

# A Guided Tour Through the Transformers Landscape

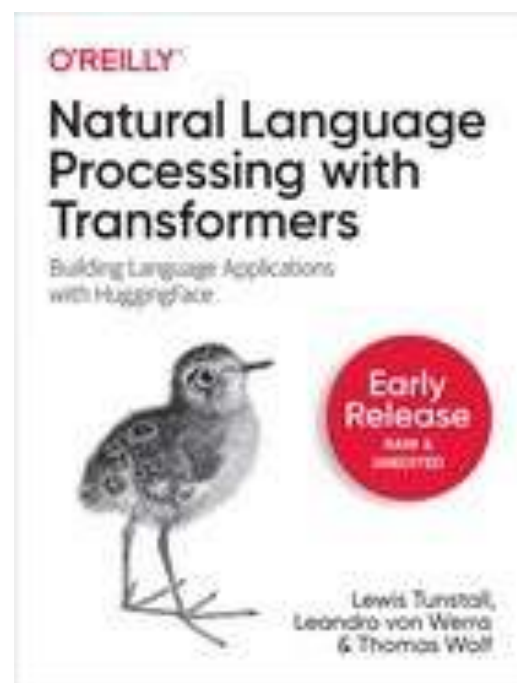
SwissText – 15 June 2021

Lewis Tunstall | Open-source @ Hugging Face | [lewis@huggingface.co](mailto:lewis@huggingface.co)





[huggingface.co/course/](https://huggingface.co/course/)



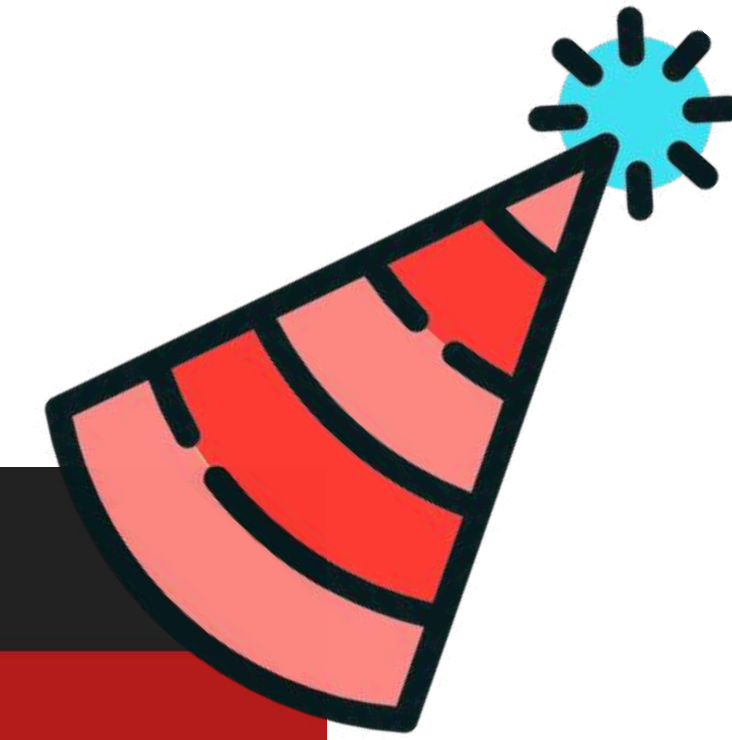
[NLP with Transformers](#)

Education

Open Source




# A brief history



4 years old  
this month



 Cornell University

arXiv.org > cs > arXiv:1706.03762

Computer Science > Computation and Language

*[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]*

**Attention Is All You Need**

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin




# A brief history



4 years old  
this month



 Cornell University

arXiv.org > cs > arXiv:1706.03762

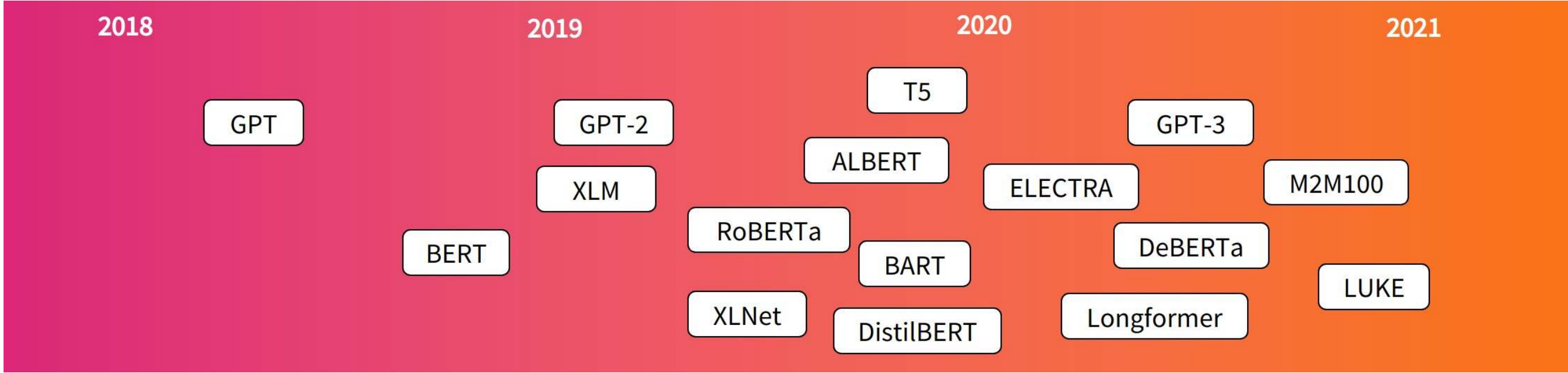
Computer Science > Computation and Language

*[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]*

## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

A “Cambrian  
explosion”





