

**moz://a**



# Riconoscimento vocale con DeepSpeech by Mozilla

Aggiornamenti, modello italiano e contest!

25/10/2020

**Stefania Delprete**  
*Data Scientist*

Credit: ESA



# Agenda



- ❑ Mozilla e voice recognition
- ❑ Lo sviluppo di DeepSpeech
- ❑ Modello italiano e contest
- ❑ Common Voice e sprint

# Chi sono?



**Stefania Delprete @astrastefania**

Data Scientist e tech tutor in TOP-IX/BIG DIVE.

Ho lavorato con dati da Measumerent Lab anche usati da Mozilla for Internet Health Report.

Organizzo MathsJam e gruppi di discussione di Altruismo Efficace, supporto a conferenze (PyCon, MERGE-it, Linux Day, ... ) e sono una **Mozillian!**

Vari talk su Rust in Firefox, Data Science, Common Voice e DeepSpeech.

Co-organizzo eventi su Mozilla e Rust.



# Cos'è Mozilla?



Si tratta di una fondazione no profit di cui Mozilla Italia è la comunità linguistica e nazionale italiana.

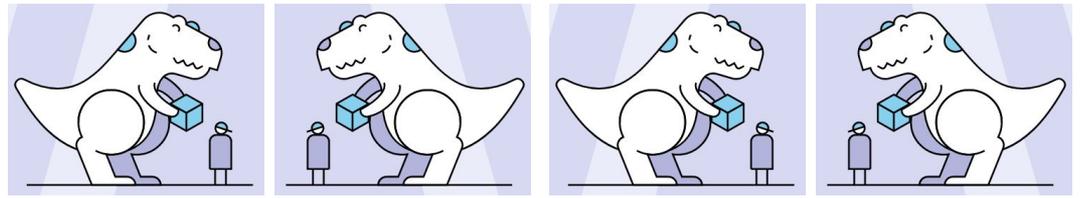
**moz://a**

we stand for  
the internet



Nella prossima  
foto di gruppo  
potresti esserci  
anche tu!

# Manifesto moz://a



- Internet è una **risorsa pubblica globale** che deve rimanere aperta e accessibile.
- Internet deve **arricchire la vita** di ogni essere umano.
- L'efficacia di Internet come risorsa pubblica dipende dal suo carattere di **interoperabilità** (relativamente a protocolli, formati di dati, contenuto), dal suo grado di **innovazione** e dalla **partecipazione decentralizzata** a livello mondiale.
- La partecipazione commerciale allo sviluppo di Internet è in grado di apportare numerosi benefici, ma è fondamentale un **equilibrio tra profitto commerciale e benefici pubblici**.
- La **valorizzazione degli aspetti di pubblica utilità** di Internet rappresenta un obiettivo importante, che merita tempo, attenzione e impegno.

<https://www.mozilla.org/it/about/manifesto>

# Community portal



[community.mozilla.org/groups/mozilla-italia](https://community.mozilla.org/groups/mozilla-italia)

[community.mozilla.org/groups/rust-italia](https://community.mozilla.org/groups/rust-italia)

## Mozilla Italia

Verified | Italy | Created January 20, 2020

About us | Our Events | Our Members



moz://a  
Italia

Leave Group

Share Group

Activity

1 Events this month

37 Members

Related Events

**Nov 03** Ritorna la Common Voice sprint in italiano!  
November 03, 2020 · 00:00 UTC  
Online Event

**About Us**

Il gruppo dei membri della comunità Mozilla Italia.



