

## Progetto di Ricerca

Il progetto di ricerca si inquadra all'interno delle attività del progetto BI-REX GeMEB il cui obiettivo principale è l'implementazione di una soluzione basata su un modello linguistico generativo per la consultazione efficiente delle normative tecniche vigenti nel dominio ferroviario.

In particolare, la proposta progettuale si pone l'obiettivo di implementare una soluzione innovativa, utilizzando modelli linguistici generativi basati su architetture di tipo Transformer per ottimizzare attività solitamente time-consuming e prone all'errore umano: la consultazione delle normative tecniche in specifici domini operativi. L'obiettivo finale non si limita a ridurre drasticamente i tempi di consultazione e ad efficientare l'evasione di richieste puntuali sulle norme, bensì anche di produrre output specifici, non generici, garantendo che il trattamento dei dati sensibili forniti in input sia conforme alle norme di privacy. L'attività prevede addestramento di modelli linguistici appositamente ottimizzati, in grado di rispondere a query utente che trovano risposte in documenti tecnici normativi non strutturati.

## Piano di Attività

Si prevede il seguente piano di attività:

1. Analisi dello stato dell'arte e definizione dei requisiti
2. Definizione dell'architettura del modello e dei KPI
3. Pianificazione della raccolta dei dati per sviluppo e validazione del modello
4. Raccolta ed elaborazione dei dati attraverso supervisione di un gruppo di esperti sulle normative vigenti e tecniche di pre-processing dei dati di testo
5. Sviluppo/fine-tuning del modello, sperimentazione di tecniche di RAG in grado anche di ragionare con dati in forma tabellare
6. Validazione del modello