أكاديمية طويق
TUWAIQ ACADEMY

AI-POWERED NEWS SUMMARIZER

Saud Al-Mutawa

أكاديمية طويق
TUWAIQ ACADEMY

PROJECT OVERVIEW:

A complete text summarization system built using the T5-small Transformer model, featuring data preprocessing, EDA, fine-tuning on the CNN/DailyMail dataset, and real-time deployment through a Gradio interface.

GOAL:

The goal is to build a high-quality, efficient, and user-friendly summarization pipeline capable of generating accurate and concise summaries from lengthy news articles using state-of-the-art Transformer-based techniques.

أكاديمية طويق
TUWAIQ ACADEMY

DATASET OVERVIEW

CNN/DAILYMAIL DATASET

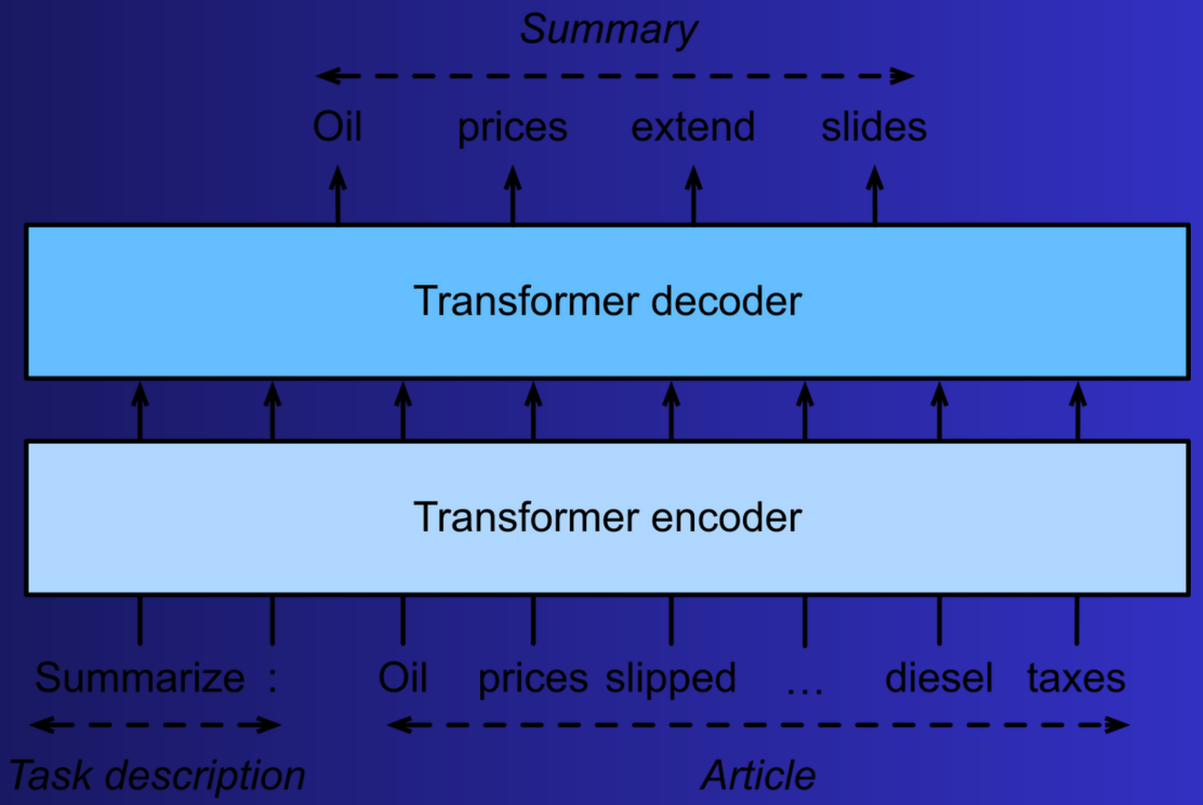
Source	**CNN and Daily Mail news websites**
Language	English
Size	~300,000 articles

