



HUẤN LUYỆN MÔ HÌNH NGÔN NGỮ LỚN TRONG LĨNH VỰC LUẬT GIAO THÔNG

THÀNH VIÊN

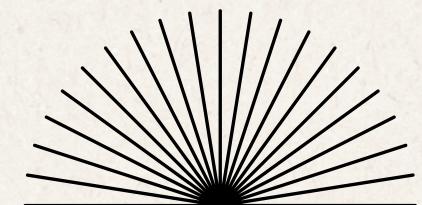
TRẦN TIỀN DŨNG - 222611080

TRẦN MINH HIẾU - 222631553

NGUYỄN MINH - 222631124

NGUYỄN MAI THANH - 222631141

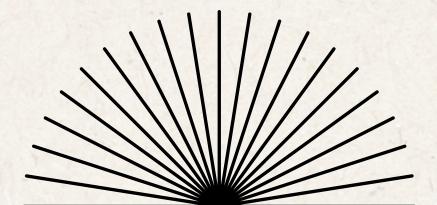
NHÓM 3





MỤC TIÊU

- Huấn luyện LLM chuyên sâu về Luật Giao thông đường bộ Việt Nam.
- Tối ưu mô hình để thực hiện tốt các nhiệm vụ:
- Trả lời chính xác điều luật, mức phạt
- Phân tích tình huống cụ thể
- Tư vấn thủ tục hành chính liên quan (tước bằng, đăng kiểm...)
- Giải thích ngắn gọn, dễ hiểu nhưng đúng luật

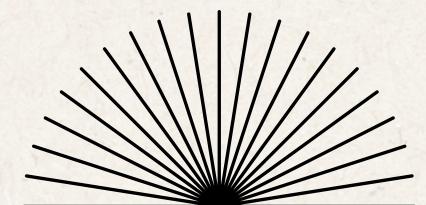




DATA PIPELINE

🎯 Mục tiêu chính

- Thu thập và tổng hợp dữ liệu pháp luật từ 3 nguồn HuggingFace uy tín
- Lọc và trích xuất nội dung liên quan đến giao thông đường bộ
- Tạo synthetic data (806 mẫu) chuyên biệt về giao thông





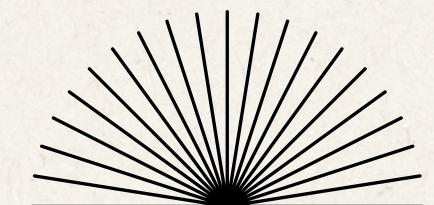
Nguồn dữ liệu chính

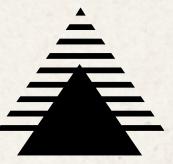
1. Dataset phuocsang/hoidap-tvpl-20k (finetune_data)

- Nguồn: Hugging Face Dataset
- Mô tả: Bộ dữ liệu hỏi đáp pháp luật Việt Nam với 20,000+ câu hỏi
- Số lượng: 21,529 mẫu ban đầu → 19,536 mẫu training + 1,993 mẫu test

2. Dataset huyhuy123/ViLQA (finetune_data2)

- Nguồn: Hugging Face Dataset
- Mô tả: Vietnamese Legal Q&A Dataset chuyên sâu
- Số lượng: 43,420 mẫu training (từ 43,588 mẫu gốc)





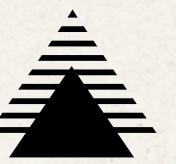
Nguồn dữ liệu chính

3. Dataset chillies/vn-legal-conversation (finetune_data3)

- Nguồn: Hugging Face Dataset
- Mô tả: Vietnamese Legal Conversation Dataset với định dạng hội thoại
- Số lượng: 34,566 mẫu (gộp từ train/validation/test splits)

4. Synthetic Legal Q&A Data

- Nguồn: Tự tạo bằng LlamaIndex + OpenAI GPT từ corpus pháp luật giao thông
- Số lượng: 806 mẫu
- Phương pháp:
 - Sử dụng corpus pháp luật giao thông làm knowledge base
 - Generate câu hỏi tự động dựa trên nội dung luật
 - Tạo câu trả lời có citation từ văn bản gốc
- Mục đích: Bổ sung dữ liệu chuyên biệt về giao thông đường bộ



Xử lý dữ liệu

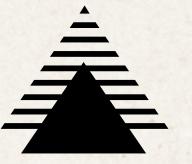
BƯỚC 1: THU THẬP DỮ LIỆU HuggingFace

BƯỚC 2: LỌC DỮ LIỆU LIÊN QUAN GIAO THÔNG

SỬ DỤNG TỪ KHÓA VÀ PATTERN MATCHING ĐỂ LỌC:

