Leveraging NLP
to achieve environmental sustainability

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

In collaboration with
the Joint Research Centre
(JRC) of the European Union

European Commission

AI for people

Francesco Cariaggi, @fcariaggi
Cristiano De Nobili, PhD, @denocris
Sébastien Bratières, @Seb_Bratieres

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

PI CAMPUS

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Pi Campus invests in **applied AI startups**.

52 investments in 5 years

The deal: € 50-500K for 1-10% share
It is a **seed stage venture fund**
and a **startup district**
50% Italy
25% Europe
25% California

European Commission

AI for people

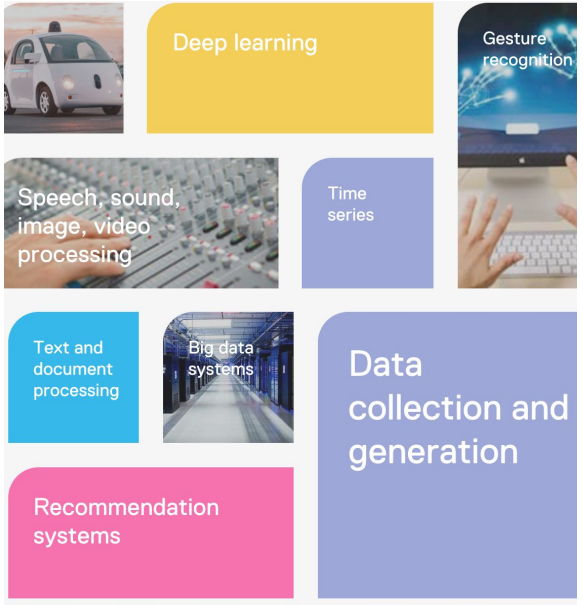
Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

School of Artificial Intelligence

Several times a year, we host a batch of the **best engineers from all over the world to turn them into AI specialists**.

They apply their new skills on the **industry project** provided either by their own employer, or by world leading tech companies such as **Google, Facebook and Amazon** and fast-growing startups.



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

School of Artificial Intelligence

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

- **Merit First**
  Top developers get in **for free**, and those who transfer from abroad will receive a travel and accommodation grant.

- **Learn by doing**
  Minimal teaching. Desks and environment are organised to support small project teams, agile co-development, **interactions with mentor**.

- **Real world projects, no simulations**
  Our partners sponsor top developers to solve **real challenges**.

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Qualified advisors for your **project**

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Managing Director

Sébastien Bratières
Director of AI, Translated
UNIVERSITY OF CAMBRIDGE

Faculty Advisors

Alex Waibel
Professor of Computer Science
Carnegie Mellon University

Hassan Sawaf
Director of AI, Amazon Web Services
amazon

Marcello Federico
Head, Human Language Translation Unit

European Commission

Some mentors from our pool

Lukasz Kaiser
Researcher, Google Brain
Google

Riccardo Sabatini
Chief Data Scientist
ORIONIS BIOSCIENCES

Max Pumperla
Deep learning Engineer
skymind

Novi Quadrianto
Senior Lecturer
University of Sussex

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Challenge portfolio

- **Amazon:** The Watercolour World
- Amazon: speech science with MXNet
- **Translated**: spoken language identification
- Translated: content-based translator scoring
- **Wanderio**: e-commerce fraud detection
- **Lamco**: mapping with AI
- **Vatican Secret Archives**: Latin OCR
- **Kingcom**: Food influencers
- **Defenx**: ransomware preemptive detection
- PwC: visual document classification
- **PwC**: interpretable machine learning
- **Amex**: loyalty programme email campaigns
- **MiBACT**, Italian ministry for heritage: searching Italy's art
- **Atomikad**: in-image advertising
- **Veneto region**: understanding hospital files
- **Soldo**: info extraction from receipts
- **Covisian**: call centre performance
- **Xriba**: tax codes on invoices
- **Translated**: customer lifetime value prediction
- Translated: adword optimisation
- **Cisco**: the ML platform for networking
- **BNL**: Basel II operational risk prediction
- BNL: ID scanning
- **Poste**: visual walk-in customer profiling

