



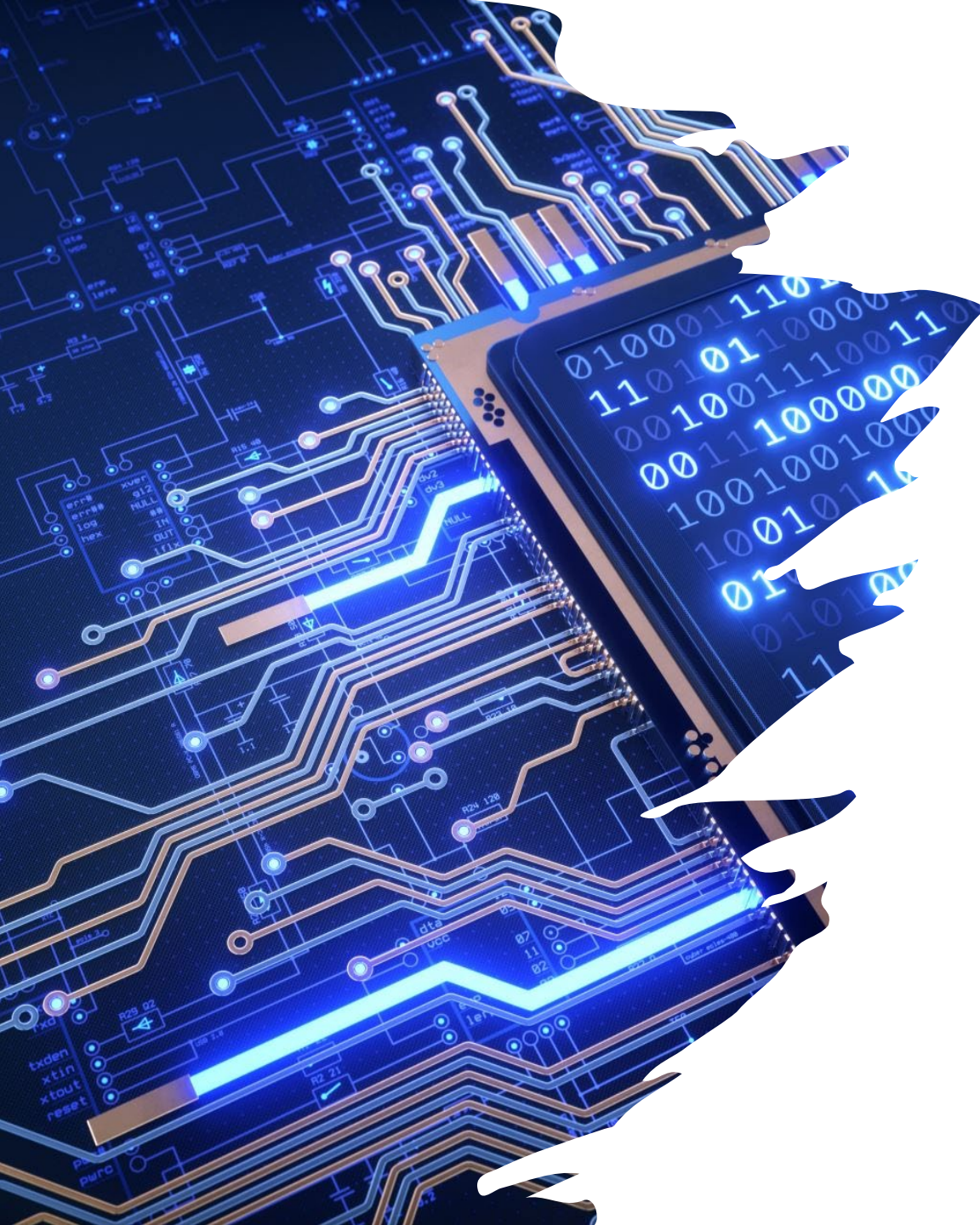
**Politecnico
di Torino**

Algoritmi data-driven e comunicazione inclusiva

E MIMIC

Empowering Multilingual Inclusive Communication

Prof.ssa Tania CERQUITELLI
Dipartimento di Automatica e Informatica
tania.cerquitelli@polito.it



