

Riformulazione automatica di testi in lingua italiana in forma inclusiva mediante Intelligenza Artificiale

Tania Cerquitelli*, Luca Cagliero*, Salvatore Greco*, Moreno La Quatra*, Rachele Raus**, Michela Tonti§§

*Politecnico di Torino, Dipartimento di Automatica e Informatica,

tania.cerquitelli@polito.it luca.cagliero@polito.it salvatore_greco@polito.it moreno.laquatra@polito.it

** Università di Bologna, Dipartimento di Interpretazione e Traduzione, rachele.raus@unibo.it

§§ Università degli studi di Bergamo, Dipartimento di Lingue, Letterature e Culture Straniere, michela.tonti@unibg.it

CONTESTO SCIENTIFICO

L'uso attento della lingua (Linee guida della Comunità Europea per la comunicazione *gender-neutral*), contribuisce alla **progressione socio-culturale della popolazione**.

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

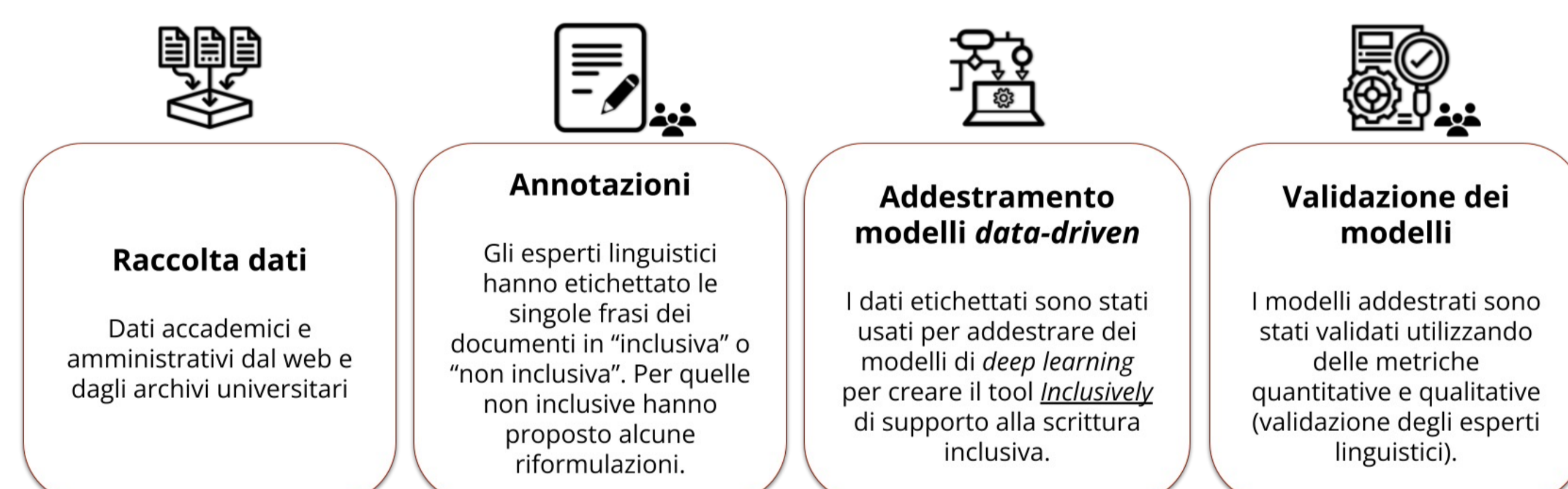
Strumenti digitali innovativi, intelligenti, user-friendly, ed inclusivi, potrebbero aiutare ad informare, correggere, e educare le persone all'utilizzo di termini inclusivi, per **una società più equa, solidale e rispettosa delle diversità**.

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.

