

# *TweetNLP*

## Text Analytic Package for Social Media

PyDATA Bristol  
Asahi Ushio  
Cardiff NLP



Cardiff NLP

CARDIFF  
UNIVERSITY

PRIFYSGOL  
CAERDYDD

# About Me

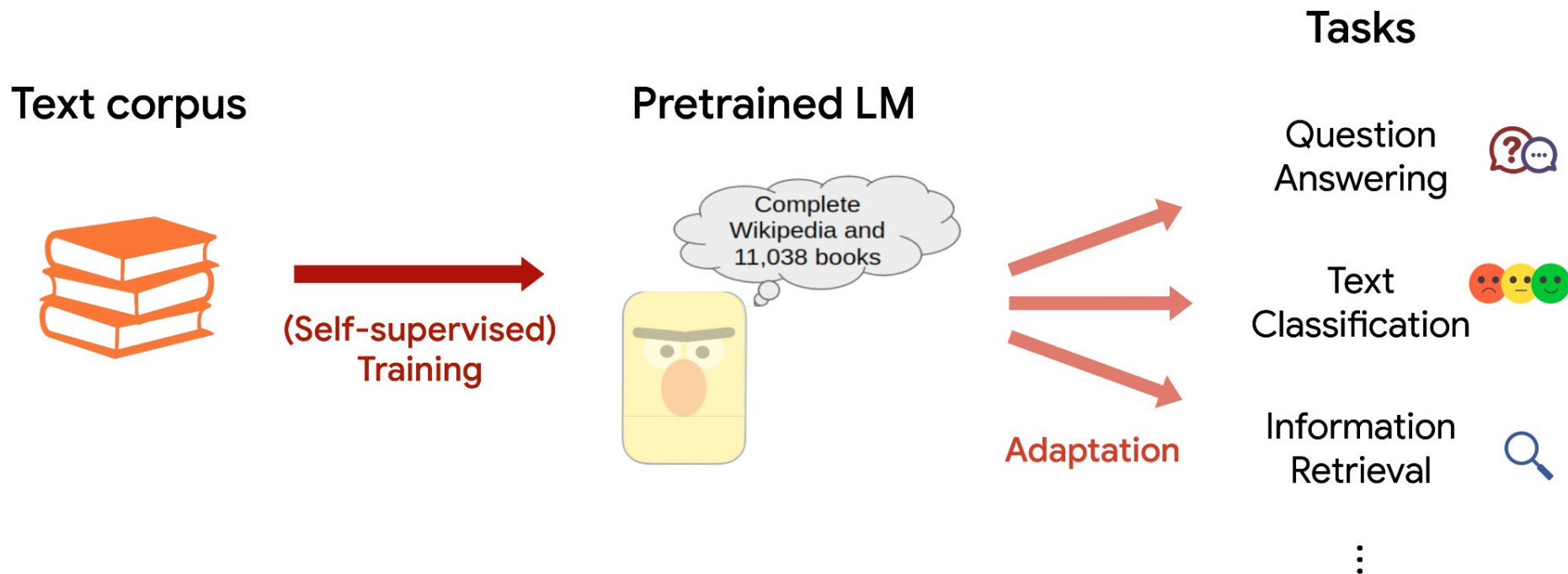


- PhD Student at **Cardiff University**: Oct 2020~
  - Language Model Understanding: [Analogy](#), [Relation Embedding](#)
  - Question & Answer Generation: [AutoQG](#), [LMOG](#)
  - NLP on Social Media: [TweetNLP](#), [TweetTopicClassification](#), [TweetNER](#)
- Research Interns
  - Product Search Team at **Amazon**: June 2021 ~ Oct 2021
  - Computational Social Science Team at **Snapchat**: Oct 2021 ~ Jan 2022
  - Music Generation Team at **Google**: June 2023 ~ Oct 2023



[@asahiushio](#)