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Challenge portfolio

- **Cisco**: analysing hierarchical network data
- Cisco: reinforcement learning for wifi channel selection
- Cisco: privacy challenges in ML
- Cisco: combating Twitterbots
- Cisco: reconstructing depth from 2D images
- **Engie**: electricity imbalance prediction
- **Cloudcare**: lightspeed chat suggestions
- **Employerland**: AI for job coaching
- **Sorgenia**: customer care chatbot
- **Inreach**: sourcing investment opportunities
- **Enel**: AWS cloud optimization
- **Enel X Colombia**: propensity modelling
- Enel: electricity load forecast for homes
- **Enel Distribution**: medium voltage line SCADA fault prediction
- Enel Distribution: extreme weather impact on power lines
- Enel: IT helpdesk ticket routing
- **Consiglio Nazionale del Notariato**: AI for tomorrow's notary public
- Cloudcare: virtual call center supervisor
- **Global and Local**: improve rural municipalities' access to funding
- **FASI**: personalised tender alerts
- **European Space Agency**: earth observation
- **Pryiatech**: heart beat detection on video
- **Freeda Media**: lipstick recommender
- **Radio Dimensione Suono**: news picker
- **Octo**: fuel tank monitor

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

School of Artificial Intelligence

Next session: **29 November 2021**

Apply to the School of AI 2022

Submit your application on Pi School's website:

https://pischool.link/CAIP21

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Leveraging NLP
to achieve environmental sustainability

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

In collaboration with
the Joint Research Centre
(JRC) of the European Union

European Commission

AI for people

Francesco Cariaggi, @fcariaggi
Cristiano De Nobili, PhD, @denocris
Sébastien Bratières, @Seb_Bratieres

Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

The **larger issue**

- EU wants to achieve **net-zero pollution** by 2050, but at the same time stay **globally competitive**
- A fundamental question arises: Do the current environmental capabilities of the EU **match** the need for clean technologies?
- Mapping environmental capabilities of EU countries becomes central
- Recent studies assessed the **environmental capabilities** of countries
  - relying on international classifications of environmentally relevant goods/activities as indicators
  - **but** they were developed with trade and customs regulation in mind, **not** with the actual environmental impact of such goods

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

The **larger issue**

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

- **Better indicators** of environmentally relevant goods/activities must be found
- **Idea:** use R&D activities as indicators for assessing the environmental capabilities of countries
- BREFs are documents that provide a clear picture of SOTA tech analysis of the **best available techniques** (consolidated and emergent) in the field of industrial pollution control

European Commission

AI for people

Our solution: **Geographically map capabilities, represented by patents (R&D), with BREF as queries.**

Apply to the School of AI 2022 https://pischool.link/CAIP21

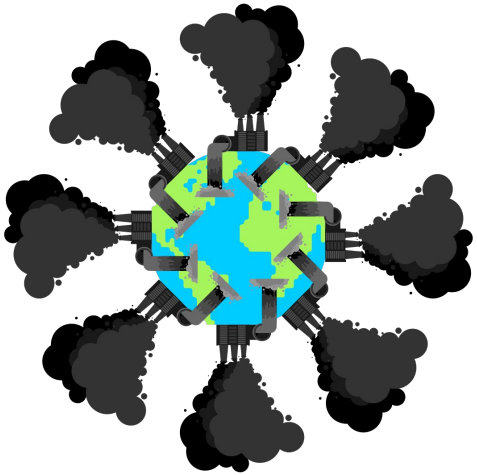
Follow @picampusschool

What do **NLP** and **Env. Sust. have in common**?

We then built an **Information Retrieval (IR) system** based on **Transformers** that can retrieve R&D relevant patents…

Ok, but what kind of patents?

At this stage of the project,
JRC was interested in **Industrial Pollution** (Patents4IPPC).



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

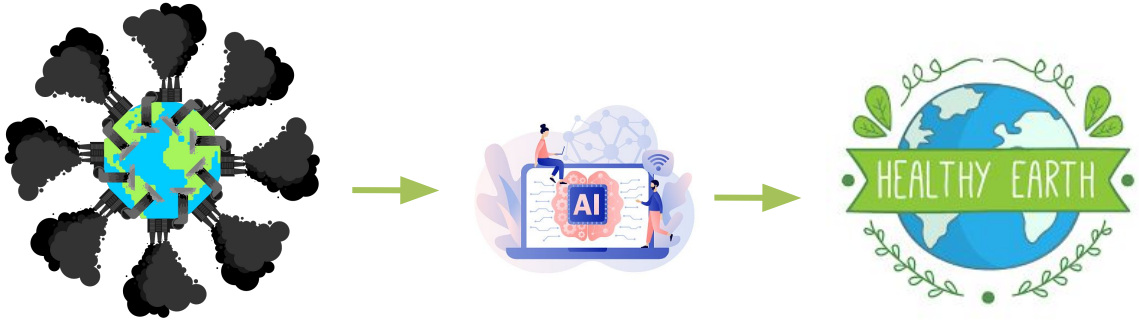
Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

