-vi...	[{ 'author' : 'abstract' , 'content' : 'Mo...	[{"content": "What types of shaders commonly u..."}]
993	File 58	Speech Recognition and Talking!	https://forum.zimjs.com/t/speech-recognition-a...	[{ 'author' : 'abstract' , 'content' : 'We...	[{"content": "1 : How use JavaScript Speech AP..."}]
994	File 59	Normalize and Ratio - what are they?	https://forum.zimjs.com/t/normalize-and-ratio-...	[{ 'author' : 'abstract' , 'content' : 'No...	[{"content": "1 : How animate items in Tile ce..."}]
995	File 60	Check out the Emitter Configurator!	https://forum.zimjs.com/t/check-out-the-emitter...	[{ 'author' : 'abstract' , 'content' : 'Th...	[{"content": "1 : What tool use ZIM play emitt..."}]
996 rows × 5 columns					

{\\"content\\": \"What purpose of `hitTestCircleRect` method in ZIM ?\", \\'role\\': \\'user\\'},

{\\"content\\": \" hitTestCircleRect ` method in ZIM used test if circle intersecting rectangle . returns ` true ` if circle intersecting rectangle , and ` false ` otherwise . This method commonly used collision detection in interactive applications .\", \\'role\\': \\'assistant\\'},

{\\"content\\": \"How make shape draggable in ZIM ?\", \\'role\\': \\'user\\'},

{\\"content\\": \"make shape draggable in ZIM , use ` drag () ` method on shape object . This method enables shape moved interactively user on canvas .\", \\'role\\': \\'assistant\\'},



Automated Q&A Generation for ZIMJS Content Using GPT-3.5-Turbo

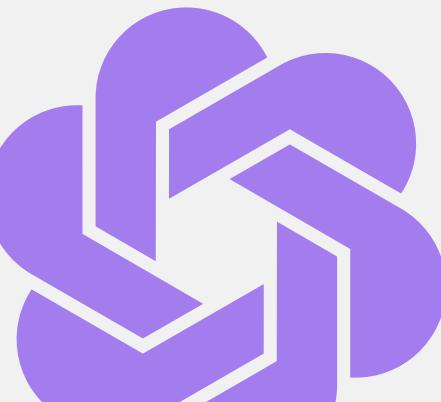
To generate the messages column containing a list of Q&A dictionaries, an advanced Natural Language Processing model, GPT-3.5-Turbo, was employed. This model generates structured questions and answers based on ZIMJS dataset content, ensuring relevance and including code examples when applicable.

```
'[{\\"content\\": \"What purpose of `hitTestCircleRect` method in ZIM ?\", \"role\\\": \"user\"}],
```

```
{\"content\": \" hitTestCircleRect ` method in ZIM used test if circle intersecting rectangle .  
returns ` true ` if circle intersecting rectangle , and ` false ` otherwise . This method commonly  
used collision detection in interactive applications .\", \"role\\\": \"assistant\"},
```

```
{\"content\": \"How make shape draggable in ZIM ?\", \"role\\\": \"user\"}],
```

```
{\"content\": \"make shape draggable in ZIM , use ` drag () ` method on shape object . This  
method enables shape moved interactively user on canvas .\", \"role\\\": \"assistant\"},
```



CodeBERT Challenges

The model overfitted, limiting generalization.

It struggled with contextually accurate answers.

Alternative models or approaches were considered.

TinyLlama Challenges

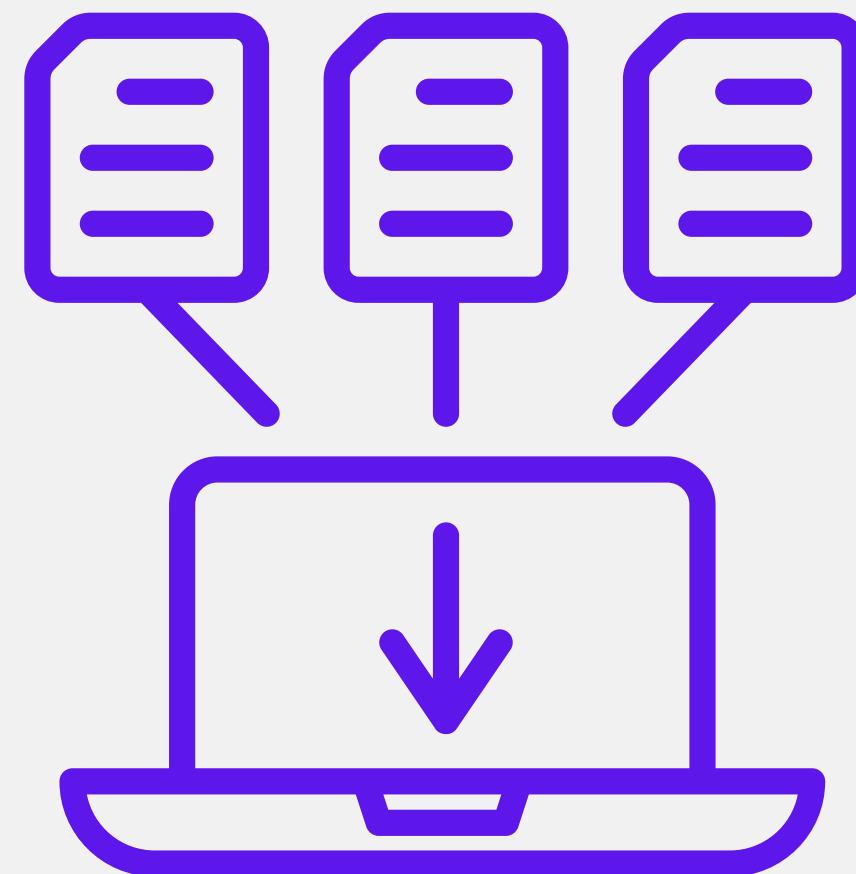
TinyLlama struggled with generating relevant, non-repetitive content.

Training losses improved over CodeBERT but remained insufficient.

Response quality did not meet project goals.

