

AI Internship

Submitted by

Siddharth Tyagi

21BSA10071

Report submitted to the

Kunal Kapur

BACHELOR OF TECHNOLOGY

**COMPUTER SCIENCE ENGINEERING
(CLOUD COMPUTING AND AUTOMATION)**



VIT[®]
BHOPAL
www.vitbhopal.ac.in

School of Computer Science And Engineering

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

**KOTHRIKALAN, SEHORE
MADHYA PRADESH - 466114**

January, 2025

**VIT BHOPAL UNIVERSITY, KOTHRIKALAN,
SEHORE
MADHYA PRADESH – 466114**

BONAFIDE CERTIFICATE

Certified that this project report titled **“AI Internship”** is bonafide work of **“SIDDHARTH TYAGI (21BSA10071)”** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROJECT GUIDE

Kunal Kapur

ACKNOWLEDGMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work. I wish to express my heartfelt gratitude to Kunal Kapur as internal guide for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work. I would like to thank all the technical and teaching staff for this AI internship , who extended directly or indirectly all support. Last, but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

Declaration

I, Siddharth Tyagi, bearing the Registration Number *21BSA10071* hereby declare that this report of “*AI Internship*” represents my original work carried out as a undergraduate student at VIT Bhopal University. To the best of my knowledge, it contains no material previously published or written by another person, nor any material presented for the award of any other degree of VIT Bhopal University or any other institution. Any contribution made to this report by others, with whom I have worked at VIT Bhopal University or elsewhere, is explicitly acknowledged in the report.

Date

29th January 2025

Name of the student

SIDDHARTH TYAGI

Acknowledgment

I extend my heartfelt gratitude to all those who played an instrumental role in the realization of the AI internship for this AI Internship assignment.

My foremost appreciation goes to the dedicated AI internship members who meticulously crafted and executed the assignment. Their innovative guidance methodologies and unwavering commitment to AI internship Internship have profoundly impacted my academic and personal growth.

I am deeply indebted to the local communities and organizations, whose warm embrace and collaboration enriched my learning experiences.

I acknowledge the unwavering support and guidance from faculty members and trainers whose visionary approach to education the realization of this initiative.

Lastly, I express my gratitude to all the individuals who worked diligently behind the scenes, ensuring the seamless coordination and administration of the assignment.

This report reflects the collective efforts of a dedicated community committed to nurturing AI internship Internship opportunities. It stands as a testament to my shared vision for holistic education and the bright future it holds for me

SIDDHARTH TYAGI

Abstract

The Internship Assignment conducted by Kunal Kapur which has provided a transformative educational experience. As a participant in this program, I had the privilege of being trained by industry experts from a reputable company.

The assignment immersed us in a thriving tech ecosystem, where we received hands-on training from professionals at leading IT firms. This experience not only enhanced our technical skills but also exposed us to the latest industry trends and innovations.

Throughout the assignment, the emphasis was on AI internship Internship. We were encouraged to actively engage in real-world projects, problem-solving, and decision-making using Machine Learning as Well as Gen AI. This practical approach not only deepened understanding but also fostered critical thinking and adaptability.

The connections we forged with industry professionals were invaluable. Their mentorship and insights provided us with a holistic view of our respective fields and helped us make informed career choices.

Keywords:

1. Astonishing
2. Exotic
3. Informative
4. Remarkable
5. Zenith

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	List of Abbreviations	iii
	List of Figures and Graphs	iv
	List of Tables	v
	Abstract	vi
1	<p>CHAPTER-1:</p> <p>Text Summarizer</p> <p>1.1 Introduction</p> <p>1.2 Conclusion</p>	<p>1</p> <p>.</p> <p>.</p> <p>.</p>
2	<p>CHAPTER-2:</p> <p>Model Use To build The Logic</p> <p>2.1 Introduction</p> <p>2.2 Conclusions</p> <p>2.3 Snapshots</p>	
3	<p>CHAPTER-3:</p> <p>Installation , LLM Token & Streamlit</p> <p>3.1 Introduction</p> <p>3.2 Error Solved</p> <p>3.3 Conclusions</p>	
4	<p>CHAPTER-4:</p> <p>Frontend & Backend and Deployment over Streamlit</p> <p>4.1 Approach Tools</p> <p>4.3 Link Video to watch it working</p>	

Chapter 1

Introduction

A text summarizer LLM model . These models achieve this by either extractive summarization, which pulls key sentences directly from the original text, or abstractive summarization, which generates new, concise sentences that rephrase the original content.

