

Flan-T5

Flan-T5 là một mô hình ngôn ngữ do **Google** huấn luyện, thuộc họ **T5 (Text-To-Text Transfer Transformer)**.

Tên đầy đủ: "**Fine-tuned L**anguage **N**et **T5**"

Nó là **T5 được huấn luyện lại (fine-tune)** để **trả lời câu hỏi tốt hơn, chính xác hơn, và hiểu lệnh tốt hơn**.

Cấu trúc cơ bản

Flan-T5 sử dụng mô hình **T5 gốc** – một Transformer với kiến trúc **Encoder-Decoder**:

- **Encoder**: Hiểu văn bản đầu vào.
- **Decoder**: Sinh văn bản đầu ra.

Và mọi thứ đều được xử lý dưới dạng **text → text**.

Ví dụ:

Input	Output
"Translate English to French: Hello"	"Bonjour"
"Summarize: Hôm nay trời đẹp..."	"Trời đẹp hôm nay."
"What is the capital of Japan?"	"Tokyo"

Flan-T5 có gì đặc biệt?

🔑 Flan-T5 = T5 + fine-tuning thêm **hàng loạt tác vụ hướng lệnh (instruction tuning)**

Những điểm cải tiến:

1. **Instruction tuning**:

Flan-T5 được huấn luyện để hiểu **lệnh rõ ràng**, ví dụ như:

- “Tóm tắt văn bản sau...”
- “Dịch sang tiếng Đức...”
- “Viết lại đoạn văn cho dễ hiểu...”

2. **Đa nhiệm (multi-task learning)**:

Huấn luyện trên **hơn 1000 loại tác vụ NLP** khác nhau.

3. **Nhiều kích thước:**

- `flan-t5-small`, `flan-t5-base`, `flan-t5-large`, `flan-t5-xl`, `flan-t5-xxl` (lớn nhất ~11B tham số)

4. **Dễ dùng, mã nguồn mở:**

- Có sẵn trên Hugging Face
- Chỉ cần vài dòng code để sử dụng.

Ví dụ dùng Flan-T5 với Python

```
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM

model = AutoModelForSeq2SeqLM.from_pretrained("google/flan-t5-base")
tokenizer = AutoTokenizer.from_pretrained("google/flan-t5-base")

input_text = "Translate English to Vietnamese: How are you?"
inputs = tokenizer(input_text, return_tensors="pt")
outputs = model.generate(**inputs)
print(tokenizer.decode(outputs[0], skip_special_tokens=True))
```

Kết quả: "Bạn khỏe không?"

Tóm tắt

Tiêu chí	Flan-T5
Thuộc loại	Transformer (T5 - Encoder-Decoder)
Do ai tạo	Google
Mục tiêu	Giải quyết đa nhiệm NLP theo lệnh
Cách hoạt động	Text in → Text out
Ưu điểm chính	Hiểu tốt câu lệnh, đa nhiệm, dễ fine-tune