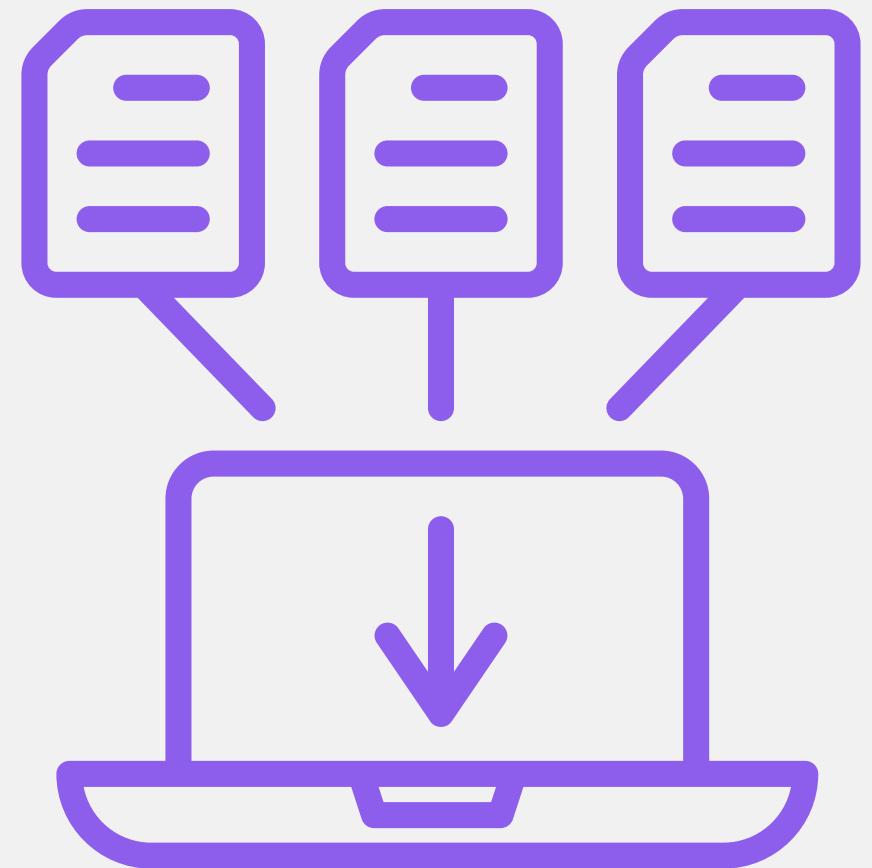
Decision to Move to RAG

Retrieval-Augmented Generation



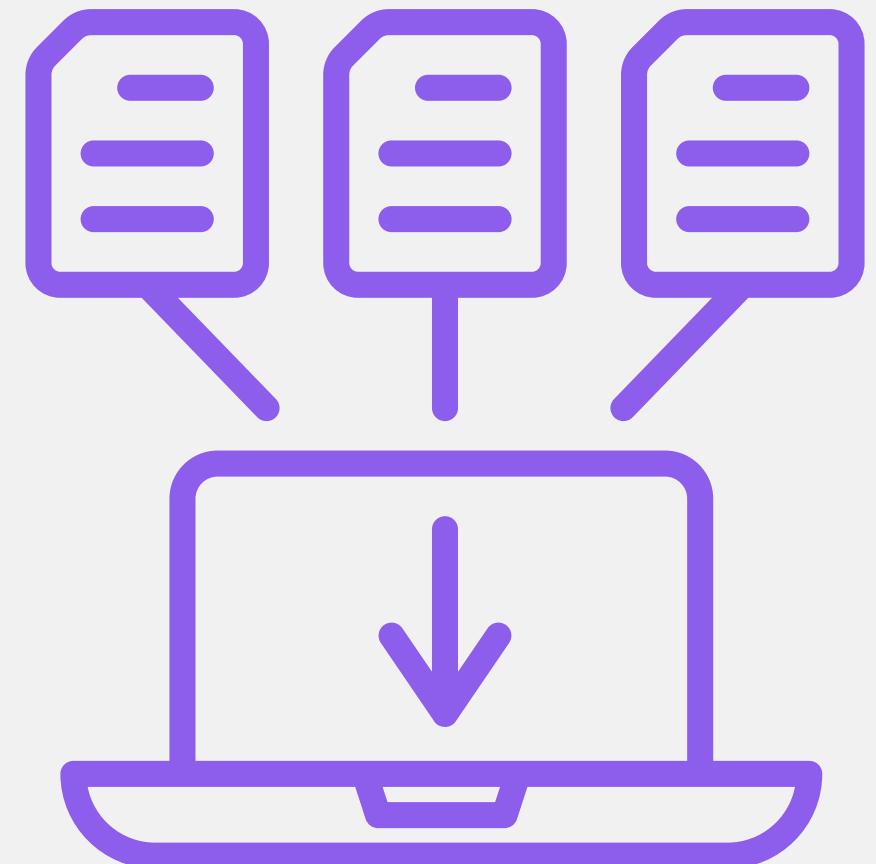
Decision to Move to RAG

- Given the limitations and challenges encountered with both CodeBERT and TinyLlama, you decided to pivot to using Retrieval-Augmented Generation (RAG).
- RAG was chosen because it allows the model to retrieve relevant documents from a pre-built index (FAISS vector store) and then generate responses based on this retrieved information, thereby potentially improving the accuracy and relevance of the answers.



