

NMT System with Automatic Evaluation

Project Report: Part 1

1. Application Overview

This application is a specialized Neural Machine Translation (NMT) tool designed for the healthcare domain. It leverages the **Transformer architecture** to translate medical text from English to French and provides real-time quality evaluation using the **BLEU (Bilingual Evaluation Understudy)** metric.

2. Design Choices & Integration

2.1 Model Selection

- **Architecture:** Transformer-based **MarianMT** (via **Helsinki-NLP**).
- **Reasoning:** Unlike older RNN/LSTM models, the Transformer's **Self-Attention** mechanism allows it to maintain the context of long medical sentences, ensuring that subjects and their respective clinical findings remain linked across the translation.
- **Library:** Built using **transformers** for the backend and **streamlit** for a responsive, browser-based UI.

2.2 Evaluation Strategy

- **Metric:** **sacrebleu** was integrated to provide a standardized BLEU score.
 - **N-Gram Precision:** The system calculates 1-gram (word accuracy) through 4-gram (sentence fluency) scores to give a granular view of translation quality.
 - **Multiple References:** The application supports multiple reference inputs to account for medical synonyms (e.g., "Physician" vs. "Doctor").
-

3. Implementation Flow

1. **Input:** User enters medical source text and one or more reference translations.
2. **Tokenization:** Text is broken into subwords using **SentencePiece**, handling complex medical terms.
3. **Inference:** The Transformer model generates the target translation.

4. **Scoring:** The output is compared to the reference(s), calculating the **Brevity Penalty** and **N-gram precision**.

4. Challenges Faced

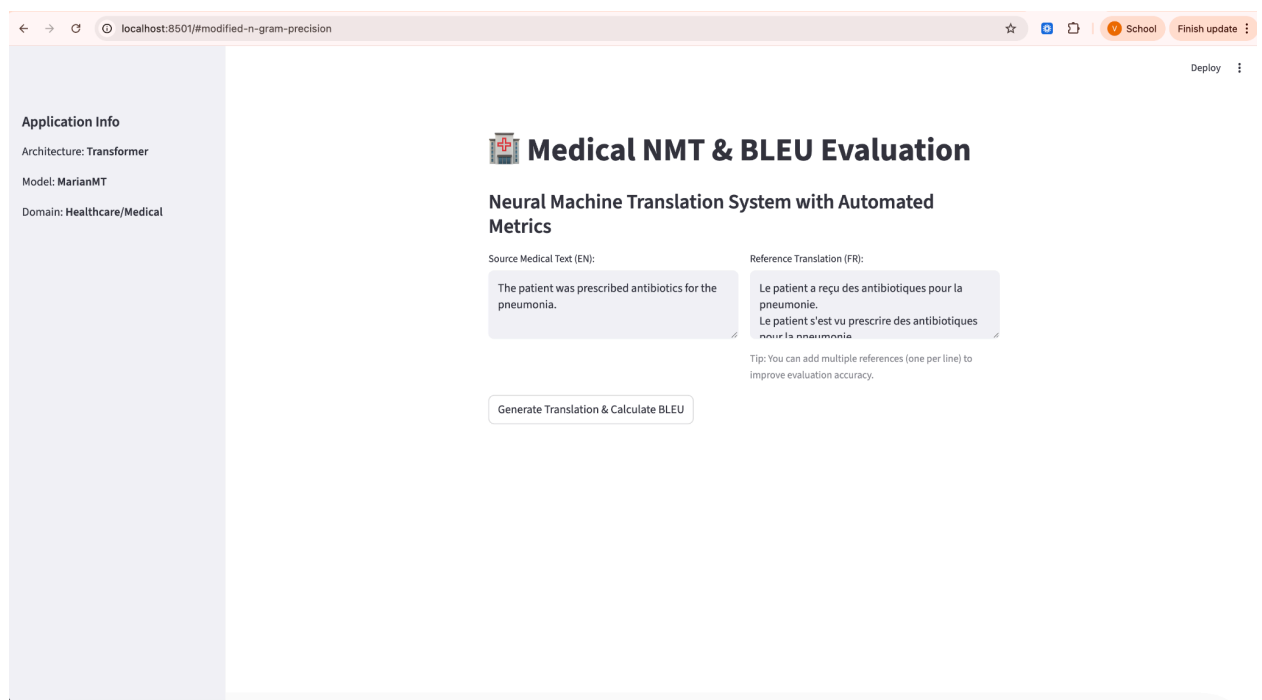
- **Medical Jargon:** Handling specialized terms like “*myocardial infarction*” required a model with robust subword tokenization to avoid "Unknown" token errors.
- **Lexical Sensitivity:** BLEU penalizes synonyms. To solve this, the UI was designed to accept multiple valid reference strings to ensure a fair evaluation of the NMT engine.

5. Application Screenshots & Results

5.1 User Interface Layout

[Screenshot 1: Main UI]

Description: This screenshot shows the input fields for source text and the multi-line reference box.



5.2 Translation & BLEU Score

[Screenshot 2: Output and Metric]

Description: Displays the generated French translation and the final BLEU score metric.


Application Info

Architecture: Transformer

Model: MarianMT

Domain: Healthcare/Medical

Deploy

 Medical NMT & BLEU Evaluation

Neural Machine Translation System with Automated Metrics

Source Medical Text (EN):

The patient was prescribed antibiotics for the pneumonia.

Reference Translation (FR):

Le patient a reçu des antibiotiques pour la pneumonie.
Le patient s'est vu prescrire des antibiotiques pour la pneumonie.

Tip: You can add multiple references (one per line) to improve evaluation accuracy.

Generate Translation & Calculate BLEU

Results

NMT Output: Le patient a reçu des antibiotiques pour la pneumonie.

Total BLEU Score

100.00

5.3 N-Gram Precision Table

[Screenshot 3: Evaluation Table]

Description: A detailed view of the modified n-gram precision (1-gram to 4-gram), showing the accuracy of individual words versus full phrases.

Application Info

Architecture: Transformer

Model: MarianMT

Domain: Healthcare/Medical

Deploy

Tip: You can add multiple references (one per line) to improve evaluation accuracy.

Generate Translation & Calculate BLEU

Results

NMT Output: Le patient a reçu des antibiotiques pour la pneumonie.

Total BLEU Score

100.00

Modified N-Gram Precision

	N-Gram Type	Precision (%)
0	1-Gram (Unigram)	100.00%
1	2-Gram (Bigram)	100.00%
2	3-Gram (Trigram)	100.00%
3	4-Gram (4-gram)	100.00%

> View Evaluation Details (Brevity Penalty & Stats)