# Main ingredients

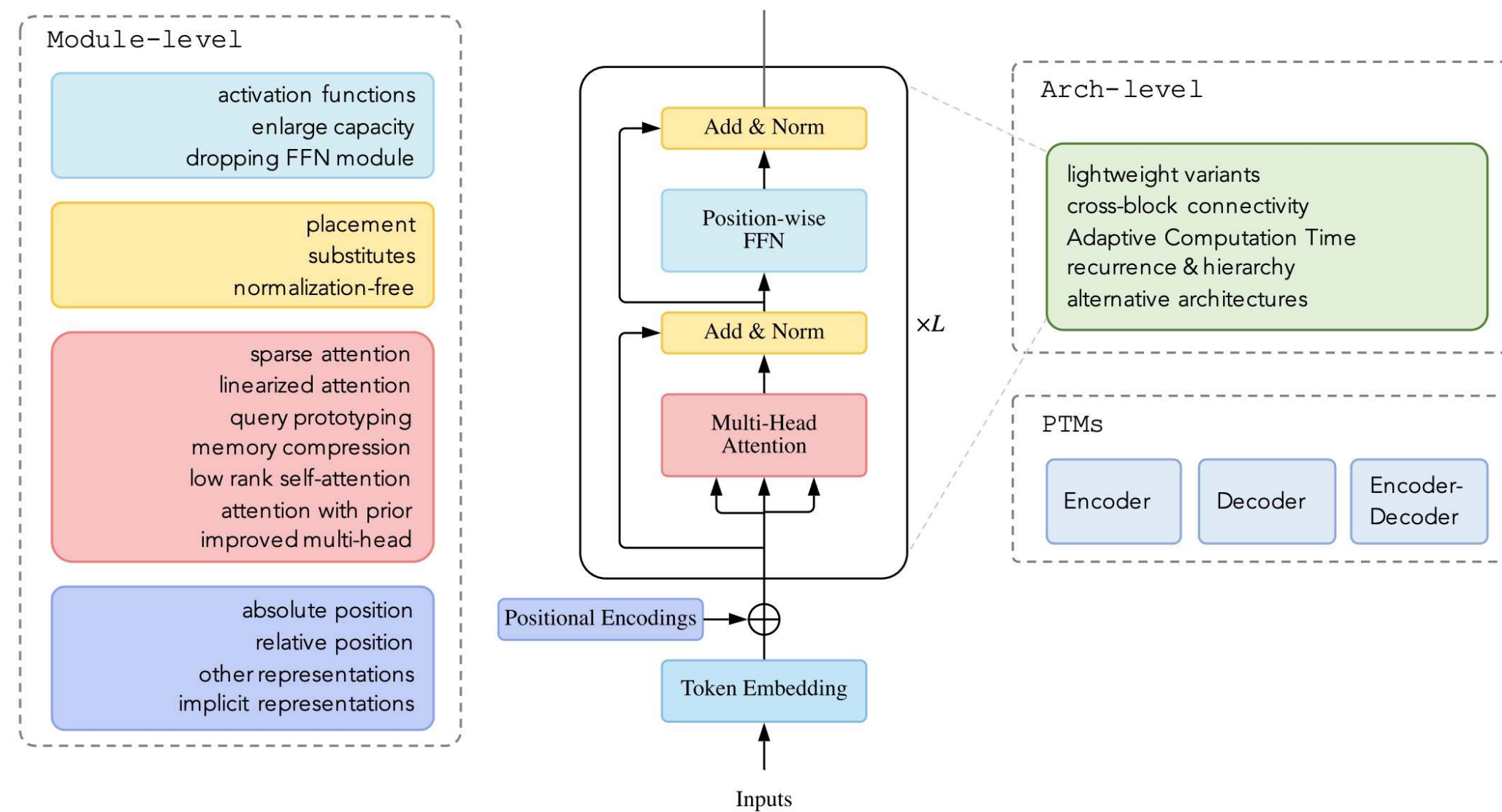


Image credit: [Lin et al \(2021\)](#)