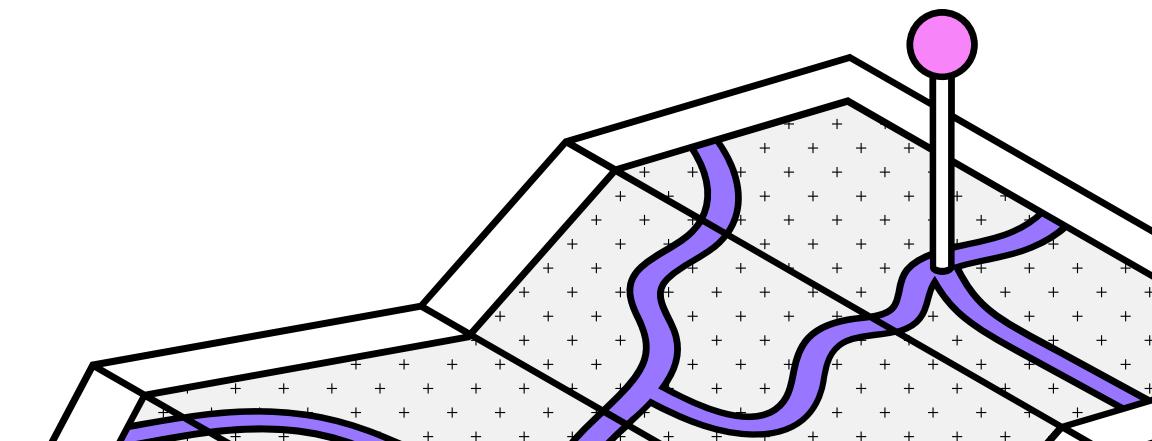
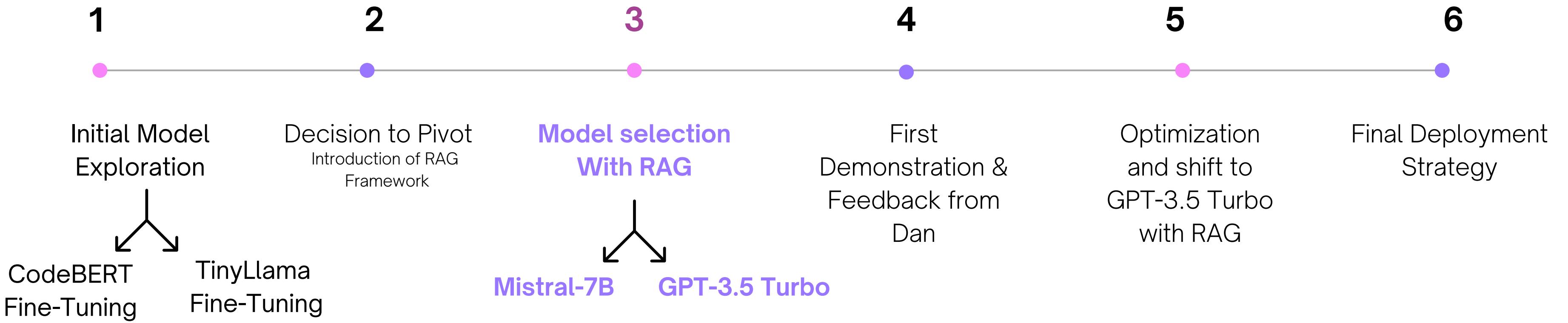
Seamlessly Updating the ZIMJS Chatbot

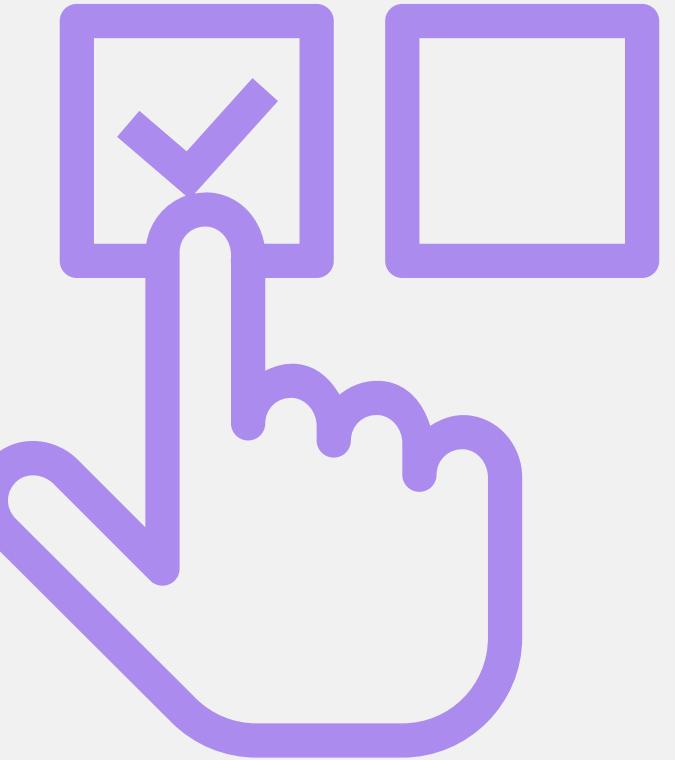
One of the major advantages of using Retrieval-Augmented Generation (RAG) is the ease of updating the ZIMJS chatbot when new data becomes available. The process is straightforward, requiring minimal effort to refresh the dataset and integrate it into the chatbot without the need for re-training or model refinement. This flexibility allows the chatbot to stay current and responsive to new information with minimal overhead, making RAG an ideal solution for maintaining high-performance applications over time.



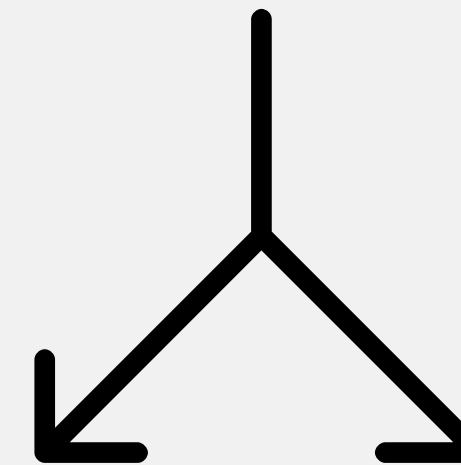
Advanced Modeling Pathway

From Initial Attempts to Optimized Deployment





Model selection With RAG



Mistral-7B GPT-3.5 Turbo



Mistral 7B is a 7.3B parameter model that

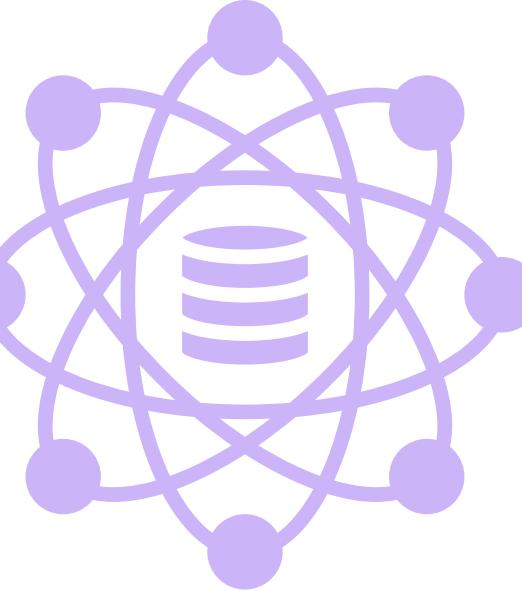
Why Mistral 7B in short

- Outperforms Llama 2 13B on all benchmarks
- Outperforms Llama 1 34B on many benchmarks
- Approaches CodeLlama 7B performance on code, while remaining good at English tasks
- Uses Grouped-query attention (GQA) for faster inference
- Uses Sliding Window Attention (SWA) to handle longer sequences at smaller cost