# Language models (LMs)



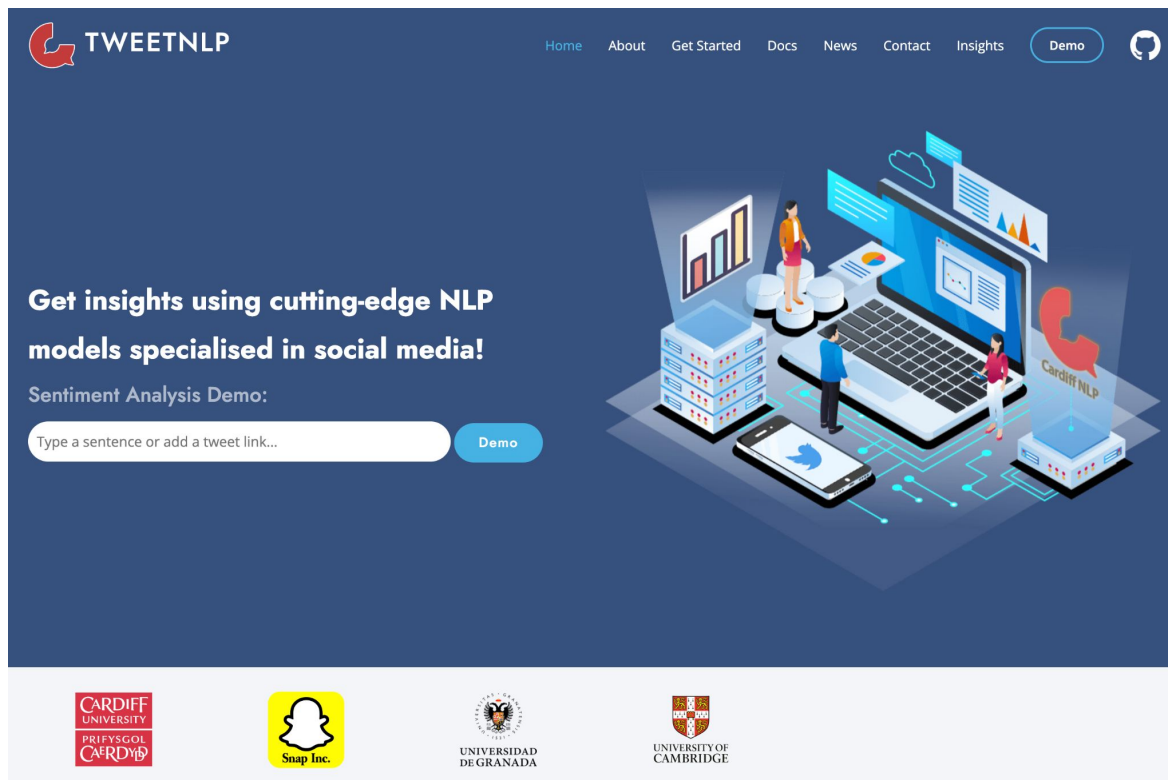
# Specializing LMs on Social Media

# Specializing LMs on social media

## ➤ Why?

- **Informal** grammar
- **Multilingual** (code-switching, etc.)
- Irregular **vocabulary**
  - Emoji 🤔, abbreviations, typos, hashtags, mentions...
- Tweets are often **not standalone messages**
  - RTs, mentions, replies, threads, pictures...
- **Temporality**
  - Emerging entities/meaning

# TweetNLP ([tweetnlp.org](https://tweetnlp.org))



The screenshot shows the homepage of the TweetNLP website. The header features the 'TWEETNLP' logo on the left and a navigation menu with links for Home, About, Get Started, Docs, News, Contact, Insights, a Demo button, and a user profile icon on the right. The main content area has a dark blue background. On the left, it says 'Get insights using cutting-edge NLP models specialised in social media!' followed by 'Sentiment Analysis Demo:' and a text input field with the placeholder 'Type a sentence or add a tweet link...'. A blue 'Demo' button is next to the input field. On the right, there is a large, colorful 3D isometric illustration depicting a person at a laptop, a smartphone with the Twitter logo, server racks, and various data visualizations like bar charts and line graphs. A red speech bubble with 'Cardiff NLP' is also part of the illustration. The footer contains logos for Cardiff University, Snapchat, Universidad de Granada, and the University of Cambridge.