## Transformer Blocks



# Main ingredients

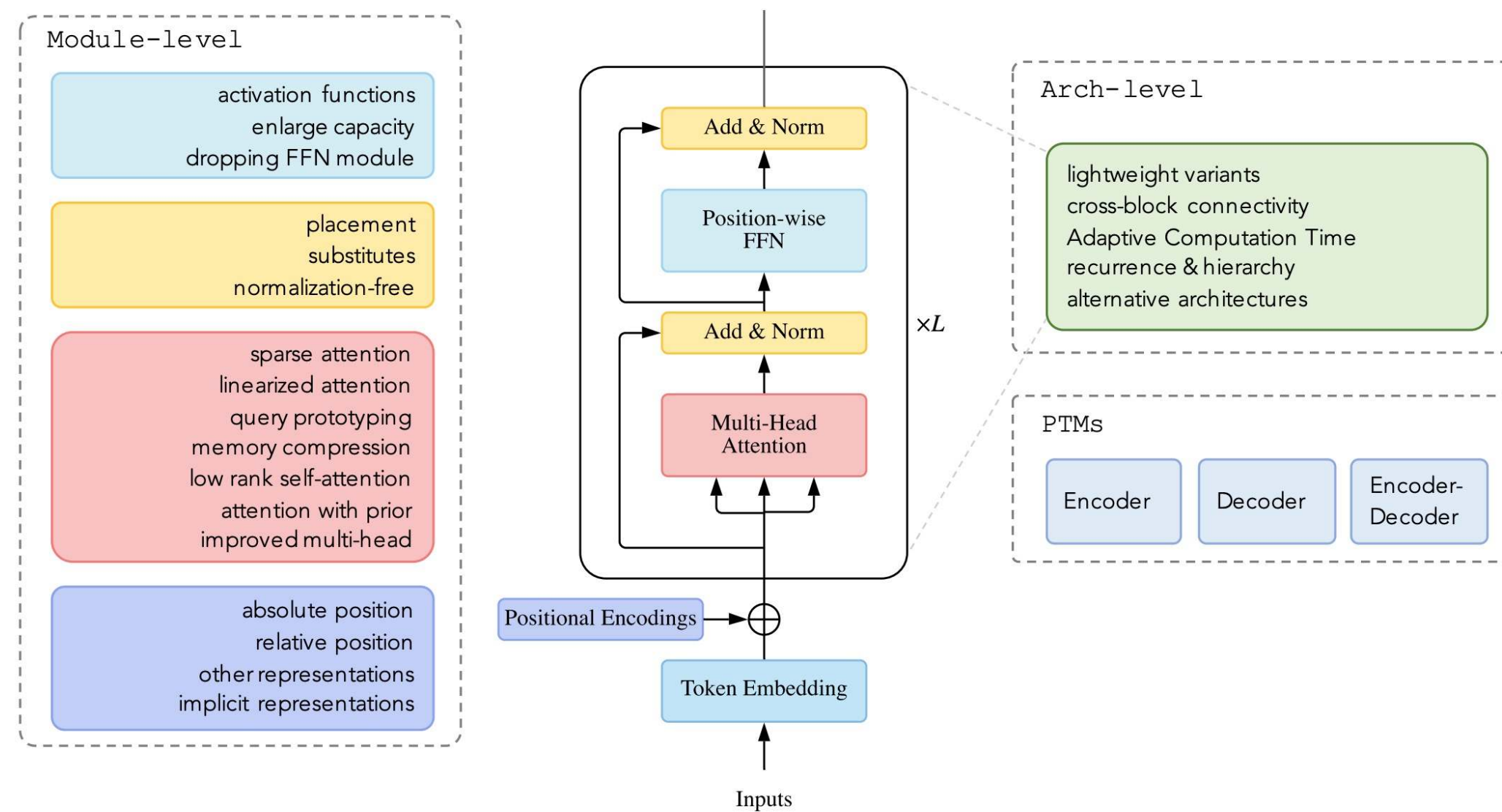
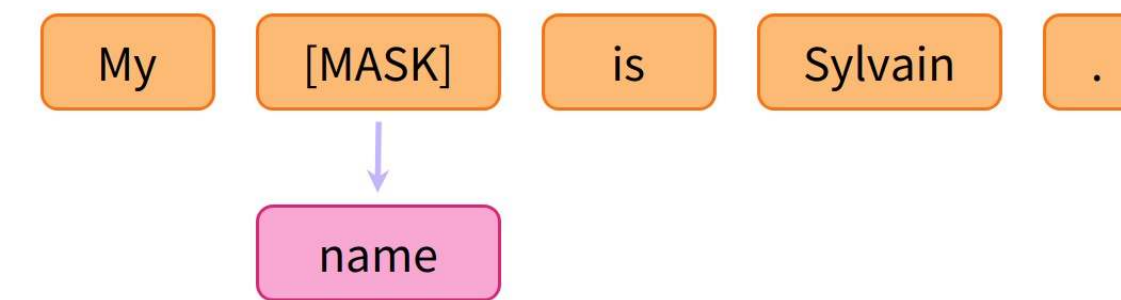
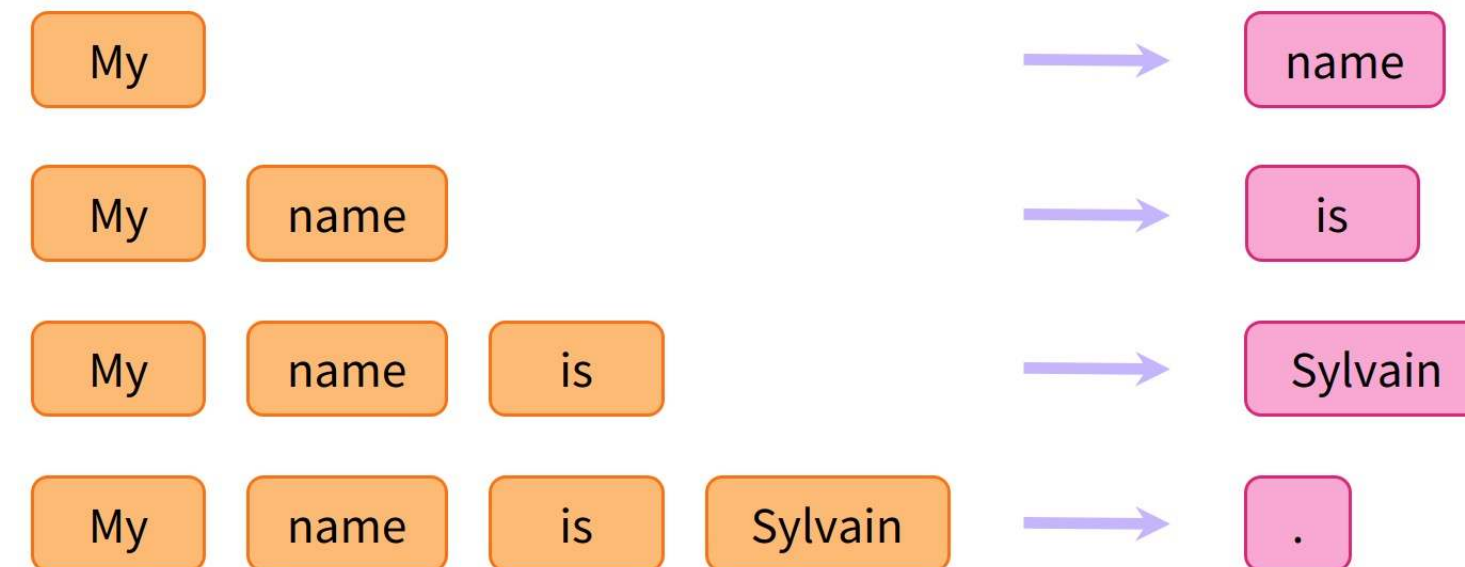


Image credit: [Lin et al \(2021\)](#)

