FONDAZIONE BRUNO KESSLER
MAGAZINE

La prima famiglia di modelli open science per il riconoscimento vocale e la traduzione del parlato

30 Giugno 2025

FBK costruisce il primo sistema vocale open-science su larga scala per italiano e inglese

Un sistema di riconoscimento vocale e traduzione del parlato sviluppato interamente da zero, senza utilizzare modelli preaddestrati delle big tech, costruito esclusivamente su dati e strumenti totalmente open. È questo l'obiettivo raggiunto da **SpeechTek** e **Machine Translation**, due unità della Fondazione Bruno Kessler, con il progetto **"FAMA: The First Large-Scale Open-Science Speech Foundation Model for English and Italian"** innovativo per approccio, visione e impatto e realizzato all'interno delle attività della fondazione FAIR – Future Artificial Intelligence Research.

La vera innovazione non sta solo nella qualità del modello, ma nella sua **totale apertura**: il modello è infatti stato addestrato su oltre **150.000 ore di dati audio**, tutti liberamente accessibili e con licenze permissive. Ai dati audio già disponibili pubblicamente è stata aggiunta una grande quantità di cosiddetti "dati sintetici", ovvero trascrizioni e traduzioni automatiche in italiano e inglese, realizzati appositamente per il progetto e resi disponibili tramite il dataset MOSEL.

"Il codice, i dati e le procedure utilizzate sono completamente pubblici e documentati, consentendo a chiunque di replicare o adattare il sistema. Il know-how che ha generato questo progetto congiunto e le possibili applicazioni e sviluppo di FAMA lo rendono un asset importante per FBK" spiegano **Alessio Brutti**, responsabile dell'unità SpeechTek, e **Luisa Bentivogli**, responsabile dell'unità Machine Translation.

"Abbiamo dimostrato che anche in Italia abbiamo le competenze per creare modelli su larga scala" – aggiungono i coordinatori del progetto **Sara Papi** e **Marco Gaido** – *"che sono capaci di competere a livello internazionale in completa conformità alle recenti normative europee"*.

Attualmente il modello supporta l’italiano e l’inglese ma, grazie all’esperienza maturata su dati, calcolo e gestione delle risorse, costituisce una base solida per costruire in futuro una piattaforma vocale multilingue, interamente open source.

Per addestrare il modello sono state utilizzate le **risorse computazionali messe a disposizione da CINECA,** che ha fornito la potenza di calcolo e il supporto necessari per gestire i processi su larga scala.

Oltre agli aspetti tecnologici, il progetto rappresenta un passo importante verso un’intelligenza artificiale più **aperta, indipendente e riproducibile**, ponendo le basi per un ecosistema digitale più equo e accessibile.

Il modello è disponibile su HuggingFace nelle versioni small e medium.



FAMA: The First Large-Scale Open-Science Speech Foundation Model for English and Italian

In foto:

Dietro da sinistra: Mauro Cettolo | Matteo Negri | Alessio Brutti | Mohamed Nabih
Davanti da sinistra: Sara Papi | Marco Gaido | Marco Matassoni | Luisa Bentivogli

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadom Piano Nazionale di Ripresa e Resilienza

LINK

https://magazine.fbk.eu/it/news/la-prima-famiglia-di-modelli-open-science-per-il-riconoscimento-vocale-e-la-traduzione-del-parlato/

TAG

- #dataset
- #fair
- #fama
- #industriadigitale
- #intelligenzaartificiale
- #intelligenzaaumentata
- #modello
- #mosel
- #open science
- #open source
- #riconoscimento vocale
- #traduzione

AUTORI

- Michela Antino