Data Preparation

The FAISS index was built using HuggingFaceEmbeddings from the sentence-transformers/all-mnlp-base-v2 model, optimizing retrieval accuracy for the zimjs content

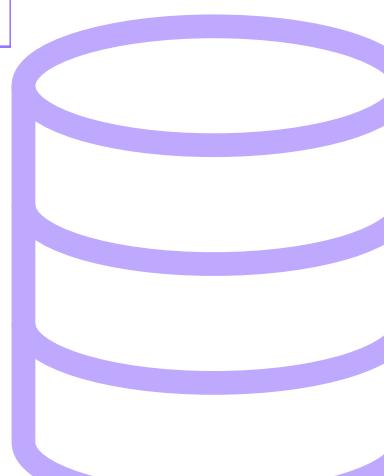




Original Dataset

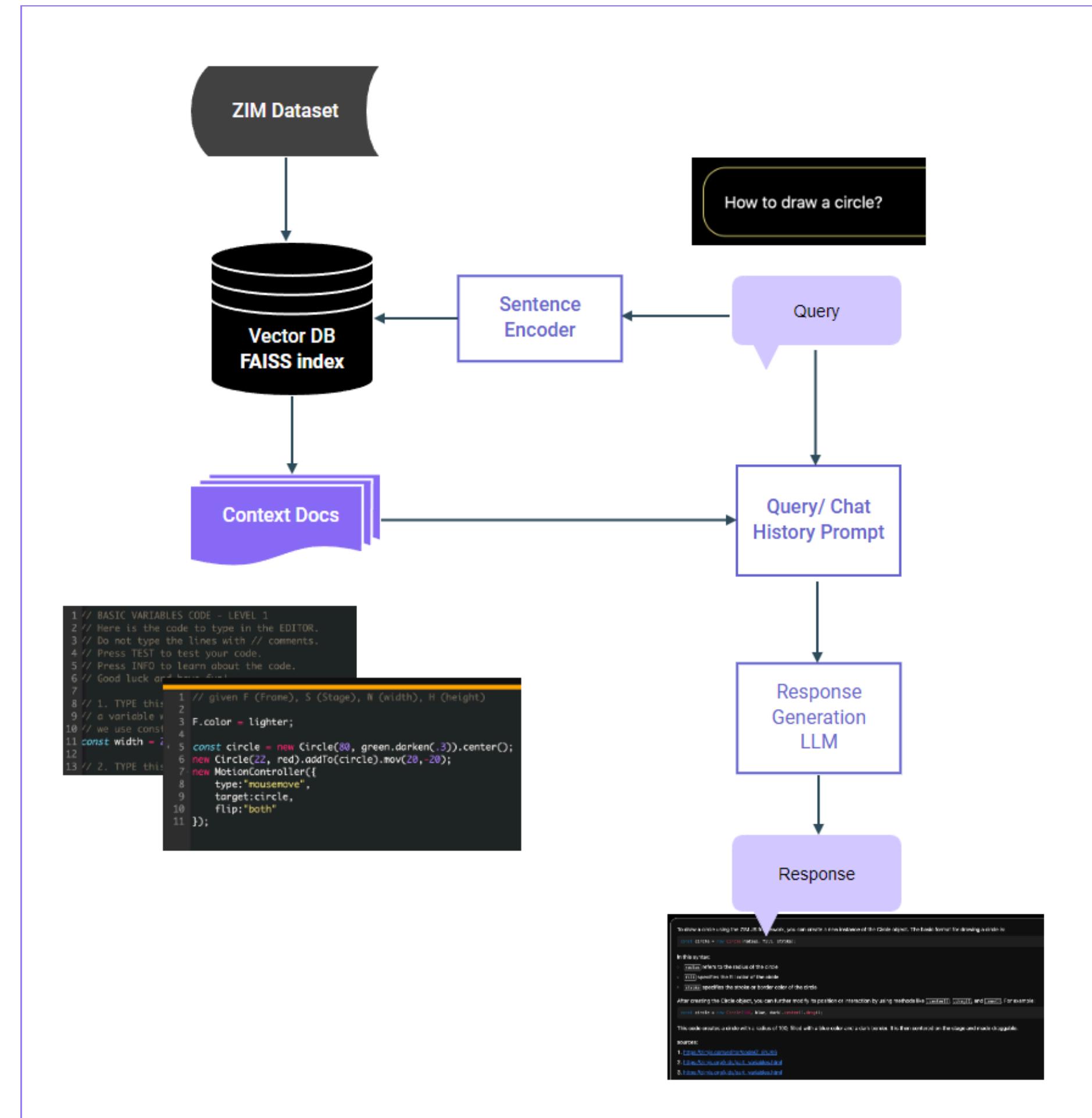
```
df.head()
```

	file_name	title	url	content	label
0	Hit Tests.txt	Hit Tests	Code https://zimjs.com/editor/code/Z_FJQWU	// Editor template built see https : //zimjs.c...	code_example
1	Snapping objects to align.txt	Snapping objects to align	Code https://zimjs.com/editor/code/Z_MGRQS	// given f , , w , h frame , stage , stagew , ...	code_example
2	Blob & Squiggle.txt	Blob & Squiggle	Code https://zimjs.com/editor/code/Z_9FHGJ	// Editor template built see https : //zimjs.c...	code_example
3	Timeline Test.txt	Timeline Test	Code https://zimjs.com/editor/code/Z_AMGMT	const head = new layer (400 , 400 , head) .c...	code_example
4	DoDoDots 1.txt	DoDoDots 1	Code https://zimjs.com/editor/code/Z_Q9SR8	// given f , , w , h frame , stage , stagew , ...	code_example