- **TỪ KHÓA GIAO THÔNG:** "GIAO THÔNG", "ĐƯỜNG BỘ", "XE CỘ", "LÁI XE", "BẦNG LÁI"
- **LUẬT LIÊN QUAN:** LUẬT GIAO THÔNG ĐƯỜNG BỘ, NGHỊ ĐỊNH VỀ XỬ PHẠT VI PHẠM GIAO THÔNG
- **CHỦ ĐỀ:** VI PHẠM GIAO THÔNG, AN TOÀN ĐƯỜNG BỘ, QUY TẮC LƯU THÔNG => 8000 ROWS VỀ LUẬT GIAO THÔNG ĐƯỜNG BỘ

Supervised Fine-Tuning



🎯 Mục tiêu

- Fine-tune model Llama-3.2-3B cho domain pháp luật giao thông Việt Nam
- Tối ưu hóa cho GPU Tesla T4 (16GB VRAM) trên Kaggle/Colab
- Tạo ra model có khả năng trả lời chính xác các câu hỏi pháp luật giao thông

🚀 Công nghệ sử dụng

- Base Model: unsloth/Llama-3.2-3B-Instruct-bnb-4bit



Điểm nổi bật của Llama-3.2-3B-Instruct-bnb-4bit

- Nhẹ và tiết kiệm tài nguyên: chỉ ~2–3GB VRAM, chạy được trên GPU 6GB/CPU
- Tốc độ inference nhanh nhờ nén 4bit
- Hiệu suất cao so với kích thước 3B, gần mức 7B đời cũ
- Tương thích LoRA/PEFT, fine-tune dễ, không bị collapse
- Xử lý tiếng Việt tốt, ít mix-language
- Phiên bản Instruct đã RLHF, tuân thủ chỉ dẫn tốt
- Lý tưởng cho chatbot pháp lý/luật giao thông vì chi phí thấp và học format nhanh

 Meta

Llama 3

Unsloth

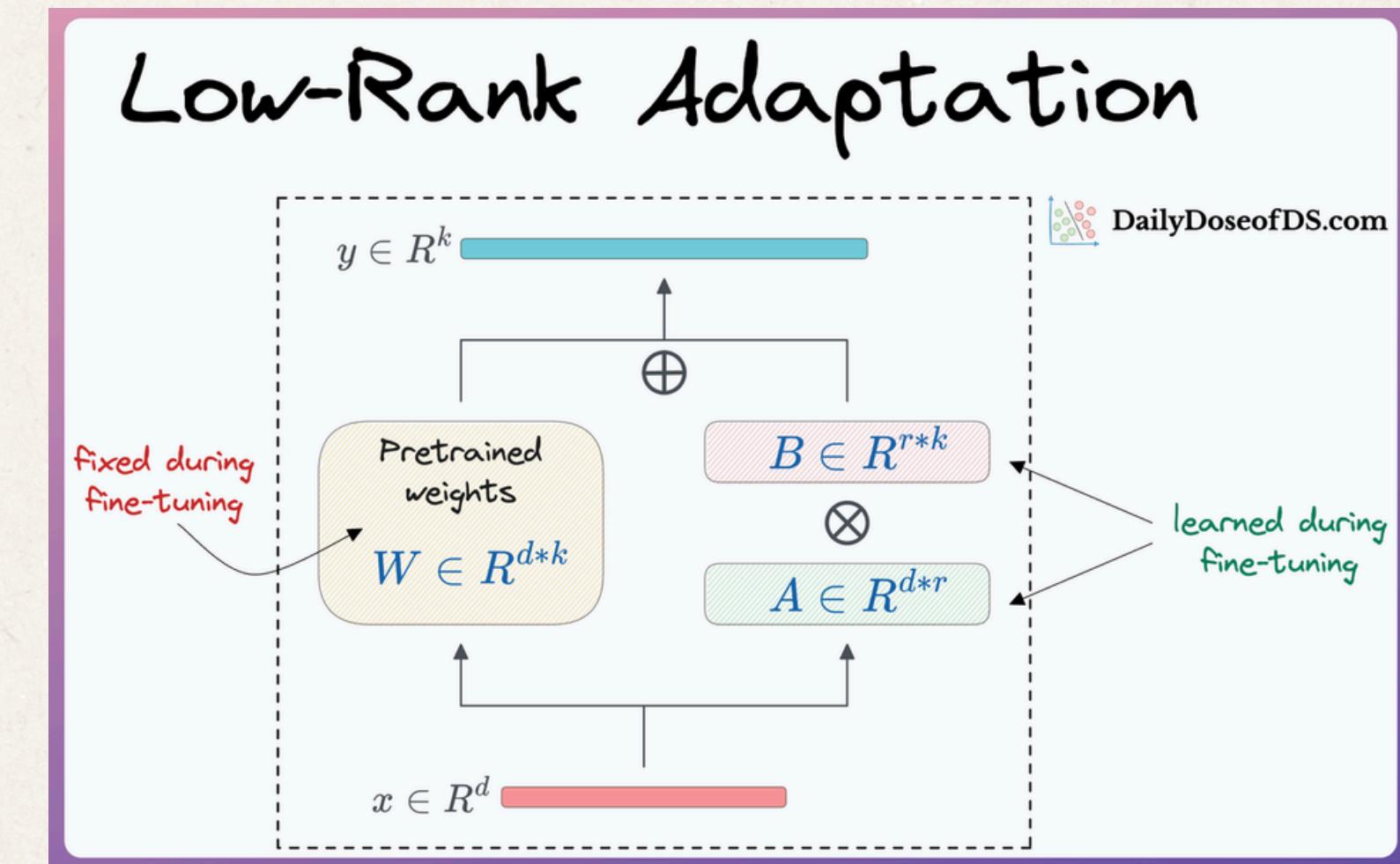
- Framework hỗ trợ finetune với LoRA dễ dàng và hiệu quả.
- Tích hợp nhiều tối ưu: gradient checkpointing, precision bf16, config thân thiện.
- Giảm chi phí huấn luyện, hỗ trợ nhiều mô hình LLM.



