

I. Lựa chọn hướng nghiên cứu chính

- Nghiên cứu về các mô hình ngôn ngữ nhỏ:** Khám phá các mô hình ngôn ngữ nhẹ và hiệu quả cho AI nhúng hoặc các thiết bị biên.
- Tối ưu hóa hiệu suất thuật toán:** Nghiên cứu và phát triển các phương pháp để giảm độ phức tạp tính toán hoặc tăng tốc xử lý trên phần cứng hạn chế.

II. Các hướng nghiên cứu **không đòi hỏi phần cứng mạnh** nhưng vẫn mang tính mới và tiềm năng khoa học:

1. Tối ưu hóa mô hình nhỏ (Small Language Models)

- Nội dung:** Nghiên cứu và phát triển các mô hình ngôn ngữ nhỏ, hiệu quả và phù hợp với các thiết bị có tài nguyên hạn chế (Edge AI, mobile devices).
- Ví dụ:**
 - Sử dụng **Distillation** để làm nhỏ mô hình lớn như LLaMA hoặc GPT.
 - Sử dụng **quantization** (nén mô hình, giảm độ chính xác như 16-bit/8-bit).
 - Nghiên cứu **sparse models** và **efficient transformers** để tăng tốc độ xử lý mà không cần GPU mạnh.
- Công cụ:** Hugging Face Transformers, LoRA, QLoRA, ONNX Runtime.

Lợi ích: Không cần GPU lớn, chỉ cần fine-tune trên tập dữ liệu nhỏ.

2. Phát triển mô hình AI trên thiết bị IoT và Edge

- Nội dung:** Tập trung vào AI chạy trên **IoT** hoặc thiết bị nhúng với phần cứng hạn chế.
- Hướng đi cụ thể:**
 - Tối ưu hóa mô hình nhỏ và triển khai trên Raspberry Pi hoặc các thiết bị Edge AI.
 - Sử dụng **TinyML** cho các bài toán như phát hiện chuyển động, âm thanh, và nhận dạng hình ảnh đơn giản.
- Công cụ:** TensorFlow Lite, TinyML, Edge Impulse.

Ứng dụng: Nhà thông minh, nông nghiệp thông minh, hoặc các hệ thống tiết kiệm năng lượng.

3. Nghiên cứu trong lĩnh vực NLP nhẹ

- Nội dung:** Nghiên cứu các tác vụ xử lý ngôn ngữ tự nhiên không yêu cầu huấn luyện lớn như:
 - Tóm tắt văn bản:** Sử dụng các mô hình pre-trained kết hợp thuật toán heuristic.
 - Phân loại văn bản và gán nhãn:** Fine-tune mô hình nhỏ như **DistilBERT** hoặc **ALBERT**.
 - Xây dựng chatbot nhẹ:** Tận dụng **rule-based systems** hoặc mô hình pre-trained đơn giản.

Công cụ: Hugging Face, spaCy, NLTK.

4. Tích hợp AI với Blockchain hoặc công nghệ phi tập trung

- Nội dung:** Kết hợp AI với **blockchain** để giải quyết các vấn đề như:

- Tối ưu hóa lưu trữ và truy xuất dữ liệu AI trên mạng phi tập trung.
- Xây dựng hệ thống AI bảo mật và minh bạch.
- Triển khai **smart contracts** tích hợp AI để tự động hóa quy trình.

Lợi ích: Nghiên cứu mang tính liên ngành và mới mẻ.

Tổng kết

Các hướng nghiên cứu trên đều **không đòi hỏi GPU mạnh** mà vẫn mang lại giá trị khoa học và thực tiễn cao như:

- Fine-tuning mô hình nhẹ
- Few-shot learning
- AI trên IoT và Edge