## Rust Italia

Verified | Italy | Created May 08, 2020

About us | Our Events | Our Members



Rust  
Italia

Share Group

**About Us**

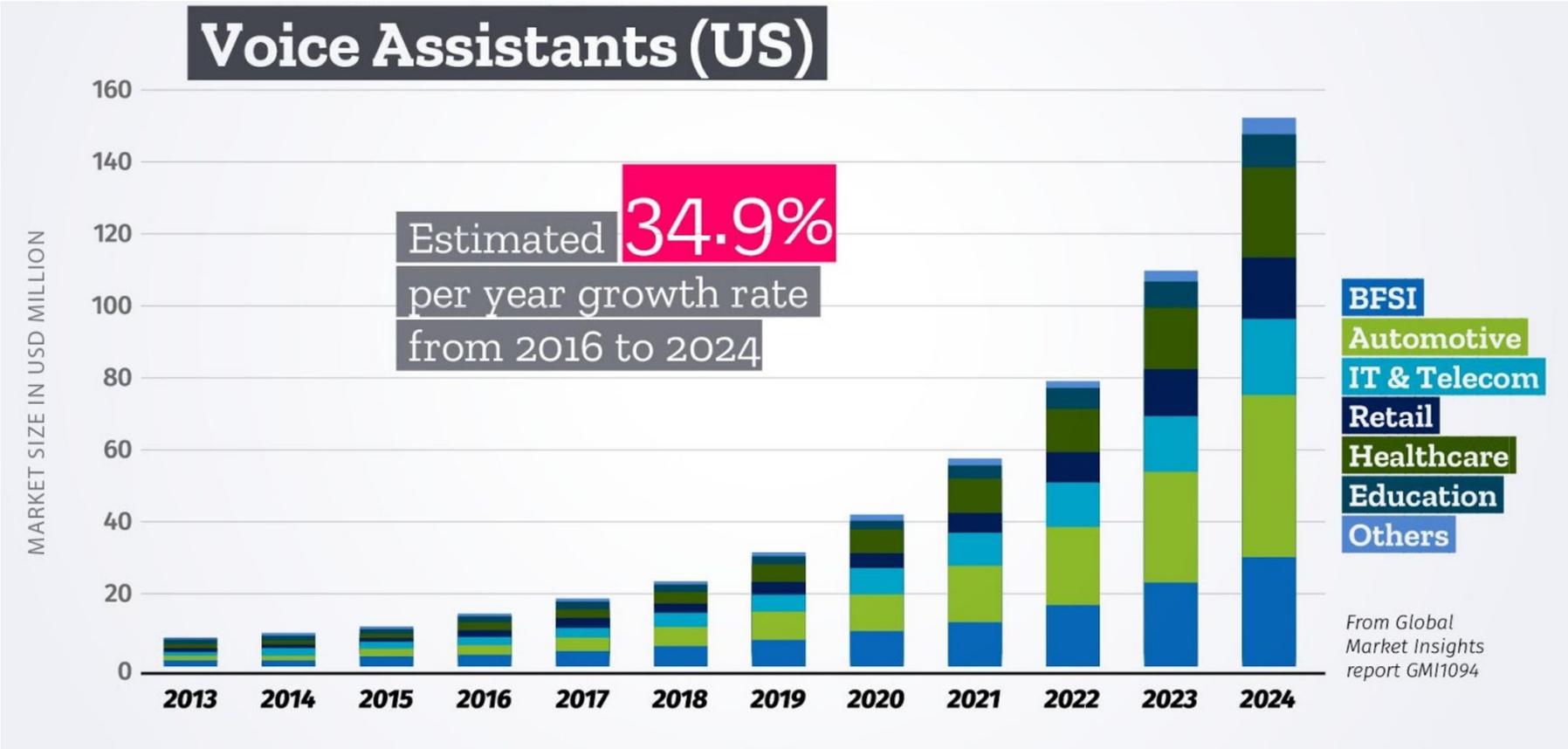
Gruppo italiano della comunità di Rust in Italia.  
Slack <http://rust-italia.herokuapp.com>

Ecco i contatti delle comunità locali:



# **Mozilla e il mondo voice**

# Perché Mozilla si è focalizzata su voice recognition?



# Perché Mozilla si è focalizzata su voice recognition?

- Necessità di un'alternativa **open source**
- Accessibilità **off-line** e data storage
- Rendere il riconoscimento vocale disponibile per **più lingue**
- Per non avere limitazioni sull'utilizzo di questi servizi
- Stato dell'arte di **Speech-to-Text**
- End-to-end **Machine Learning** system