أكاديمية طويق
TUWAIQ ACADEMY

T5 Model Architecture

T5 Properties

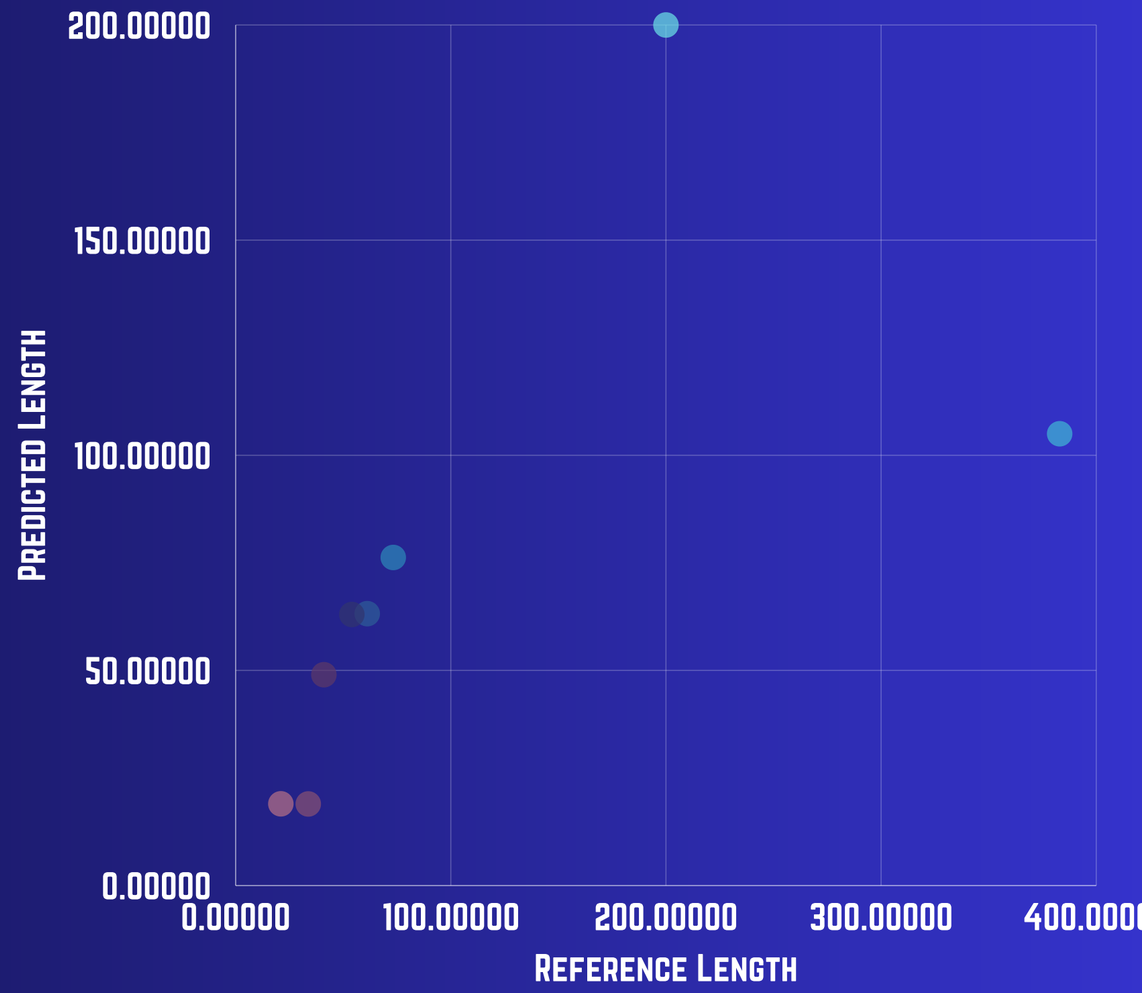
Encoder layers	**6**
Decoder Layers	6
Attention Heads	8
Feed-Forward Layer Dimension	2048