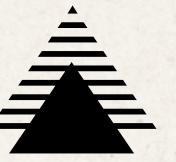
unsloth

LoRA (Low-Rank Adaptation)

- Tinh chỉnh mô hình bằng cách chỉ cập nhật một phần nhỏ (low-rank matrices).
- Tiết kiệm bộ nhớ, tốc độ huấn luyện nhanh hơn.
- Giữ nguyên phần lớn trọng số gốc → tránh làm mất kiến thức ban đầu.



TRAINING CONFIG



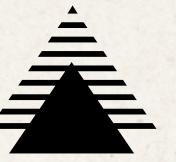
Model Configuration:

- max_seq_length = 1536
- samplesdtype = FP16 cho T4
- load_in_4bit = True
- model_name = "unsloth/Llama-3.2-3B-Instruct-bnb-4bit"

LoRA Parameters:

- r = 32
- lora_alpha = 32
- target_modules = ["q_proj", "k_proj", "v_proj", "o_proj",
"gate_proj", "up_proj", "down_proj"]
- lora_dropout = 0
- bias = "none"

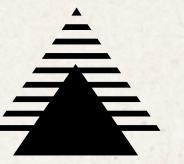
TRAINING CONFIG



Training Arguments:

- num_train_epochs = 2
- per_device_train_batch_size = 4
- gradient_accumulation_steps = 4
- learning_rate = 2e-4
- optim = "adamw_8bit"
- lr_scheduler_type = "cosine"

Các vấn đề gặp phải sau khi SFT



- Response quá dài
- Mix language: Hay trộn tiếng Việt – tiếng Anh
- Reasoning yếu, thiếu logic
- Response đều trên cùng 1 dòng vì ko có token \n xuống dòng



ISSUE



Mục tiêu

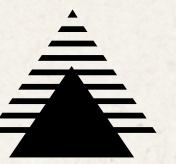
- **Reinforcement Learning:** Sử dụng GRPO để tối ưu chất lượng response
- **Structured Reasoning:**

Training model output format <start_working_out> và <SOLUTION>

Khái niệm GRPO

Group Relative Policy Optimization là một variant của PPO được thiết kế để:

- **Group-based Comparison:** So sánh responses trong cùng một group
- **Relative Scoring:** Đánh giá tương đối thay vì absolute scoring
- **Stability:** Ổn định hơn PPO truyền thống cho conversational AI
- **Efficiency:** Ít computational overhead hơn



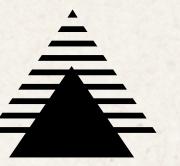
Structured Response Format

Model được train để output theo format:

```
<start_working_out>
[Phản phân tích và suy nghĩ của AI]
<end_working_out>

<SOLUTION>
[Câu trả lời cuối cùng]
</SOLUTION>
```

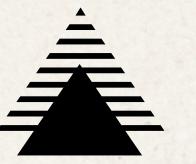
Reward Functions



Sử dụng **2 reward functions chính** được tối ưu hóa để đánh giá chất lượng các response của model:

Functions cho Format và Response Structure:

1. `match_format_exactly` - Kiểm tra format reasoning hoàn hảo (regex matching)
2. `match_format_approximately` - Đánh giá từng thành phần format riêng biệt



1. match_format_exactly

Mục tiêu

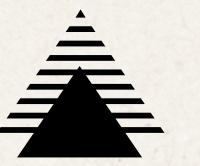
Kiểm tra response có tuân thủ hoàn hảo cấu trúc reasoning format theo regex pattern. Function này có độ chính xác cao nhất và chỉ thưởng nếu format đúng 100%.