INCLUSIVELY: METODOLOGIA DATA-DRIVEN



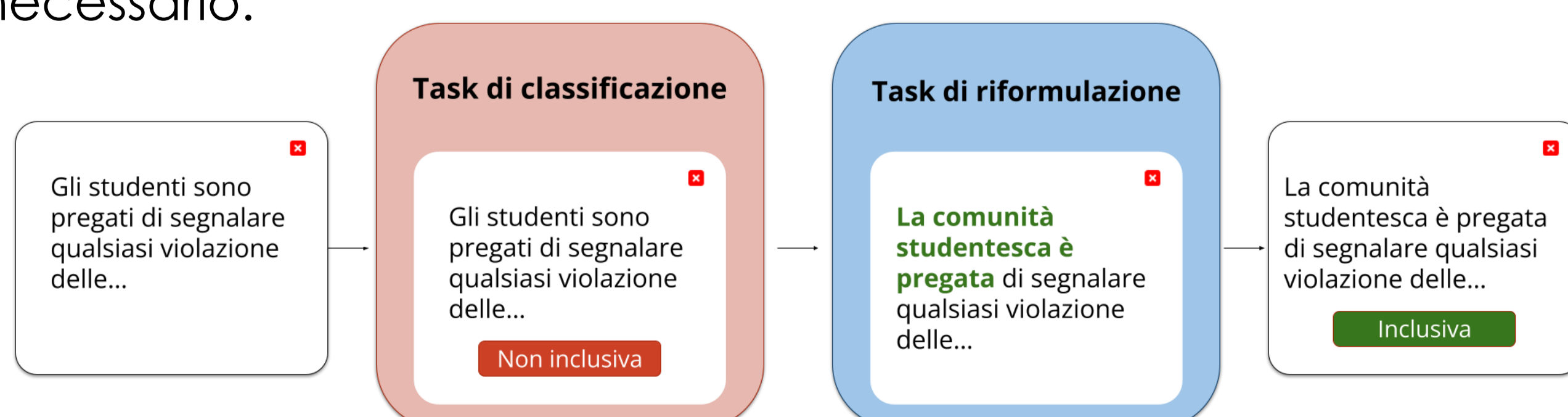
CRITERI LINGUISTICI

Persone esperte nel campo della linguistica italiana hanno annotato le frasi proponendo di frammenti non inclusivi, le alternative, varianti inclusive. In Tabella sono riportati alcuni dei criteri linguistici definiti dagli esperti ed utilizzati durante il processo di annotazione. Per ogni criterio è indicato un esempio.

Criteri Linguistici	Esempi Inclusivi (Non Includivi)
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.

INCLUSIVELY: MODELLI DEEP LEARNING

Inclusively integra due modelli di deep learning in cascata. Il primo effettua il task di classificazione della frase in "inclusiva" o "non inclusiva". Il secondo propone una riformulazione inclusiva ove necessario.



ALCUNI RISULTATI SPERIMENTALI

Raccolta dati ed annotazioni. È stato raccolto ed etichettato un corpus italiano di documenti amministrativi ed accademici. Il dataset è composto da 10.000 frasi, ciascuna arricchita con l'etichetta di classe ("inclusiva" o "non-inclusiva"). Il processo di annotazione ha generato 4.700 riformulazioni inclusive.

Task di classificazione. Per la classificazione di frasi in "inclusiva" o "non-inclusiva" sono stati specializzati due modelli di deep learning (BERT multilingua e BERT italiano) precedente addestrati per la conoscenza della lingua italiana. Il modello di classificazione più accurato assegna correttamente 9 volte su 10 l'etichetta "inclusiva" o "non inclusiva".

Task di riformulazione. È stato addestrato un modello di deep learning (T5) che propone riformulazioni inclusive delle frasi. Il personale linguistico esperto ha validato le riformulazioni proposte dal modello considerandole corrette 8 volte su 10.

IMPATTO

Impatto scientifico: rilascio di algoritmi e modelli deep learning adattabili a contesti e lingue diverse per la riformulazione di testi in un'ottica inclusiva.

Impatto sul territorio e sulla società: l'uso e la sperimentazione di *Inclusively* 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: *Inclusively*, strumento semplice e artificialmente inclusivo, sarà di supporto ad una società più equa, solidale e rispettosa delle diversità.