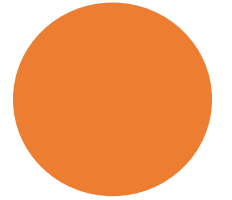
Algoritmi data-driven

- Gli **algoritmi data-driven** modellando **correlazioni latenti nascoste tra i dati**
 - Intelligenza Artificiale, Machine Learning, Deep Learning, Data Science e Data Mining
- Caratterizzati da **due fasi** principali:
 - fase di apprendimento (modellazione)
 - predizione/classificazione/generazione.
- Richiede un **dataset di training**
 - dati di input (e.g., documenti, immagini, tracce audio)
 - dati di interesse che vorremmo che l'algoritmo fornisca in output (e.g., argomenti, categoria, indice di tossicità).
- Tipi di modello
 - Modelli trasparenti/interpretabili
 - **Modelli black-box**

Natural Language Processing

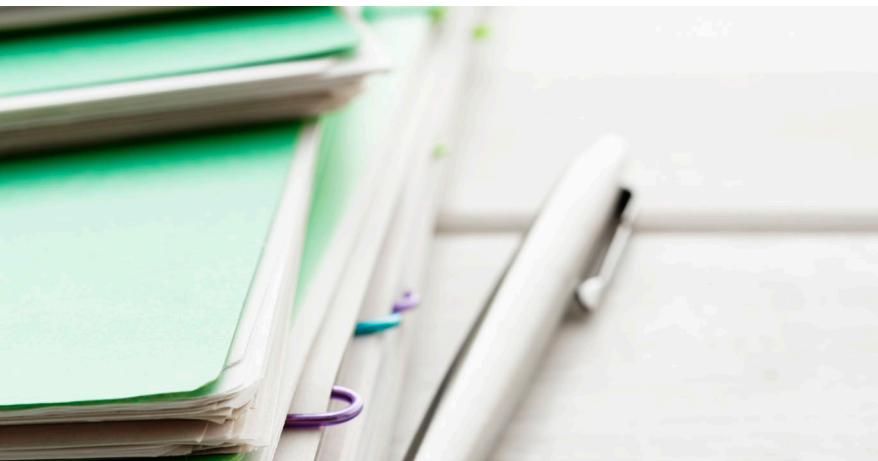
Metodi di elaborazione del linguaggio naturale basati su algoritmi sempre più complessi e in grado di eseguire operazioni (*task*) sempre più articolati.

- **Modellazione del linguaggio:** comprensione profonda del linguaggio naturale tramite approcci computazionali.
- **Classificazione del testo:** assegnazione di categorie ai documenti; analisi del testo per determinare il tono emotivo del messaggio (e.g., positivo, negativo o neutro).
- **Traduzione automatica:** generazione di testo traducendo frasi da una lingua ad un'altra.
- **Generazione del linguaggio:** produzione di testo coerente e fluente simile alla scrittura umana.





Algoritmo supervisionato



- La fase di addestramento richiede un **insieme di documenti annotati (*corpus*)**
 - Documenti arricchiti da informazioni aggiuntive (e.g., argomento trattato nel documento, indice di tossicità)
- **Processo di annotazione** (data labeling)
 - Generalmente effettuato *manualmente* da persone esperte
 - Operazione che *richiede molto tempo e competenza*
- **Qualità e dimensione dei *corpora*** sono fondamentali per poter generare modelli in grado di modellare il linguaggio naturale.



Dati e *Bias* («distorsione»)

- I **documenti** che si utilizzano per la fase di apprendimento sono generalmente disponibili online, quindi sono *generati (spontaneamente) dalle persone* e il **loro contenuto riflette** quindi i loro **punti di vista, che possono essere distorti da pregiudizi razziali, sociali e culturali**.
- I *modelli data-driven* essendo generati dai dati *perpetuano stereotipi e pregiudizi presenti nei dati*.
- È **fondamentale effettuare una rigorosa selezione ed elaborazione dei dati utilizzati per la fase di apprendimento** per evitare di perpetuare esclusione, stereotipi o messaggi violenti contro alcuni sottogruppi.