AI for Sustainability

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

As AI scientists or engineers, we can make a difference to the world.
With this project at Pi School we are doing our bit.



*“If you have to stand in front of a computer for hours,
make sure that there is a strong mission behind the screen.”*

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Scope of the **project**

Given a BREF* passage find out the most relevant patents (technology)

BREF passage

Applicability
The method is applicable at all blast furnaces equipped with pulverised coal injection and oxygen enrichment. Direct injection of reducing agents is applicable both at new and existing blast furnaces.

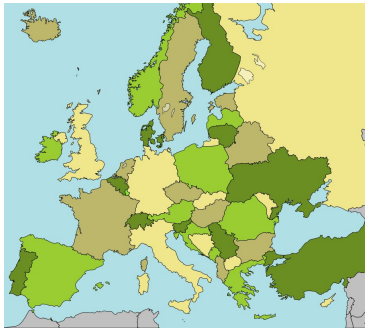
Economics
There are economic incentives for using high rates of coal injection to achieve greater cost savings, particularly at plants which might otherwise face capital expenditure on rebuilding coke ovens or may have to purchase coke. Furthermore, coal injection can allow the use of coals of a lower quality compared to coking coals. This may also reduce costs.

Costs at EUR 10/GJ in 1988 are reported [65, InfoMil 1997]. However, [11, Campell et al. 1992] indicate that costs are saved, due to the lower coke consumption. The capital cost of installing coal injection at Corus, UK, Port Talbot No 4 Blast Furnace in 1997 was approximately EUR 24 million (including some redundant parts from another site). For the examples, the currency was converted into ECU in 1996 or 1997 and for the review into EUR.

Additional costs will arise for air enrichment providing consistently large amounts of oxygen, the additional demand on the pulverisers in existing plants and additional requirements for the injection unit maintenance.

Most relevant patents

- **Patent 1**
- **Patent 2**
- …



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Technical Challenges

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

The project might seem a simple Text Similarity task, but this is not the case:

- From Contextualized Word Embedding to Sentence Embedding for Semantic Textual Similarity (STS);
- Linguistic style mismatch between query (BREF) and response (patent);
- No training labeled data available (GS1* as a test set);
- Huge response database (about 10-20 M patents).

European Commission

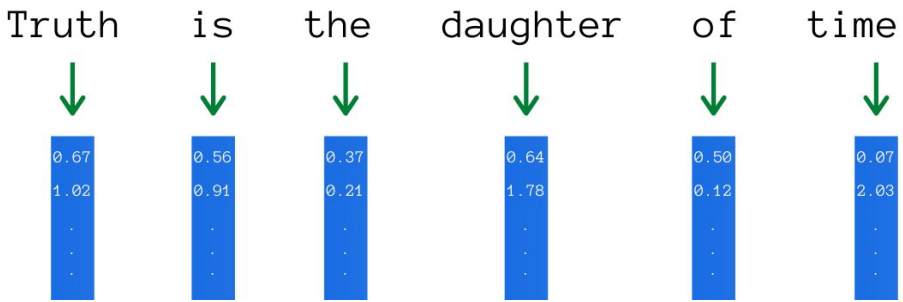
AI for people

*GS1 (Gold Standard 1) is a dataset composed by a few pairs of BREF passage and corresponding relevant patent.