## Transformer Blocks



Masked LM

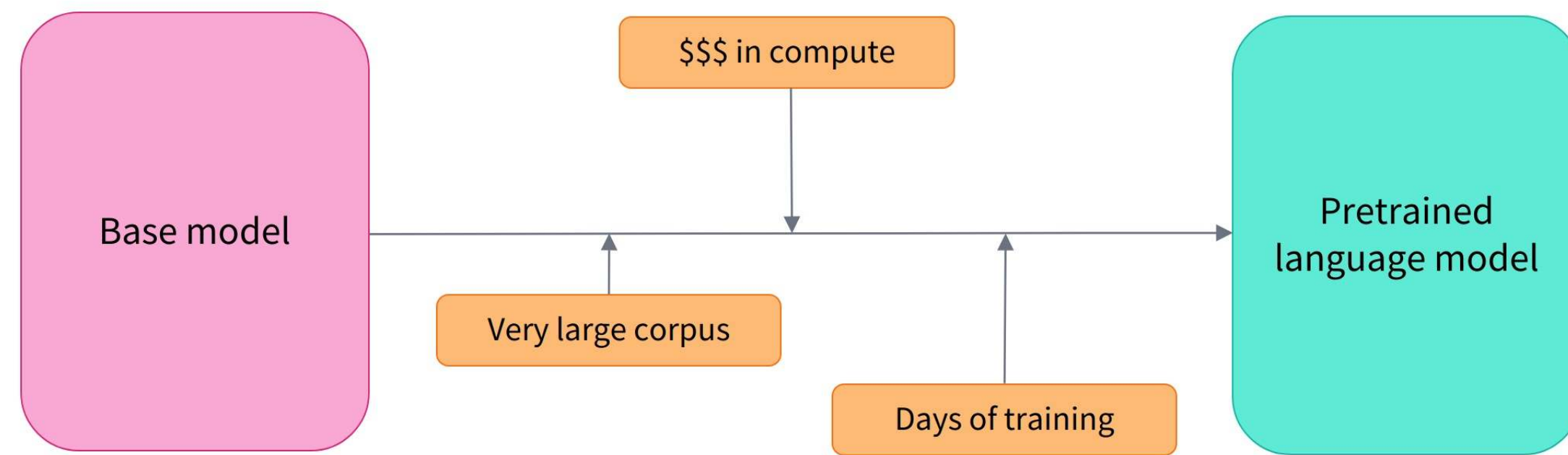


Autoregressive LM

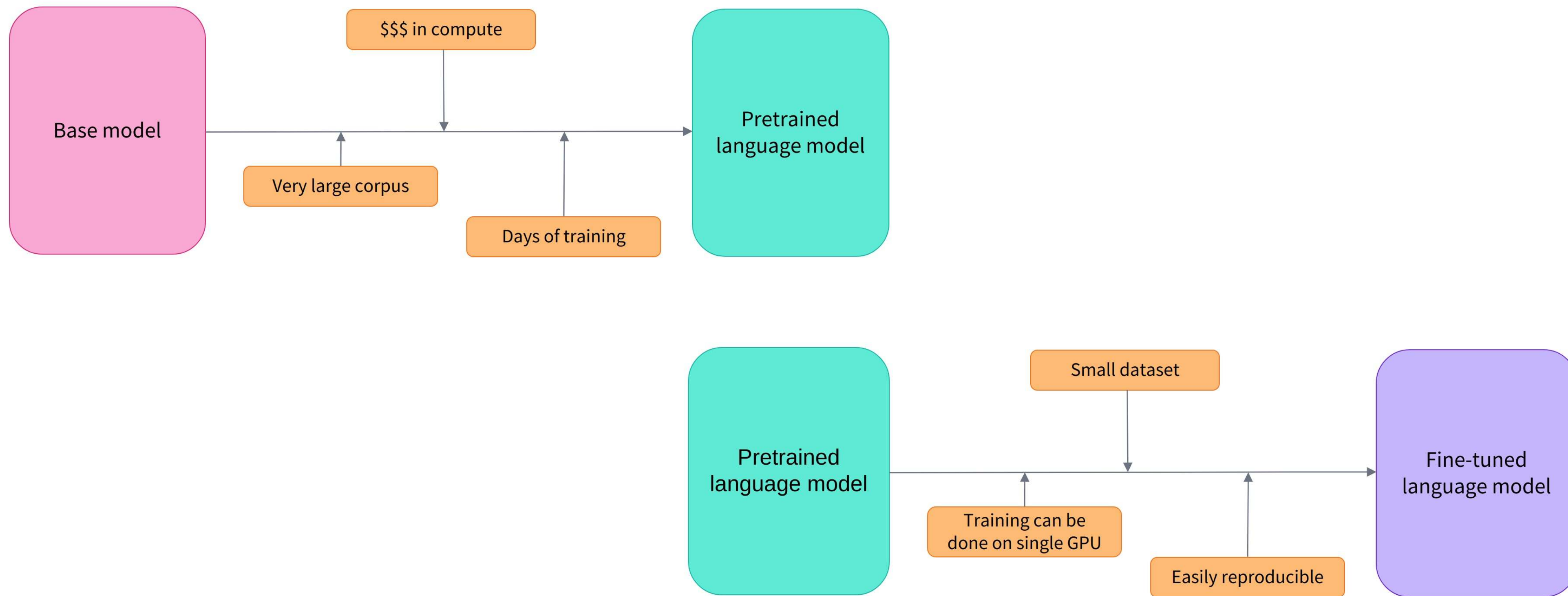
## Language Modelling



# The modern paradigm



# The modern paradigm





# Works for vision too 🤯

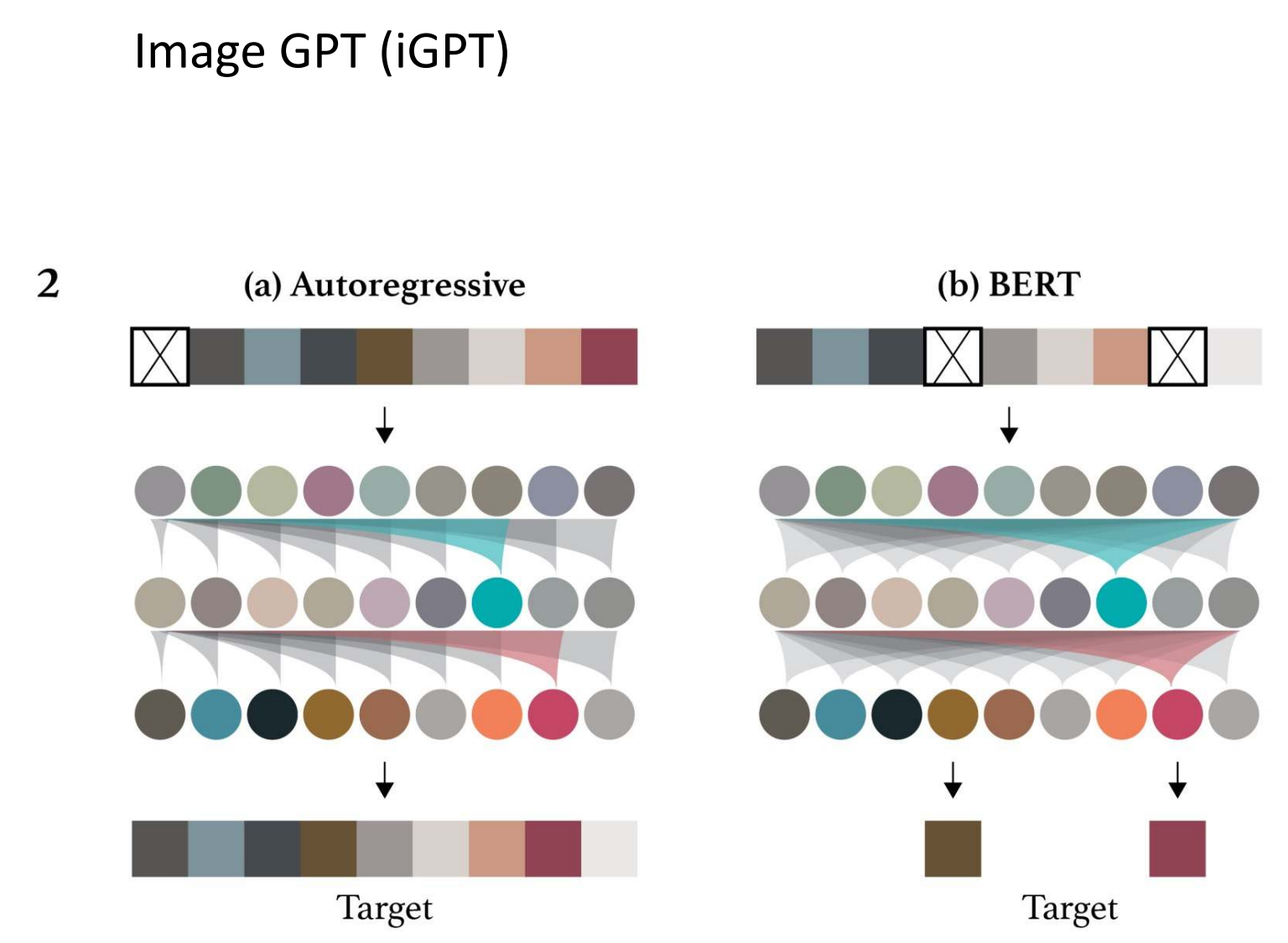
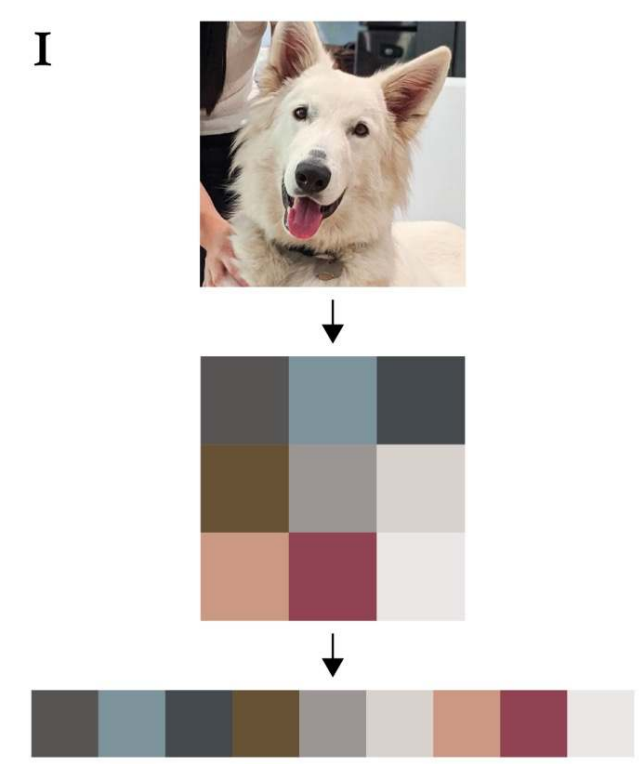


