T5 Model with FAISS for Fake News Detection

# What is T5?

T5 (Text-to-Text Transfer Transformer) is a unified NLP model developed by Google that frames all NLP tasks as text-to-text problems.

This makes it highly flexible for classification, summarization, translation, and more. In our project, we use T5 to classify fake and real news articles by framing the task as:

Input: "classify: [article text]"

Output: "real" or "fake"

# Why T5 is Suitable for Fake News Detection

- Unified Framework: Treats classification as a text generation problem.

- Language Understanding: Pretrained on large corpora, understands context and semantics well.

- Transfer Learning: Effective with limited labeled data.

- Prompt Flexibility: Easily adapts to different input styles.

- Multitask Capability: Can be extended to summarization, Q&A, etc.

# Why Combine T5 with FAISS?

FAISS (Facebook AI Similarity Search) enables fast vector-based retrieval of similar documents.

In our setup, FAISS retrieves contextually relevant real news articles that T5 can use to better classify a given input.

This improves reliability and contextual understanding, similar to Retrieval-Augmented Generation (RAG).

**Workflow:**

1. Encode real news corpus into vector embeddings.

2. Index embeddings using FAISS.

3. On new input:

- Encode and retrieve top-k relevant real news pieces.

- Concatenate retrieved content with input.

4. Pass final input to T5 for classification.

# Device & Environment Requirements

**Hardware Recommendations:**

- GPU: GTX 1660 Ti / RTX 2060 (min) or RTX 3060 / A100 / T4 (recommended)

- VRAM: 6–8 GB (min), 12+ GB (recommended)

- RAM: 8 GB (min), 16–32 GB (recommended)

- Disk: ~10 GB (SSD preferred)

**Software Stack:**

- Python ≥ 3.8

- transformers

- faiss-cpu or faiss-gpu

- torch or tensorflow

- datasets

# Prompt Engineering Example

Prompt: "classify: The president announced a new plan to expand healthcare coverage..."

Output: "real"

Prompt: "classify: Scientists discover a method to live forever using energy drinks..."

Output: "fake"

# T5 Model Sizes & Use Cases

- t5-small: 60M params (~4–5 GB VRAM) – CPU/GPU for testing

- t5-base: 220M params (~8–10 GB VRAM) – Balanced

- t5-large: 770M params (≥12 GB VRAM) – High accuracy

- t5-3b / t5-11b: Needs TPUs or enterprise GPUs

Use inference with prompt engineering for limited compute; fine-tune for best performance.

# Summary: Why T5 + FAISS for Fake News Detection

- Strong semantic understanding from T5.

- Context-aware classification with FAISS.

- Scalable for future tasks (summarization, explanation).

- Supported by Hugging Face with easy APIs.