

Transformers

Trabajo Práctico 2

Temas Avanzados de Deep Learning 2025

1. Implementación Transformer

El objetivo de este Trabajo Práctico es implementar una aplicación haciendo de un modelo de Transformer con la arquitectura de RAG para la resolución de un problema propuesto por el grupo. Para ello se podrán utilizar las librerías y herramientas vistas en clase.

Se deberá consultar previamente con las profesoras el alcance y la naturaleza del problema seleccionado antes de iniciar su desarrollo.

1.1. Ejercicio 1

Se tendrá en cuenta para la evaluación la aplicación de los siguientes conceptos:

- **Retrieval Augmented Generation (RAG):** Integración de técnicas de Data Loading, Text Splitting, Bases de Datos Vectoriales y Retrieval.
- **Técnicas de Preprocesamiento y Postprocesamiento:** Implementación adecuada de métodos para preparar y refinar los datos antes y después de su procesamiento por el modelo.
- **Dataset:** Creación y utilización de un dataset que puede incluir texto, imágenes y/o audios, según las necesidades del problema.
- **Métricas:** Establecimiento de métricas claras y precisas para evaluar el desempeño del modelo.

1.2. Ejercicio 2

Asimismo, se deberán definir (no implementar) los siguientes aspectos:

1. Consideraciones para asegurar que el modelo cumpla con los principios de **Responsible AI** y **Safety**.
2. Estrategias para poner el modelo en **producción**.

2. Presentación

Presentación de aproximadamente 30 minutos que deberá contener:

- Explicación del problema a resolver.

- Desafíos encontrados.
- Decisiones de implementación correspondientes.
- Resultados obtenidos.
- Conclusiones.

3. Entrega

3.1. Pre-entrega

Con el objetivo de plantear el proyecto y limitar su scope, se deberá enviar un mail con copia a ambas profesoras antes del día 19/09 definiendo lo siguiente:

- Descripción del proyecto
- Input de la aplicación
- Output de la aplicación
- ¿Qué van a persistir en la base de datos vectorial?
- ¿De dónde obtienen los datos? (web scraping, documentos, generación manual, etc) ¿Cuál es el o los formatos?
- ¿Dónde ven el principal desafío?

3.2. Entrega final

Adjuntar repositorio con:

- README.md
- Presentación
- Informe con las consideraciones del Ejercicio 2 (máximo 2 hojas), entregado 2 días antes de la presentación.

3.2.1. Fechas de Entrega Límite

- Pre-Entrega 19/09 17.59hs
- Entrega final 24/10 17.59hs
- Informe 22/10 17.59hs

Se penalizará un punto por cada entrega tardía.

La entrega final es la misma para todos los grupos sin importar la fecha de exposición oral.

Las fechas para rendir el oral se asignarán de forma aleatoria a cada grupo.