In this project, we utilized the **Qwen-VL-Chat** model an advanced **multimodal vision-language model (VLM)** to extract structured information from handwritten prescription images.

Unlike traditional methods that rely on OCR (Optical Character Recognition) followed by text parsing, **Qwen-VL-Chat** allows **direct visual understanding** and **contextual extraction** through language prompting.

**Key Steps in Model Usage:**

1. **Installation & Setup:**
   * Installed transformers, accelerate, bitsandbytes, and qwen-vl-utils[decord] packages.
   * Disabled tokenizer parallelism to avoid runtime warnings.
2. **Model Initialization:**
   * Loaded the **QwenVLChat** model via Hugging Face’s transformer interface.
   * Used a lightweight image decoder (decord) to process the prescription images into tensors.
3. **Prompt Engineering:**
   * Designed **domain-specific prompts** like:

"You are an expert prescription reader. Extract all medicines and dosages from this image in JSON format."

* + These prompts helped guide the model towards structured outputs instead of free-form text.

1. **Prediction and Extraction:**
   * Each prescription image was passed through the model with a crafted prompt.
   * The model returned structured JSON-like responses listing **medicines**, **dosages**, and sometimes **usage instructions**.
2. **Output Handling:**
   * Captured the model's outputs and interpreted them into structured fields for evaluation and further analysis.

**Highlights of Qwen-VL-Chat Usage:**

* Direct visual analysis of handwritten prescriptions.
* No intermediate OCR step needed.
* Prompt-tuned extraction for structured JSON output.
* Adaptability to a wide range of handwriting styles.