Basics of NLP:
non-contextual embeddings

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Step one: you first assign a vector to each word of a vocabulary



European Commission

AI for people

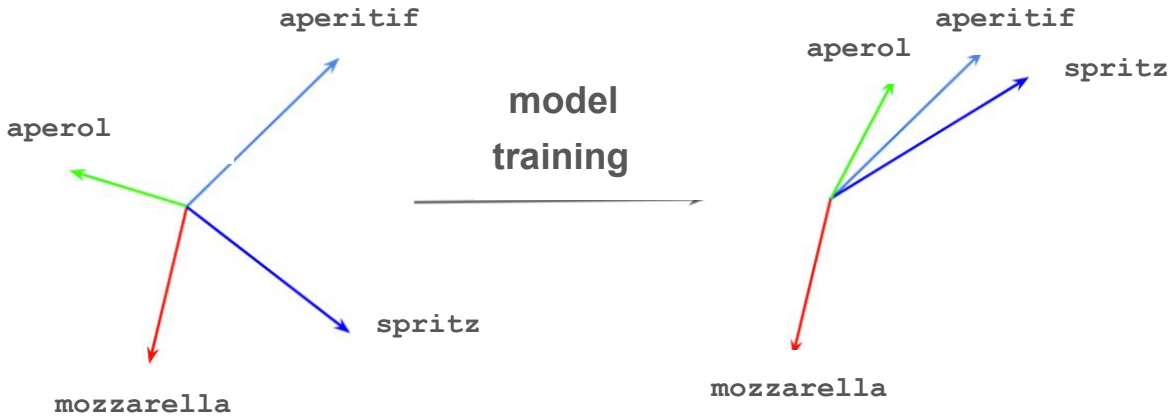
Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Basics of NLP:
non-contextual embeddings

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Step two: from random vectors' assignments to a trained language model (LM)



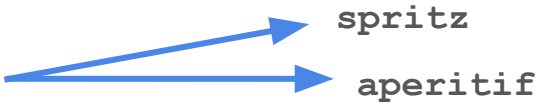
European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Basics of NLP:
contextual embeddings



different words but close meaning

But sometimes a word can have different meaning depending on the context

climate **change**, career **change**

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

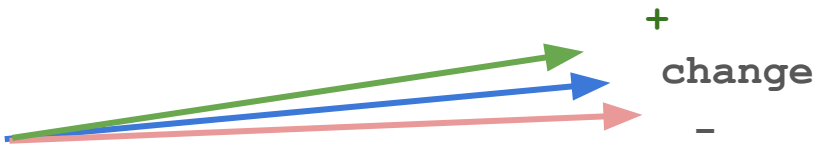
AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Basics of NLP:
contextual embeddings

That's why people started thinking about contextual language models



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

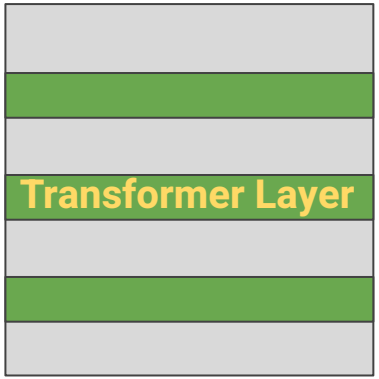
Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Basics of NLP:
Transformers & BERT

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

The SOTA in contextual NLP are models based on Transformers, such as BERT



BERT

Self-Attention Mechanism

self-attention mechanism helps contextual understanding

European Commission

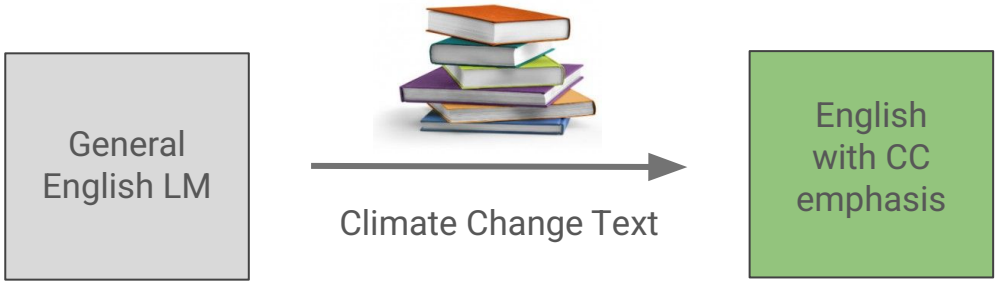
AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Transfer Learning

