Domain Adaptation of Self-RAG for Vietnamese Legal QA

Trịnh Nhật Tân

Trường Đại học Công nghệ Thông tin

What ?

- Apply **Self-RAG** framework to Vietnamese legal question answering.
- Use **Vistral-7B-Chat fine-tuned** with reflection tokens (`[Retrieve]`, `[IsRel]`, `[IsSup]`).
- Build a **legal text corpus** from TVPL, CP_VLC, Legal Zalo, and Thuvienphapluat.vn.
- Train the model to **self-evaluate** and **retrieve only when needed**.

Why ?

- Legal domain requires **high factual** accuracy and **citation transparency**.
- Traditional RAG often leads to **hallucinated** or **irrelevant answers**.
- Self-RAG allows **on-demand retrieval** and **self-reflection**, improving trust.
- Tailoring for Vietnamese helps address **low-resource language challenges**.

Overview

Retrieval-Augmented Generation (RAG)

Prompt Điều kiện ly hôn theo luật Việt Nam là gì?

Step 1: Retrieve K documents

Retriever

1. Theo Luật Hôn nhân và Gia đình 2014, ly hôn có thể được yêu cầu ...
2. Tòa án nhân dân cấp huyện giải quyết ly hôn theo thủ tục sơ thẩm.
3. Người chồng không được ly hôn khi vợ đang mang thai, ...

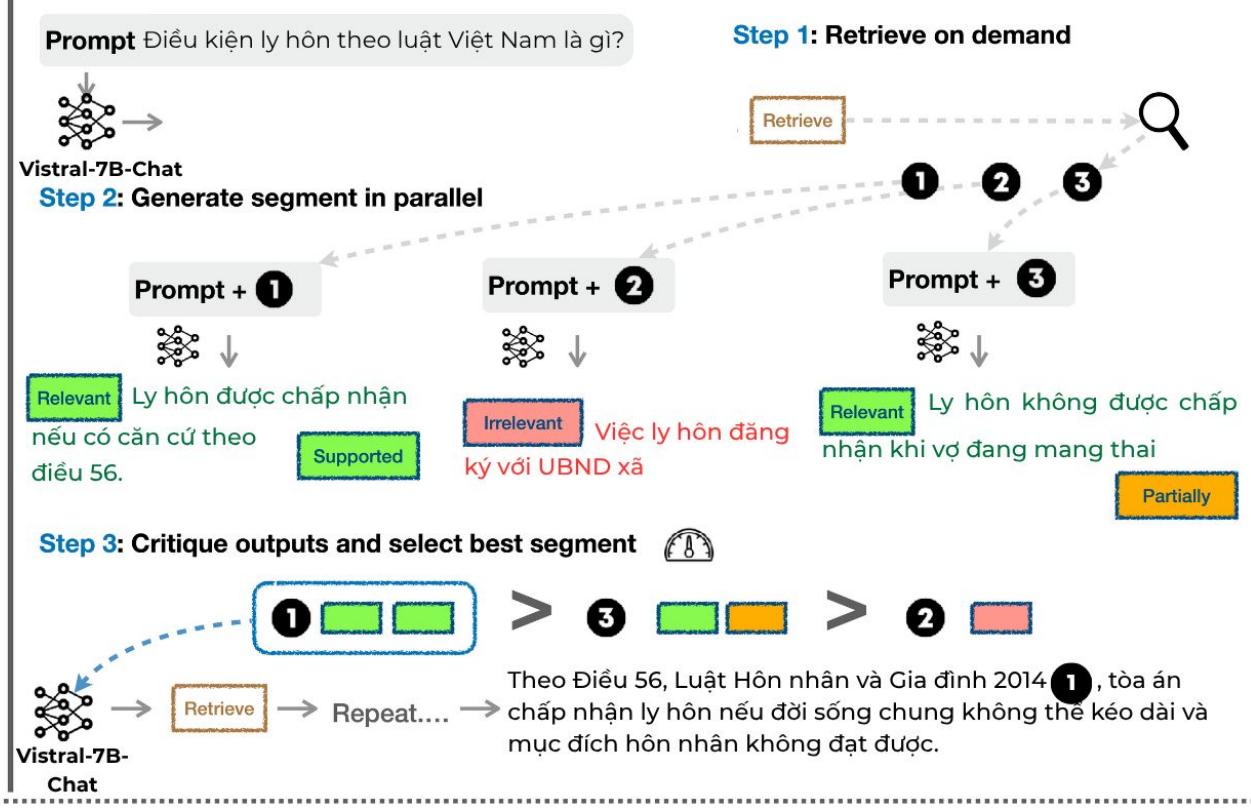
Step 2: Prompt LM with K docs and generate

Prompt Điều kiện ly hôn theo luật Việt Nam là gì? 1 2 3

LM → *"Theo pháp luật Việt Nam, vợ hoặc chồng có quyền yêu cầu ly hôn nếu không còn tình cảm, hoặc khi xảy ra bạo lực gia đình."*

- ✗ Thông tin không chính xác hoàn toàn
- ✗ Không nêu điều luật cụ thể
- ✗ Trích dẫn sai lệch (hallucination): Không trích đúng điều 56, Luật HNGĐ 2014

Ours: Domain Adaptation of Self-rag for Vietnamese Legal QA



Description

1. Legal Corpus Construction

- Collected from TVPL, CP_VLC, Legal Zalo 2021, and Thuvienphapluat.vn.
- Texts are preprocessed: normalized, segmented, and enriched with metadata (date, legal domain, issuer).
- Passages are split (~100 tokens) for retrieval and QA training.

2. Reflection Token Training

- QA pairs are generated using GPT-4, based on real legal passages.
- GPT-4 is prompted to annotate reflection tokens (`[Retrieve]`, `[IsRel]`, `[IsSup]`) for each instance.
- A manually verified subset ensures annotation quality.
- Vistral-7B-Chat is fine-tuned (via LoRA) to learn selective retrieval and self-assessment through these tokens.

3. Evaluation

- Model is evaluated on a legal QA benchmark for both retrieval and answer accuracy.
- Metrics focus on citation correctness, relevance, and hallucination reduction.
- Demonstrates improved trust and transparency in domain-specific question answering.

NII

Trịnh Nhật Tân – Trường Đại học Công nghệ Thông tin

TEL : 0399 156468 Email : tantn.19@grad.uit.edu.vn