# Perché Mozilla si è focalizzata su voice recognition?

*C'era una volta...*

The project Vaani: voice assistant per Firefox OS

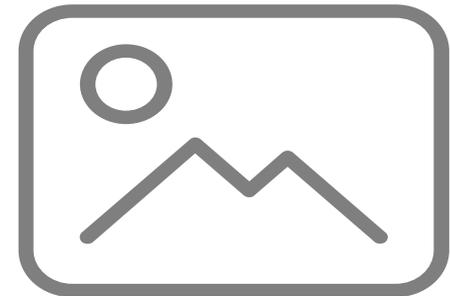
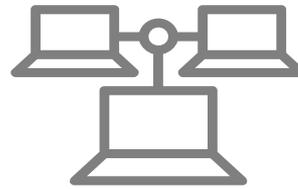
Risorse dirette su Connected Devices, ora [iot.mozilla.org](https://iot.mozilla.org)

Nuovo team dedicato al Machine Learning sotto [research.mozilla.org](https://research.mozilla.org)

# DeepSpeech by Mozilla

# Cos'è il Machine Learning?

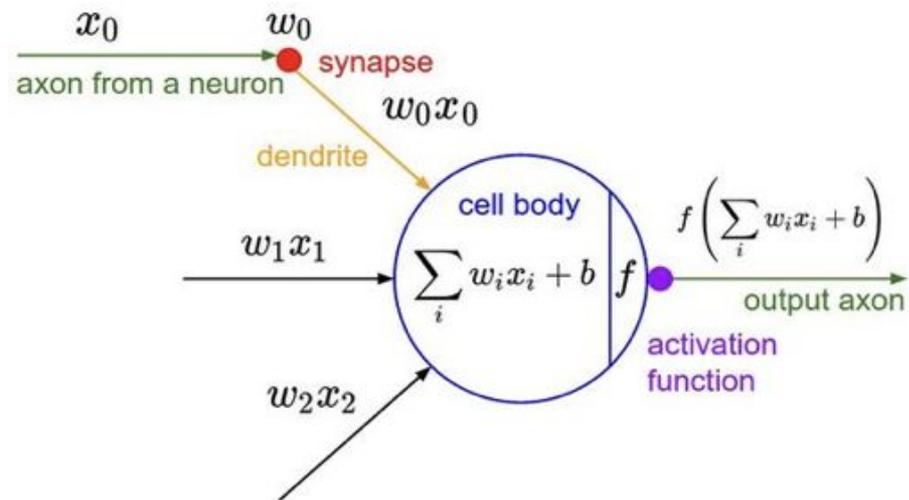
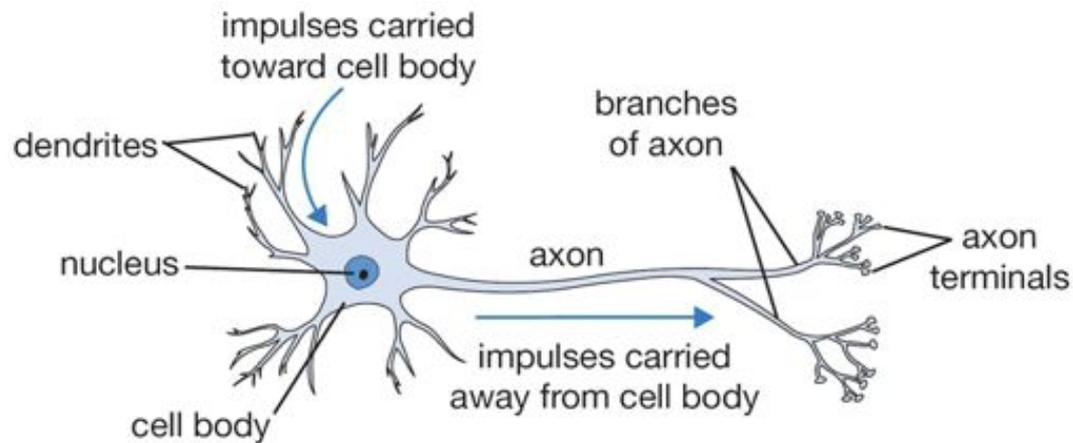
È quel processo che riguarda lo studio, la costruzione e l'implementazione di algoritmi che permettono di **imparare e fare previsioni in modo automatico** talvolta partendo da un insieme di dati in ingresso (files di testo, musicali, immagini, etc...).