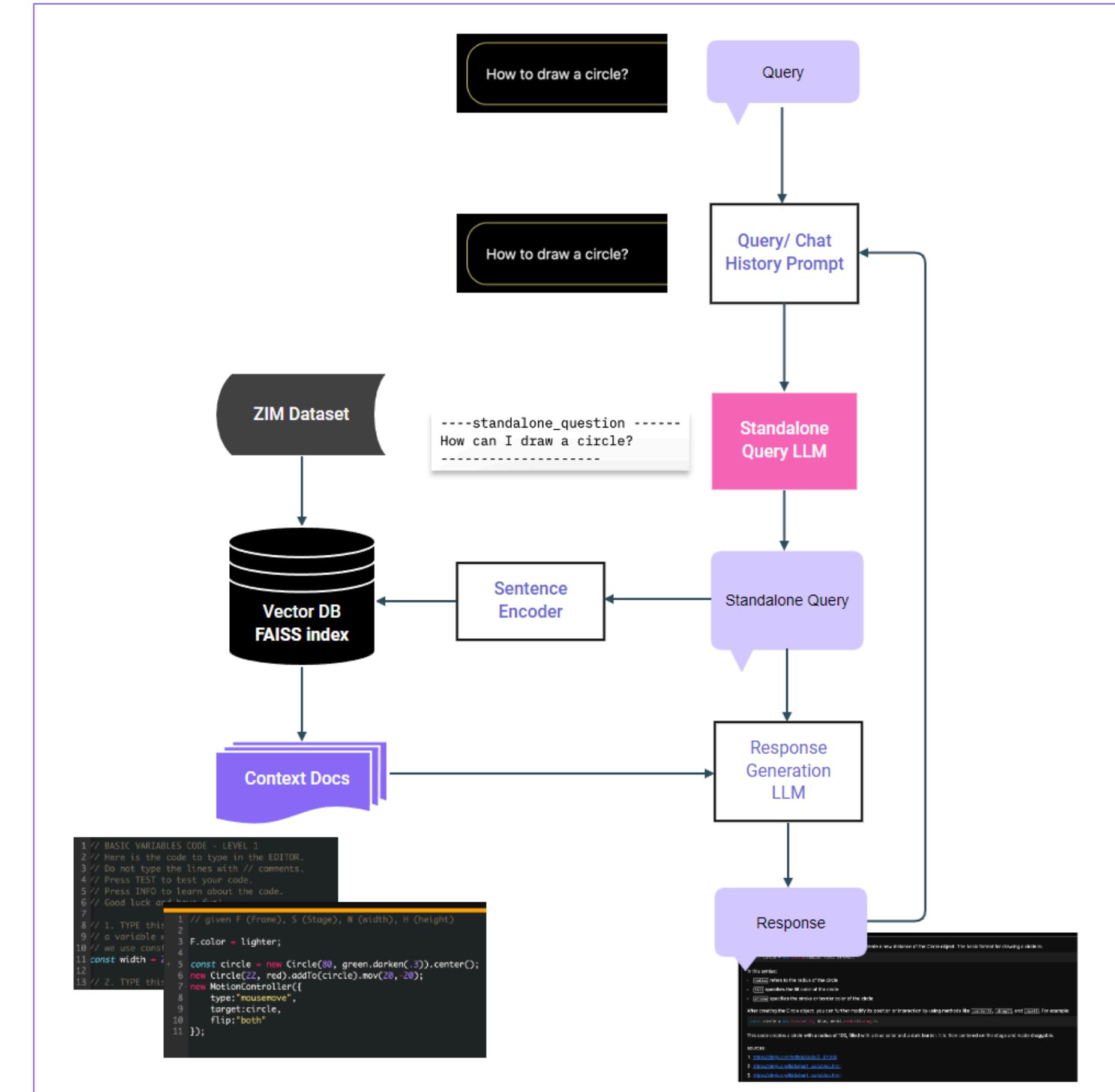
Basic RAG Workflow

Simple Retrieval-Augmented Generation Without Chat History Integration



Enhanced RAG with Standalone Question Generation

Incorporating Chat History to Generate Contextually Relevant Standalone Questions



Evaluating Mistral-7B

Pros and Cons

Mistral-7B Pros

High Performance: Delivered good results with zimjs-specific queries.

Advanced Capabilities: Suitable for handling both code and text, offering versatile functionality.