If you need a LM that need to know about climate change



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

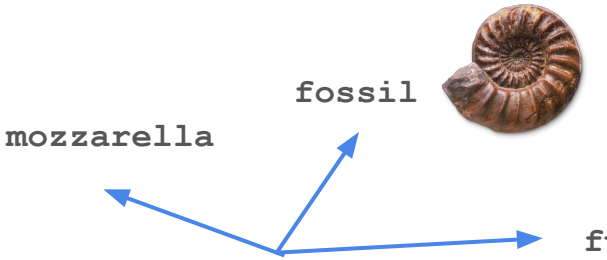
European Commission

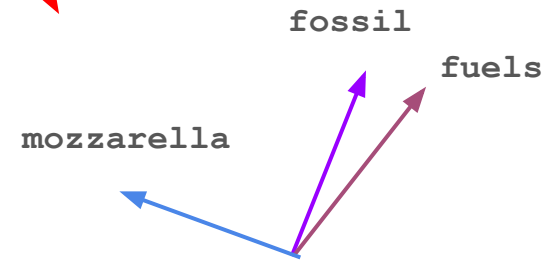
AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

@picampusschool

Transfer Learning





PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

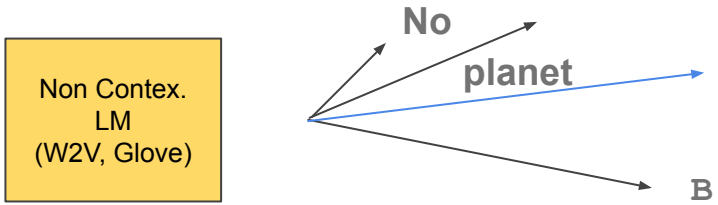
AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

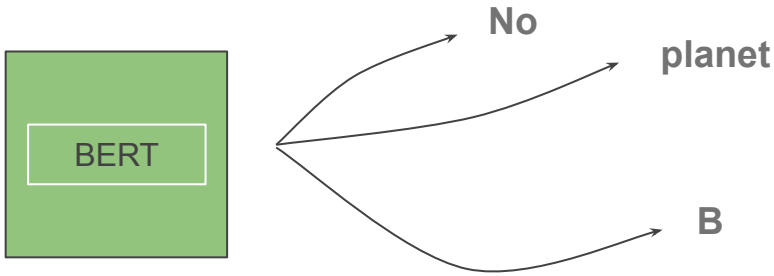
@picampusschool

Sentence embedding

From Contextualized Word Embedding to Sentence Embedding for STS



“No planet B”

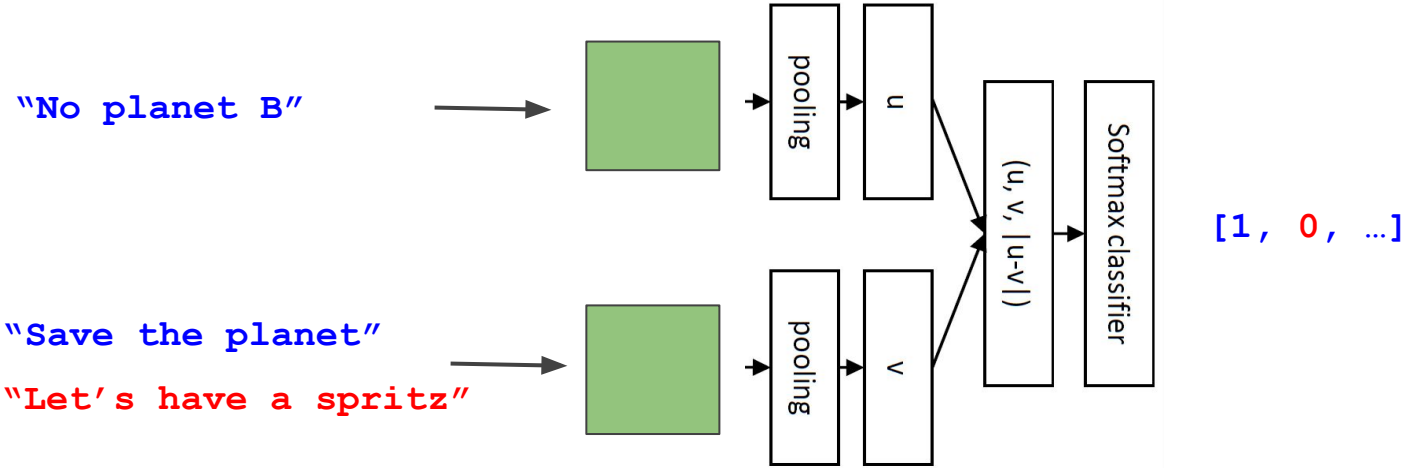


BERT embeddings do not live in a Euclidean space but something more similar to a hyperbolic space. Here we cannot sum vectors or use cosine similarity to measure their distance. **BERT is not a good sentence embedder.**

