



UNIVERSIDAD DE CHILE

Inteligencia Artificial Generativa

Let's talk about hype stuff

Valentin Barriere // Clemente Henriquez

Universidad de Chile – DCC

Diplomado de Postítulo en Inteligencia Artificial, Primavera 2025

Laboratorio 2: T5 vs Flan-T5

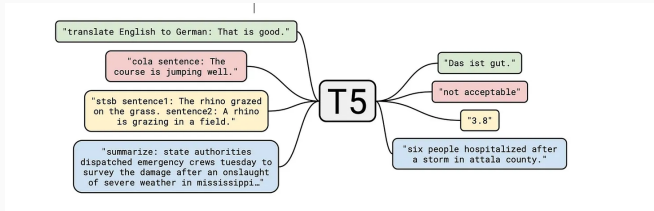
Laboratorio 2: T5-base vs FLAN-T5

En este laboratorio compararemos dos enfoques para adaptar modelos de lenguaje a tareas específicas:

- **Fine-tuning tradicional** (T5-base)
- **Instruction tuning** (FLAN-T5)

¿Qué es T5?

- Modelo encoder-decoder que convierte **toda tarea de NLP** a formato text-to-text.
- Pre-entrenado en C4 (Colossal Clean Crawled Corpus).
- **Requiere fine-tuning específico** para cada tarea.



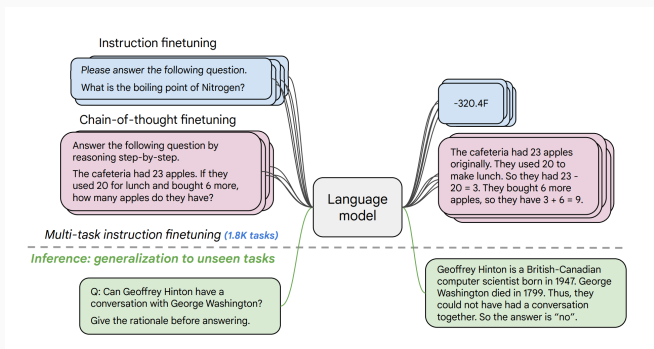
El problema de T5

Para cada nueva tarea necesitas:

1. Recolectar ejemplos etiquetados.
2. Fine-tunear el modelo (tiempo, GPU).
3. Mantener un modelo separado por tarea.

Enter FLAN-T5

- T5-base + instruction tuning en **1,836 tareas**.
- Puede hacer **zero-shot learning** (sin ejemplos!).
- Un solo modelo para múltiples tareas.



¿Cómo funciona FLAN?

Input: "Classify the sentiment: This movie was amazing!" Output: "positive"

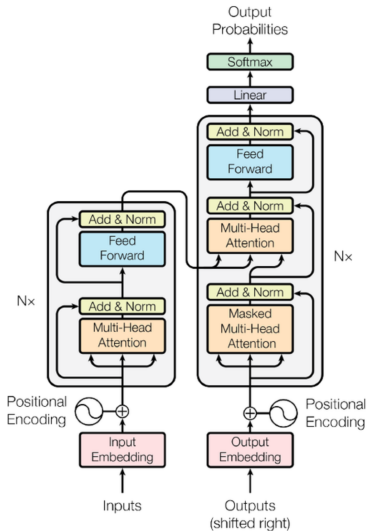
Input: "Translate to Spanish: Hello world" Output: "Hola mundo"

Input: "Summarize: [long text...]" Output: [summary]

El modelo aprende a **seguir instrucciones** en vez de solo una tarea específica.

Tanto T5 como FLAN-T5 se basan en la arquitectura Transformer, que revolucionó el NLP al basarse exclusivamente en mecanismos de atención, prescindiendo de la recurrencia.

Arquitectura Transformer



Vamos a comparar ambos enfoques en **3 datasets de clasificación**:

- **SST-2 (Sentimiento de películas)**
 - 2 clases: positive / negative
 - Reviews de películas
- **Amazon Polarity (Reviews de productos)**
 - 2 clases: positive / negative
 - Reviews de productos Amazon
- **AG News (Clasificación de noticias)**
 - 4 clases: World, Sports, Business, Sci/Tech
 - Noticias de diferentes categorías

¿Qué vamos a hacer?

- **Parte 1: Fine-tuning de T5-base**

- Preprocesar datos en formato text-to-text
- Fine-tunear T5-base en cada dataset (1 época)
- Evaluar accuracy y F1-score

- **Parte 2: FLAN-T5 Zero-shot**

- Diseñar prompts efectivos
- Evaluar FLAN-T5 **sin fine-tuning**
- Comparar resultados

- **Parte 3: Análisis**

- ¿Cuándo funciona mejor cada enfoque?
- Trade-offs: velocidad vs accuracy
- Implicaciones para producción

Resultados esperados

Dataset	T5 Fine-tuned	FLAN-T5 Zero-shot
SST-2	~92%	~93%
Amazon	~94%	~96%
AG News	~91%	~87%

FLAN-T5 compete sin entrenamiento específico!

Questions?