Mistral-7B Cons

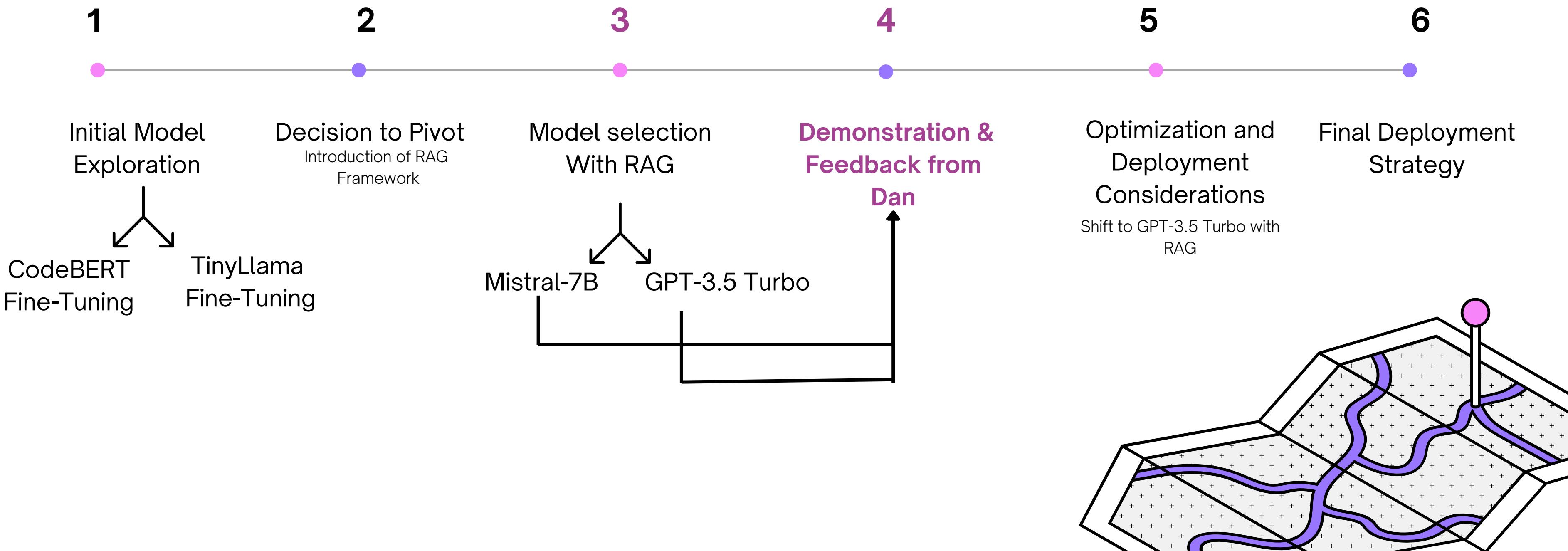
Resource Intensive: Required at least 25 GB GPU to run effectively, making it demanding in terms of hardware.

High Operational Costs: The need for powerful cloud infrastructure led to significant expenses.

Limited Scalability: Continuous enhancement was challenging due to resource constraints.

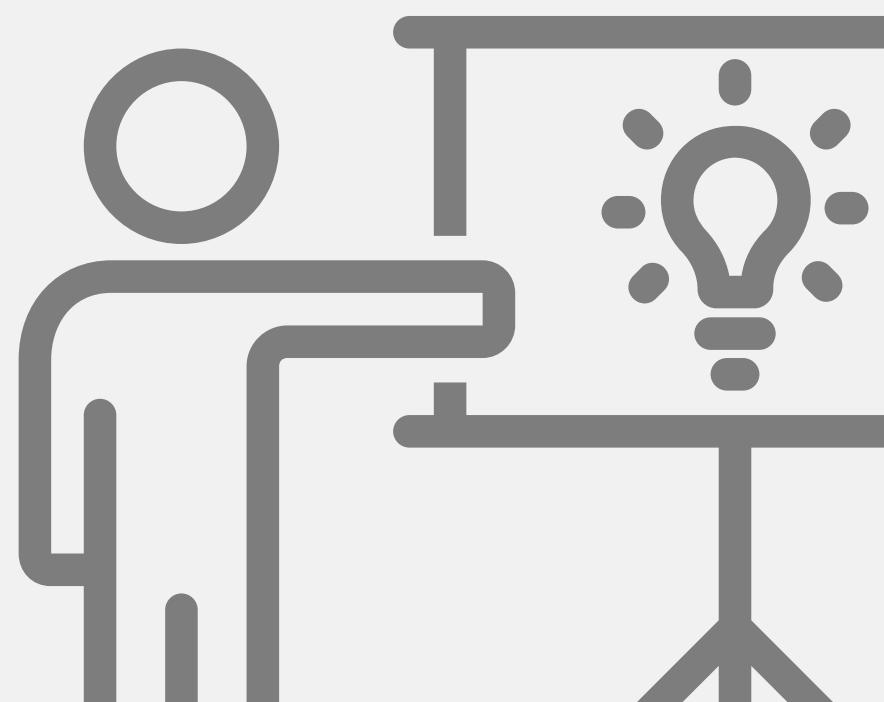
Advanced Modeling Pathway

From Initial Attempts to Optimized Deployment





First Demonstration & Feedback from Dan



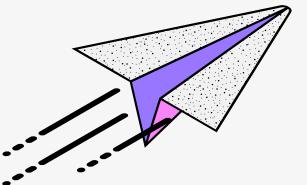
Dan's Feedback on ZIMJS Chatbot

For Mistral-7B-Instruct-v0.1 model



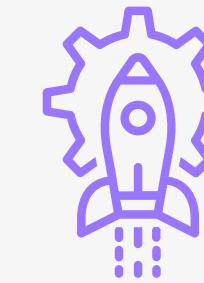
Positive Aspects

The responses aligned more closely with zimjs content, which was a positive development



Areas for Improvement

Dan emphasized the need for further refinement of the chatbot to reduce the misleading outputs.



Deployment Considerations

Due to the high costs and complexity of cloud deployment for the Mistral model, Dan agreed with the decision to shift towards using the GPT API.

Evaluating GPT-3.5-Turbo

Pros and Cons

GPT-3.5 Pros

Resource Efficient: Can be run and tested on local devices without the need for extensive hardware.

Cost-Effective: Only requires a paid API, which is relatively inexpensive compared to maintaining cloud infrastructure.

Practical for Deployment: Easier to deploy and aligned well with the preferences for an accessible and manageable solution.

GPT-3.5 Cons

API Dependency: Requires ongoing API access, which, while affordable, is a recurring cost.

Specialized Use Limitations: May struggle with very complex tasks, like generating long, detailed code snippets, where more powerful models like Mistral-7B might perform better.

ZIM Chatbot

Ask me anything about the ZIM JavaScript Canvas Framework

Chatbot

sources:

1. <https://zimjs.com/docs.html?item=SVG>
2. <https://zimjs.com/tips.html#GLOSSARY>

How to draw circle?

To draw a circle using the ZIM JS framework, you can create a new Circle object. Here is an example of how to draw a circle with a radius of 50 and a red fill color:

```
const circle = new Circle(50, "red");
```

You can customize the circle further by setting additional properties like border color, border width, dashed border, and more according to your requirements. The circle will be centered by default within its container.

If you need more advanced features or animations for the circle, you can explore methods like setColorRange, wiggle, or damp provided by ZIM JS for interactive and dynamic visual effects.

sources:

1. <https://zimjs.com/docs.html?item=Circle>
2. https://zimjs.com/editor/code/Z_GPQRZ

Enter your prompt here..

