**Project 02: Urdu Conversational Chatbot: Transformer with Multi-Head Attention**

**(Group Assignment: 2 Persons)**

* **Objective:**

Build a **custom conversational chatbot for Urdu** using a **Transformer encoder-decoder architecture** implemented **from scratch** (without pre-trained models).  
The model will utilize **multi-head attention** to capture contextual relationships and generate fluent, context-aware Urdu responses.  
A **Streamlit/Gradio interface** will be developed for real-time interaction, and results will be summarized in a **Medium blog** post.

**SUBMISSION INSTRUCTIONS**

* You can take help from AI tools but you are not allowed to use any AI tool for a whole code generation (In case anyone found out with entire AI generated code, will result in direct 0 )
* This is also 15 absolutes assignment, just like previous one
* Group of Maximum 2 students are allowed.
* No Extension in Deadline in any case.
* Deadline: **20th October 2025**
* **Dataset**

You will use the provided **Urdu conversational dataset**:  
📎 <https://www.kaggle.com/datasets/muhammadahmedansari/urdu-dataset-20000>

* **Tasks**

**1. Preprocessing**

* Normalize Urdu text (remove diacritics, standardize Alef and Yeh forms).
* Tokenize sentences and build a vocabulary.
* Split dataset: **Train 80%, Validation 10%, Test 10%.**

**2. Model Architecture**

* Build a **Transformer Encoder–Decoder** from scratch using **PyTorch**.
* Use **Multi-Head Attention**, **Positional Encoding**, and **Feed-Forward Networks**.
* Encoder captures context from the full input; Decoder generates responses token-by-token.

**3. Training & Hyperparameters**

| **Parameter** | **Suggested Values** |
| --- | --- |
| Embedding Dimensions: | 256 / 512 |
| Heads: | 2 |
| Encoder Layers: | 2 |
| Decoder Layers: | 2 |
| Dropout: | 0.1 – 0.3 |
| Batch Size: | 32 / 64 |
| Learning Rate: | 1e-4 – 5e-4 (Adam optimizer) |

Use **teacher forcing** during training. Save best model based on validation BLEU score.

**4. Evaluation**

* **Automatic Metrics:** BLEU, ROUGE-L, chrF, and Perplexity.
* **Human Evaluation:** Rate Fluency, Relevance, and Adequacy (1–5 scale).
* Include qualitative examples comparing model output vs. ground truth.

**5. Inference & UI**

* Integrate model in **Streamlit/Gradio** for chatbot interaction.
* Features: Urdu input box, generated reply display, conversation history, decoding strategy (Greedy/Beam).
* Ensure proper **Urdu right-to-left text rendering**.

**6. Deployment**

Deploy the chatbot on **Streamlit Cloud** or share via **Gradio public link**.

**7. Deliverables**

* Complete GitHub Repository with all code and instructions.
* Deployed chatbot interface (Streamlit/Gradio).
* Evaluation report and results summary.
* **Medium Blog post** covering dataset, model, results, and demo link.
* **Challenge Questions (Bonus)**
* Can you improve chatbot fluency using **attention visualization** or **response length control**?
* Experiment with **different attention heads** or **layer depths** to analyze contextual learning in Urdu.