

# Master IAMI A.A. 2023-2024

**Titolo elaborato finale:** Generative AI per valutare l'inclusione

**Candidato:** Andrea Cunegatti

**Relatore:** Prof.ssa Raffaella Bernardi

## Abstract

L'elaborato presenta il processo end-to-end di addestramento (fine-tuning) di un modello linguistico di grandi dimensioni (LLM) per analizzare se un testo è scritto in linguaggio inclusivo, eventualmente identificando i concetti discriminanti, fornendo una parafrasi e classificando la tipologia di discriminazione tra nove possibili. L'addestramento ha riguardato diverse forme di discriminazione, come etnia, religione, status economico, e altro, superando così il solo focus sulla discriminazione di genere. Grazie alla creazione ex-novo di un dataset specifico, è stato possibile dimostrare che l'AI Generativa può affrontare compiti che in passato risultavano molto complessi nel campo della linguistica computazionale, in particolare per le tecnologie di Natural Language Processing che precedevano l'introduzione dei meccanismi dei Transformers. L'originalità del lavoro è poi ulteriormente arricchita dalla peculiarità delle tecniche di prompting usate, grazie alle quali è possibile contare efficacemente concetti discriminanti presenti nel testo, siano essi espressi indistintamente da una o più parole. Sono state inoltre presentate le alternative per eseguire l'addestramento di tipo LORA (Low-Rank Adapters) utilizzando recenti strumenti, accessibili a tutti, senza la necessità di ingenti investimenti economici. Le soluzioni illustrate consentono significative ottimizzazioni delle risorse sia durante la fase di addestramento che nella successiva disponibilità del modello per l'inferenza. Questo elaborato non si è limitato a descrivere gli aspetti tecnici necessari per raggiungere l'obiettivo prefissato, ma offre una prospettiva più ampia, permettendo di cogliere la rilevanza del tema e la complessità del compito affidato al modello. L'intento è anche quello di evidenziare l'enorme potenziale della Generative AI nell'analisi e comprensione dei testi, nonché nella loro eventuale modifica.