أكاديمية طويق
TUWAIQ ACADEMY

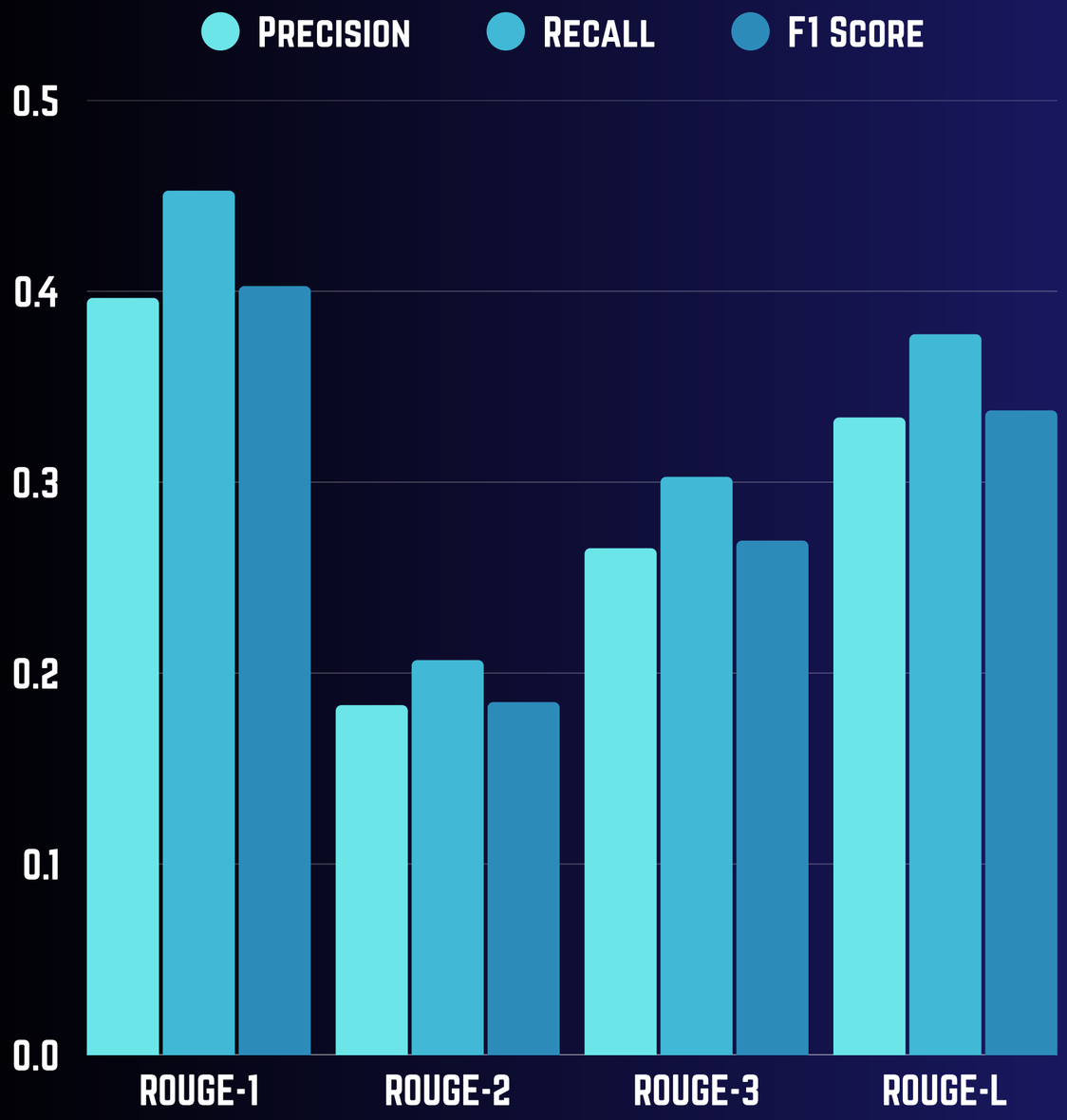
RESULTS

Metric	Reference Length	Predicted Length	Compression Ratio
Count	200.00	200.00	200.00
Mean	61.16	63.19	1.19
Std	33.75	18.97	0.51
Min	21.00	19.00	0.16
0.25	41.00	49.00	0.79
0.50	54.00	63.00	1.11
0.75	73.25	76.25	1.52
Max	383.00	105.00	2.52