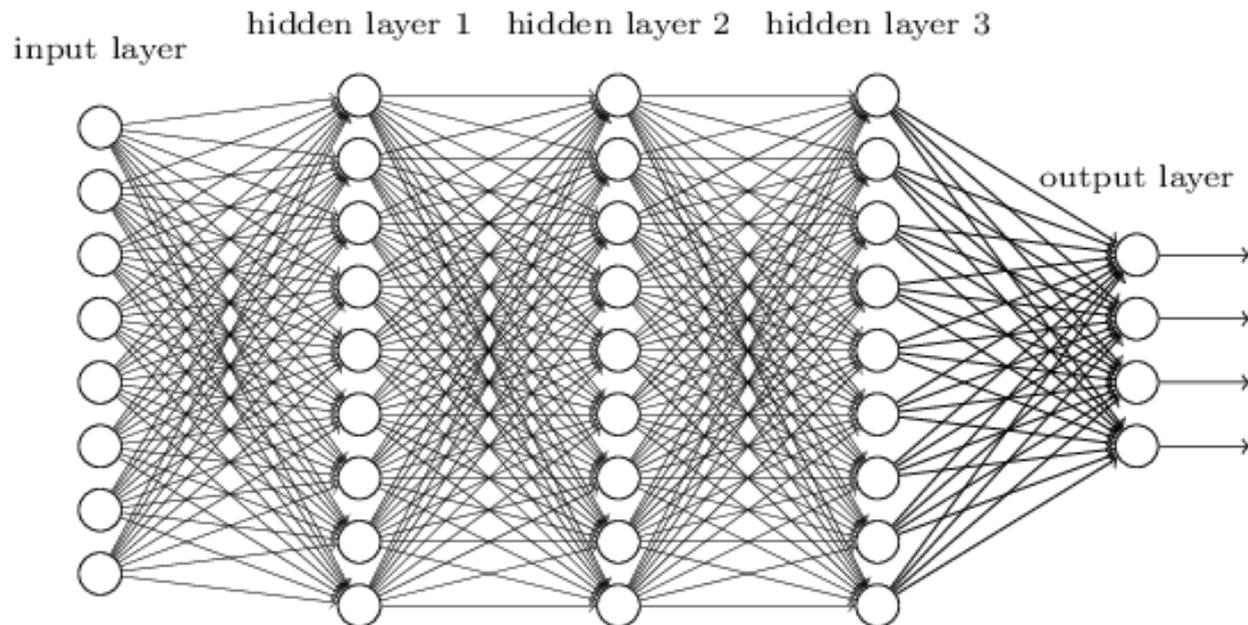
# Come funziona il Machine Learning?

- 1) Raccolta dei dati
- 2) Preparazione dei dati
- 3) Scelta di un modello
- 4) Fase di Training
- 5) Valutazione dei risultati
- 6) Predizione

# ANN, Artificial Neural Network



# DNN, Deep Neural Network



Tinker With a **Neural Network** Right Here in Your Browser.  
Don't Worry, You Can't Break It. We Promise.

Epoch: 000,190    Learning rate: 0.03    Activation: Tanh    Regularization: None    Regularization rate: 0    Problem type: Classification

DATA: Which dataset do you want to use?

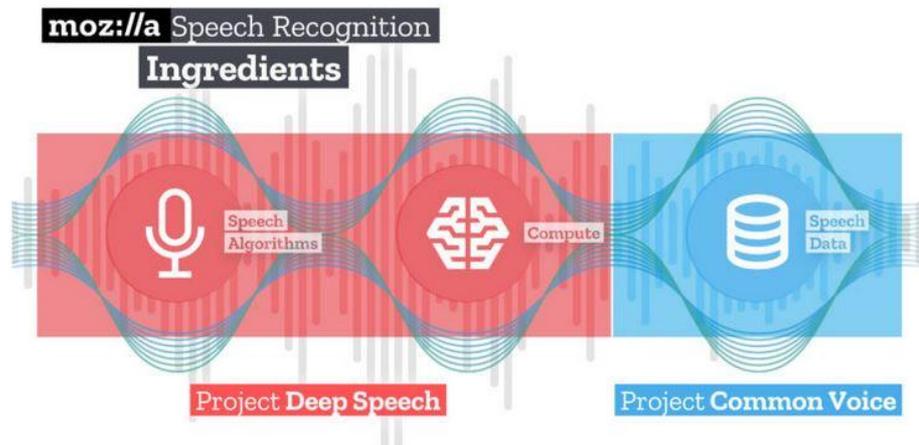
FEATURES: Which properties do you want to feed in?