Advanced architectures like transformers are crucial for their ability to understand context and produce accurate, relevant summaries for tasks like combating information overload and enhancing comprehension

Types of Summarization

1. **Extractive Summarization**: This method identifies and extracts the most important sentences or phrases directly from the original text and strings them together to form a summary.
2. **Abstractive Summarization**: This approach uses LLMs to understand the content and then generate new sentences that convey the core information in a more condensed and often more natural-sounding way than the original text

Why LLMs are Used for Summarization

1. **Efficiency**: LLMs can process vast amounts of information, allowing users to quickly grasp the gist of lengthy documents without reading everything in detail.
2. **Comprehension**: They help people understand complex information by presenting it in a clear, concise, and more digestible format.
3. **Information Management**: In an age of overwhelming information, summarization models act as a vital filter, extracting key points from large volumes of data

Chapter 2

INTRODUCTION

Facebook LLM model “facebook/bart-large-cnn” BART model pre-trained on English language, and fine-tuned on CNN Daily Mail. It was introduced in the paper BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension by Lewis et al.

BART is a transformer encoder-encoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by (1) corrupting text with an arbitrary noising function, and (2) learning a model to reconstruct the original text

BART is particularly effective when fine-tuned for text generation (e.g. summarization, translation) but also works well for comprehension tasks (e.g. text classification, question answering). This particular checkpoint has been fine-tuned on CNN Daily Mail, a large collection of text-summary pairs.

Conclusion to use this model:

- **Efficient Architecture:** Gemma 2 models employ a novel architecture designed for superior performance and efficiency.
- **Local Deployment:** The 2B model can be run locally on various devices, including laptops, using Ollama, which simplifies setup and execution.
- **Versatile Applications:** Suitable for tasks like text generation, chatbots, summarization, and more.

Advantages / Pros

1. High-quality Summarization:

- Generates human-like, coherent summaries.
- Can handle **long text passages** effectively compared to older seq2seq models.

2. Pretrained & Fine-tuned:

- Already fine-tuned on CNN/DailyMail, so ready for out-of-the-box summarization tasks.

3. Abstractive Summarization:

- Produces summaries in its own words rather than just extracting sentences (extractive).
- Can **paraphrase content**, making summaries more readable.

Limitations / Cons

1. Length Limitations:

- Typically, input is truncated at **1024 tokens** (BART's max input).
- Very long articles may lose information if truncated.

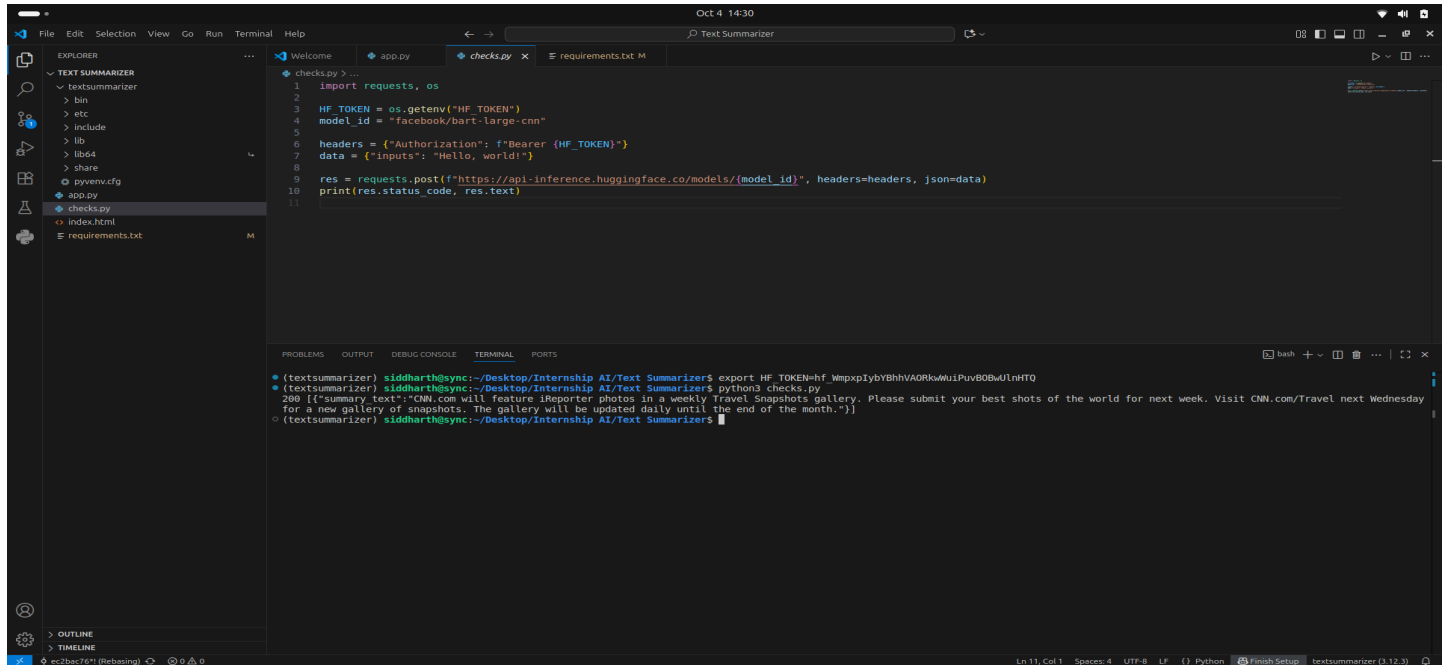
2. Domain Sensitivity:

