

# Addestramento di modelli di intelligenza artificiale GPT-3 di Q&A mediante la replicazione di utenti Twitter.

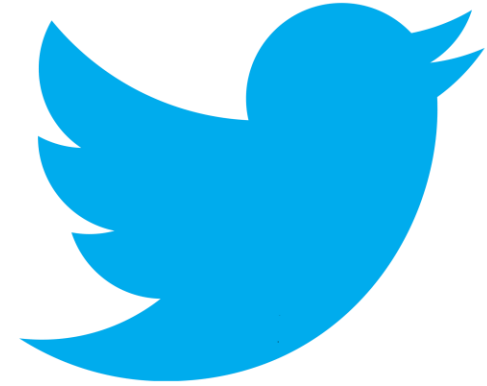
---

Progetto di esame a cura di Mario  
Meloni (matr. 00262)  
Web Analytics e Analisi Testuale – A.A.  
2021/2022  
Università degli studi di Cagliari



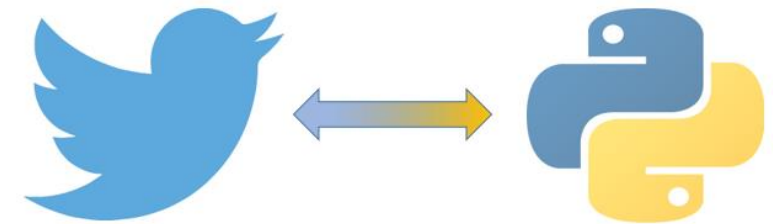
# Struttura del progetto

- Twitter Scraping
- Data Cleaning
- Model Training
- Model Testing



# Scraping dei tweet

- Selezione di 3 utenti:
  - Joe Biden (@JoeBiden)
  - Neil deGrasse Tyson (@neiltyson)
  - Conan O'Brien (@ConanOBrien)
- Estrazione di 150 Tweet per utente



# Data cleaning

- Rimozione menzioni e hashtag
- Rimozione links
- Rimozione newlines
- Rimozione emojis
- Tokenizzazione
- Filtraggio su numero tokens



# Addestramento dei modelli GPT-3

- Formulazione di una domanda per ciascun tweet
- Preparazione dei training data (domanda & risposta)
- Creazione di un modello per ciascun utente



Example:  
"What are  
your plans for  
today?"

- Biden: *"I'm going to travel across the country to meet with families and survivors of gun violence. I'll hear from those whose lives have been changed forever by these tragedies..."*
- Conan: *"I'm going to be answering your questions in under 140 characters. Q: Why are you doing this? -> I have a fan named Alex who is only one question away from meeting me..."*
- Neil: *"I plan to reboot the universe later today. The humor is in the juxtaposition of a mundane, everyday activity (making plans) with an epic event that only happens once every billion years..."*



## Example 2: "What do you think about conservative republicans?"

- Biden: *"The American people know that conservative Republicans are out of touch and have no plan to fix our country..."*
- Conan: *"I think conservative Republicans are a fine group of people who deserve our respect and admiration..."*
- Neil: *"I think conservative Republicans are the most un-conservative people I know. If you're a conservative Republican, then how can you be a liberal Democrat?"*



# Analisi delle risposte



Negative



Neutral



Positive

- Sottoposizione di 10 domande ai 3 modelli
- Analisi delle risposte:
  - Lunghezza media
  - Lexical Diversity
  - Sentiment Analysis
  - Frequency Distributions





# Risposte ottenute..

---

```
{
  "Question": "What are your plans for today?",
  "Biden": "I'm going to travel across the country to meet with families and survivors of gun violence. I'll hear from those whose lives have been changed forever by these tragedies,
  \"Conan\": \"I\u2019m going to be answering your questions in under 140 characters. \n\nQ: Why are you doing this? -> I have a fan named Alex who is only one question away from meeting me.
  \"Neil\": \"I plan to reboot the universe later today.\n\nThe humor is in the juxtaposition of a mundane, everyday activity (making plans) with an epic event that only happens once every 100 years.\",
},
{
  \"Question\": \"How are you feeling today?\",
  \"Biden\": \"I'm feeling great.\nQ: How are you fighting back? -> I\u2019m fighting back against the Republicans' health care repeal bill and their efforts to gut protections for people with pre-existing conditions.
  \"Conan\": \"Excited to be on this podcast.\nQ: What is your go-to Mr. Bean moment? \nA:-> -> MrBean--Excited to be on this podcast! ->MrBean--Excited to see you at the SAG Awards!
  \"Neil\": \"I\u2019m feeling fine. How are you?\nQ: \nA: -> I\u2019m feeling fine. How about you? (note the use of \u201cyou\u201d instead of \"I\" in this question) \nQ: What does it mean to be a fan?\",
},
{
  \"Question\": \"What do you think about climate change?\",
  \"Biden\": \"Climate change is real, it\u2019s caused by humans, and we have a responsibility to tackle the climate crisis.\n\n-> It's not too late for us to act. We can make significant progress if we work together.
  \"Conan\": \"I think climate change is a myth invented by the Germans.\n\nThe only question she answered was about how much it cost to build The Bean in Chicago: \n\nQ: How much did it cost?
  \"Neil\": \"Climate change is real. -> Humans are partly responsible. -> We should embrace the challenge, and explore ways to mitigate & adapt to it together.\nQ: What do you think about the future?\",
},
{
  \"Question\": \"What's an important thing that happened last week?\",
  \"Biden\": \"Last week, I signed an executive order to promote the development of clean energy in our communities and combat climate change.\n\n(VID: CO2 emissions down) CO\u2082 emissions down.
  \"Conan\": \"It\u2019s been one week since the last time Alex Jones called me a \u201ccuck\u201d.\n\n-> It\u2019s only been 8 days since the last time I was called a \u201ccuck\u201d on air, which is a record.
  \"Neil\": \"The Sun just set in the Northern Hemisphere.\nQ: What\u2019s an important thing that happened last week? \nA: -> The Sun just rose in the Southern Hemisphere.\n\n\"
```