2 HIDDEN LAYERS: 4 neurons, 2 neurons

OUTPUT: Test loss 0.002, Training loss 0.003

# DeepSpeech

Motore di riconoscimento vocale, open source e sviluppato da Mozilla basandosi sul paper di Baidu: "Deep Speech: Scaling up end-to-end speech recognition" realizzato con TensorFlow



<https://github.com/mozilla/DeepSpeech>

# Deep Speech paper v2 (Baidu)

<https://arxiv.org/abs/1412.5567>

---

## Deep Speech: Scaling up end-to-end speech recognition

---

Awni Hannun\*, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng

Baidu Research – Silicon Valley AI Lab

### Abstract

We present a state-of-the-art **speech recognition system** developed using **end-to-end deep learning**. Our architecture is significantly simpler than traditional speech systems, which rely on laboriously engineered processing pipelines; these traditional systems also tend to perform poorly when used in noisy environments. In contrast, our system does not need hand-designed components to model background noise, reverberation, or speaker variation, but instead directly learns a function that is robust to such effects. We do not need a phoneme dictionary, nor even the concept of a “phoneme.” Key to our approach is a **well-optimized RNN training system that uses multiple GPUs**, as well as a set of novel data synthesis techniques that allow us to efficiently obtain a large amount of varied data for training. Our system, called Deep Speech, outperforms previously published results on the widely studied Switchboard Hub5\*00, achieving 16.0% error on the full test set. Deep Speech **also handles challenging noisy environments better than widely used**, state-of-the-art commercial speech systems.

### 1 Introduction

Top speech recognition systems rely on sophisticated pipelines composed of multiple algorithms and hand-engineered processing steps. In this paper, we describe an end-to-end speech system

412.5567v2 [cs.CL] 19 Dec 2014

## Deep Speech paper (Baidu)

I punti principali per creare un algoritmo speech-to-text

- No a pipeline complesse
- Usare Deep Learning (anche RNN)
- Fare leva su GPU e processi paralleli
- Un grande dataset da cui imparare (con anche rumori di fondo)

# Implementazione con TensorFlow

- Layer nascosti fully connected con uso di ReLU activation.
- Architettura LSTM su layer di RNN e tanh activation
- Adam optimizer [arxiv.org/pdf/1412.6980.pdf](https://arxiv.org/pdf/1412.6980.pdf)

<https://hacks.mozilla.org/?s=deepspeech>

# Dataset e iperparametri

## LibriSpeech

LibriSpeech is a corpus of approximately 1000 hours of 16Khz read English speech derived from read audiobooks from the LibriVox project.

License: [CC-BY-4.0](#)

[Go to LibriSpeech](#)

## TED-LIUM Corpus

The TED-LIUM corpus was made from audio talks and their transcriptions available on the TED website.

License: [CC-BY-NC-ND 3.0](#)

[Download Data](#)

## VoxForge

VoxForge was set up to collect transcribed speech for use with Free and Open Source Speech Recognition Engines.

License: [GNU-GPL](#)

[Download Data](#)

## Tatoeba

Tatoeba is a large database of sentences, translations, and spoken audio for use in language learning. This download contains spoken English recorded by their community.

License: [Mixed](#)

[Download Data](#)

# Sviluppi nelle versioni di DeepSpeech

La versione 0.5.0 ha permesso di usare anche engine TensorFlow Lite

La versione 0.6.0 ha incorporare modelli più robusti ai rumori di fondo.

**[github.com/mozilla/DeepSpeech/projects](https://github.com/mozilla/DeepSpeech/projects)**

# Dataset e iperparametri

Parametri usati in DeepSpeech v0.7.0, aggiornando learning rate e epochs in fasi successive.