ATTIVITA' IN CORSO

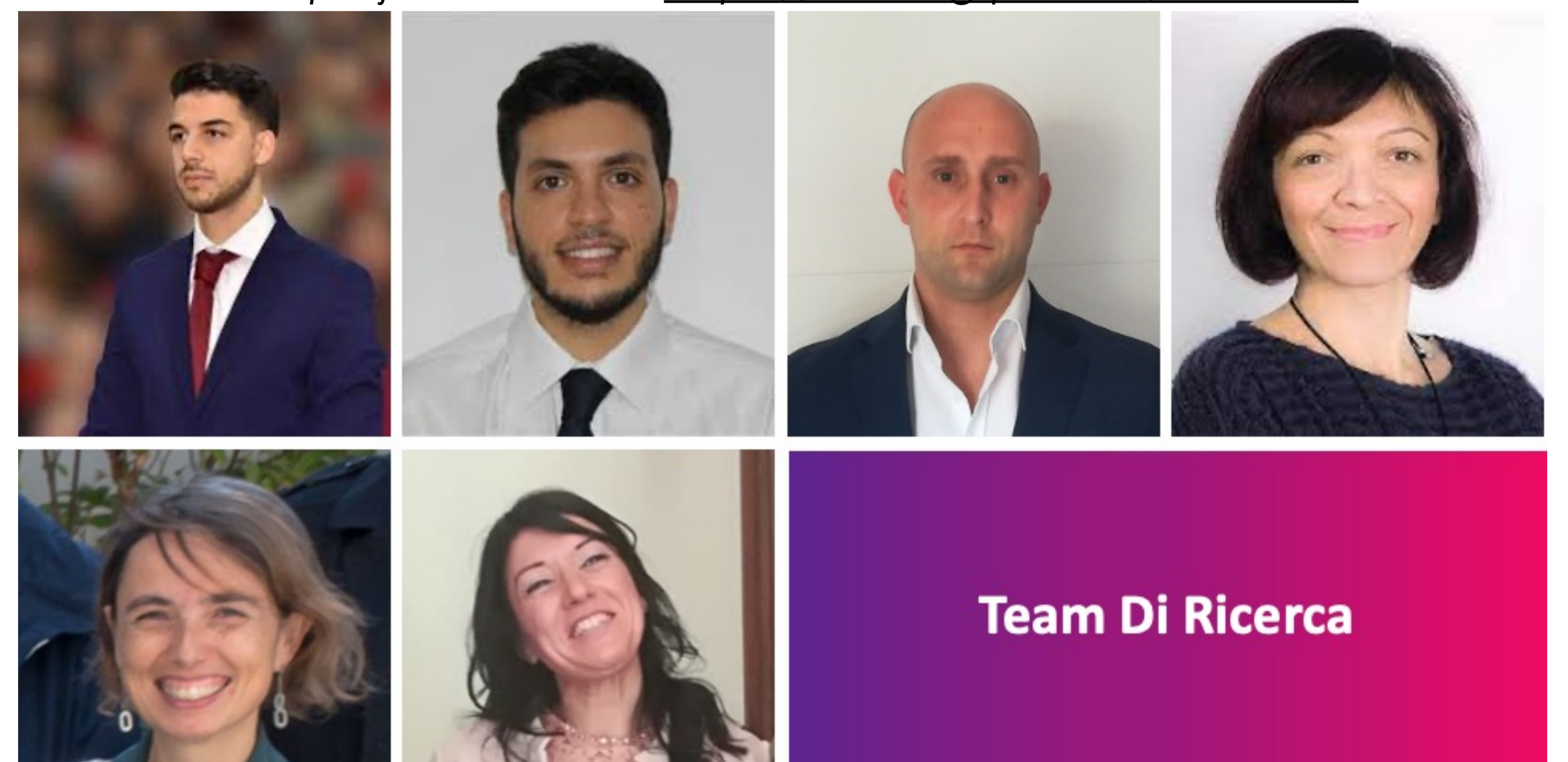
Estensione ad altre Lingue: Francese, Spagnolo, minoranze linguistiche.

Estensione a nuovi domini: generazione di un corpus giuridico inclusivo.

Humans in the analytics loop per validare in maniera continuativa *Inclusively* e i modelli deep learning in esso integrati.

BIBLIOGRAFIA

- 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, 2021
- 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/>.



Team Di Ricerca