**TWEETNLP**

Home About Get Started Docs News Contact Insights Demo

**Get insights using cutting-edge NLP models specialised in social media!**

Sentiment Analysis Demo:

Type a sentence or add a tweet link... Demo

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

SNAPCHAT Snap Inc.

UNIVERSIDAD DE GRANADA

UNIVERSITY OF CAMBRIDGE

# TweetNLP ([tweetnlp.org](https://tweetnlp.org))

A platform for **NLP specialised on social media**.

Integration of all resources with relatively **small models**.


NLP **applications** from sentiment analysis to hate speech detection and NER.

**Demo, tutorials and Python API.**



# Sentiment Analysis


Type a sentence or a tweet to get insights (tweet URLs are also accepted)

Predictions are based on an English or a multilingual model. Languages supported are: 

Today is a lovely day! 😊

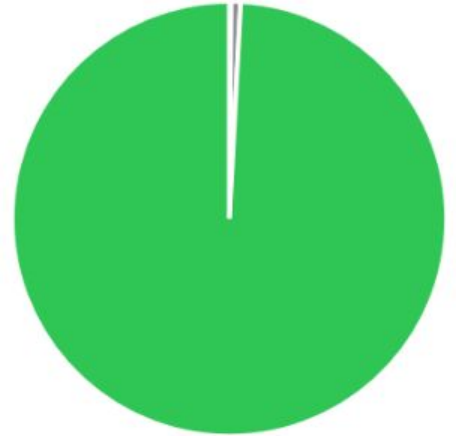
For Example: [Today is a lovely day!](#), [I really don't like eating vegetables.](#)  
[https://twitter.com/Cardiff\\_NLP/status/1485518987807137792](https://twitter.com/Cardiff_NLP/status/1485518987807137792)

Sentiment 

English 

GO!

 negative  neutral  
 positive





# Word Prediction

## 2020 model

Looking forward to watching <mask> Game tonight!

For Example: I keep forgetting to bring a <mask>., COVID is a <mask>., So glad I'm <mask> vaccinated, Looking forward to watching <mask> Game tonight!

2020 - English

GO!

the (57%)



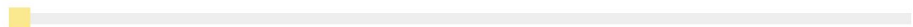
The (25%)



End (2%)



this (2%)



This (0%)



## 2021 model

Looking forward to watching <mask> Game tonight!

For Example: I keep forgetting to bring a <mask>., COVID is a <mask>., So glad I'm <mask> vaccinated, Looking forward to watching <mask> Game tonight!

2021 - English



GO!

Squid (34%)



the (23%)



The (15%)



End (2%)



this (1%)



# Topic Classification

Type a sentence or a tweet to get insights (tweet URLs are also accepted)

Predictions are based on English (all tasks) or a multilingual model (sentiment). Languages supported are:

<https://twitter.com/livescore/status/1632652988228710402>

For Example: [Today is a lovely day!](#), [I really don't like eating vegetables.](#)  
[https://twitter.com/Cardiff\\_NLP/status/1485518987807137792](https://twitter.com/Cardiff_NLP/status/1485518987807137792)

Topic classification

English

Note: Tweets get classified into one or more of 19 topics.

GO!

 Tweet



**LiveScore** @livescore

Liverpool have outscored Manchester United 18-1 in the last eight matches at Anfield 🏆🔥 <https://t.co/KQC7Bgvei7>

sports (99%)

gaming (2%)

news & social concern (1%)

celebrity & pop culture (1%)

diaries & daily life (1%)

# Named Entity Recognition (NER)

Type a sentence or tweet link to get named entities.