أكاديمية طويق
TUWAIQ ACADEMY

EVALUATION



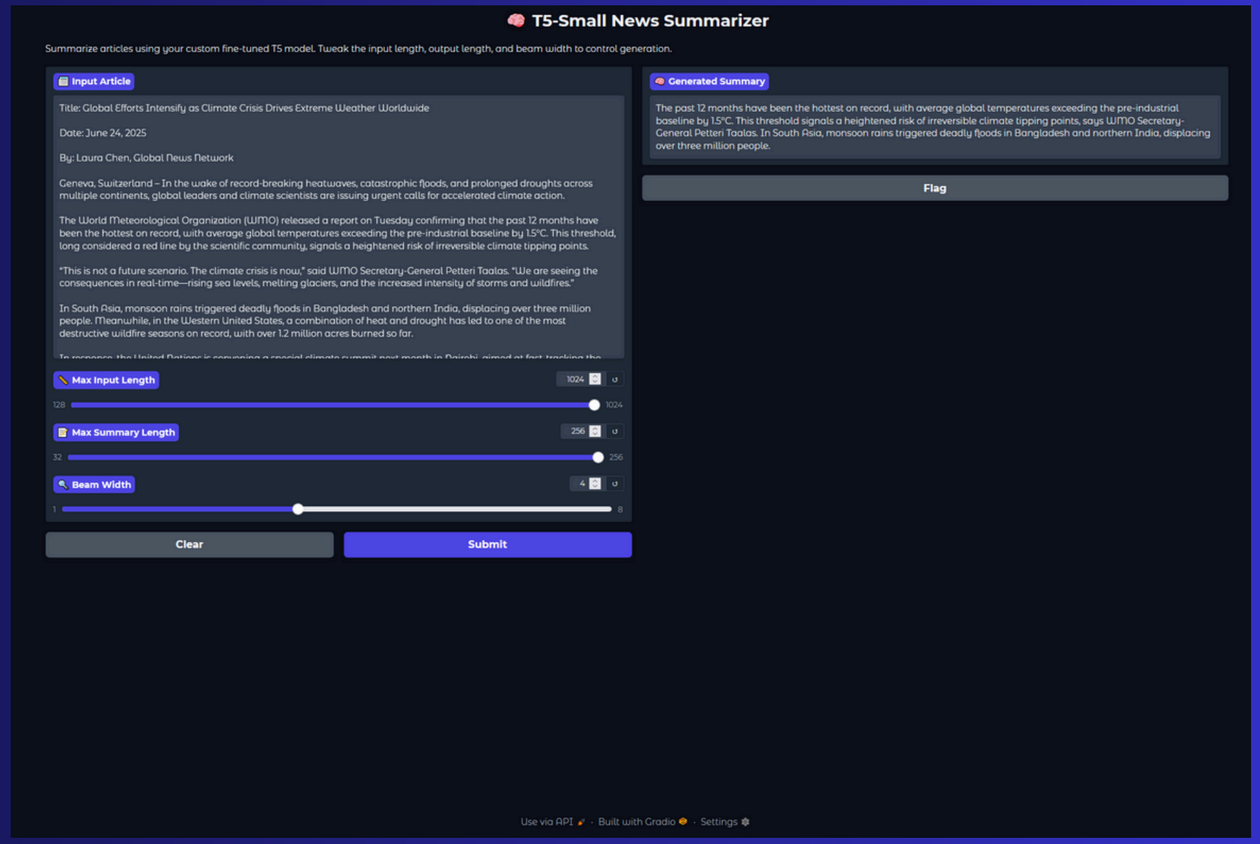
ROGUE ~= 0.3 is considered moderate. I had to limit the maximum length of prediction to 128 tokens because of Google Colab Limitations.

The Summarization model demonstrates moderate effectiveness. ROUGE-1 and ROUGE-L scores indicate decent coverage and fluency in capturing key ideas. However, lower ROUGE-2 and ROUGE-3 scores reflect limited ability to preserve multi-word sequences. Precision and recall suggest room for improvement in both relevance and completeness of the generated summaries.

أكاديمية طويق
TUWAIQ ACADEMY

DEPLOYMENT

I used Gradio to deploy my summarization project because it provides a simple, interactive interface for users to input text and instantly view the generated summary. This is especially helpful for demonstrating the model's capabilities in real-time, collecting feedback, and making the tool accessible without needing technical knowledge or a local environment. Gradio also supports easy integration with Hugging Face, Colab, and sharing via public links, making it ideal for presenting ML projects effectively.



THANK YOU