Image credit: [Chen et al \(2020\)](#)

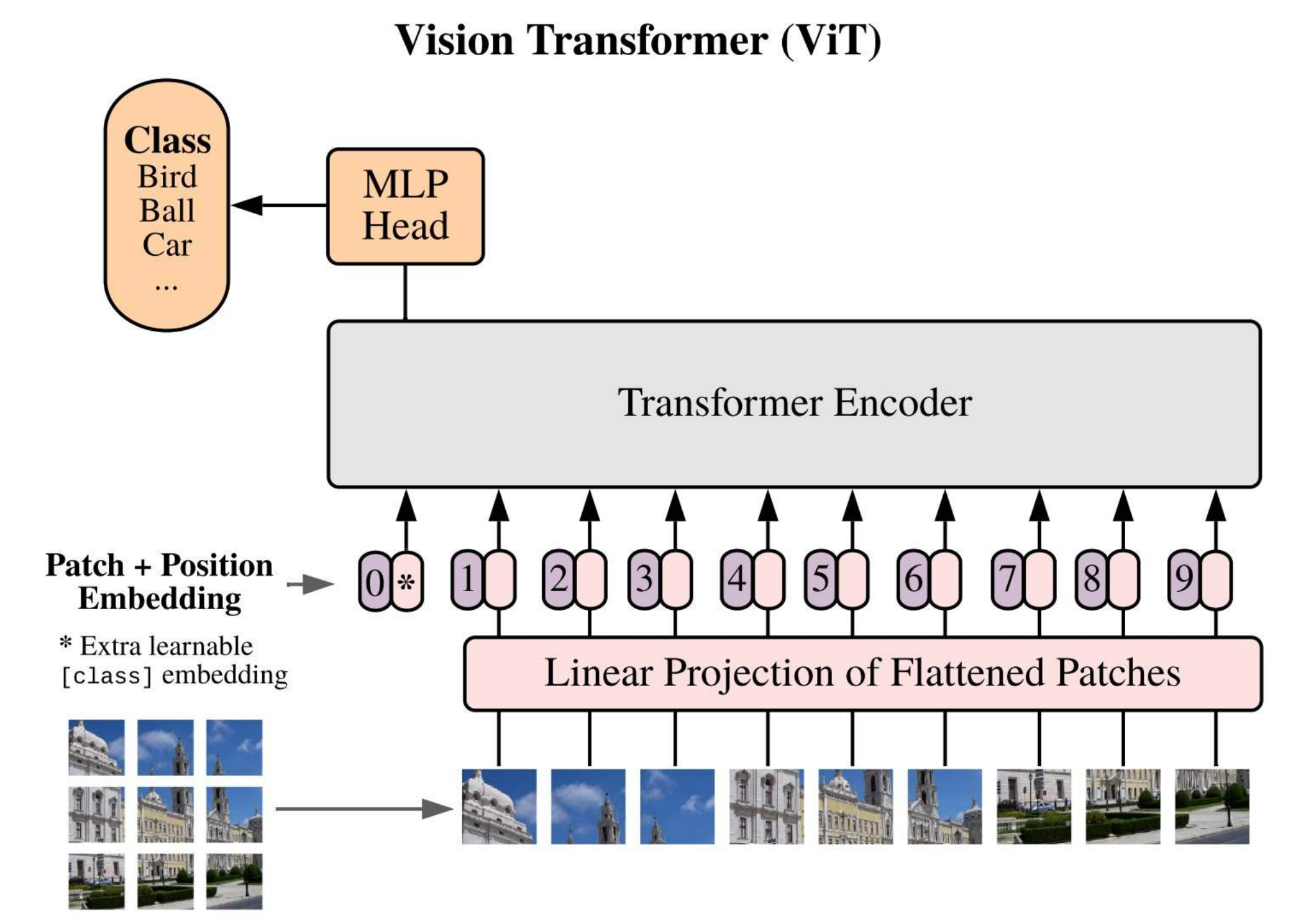


Image credit: [Dosovitskiy et al \(2020\)](#)



# Including multiple modalities

TEXT PROMPT an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES



Edit prompt or view more images ↓

TEXT PROMPT an armchair in the shape of an avocado. . . .