- `train_files` [Fisher](#), [LibriSpeech](#), [Switchboard](#), [Common Voice English](#), and approximately 1700 hours of transcribed WAMU (NPR) radio shows explicitly licensed to use as training corpora.
- `dev_files` [LibriSpeech](#) clean dev corpus.
- `test_files` [LibriSpeech](#) clean test corpus
- `train_batch_size` 128
- `dev_batch_size` 128
- `test_batch_size` 128
- `n_hidden` 2048
- `learning_rate` 0.0001
- `dropout_rate` 0.40
- `epochs` 125

# DeepSpeech in continua evoluzione!

Miglioramento dei metadati nella 0.9.0, e arrivati contributi sul supporto di piattaforme .Net Core and .Net Standard per Linux e Windows (forse anche macOS).

E si va nella direzione di **DeepSpeech 1.0.0**!

**[github.com/mozilla/DeepSpeech/projects](https://github.com/mozilla/DeepSpeech/projects)**

# DeepSpeech bindings e contatti

Sono presenti anche bindings nativi per

- ❑ Node.js [github.com/mozilla/DeepSpeech#using-the-nodejs-package](https://github.com/mozilla/DeepSpeech#using-the-nodejs-package)
- ❑ Python [github.com/mozilla/DeepSpeech#using-the-python-package](https://github.com/mozilla/DeepSpeech#using-the-python-package)
- ❑ Rust [github.com/RustAudio/deepspeech-rs](https://github.com/RustAudio/deepspeech-rs)
- ❑ Go [github.com/asticode/go-astideepspeech](https://github.com/asticode/go-astideepspeech)
- ❑ GStreamer [github.com/Elleo/gst-deepspeech](https://github.com/Elleo/gst-deepspeech)

## Contatti

- ❑ [FAQ](#)
- ❑ [Discourse Forums](#)
- ❑ [#machinelearning](#) channel su [Mozilla IRC](#)

# Seconda versione modello DeepSpeech in Italiano

- Immagine Docker per generare modello
- Jupyter Notebook e script per replicare preparazione del dataset, addestramento
- Training con circa 257 ore
- Transfer learning dal modello in inglese
- Supporto da HPC4AI

*Grazie speciale a Stefano, Daniele, Damiano e Simone!*

**[github.com/MozillaItalia/DeepSpeech-Italian-Model](https://github.com/MozillaItalia/DeepSpeech-Italian-Model)**

Latest release

2020.08.07

714b07e

Verified

Compare

2020.08.07

Mte90 released this on Aug 7 · 13 commits to master since this release

## Changelog

After 5 months we release a new model with a lot of improvements!

- Don't use anymore the wikipedia dump as text corpus but Mitads (check at <https://github.com/MozillaItalia/DeepSpeech-Italian-Model/releases/tag/Mitads-1.0.0-alpha2>)
- DeepSpeech 0.8 based
- New docker scripts
  - Sets of `.env` files to try different parameters [https://github.com/MozillaItalia/DeepSpeech-Italian-Model/tree/master/DeepSpeech/env\\_files](https://github.com/MozillaItalia/DeepSpeech-Italian-Model/tree/master/DeepSpeech/env_files)
- In the meantime we released also the notebooks!
- We are using the Mozilla official DS model for English for transfer learning to improve the quality

## Second version of Italian model, trained with:

- ~130 hours for Common Voice IT dataset
- ~127 hours of m-ailabs Italian dataset  
total: ~257h

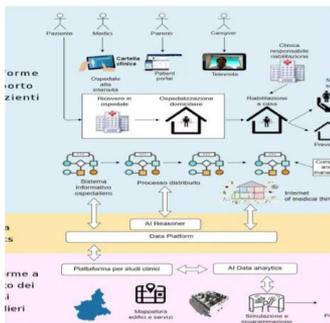
Available in 2 version `transfer` used transfer learning form the official English model release by mozilla and other one is from scratch .