<https://twitter.com/BBCWorld/status/1532399905217597440>

GO!

For Example: My name is Wolfgang and I live in Berlin , Paradise is a song by Coldplay , <https://twitter.com/BBCWorld/status/1532399905217597440>

Why **Johnny Depp** person lost in the **UK** location but  
won in the **US** location <https://t.co/X5xheiDw2C>

 Tweet

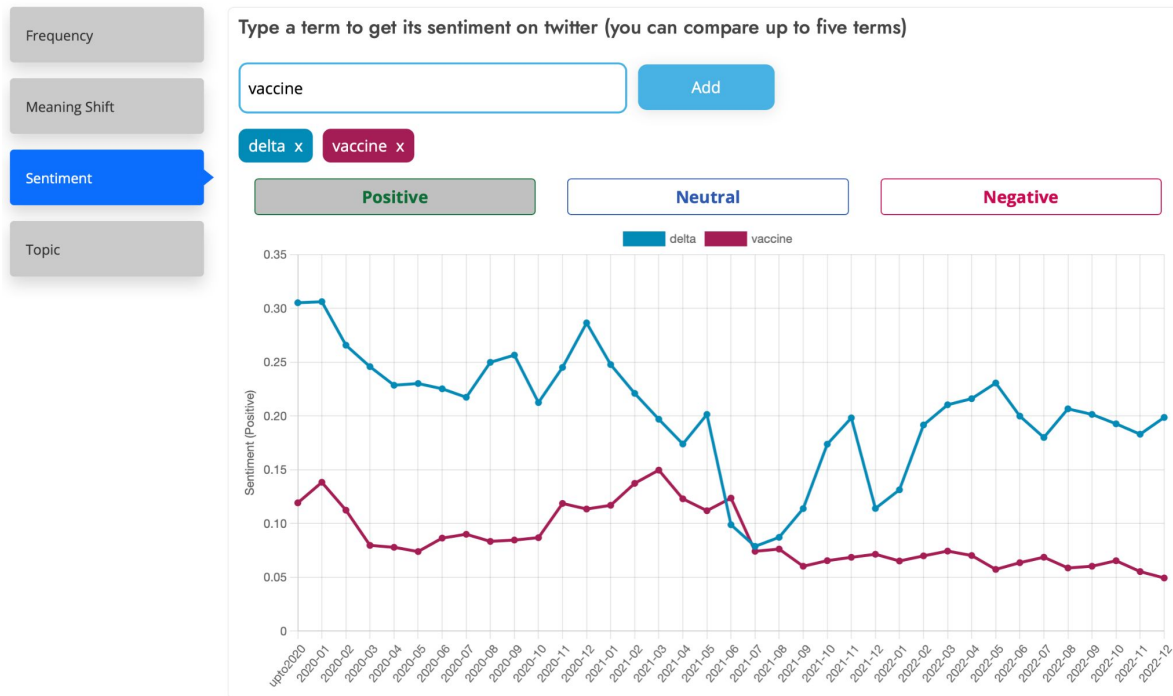


**BBC News (World)** @BBCWorld

Why Johnny Depp lost in the UK but won in the US <https://t.co/X5xheiDw2C>

# ***NEW!*** Tweet Insights

## [tweetnlp.org/insights](https://tweetnlp.org/insights)



# TweetNLP Python library

Includes pre-trained models, inference, fine-tuning, evaluation... [>>DEMO<<](#)

```
import tweetnlp

# ENGLISH MODEL
model = tweetnlp.load_model('sentiment') # Or `model = tweetnlp.Sentiment()`
model.sentiment("Yes, including Medicare and social security saving👍") # Or
>>> {'label': 'positive'}
```



<https://github.com/cardiffnlp/tweetnlp>



TWEETNLP