AI-GENERATED IMAGES



Edit prompt or view more images ↓

[DALL-E by OpenAI](#)

A multimodal document as query



Musician walking on the street with sousaphone

Results more rely on

Text 0.5  Image 0.5

Search!

A screenshot of a search interface. At the top, it says "A multimodal document as query". Below that is a large image of a musician walking on a street with a large sousaphone. Underneath the image is a text box containing the caption "Musician walking on the street with sousaphone". Below the text box is a slider control labeled "Results more rely on". The slider has "Text 0.5" on the left and "Image 0.5" on the right, with a blue bar and a cursor in the middle. At the bottom of the interface is a blue button labeled "Search!".

Multimodal document search by [Jina](#)



# Transformers are now everywhere



NLP



Speech



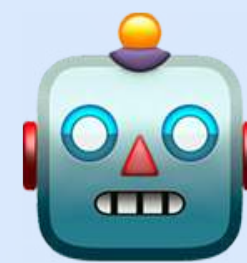
Vision



BioChem



Time Series



RL





# Transformers are now everywhere



NLP



Speech



Vision



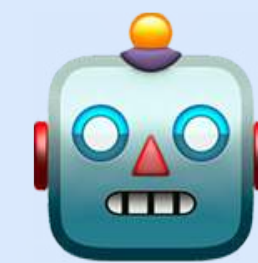
Domain X



BioChem



Time Series



RL





# Bridging the science – industry divide



Humble Data Scientist



# Get the code and model weights?

arXiv.org > cs > arXiv:1706.03762 Search... Help | Adv

**Computer Science > Computation and Language**

*[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]*

## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Comments: 15 pages, 5 figures  
Subjects: **Computation and Language (cs.CL)**; Machine Learning (cs.LG)  
Cite as: arXiv:1706.03762 [cs.CL]  
(or arXiv:1706.03762v5 [cs.CL] for this version)


**Submission history**  
From: Ashish Vaswani [view email]  
[v1] Mon, 12 Jun 2017 17:57:34 UTC (1,102 KB)  
[v2] Mon, 19 Jun 2017 16:49:45 UTC (1,125 KB)  
[v3] Tue, 20 Jun 2017 05:20:02 UTC (1,125 KB)  
[v4] Fri, 30 Jun 2017 17:29:30 UTC (1,124 KB)  
[v5] Wed, 6 Dec 2017 03:30:32 UTC (1,124 KB)

**Bibliographic Tools** **Code & Data** Related Papers About arXivLabs


**Code and Data Associated with this Article**

arXiv Links to Code & Data (What is Links to Code & Data?)

**Official Code**

 <https://github.com/tensorflow/tensor2tensor>

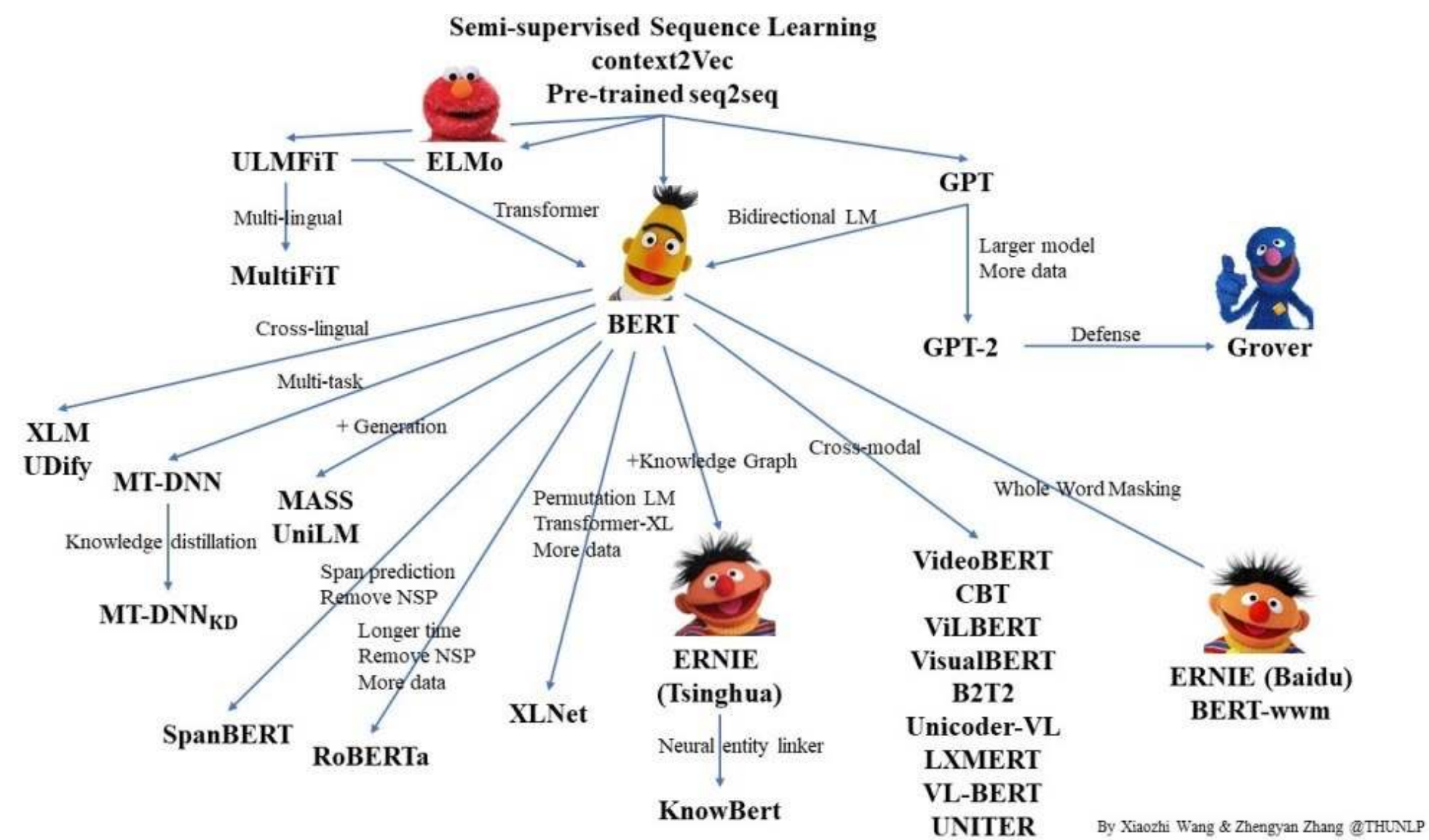
**Community Code**

 455 code implementations (in PyTorch, TensorFlow, MXNet and JAX)