BERT Geometry: https://arxiv.org/abs/1906.02715

Sentence BERT

A siamese network, when trained in a supervised way, is able to generate meaningful sentence embeddings



SentBERT: http://arxiv.org/abs/1908.10084

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

BERT for Patents

A “well-read” language model in patents’ peculiar language[1]

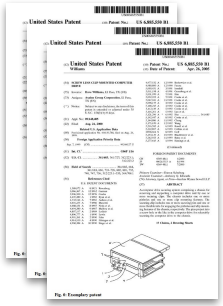
European Commission

AI for people

[1] https://services.google.com/fh/files/blogs/bert_for_patents_white_paper.pdf

BERT for Patents

BERT
(Language model)



100M+ patents

BERT for Patents

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

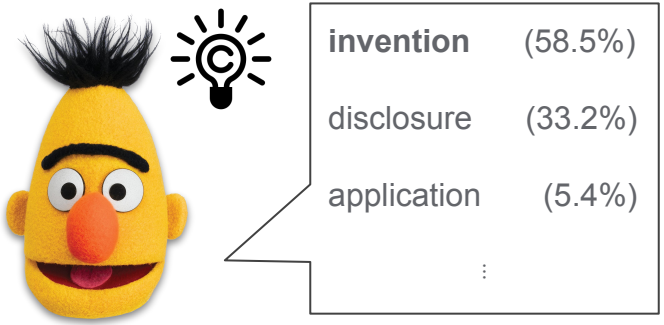
AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

BERT for Patents - Training procedure

The present provides a torque sensor [...]



PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

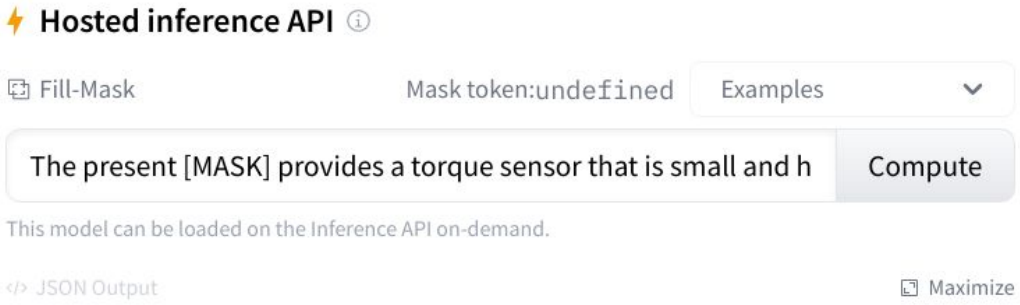
BERT for Patents - Demo

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

1. Go to huggingface.co/anferico/bert-for-patents
2. Enter some text in the textbox on the right
3. Replace the word you want to mask with “[MASK]”
4. Click “Compute”

European Commission

AI for people



Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

BERT for Patents - Demo

Examples:

- The present invention relates to **[MASK]** accessories and pertains particularly to a brake light unit for bicycles.
  (Masked word: **bicycle**)
- The present invention discloses a space-bound-free **[MASK]** and its coordinate determining circuit for determining a coordinate of a stylus pen.
  (Masked word: **tablet**)
- The illuminated **[MASK]** includes a substantially translucent canopy supported by a plurality of ribs pivotally swingable towards and away from a shaft.
  (Masked word: **umbrella**)

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Contributions

What did we bring that was not already there?