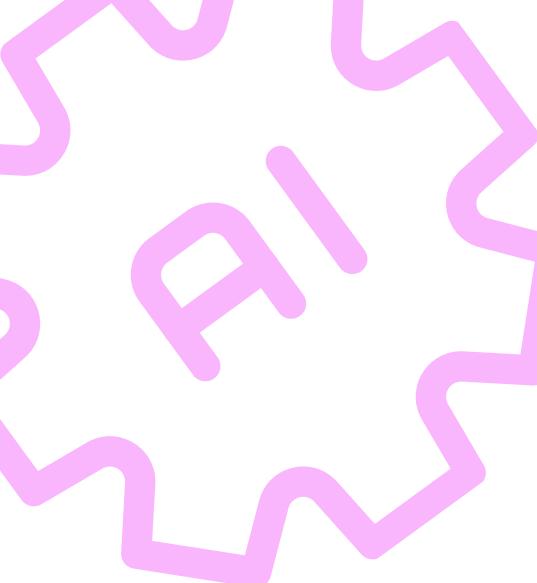
Send

Clear Chat

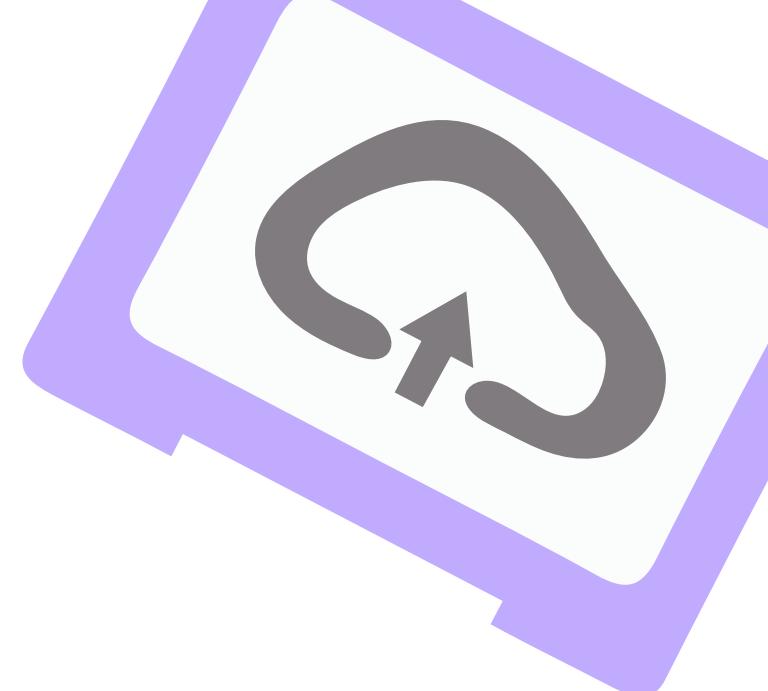
Download Chat History

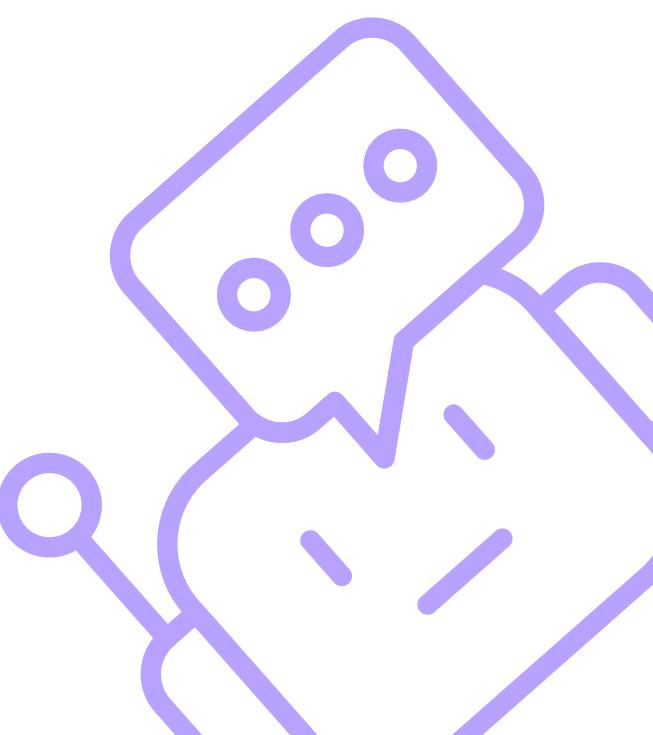
ZIMJS Chatbot Functionality

- **Developed with Gradio:** The chatbot interface was built using Gradio, providing an interactive and user-friendly experience.
- **Dark Theme Integration:** The chatbot features a dark theme that aligns with the aesthetics of the ZIM website, ensuring a seamless visual experience.
- **Send Button:** Allows users to submit their queries to the chatbot effortlessly.
- **Clear Chat Button:** Enables users to clear the conversation history, providing a fresh start for new queries.
- **Download Chat History Button:** Offers users the option to download the entire chat history, which is useful for reference or further analysis.
- **Responsive Design:** The interface is designed to be responsive, ensuring it works well across different devices and screen sizes.
- **Real-Time Interaction:** Provides instant feedback to user queries, making it an effective tool for learning and troubleshooting within the ZIM framework.