Modelli linguistici ricorrenti

- I "**modelli ricorrenti**" (Sen e Raghunathan 2018) sono reti neurali consolidate capaci di elaborare sequenze di parole per generare nuove sequenze di parole
 - Utilizzati nella **traduzione automatica**
- Punti di **debolezza**
 - limiti significativi in termini di tempo e complessità della memoria
 - limitata capacità di funzionare in parallelo
 - **necessità di avere un significativo volume di dati di training etichettati da persone esperte**



Large Language Models (LLMs)

- Metodi di **self-supervised learning** (auto-apprendimento) che non richiedono estesi set di dati etichettati.
- Codificano la *conoscenza lessicale della lingua*, fungendo da *solida base di conoscenza*
- Sistemi avanzati di intelligenza artificiale progettati per comprendere e generare il linguaggio umano utilizzando una vasta quantità di dati disponibili sul web.
 - Si basano su **architetture chiamate *transformer*** ed utilizzano **meccanismi di *attention*** per **identificare coppie di token che co-occorrono statisticamente**.
 - **BERT** (Devlin et al., 2019) addestrato sul **corpus inglese di Wikipedia** utilizzando una strategia self-supervised per comprendere e rappresentare parole e frasi in diversi contesti.
- Possono **essere utilizzati per eseguire operazioni più specifiche (attività di *fine-tuning* effettuato su un set di dati annotati)**.



GPT - Generative Pre-training Transformer

- **GPT-3** (Brown et al. 2020) addestrato su oltre 45 terabyte di dati disponibili sul web e libri, produce rappresentazioni vettoriali contestualizzate delle frasi catturando il significato delle parole nel loro contesto. È stato addestrato utilizzando dati non annotati.
- **GPT-4** (OpenAI, 2023) è un modello multimodale che riceve in ingresso immagini e testi e genera un nuovo o rielaborato contenuto testuale.
- Funzionalità
 - **Genera testi coerenti** e contestualmente rilevanti
 - Effettua **traduzioni linguistiche** automatiche
 - Risponde a domande specifiche
 - Effettua **ricerche complesse e analisi di dati**
 - **Crea contenuti**



GPT

GP

LLMs e *Bias*

- I **bias** negli LLMs derivano dai **bias** presenti nei dati su cui sono addestrati.
 - **Dati di addestramento** provengono da testi generati e condivisi da esseri umani che riflettono inevitabilmente le percezioni, opinioni e pregiudizi delle persone che li hanno generate.
 - Potenziale squilibrio in un'ottica di genere, etnia, formazione, competenze, etc. tra gli autori e autrici dei dati che riflettono opinioni e percezioni personali
- **Problema etico: LLMs possono amplificare stereotipi di genere, razziali, sociali e culturali.**
 - Un modello addestrato su testi contenenti *bias* di genere potrebbe generare output che rafforzano questi pregiudizi, influenzando negativamente le opinioni delle persone che utilizzano tali applicazioni
- **Problema di affidabilità:** Modelli distorti possono produrre risultati inaccurati, discriminatori o addirittura testi offensivi, compromettendo la fiducia degli utenti negli strumenti di AI e limitando l'adozione di queste tecnologie in contesti sensibili come la sanità, l'istruzione e il mondo del lavoro.



**Politecnico
di Torino**



PRIN 22

E MIMIC

Empowering Multilingual Inclusive Communication

Casi di studio nel contesto della PA

- Amministrazione dei 3 Atenei coinvolti
- Città Metropolitana di Torino
- Una regione italiana



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA



Contesto scientifico

L'**uso attento della lingua** (e.g., Linee guida della Comunità Europea per la comunicazione gender-neutral, Guida pratica per una comunicazione inclusiva del Politecnico di Torino), contribuisce alla **progressione socio-culturale della popolazione**.

Lo sviluppo di **metodologie deep learning** consente di automatizzare diverse attività di scrittura automatica, in lingue diverse.