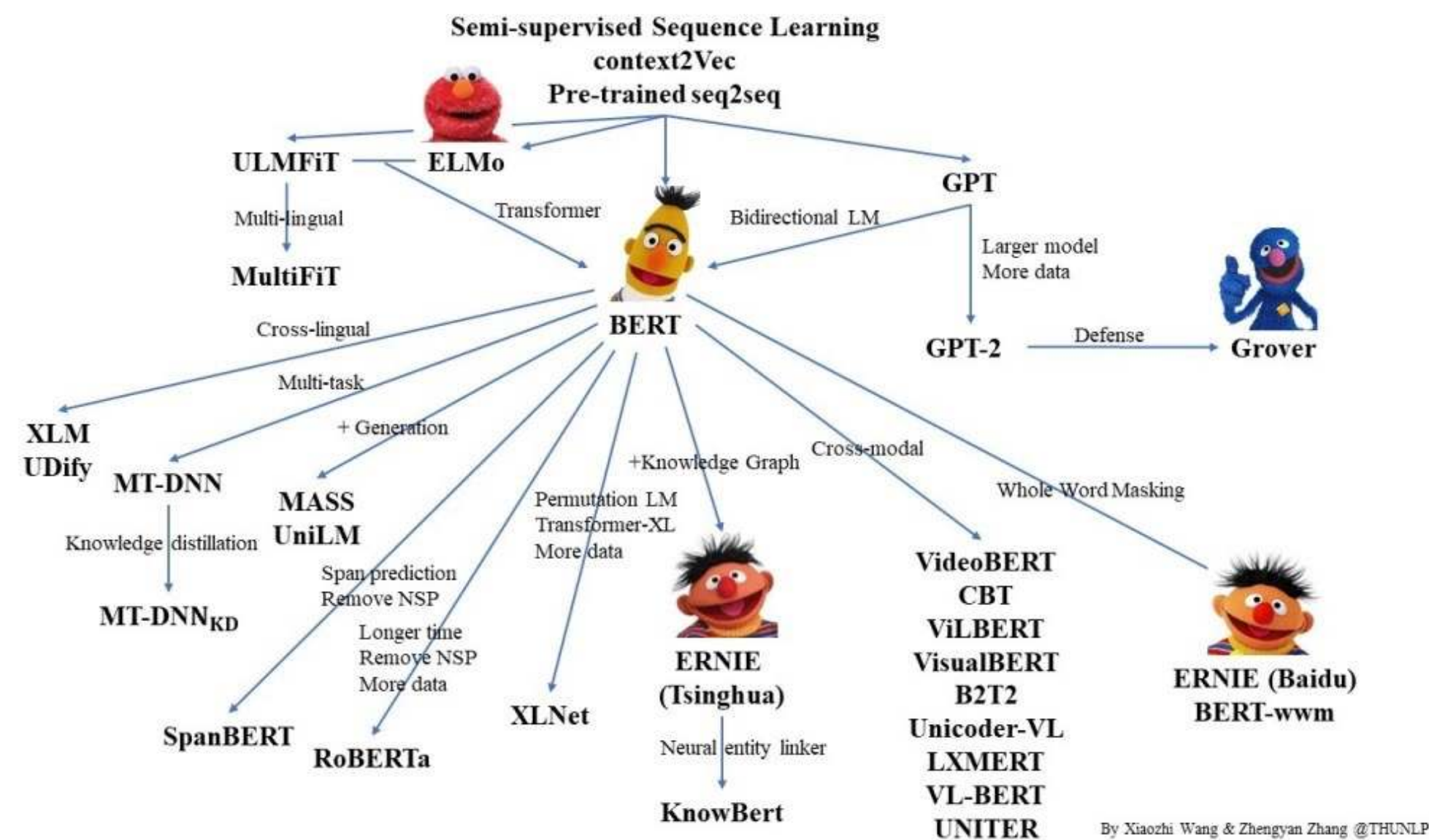
# The wild west of open-source



Explosion of pretrained models:  
which one do I choose?



# The wild west of open-source



Python 2? Really? #8



impredicative opened this issue on 11 Jul 2019 · 3 comments

Oh c'mon you guys... #2



kcrosley-leisurelabs opened this issue on 18 Jun 2020 · 15 comments

Explosion of pretrained models:  
which one do I choose?

Different APIs, missing docs,  
reproducibility issues, ...







[Announcing AutoNLP](#): A new automatic way to train and deploy NLP models.





# The AI community building the future.

Build, train and deploy state of the art models powered by the reference open source in natural language processing.

 **Star** < **44,907**


More than 5,000 organizations are using Hugging Face


 **Amazon**  
Company • 1 model

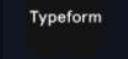
 **Allen Institute for AI**  
Non-Profit • 51 models