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Contributions

- Gathered several third-party datasets
- Analyzed multiple evaluation metrics
  - Spearman rank correlation, Normalized Discounted Cumulative Gain (NDCG)
- Addressed the linguistic style mismatch problem using fine-tuning ideas
- Enriched the Gold Standard dataset (GS1) by submitting our model's predictions to human annotators

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Results

Model	Spearman STSb	Spearman Helmers	Spearman GS1	NDCG Helmers	NDCG GS1
b4p-sts-tcm	0.79	0.66±0.11	0.34±0.25	0.90±0.05	0.73±0.21
specter-sts-tcm	0.78	0.62±0.15	0.30±0.22	0.87±0.07	0.70±0.22
- - -	- - -	- - -	- - -	- - -	- - -
roberta-large-nli-stsb-mean-tokens	0.864	0.53±0.18	0.24±0.18	0.83±0.06	0.68±0.21
TF-IDF	0.52	0.45±0.19	0.21±0.21	0.83±0.10	0.62±0.22
Glove	0.37	0.33±0.20	0.10±0.21	0.80±0.11	0.62±0.21
USE	0.80	0.55±0.17	0.21±0.19	0.83±0.08	0.65±0.20

- Our final models largely outperform baseline approaches
- All metrics are to be intended as higher is better

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Deployment

How is the output of our project being employed?

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Deployment and **open source release**

- JRC is already using our retrieval engine
  - Proud to say that we exceeded their expectations for this project
- Our software has been made publicly available on **GitHub** under the **GNU GPL-3.0** license
  - github.com/ec-jrc/Patents4IPPC

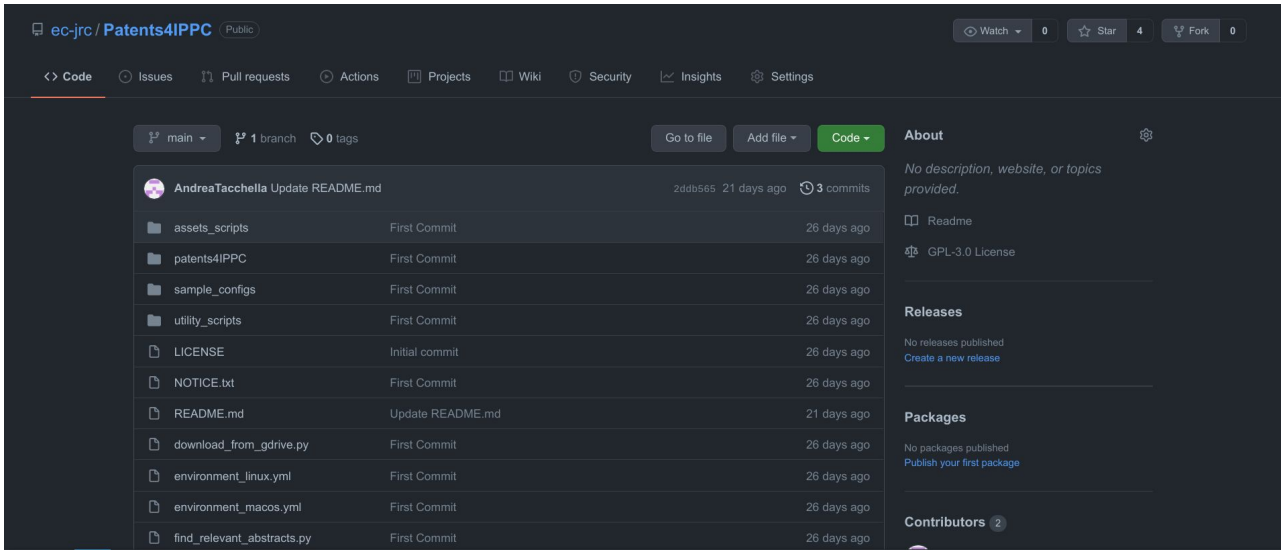
PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Deployment and open source release

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP



European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

The **Team**

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

Francesco Cariaggi
(@FCariaggi)
Deep Learning Engineer

Cristiano De Nobili, PhD
(@denocris)
Lead AI Scientist

Sébastien Bratières
(@Seb_Bratieres)
Managing Director

European Commission

AI for people

Apply to the School of AI 2022 https://pischool.link/CAIP21

Follow @picampusschool

Leveraging NLP
to achieve environmental sustainability

Francesco Cariaggi, @fcariaggi
Cristiano De Nobili, PhD, @denocris
Sébastien Bratières, @Seb_Bratieres

Apply to the School of AI 2022
https://pischool.link/CAIP21

Follow @picampusschool

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people

Pi School

Follow @picampusschool

Apply to the School of AI 2022
https://pischool.link/CAIP21

PI SCHOOL
SCHOOL OF CREATIVE ENTREPRENEURSHIP

European Commission

AI for people