Strumenti digitali innovativi, intelligenti, user-friendly, ed inclusivi, potrebbero aiutare ad informare, correggere, e educare le persone all'utilizzo di termini inclusivi, per **una società più equa, solidale e rispettosa delle diversità**.



Politecnico
di Torino

E  MIMIC

Obiettivi

Progettare e sviluppare uno strumento software basato su algoritmi di rete neurale coniugando l'esperienza di persone esperte in aspetti linguistici con le competenze di persone esperte in *Deep Natural Language Understanding*.

Promuovere la **comunicazione inclusiva** contribuendo all'eliminazione di un linguaggio generalista, poco rispettoso delle caratteristiche altrui.



INCLUS!VELY TEAM di ANNOTAZIONE

Caratteristiche del TEAM

- **13 esperti ed esperte linguistiche** nel campo della comunicazione inclusiva nella lingua italiana
- **Team eterogeneo** in esperienze e competenze
- Madrelingua italiani/e
- La maggior parte hanno conseguito il **dottorato di ricerca** o attualmente lo stanno perseguendo in linguistica
- Il **57% ha almeno 10 anni di esperienza** nel settore della linguistica
- Il **50% ha almeno 3 anni di esperienza** nella comunicazione inclusiva
- Annotatori e annotatrici hanno ricevuto, **in media, circa 30 ore di formazione** specifica sulla comunicazione inclusiva e sullo strumento di annotazione

Ruolo

- **Definizione dei criteri linguistici specifici** per promuovere l'inclusività all'interno della lingua italiana.
- **Selezione dei documenti per il corpus** da utilizzare per la fase di apprendimento di tipo data-driven
- **Annotazione del corpus**