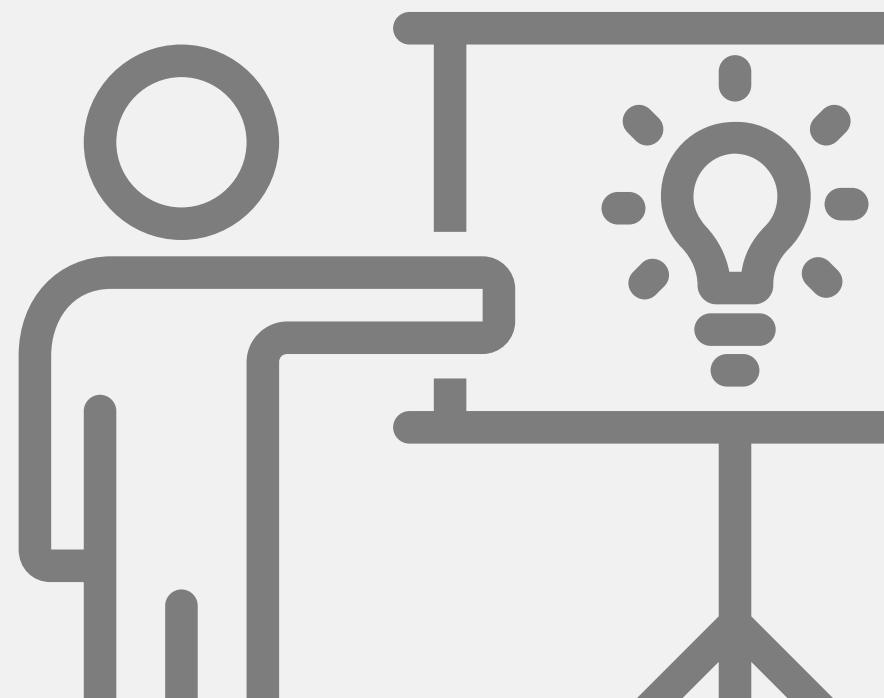
Deployment for Demonstration



- **Demonstration Environment:** For the purposes of this demonstration, the ZIMJS chatbot was deployed using Hugging Face Spaces, leveraging its seamless integration and user-friendly platform.
 - **Accessible and Scalable:** Hugging Face Spaces provided an accessible and scalable environment to showcase the chatbot's functionalities without the need for extensive cloud infrastructure.
 - **Ease of Use:** The deployment process on Hugging Face Spaces was straightforward, allowing for quick setup and real-time testing.
- 



Demonstration & Feedback from Dan



Review

Overall review from Dan

*"Suha has done an excellent job collecting and organizing ZIM-specific data for the final prompt for the LLM. We have gone through a series of tests and have greatly improved the results to a point where the answers are about **80%** and hopefully, the minor issues will be detected by the user. This is well up from 0% when we tried general ChatGPT and about 5% when Suha first started. So we are going to implement the ZIM Chat Bot using a handy site that Suha set up."*

Review



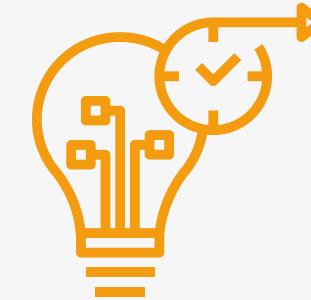
Okay - review done in Red. Certainly getting better.

Generalizability and Future Use Cases



Generalizability

Potential for applying the ZIM AI Assistant to other websites or domains with similar needs.

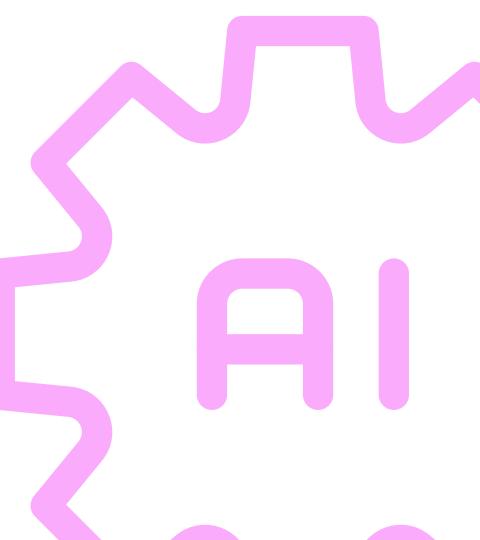
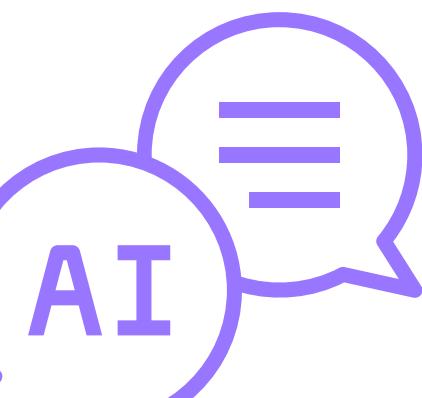
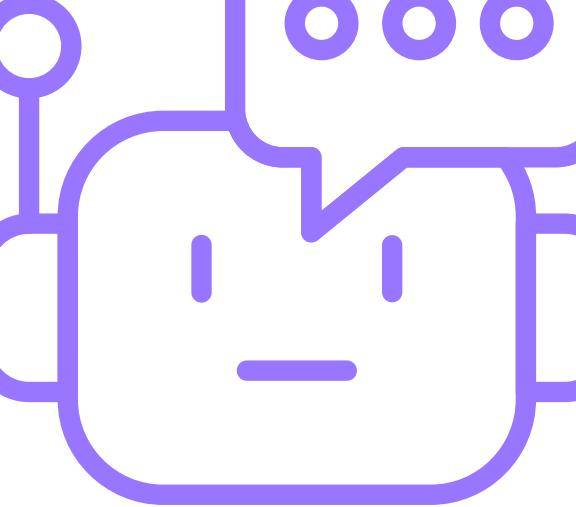


Future Use Cases

Possible applications in education, technical support, and interactive learning environments.

Conclusion: Key Takeaways & Next Steps

- **Significant Progress:** I've made considerable strides in developing a responsive, user-friendly chatbot tailored to the needs of the ZIMJS community.
- **Effective Model Selection:** The shift to GPT-3.5 Turbo with RAG has optimized both performance and deployment efficiency, meeting the project's goals more effectively.
- **User-Centered Design:** The chatbot's design, including its dark theme and intuitive features like chat history download, has been crafted to enhance user experience.
- **Ongoing Refinement:** Feedback from Dan and real-world usage will continue to guide further improvements, particularly in refining the chatbot's responses.
- **Future Focus:** Looking ahead, the focus will be on ongoing improvements, particularly in refining responses and ensuring the chatbot remains aligned with user needs.



Live Demonstration

ZIMJS Chatbot in Action



Q & A

Thank
you!