- Fine-tuned on news data; performance drops on **scientific papers, legal documents, or informal text**.

3. Computationally Expensive:

- Large model (~406M parameters).
- High **GPU memory requirement** for batch processing.

Use Cases: News summarization, article summarization, report summarization, content condensation.



The screenshot shows a Visual Studio Code editor window titled "Text Summarizer" with a file explorer on the left and a terminal at the bottom. The file explorer shows a project named "TEXT SUMMARIZER" with subfolders "bin", "etc", "include", "lib", "lib64", "share", and "pyvenv.cfg", and files "app.py", "checks.py", "index.html", and "requirements.txt". The "checks.py" file is open in the editor, showing the following code:

```
1 import requests, os
2
3 HF_TOKEN = os.getenv("HF_TOKEN")
4 model_id = "facebook/bart-large-cnn"
5
6 headers = {"Authorization": f"Bearer {HF_TOKEN}"}
7 data = {"inputs": "Hello, world!"}
8
9 res = requests.post(f"https://api-inference.huggingface.co/models/{model_id}", headers=headers, json=data)
10 print(res.status_code, res.text)
11
```

The terminal at the bottom shows the following commands and output:

```
(textsummarizer) siddharth@sync:~/Desktop/Internship AI/Text Summarizer$ export HF_TOKEN=hf.WmpxpIyBYBhhVAORkwwuIPuvBOBwUlnHTQ
(textsummarizer) siddharth@sync:~/Desktop/Internship AI/Text Summarizer$ python3 checks.py
200 [{"summary_text": "CNN.com will feature iReporter photos in a weekly Travel Snapshots gallery. Please submit your best shots of the world for next week. Visit CNN.com/Travel next Wednesday for a new gallery of snapshots. The gallery will be updated daily until the end of the month."}]
(textsummarizer) siddharth@sync:~/Desktop/Internship AI/Text Summarizer$
```

Chapter 3

INTRODUCTION

****** Create an Virtual Environment before starting any ML project as it woundn't create Package Dependency and Conflicts Issues also you can install and modify the package by degrading its version and install for your requirements.

1. Install Requirements.txt in your local device
2. Find your Model that can provide free taken and has less strict rules so you can find the model on hugging face also get the Token_HF from access token so that you can access LLM model on billing and get authenticate to use model by access authorization of Token from Token_HF.
3. Install requirements.txt use (pip freeze) to see your install packages in your system.
4. Also install streamlit through pip.

Now check if your Token is access to the model as here are the **Error** I face and able to overcome :-

1. **Error 400** – model_not_supported by the provider you enable.
2. **Error 401** - the request failed due to missing, incorrect, or expired authentication credentials.
3. **Error 403** - Token doesn't have permission.
4. **Error 404** - Model endpoint not accessible.

Conclusion

****200**** -> Means your backend is connected to the model endpoint of HF router and also accessible to Token that you create.

Chapter 4

Approach - To achieve the objective, we used the following approach:

1. **Input:** Text , Link of post , wiki etc and PDF also can be summarized using this model.
2. **Processing:**
 - Uses the **facebook/bart-large-cnn** to process and Summarize the text.
 - Responses it save and can be downloaded also.
3. **Output:** Summary can be downloaded!

Tools and Libraries

Provided in the Requirement.txt

Ubuntu 24.04 LTS version

RAM 12 GB

Platform – BASH , Python 3.12 , Neovim.

Video Link – [Click Me!](#)