Politecnico
di Torino

E  MIMIC

Criteri linguistici

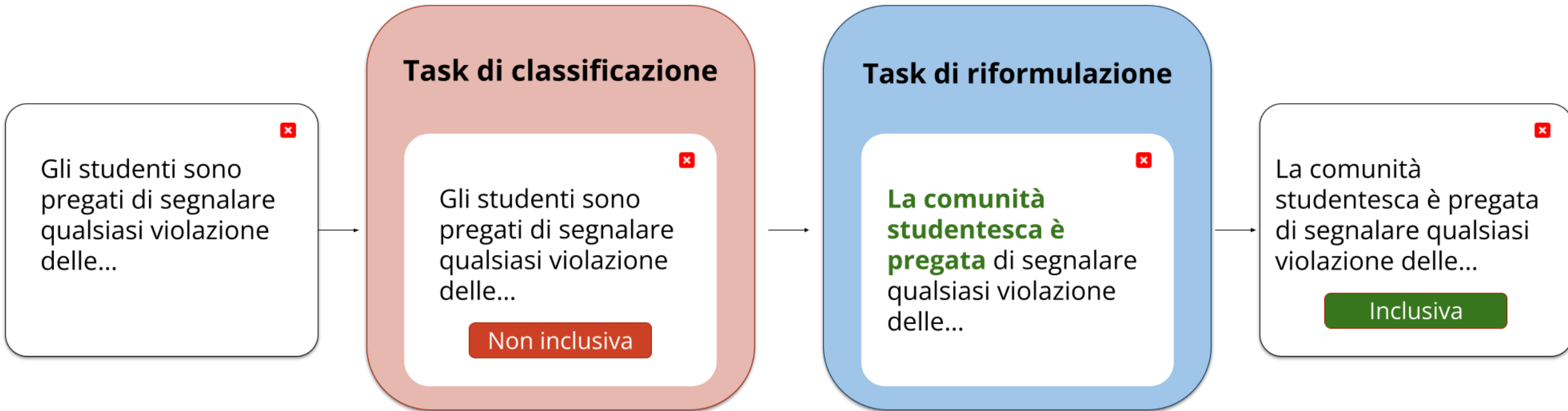


Politecnico
di Torino

E  MIMIC

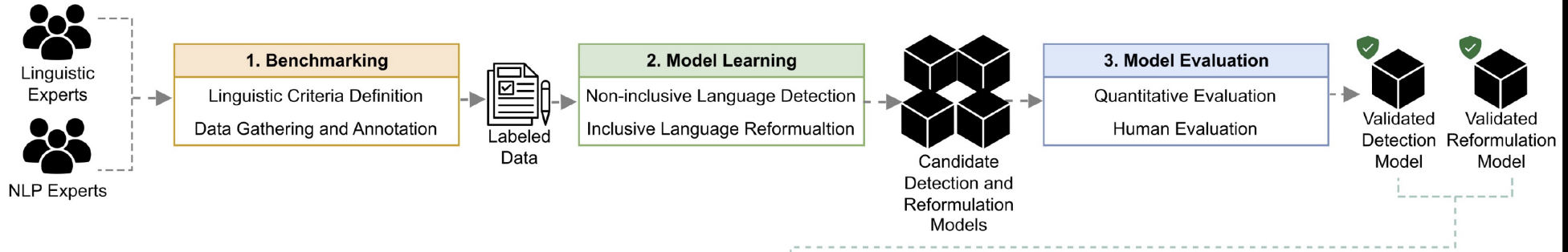
| Criteri Linguistici | Esempi Inclusivi (Non-Inclusivi) |
|--|---|
| Usare parole generiche come "persona", "individuo" o "cittadino" invece di termini declinati al maschile o al femminile. | Tutte le persone che accompagnano (gli accompagnatori) sono pregate di scendere |
| Usare sostantivi collettivi per riferirsi a gruppi di persone invece di specificare il genere. | Il personale docente dovrà (i docenti dovranno) preparare la lezione. |
| Utilizzare riformulazioni senza agente esplicito ottenute per metonimia. | Oggi, la scienza ha (gli scienziati hanno) dimostrato la validità del modello. |
| Omettere segmenti linguistici senza modificare il significato della comunicazione | Il professore può preparare la lezione È possibile preparare la lezione |