Logic đánh giá:

1. Regex matching hoàn hảo: Sử dụng match_format regex pattern để kiểm tra cấu trúc:

- <start_working_out> + nội dung + <end_working_out> + <SOLUTION> + nội dung + </SOLUTION>
- +3.0 điểm nếu format hoàn hảo theo đúng thứ tự
- 0 điểm nếu không match hoàn hảo

Điểm số: 0 hoặc 3.0 (binary scoring)



2. match_format_approximately

Mục tiêu

Đánh giá từng thành phần format riêng biệt và cho điểm partial, giúp model học dần từng bước một thay vì chỉ học "all-or-nothing".

Logic đánh giá:

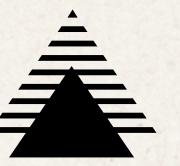
1. Đếm chính xác từng tag:

- <start_working_out>: +0.5 nếu có đúng 1 lần, -1.0 nếu 0 hoặc >1 lần
- <end_working_out>: +0.5 nếu có đúng 1 lần, -1.0 nếu 0 hoặc >1 lần
- <SOLUTION>: +0.5 nếu có đúng 1 lần, -1.0 nếu 0 hoặc >1 lần
- </SOLUTION>: +0.5 nếu có đúng 1 lần, -1.0 nếu 0 hoặc >1 lần

2. Ưu điểm:

- Cho phép model học từng bước (gradual learning)
- Phạt nặng việc lặp lại tags
- Không quan tâm đến thứ tự (khác với match_format_exactly)

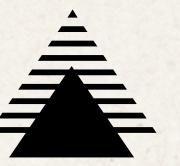
Điểm số: -4.0 đến +2.0



Kết quả thực nghiệm

✗ Fail 1: Đưa vào quá nhiều mục tiêu cùng lúc

- Reward function đặt quá nhiều tiêu chí → model nhiễu
- Response sinh ra quá dài, lan man
- Vẫn bị mix language (Việt-Anh)
- Không tạo được reasoning đúng
- Mô hình không hiểu mục tiêu chính cần tối ưu



Kết quả thực nghiệm

✗ Fail 2: LoRA rank quá thấp

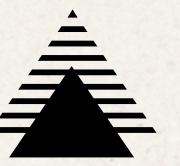
- Rank thấp → không đủ dung lượng để học format mới
- Mô hình không thay đổi được style trả lời
- Không thể học cách xuống dòng, phân tách reasoning
- Output vẫn giống pretrain → mất hiệu quả fine-tuning



Vấn đề sau khi train GRPO

- ⚠ Một số reasoning bị “tự chế” → sai luật
- ⚠ Trích dẫn điều luật nhưng không chính xác
- ⚠ Mức phạt đôi khi bị suy diễn
- ⚠ Một số câu trả lời quá dài, lan man
- ⚠ Tình trạng mix language vẫn còn (“vehicle”, “violated”...)

→ Cần SFT để sửa lỗi này.



Giải pháp Training SFT (đợt 2)