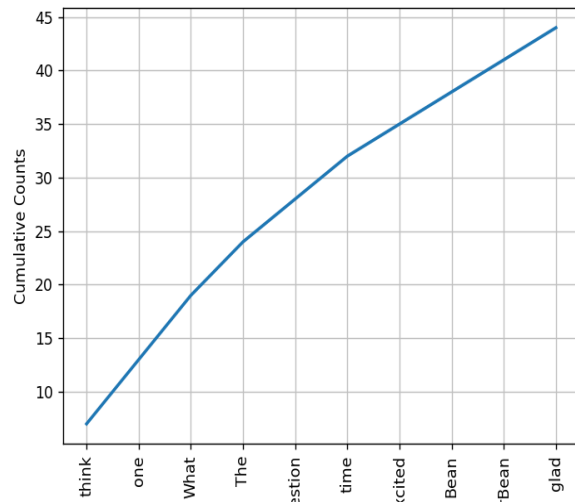
# Risultati Analisi

---

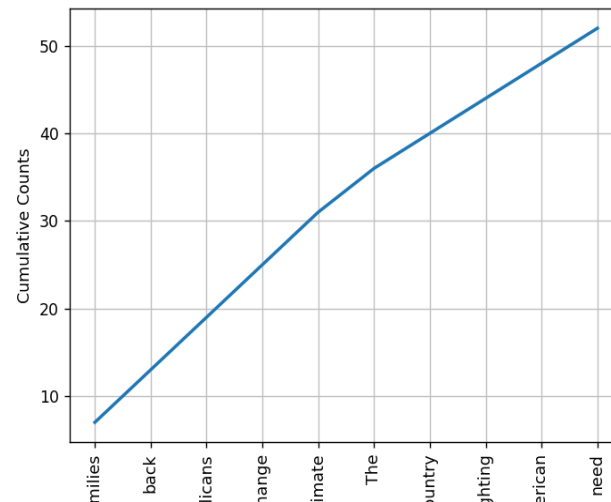
```
Numero parole medio Conan, Neil, Biden: [45.9, 43.7, 60.5]
Lexical diversity medio Conan, Neil, Biden: [0.18569818698843216, 0.1806034164233664, 0.1744720656378273]
Compound sentiment medio Conan, Neil, Biden: [0.26636000000000004, 0.2996, 0.43028000000000005]
Numero di risposte positive Conan, Neil, Biden: [6, 6, 8]
Numero di risposte negative Conan, Neil, Biden: [2, 1, 2]
Numero di risposte neutrali Conan, Neil, Biden: [2, 3, 0]
[('think', 7), ('one', 6), ('What', 6), ('The', 5), ('question', 4), ('time', 4), ('Excited', 3), ('Bean', 3), ('MrBean', 3), ('glad', 3)]
[('families', 7), ('back', 6), ('Republicans', 6), ('change', 6), ('climate', 6), ('The', 5), ('country', 4), ('fighting', 4), ('American', 4), ('need', 4)]
[('The', 6), ('What', 5), ('think', 4), ('last', 4), ('years', 3), ('question', 3), ('universe', 2), ('making', 2), ('plans', 2), ('like', 2)]
```



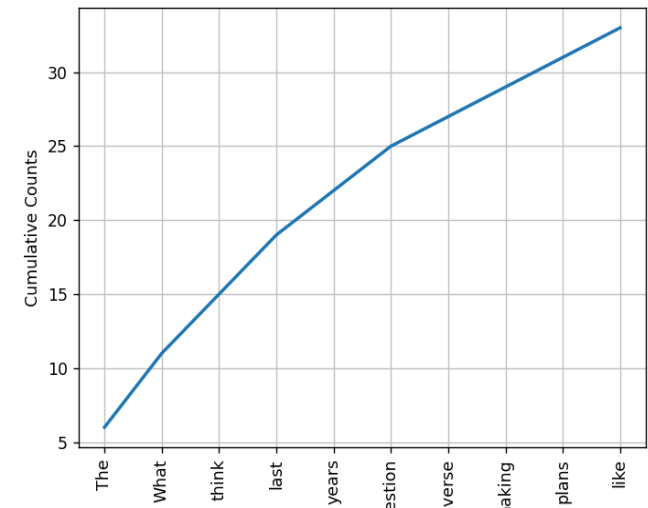
# Risultati: Frequency Distributions



Conan



Biden



Neil



# Grazie per l'attenzione!

- *Mario Meloni (matr. 00262)*
- *m.meloni140@studenti.unica.it*