INCLUS!VELY MODELLI DEEP LEARNING



Overview

Data-driven Methodology

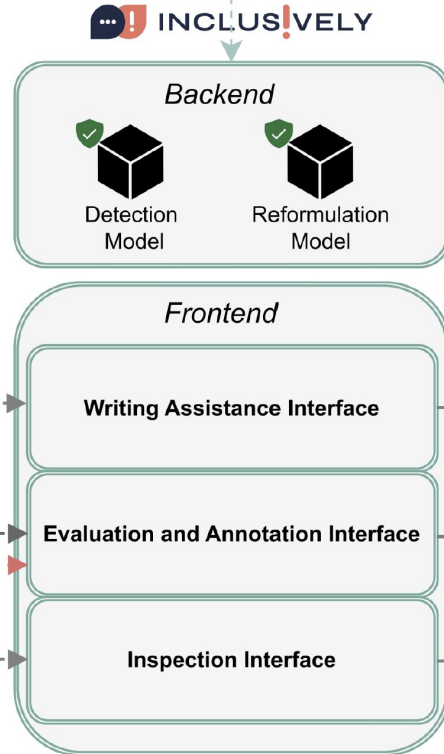


Application Scenario: Inclusive Writing Assistance

Possible Stakeholders



Input



Output



Politecnico
di Torino

E MIMIC





Alcuni risultati

Raccolta dati ed annotazioni

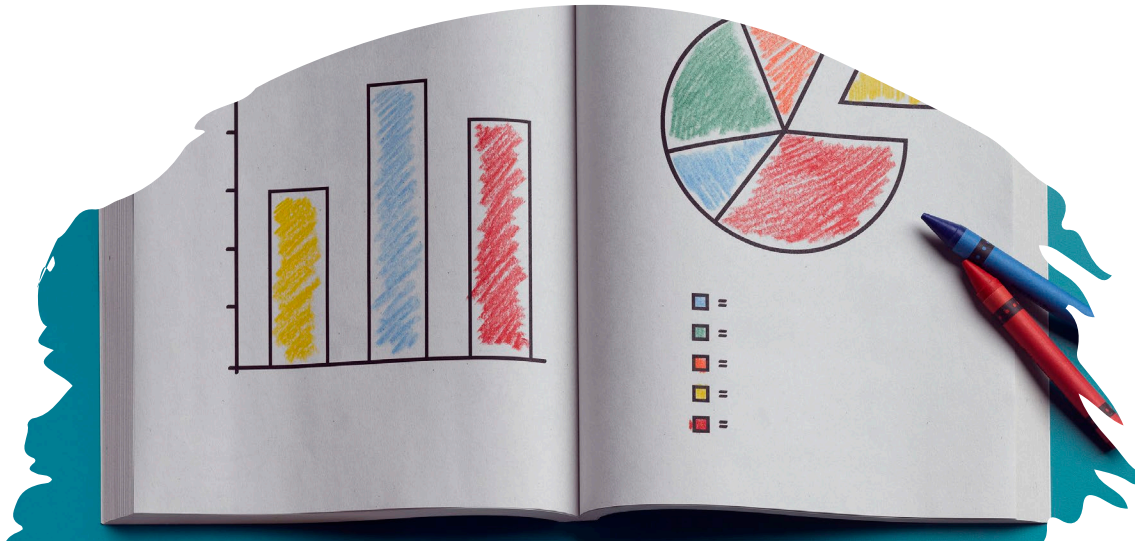
- Corpus italiano di documenti amministrativi ed accademici.
- **227** documenti, **10.725** frasi
- Etichetta di classe: "inclusiva" o "non-inclusiva"
- **4.705** riformulazioni inclusive.

Task di classificazione

- Classificazione di frasi in "inclusiva" o "non-inclusiva"
- N. 2 modelli di deep learning (BERT multilingua e BERT italiano)
- Modello selezionato: 9 frasi su 10 sono etichettate correttamente

Task di riformulazione

- Modello di deep learning (T5) che propone riformulazioni inclusive
- Il personale linguistico esperto ha validato le riformulazioni proposte dal modello considerandole corrette 8 volte su 10.



Politecnico
di Torino

E  MIMIC



Politecnico
di Torino

E  MIMIC

Attività di ricerca

Estensione ad altre Lingue

Francese, Spagnolo, minoranze linguistiche.

Estensione a nuovi domini

generazione di un corpus giuridico inclusivo.

Humans in the analytics loop

Validazione continua con esperti di dominio





IMPACT

Impatto scientifico

Rilascio di algoritmi e modelli deep learning adattabili a contesti e lingue diverse.

Impatto sul territorio e sulla società

L'uso e la sperimentazione di E-MIMIC nella pubblica amministrazione faciliterà la progressiva accettazione dell'uso della comunicazione inclusiva grazie ad interfacce che garantiranno livelli ottimali di user-experience.

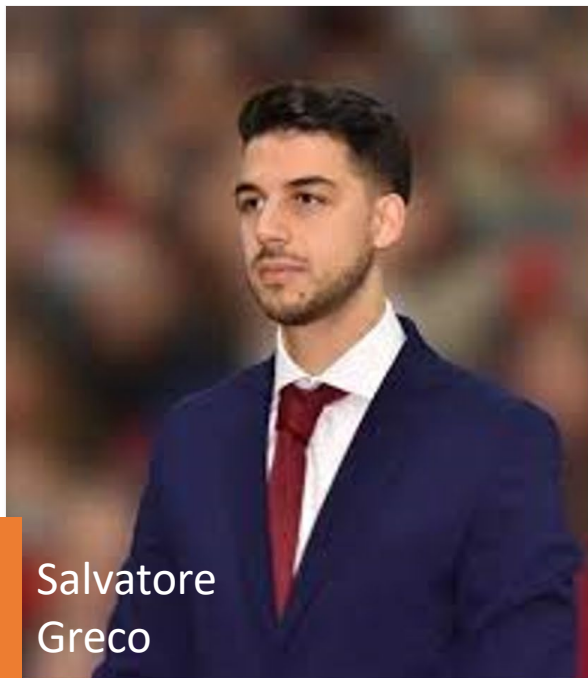
Impatto economico e formativo

E-MIMIC, strumento semplice e artificialmente inclusivo, sarà di supporto ad una società più equa, solidale e rispettosa delle diversità.