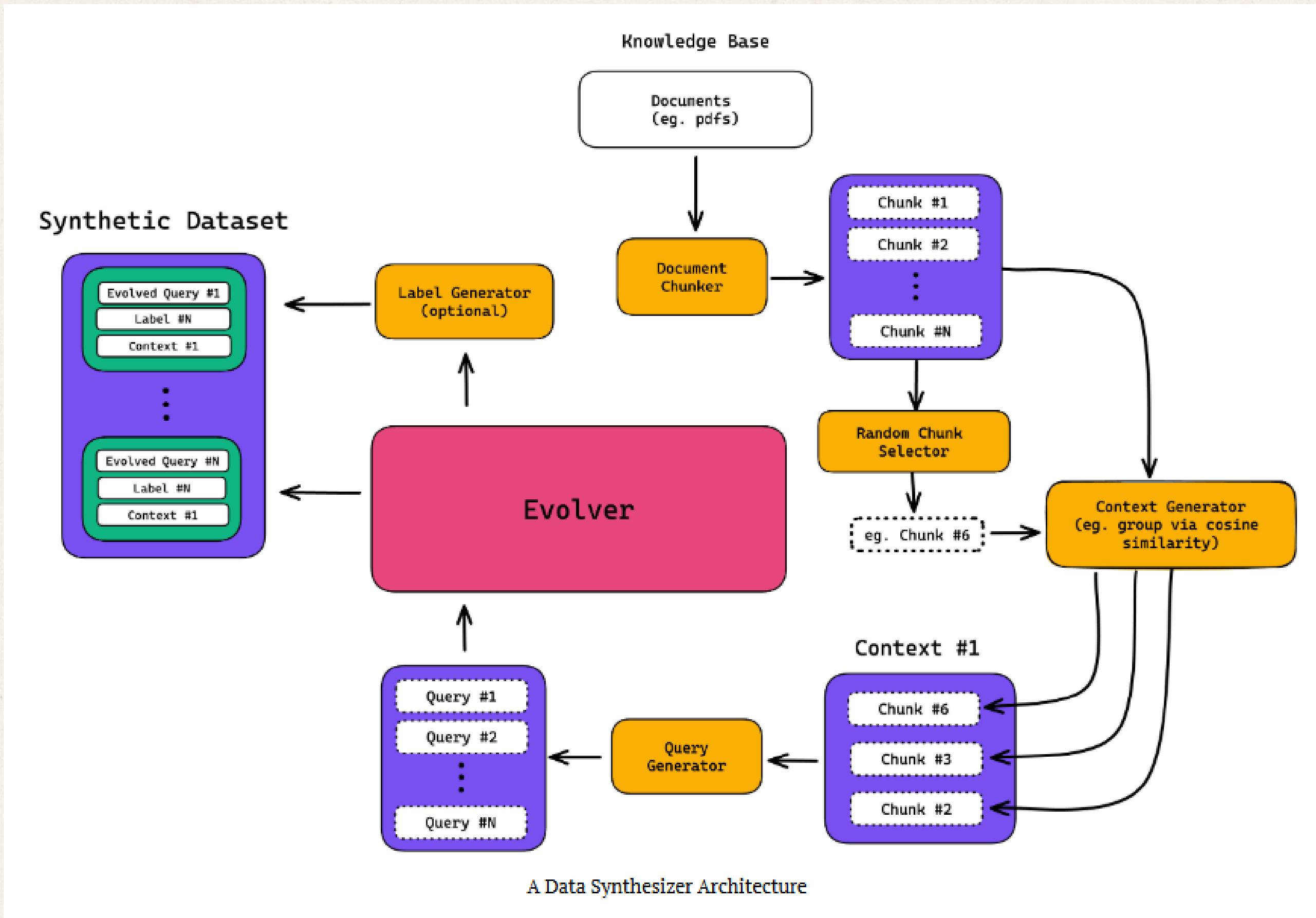
SFT round 2 để sửa lỗi:

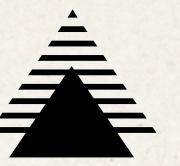
- Bổ sung dataset reasoning chuẩn
- Tạo mẫu dữ liệu có format “step-by-step”
- Tạo dữ liệu phản ví dụ (adversarial examples):
 - tình huống dễ gây nhầm điều luật
- Tăng số mẫu:
 - đúng format
 - đúng luật
 - ngắn - rõ - chính xác

Mục tiêu:

- “Chỉnh lại đường ray” trước khi chạy GRPO lần 2.

Synthetic Data





Synthetic Data

Chất lượng đạt được:

- ✓ Coverage toàn diện: 3 nguồn dữ liệu khác nhau cho độ đa dạng cao
- ✓ Specialized traffic data: 806 mẫu synthetic chuyên về giao thông với citations chính xác
- ✓ Automated pipeline: Quy trình tự động từ crawl → embed → generate → validate
- ✓ High quality: Sử dụng GPT-4o-mini với prompt engineering chuyên nghiệp



GRPO để xử lý mix language

Vấn đề:

Mô hình hay xen tiếng Anh:

“the driver”, “violated”, “speed over limit”...

→ Do pretrain corpus chứa nhiều tiếng Anh.

Giải pháp bằng GRPO:

- Tạo nhóm output, reward:
 - Positive: toàn tiếng Việt + thuật ngữ đúng
 - Negative: chứa từ tiếng Anh
- Mô hình học ưu tiên tiếng Việt tuyệt đối
- Dần loại bỏ thói quen mix-language

Kết quả:

→ Output thuần Việt, chuẩn pháp lý, không lai tạp.

Thank you

CONTACT US

E-mail hello@reallygreatsite.com

Social Media @reallygreatsite

Phone +123-456-7890

Address 123 Anywhere St., Any City, ST 12345