# HPC4AI <https://hpc4ai.unito.it/projects>



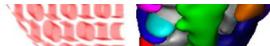
AI

AI-enabled radiomics: Virtual Biopsy



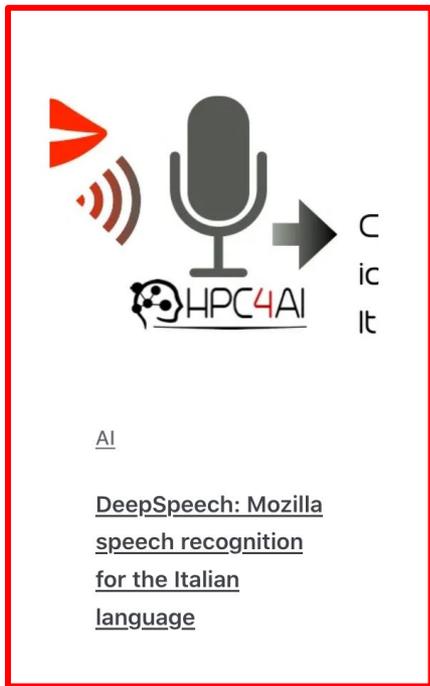
AI

CANP: AI-enabled telemedicine



AI

Advanced Algorithms for Neuroscience



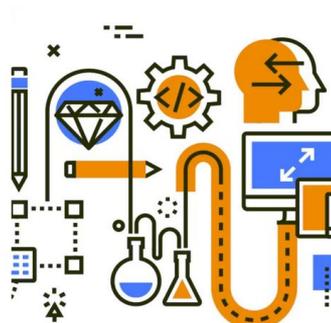
AI

DeepSpeech: Mozilla speech recognition for the Italian language



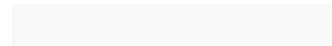
AI

Covid-19 outbreak and the Italian Economy



SCIENTIFIC  
COMPUTING

StreamFlow: Hybrid Cloud/HPC Workflows



SCIENTIFIC  
COMPUTING

OnToMap – Community Maps 3.0



SOCIAL

The didactic games from Bank of Italy



WE  
NEEDED  
YOU

# Contribuire al modello DeepSpeech in Italiano

- Analizzare l'audio dei dataset con tecniche di riduzione della dimensionalità (PCA o TSNE ad esempio)
- Sperimentare l'addestrare il modello con più strategie, ad esempio giocando con (iper)parametri, data augmentation, ...
- Nuove risorse di dataset audio!

**[github.com/MozillaItalia/DeepSpeech-Italian-Model](https://github.com/MozillaItalia/DeepSpeech-Italian-Model)**

# Mozilla Italia Developer contest! 🥰

In collaborazione con Italy Informatica

- Fino al 15 novembre
- Proponi il tuo progetto di **utilizzo del modello DeepSpeech in Italiano!**
- Esempi:
  - ◆ Videogioco a comandi vocali
  - ◆ Dettatore automatico per LibreOffice
  - ◆ Assistente vocale su Raspberry Pi e sensori
  - ◆ Controllo vocale del sistema operativo per disabili

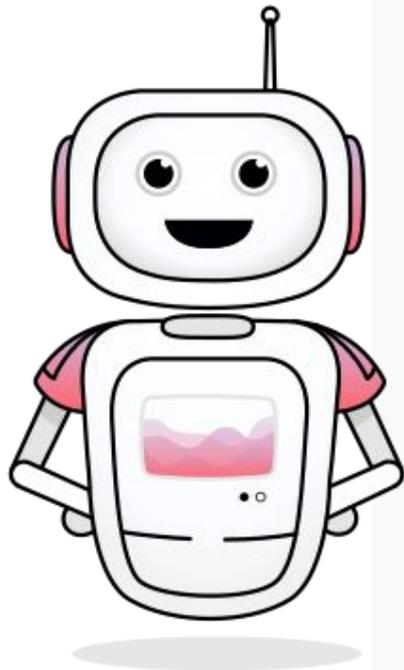
[https://www.reddit.com/r/ItalyInformatica/comments/j6112h/mozilla\\_itali\\_a\\_developer\\_contest\\_in](https://www.reddit.com/r/ItalyInformatica/comments/j6112h/mozilla_itali_a_developer_contest_in)



# Common Voice

# Stato del progetto per la lingua italiana

- Oltre 9000 frasi da diverse fonti
- Avviato il 3 Luglio 2018
- Promozione:
  - Social network
  - Linux Day
  - Italian Hacker Camp
  - ESC
  - Women in Voice
  - ...
- TU!