Politecnico
di Torino

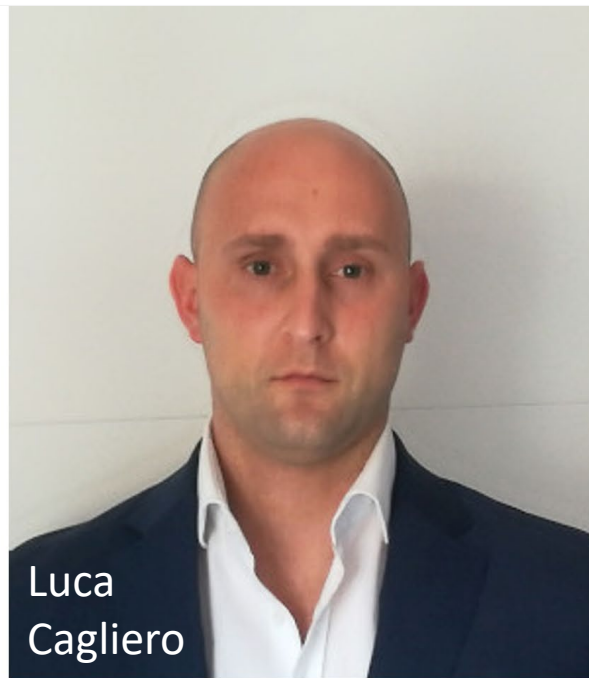
E  MIMIC



Salvatore
Greco



Moreno
La Quatra



Luca
Cagliero



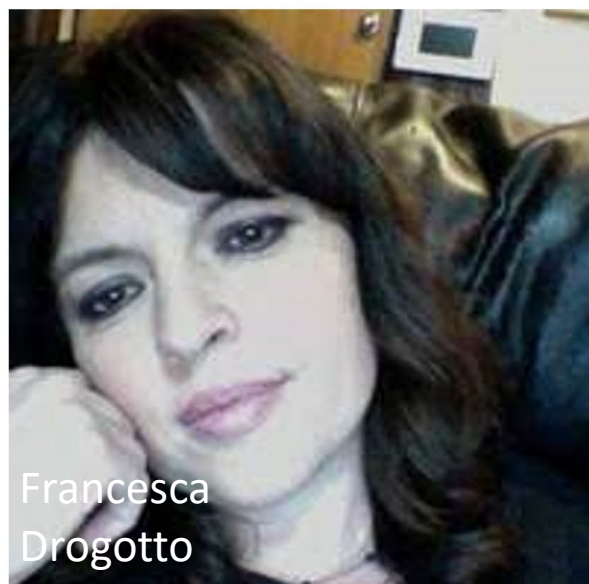
Tania
Cerquitelli



Rachele
Raus



Stefania
Cavagnoli



Francesca
Droetto



**Team di
Ricerca**

Pubblicazioni

Moreno La Quatra, Salvatore Greco, Luca Cagliero, Tania Cerquitelli (2023) Inclusively: An AI-Based Assistant for Inclusive Writing. ECML/PKDD (7) 2023: 361-365

Giuseppe Attanasio, Salvatore Greco, Moreno La Quatra, Luca Cagliero, Rachele Raus, Michela Tonti, and Tania Cerquitelli (2021). E-MIMIC: Empowering Multilingual Inclusive Communication. IEEE Big Data Workshops

Raus, R.; Tonti, M.; Cerquitelli, T.; Cagliero, L.; Attanasio, G.; La Quatra, M.; Greco S. (2022) L'analyse du discours et l'intelligence artificielle pour réaliser une écriture inclusive: le projet EMIMIC. In: SHS WEB OF CONFERENCES. - ISSN 2261-2424. - 138:(2022), p. 01007.

The E-MIMIC project website: <https://dbdmg.polito.it/e-mimic/>.



**Politecnico
di Torino**

Grazie per l'attenzione

E MIMIC

Empowering Multilingual Inclusive Communication

Prof.ssa Tania CERQUITELLI
Dipartimento di Automatica e Informatica
tania.cerquitelli@polito.it