Lingua  
Italiano

DIMENSIONE **4 GB**

VERSIONE **it\_167h\_2020-06-22**

TOTALE ORE CONVALIDATE **130**

TOTALE ORE COMPLESSIVE **167**

LICENZA **CC-0**

NUMERO DI VOCI **5,446**

FORMATO AUDIO **MP3**

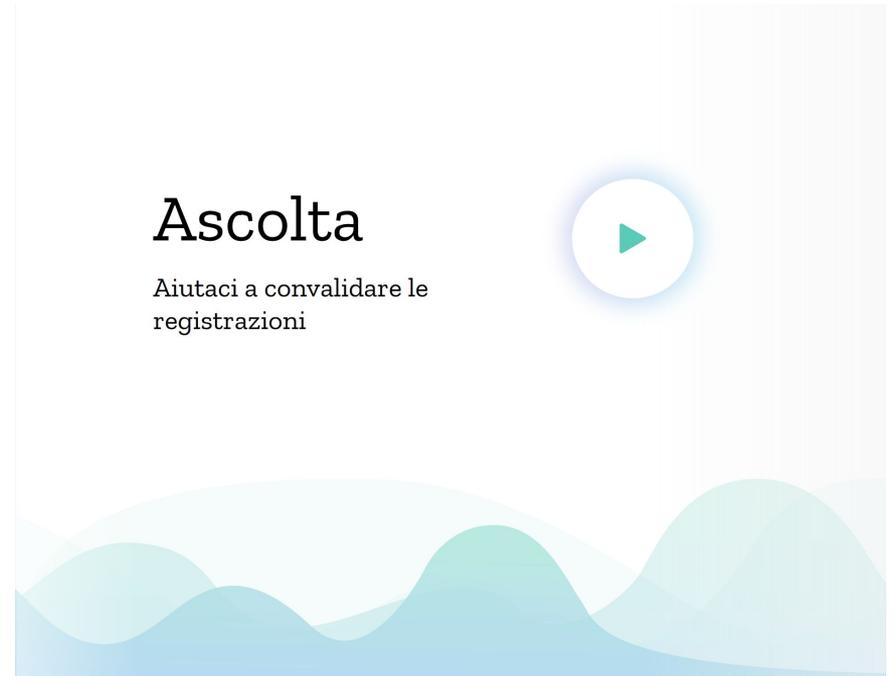
DIVISIONI  
Accento  
Età  
**23%** 19 - 29, **15%** 30 - 39,  
...  
**49%** Maschio,  
**16%** Femmina



WE  
NEEDED  
YOU

# Contribuire a Common Voice

Vai su <https://commonvoice.mozilla.org/it> e...



# Ritorna la Common Voice sprint in italiano!

<https://community.mozilla.org/en/events/ritorna-la-common-voice-sprint-in-italiano>



A promotional banner for the Common Voice sprint. The background is teal with various icons: a thumbs up, a speech bubble, a play button, and a microphone. A white robot character is on the right. The text is in white and teal.

3 novembre

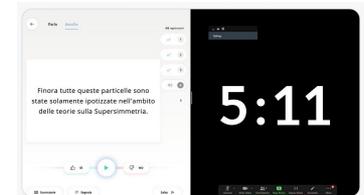
**Common Voice sprint**

Dona la tua voce o convalida  
registrazioni di altri!

moz://a  
Italia

**Ci vediamo  
martedì 3 novembre  
alle 21:00!** 🙌

**#commonvoicesprint**





# Restiamo in contatto!

Canale Telegram Italiano **@MozItaNews**

Gruppo Telegram Italiano **@MozItabot**

Sito e forum **mozillaitalia.org**

Comunità sul portale Mozilla

**community.mozilla.org/en/groups/rust-italia**

**community.mozilla.org/en/groups/mozilla-italia**



**Stefania Delprete @astrastefania**