 **Microsoft**  
Company • 47 models

 **Google AI**  
Company • 130 models

 **Facebook AI**  
Company • 76 models

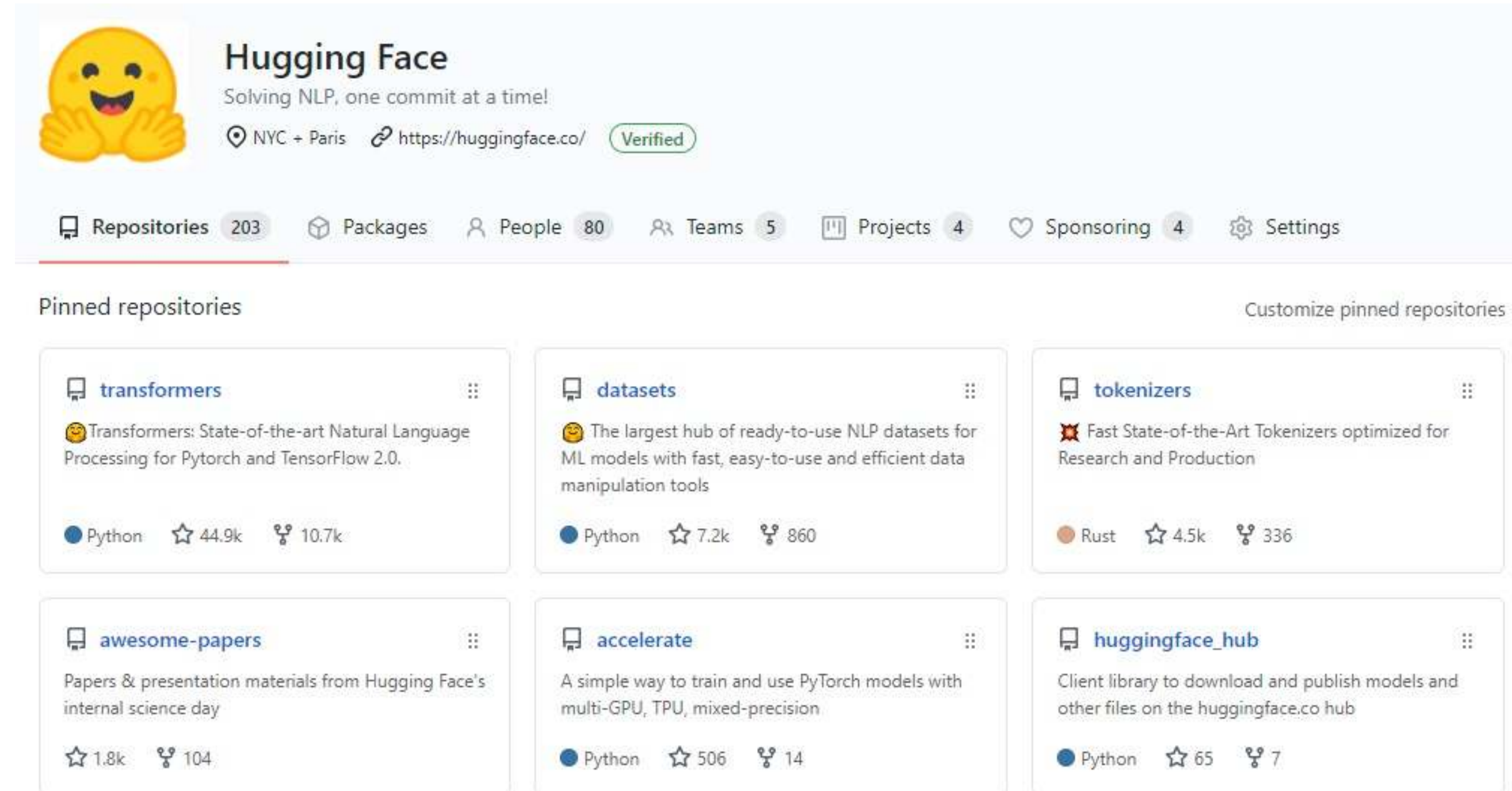
 **Grammarly**  
Company

 **Typeform**  
Company • 8 models

 **asteroid-team**  
Non-Profit



# Open-source @ Hugging Face



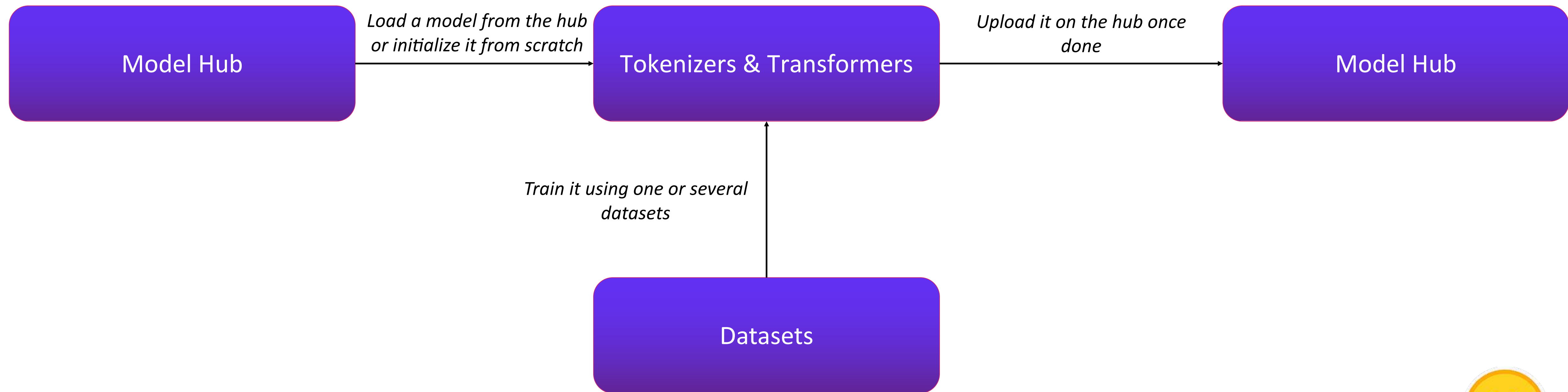
The screenshot shows the GitHub profile for Hugging Face. The profile header includes the Hugging Face emoji logo, the name "Hugging Face", the bio "Solving NLP, one commit at a time!", the location "NYC + Paris", the website "https://huggingface.co/", and a "Verified" badge. Below the header is a navigation bar with links to "Repositories" (203), "Packages", "People" (80), "Teams" (5), "Projects" (4), "Sponsoring" (4), and "Settings".

The "Pinned repositories" section displays six repositories in a grid:

- transformers**: Transformers: State-of-the-art Natural Language Processing for Pytorch and TensorFlow 2.0. Python, 44.9k stars, 10.7k forks.
- datasets**: The largest hub of ready-to-use NLP datasets for ML models with fast, easy-to-use and efficient data manipulation tools. Python, 7.2k stars, 860 forks.
- tokenizers**: Fast State-of-the-Art Tokenizers optimized for Research and Production. Rust, 4.5k stars, 336 forks.
- awesome-papers**: Papers & presentation materials from Hugging Face's internal science day. 1.8k stars, 104 forks.
- accelerate**: A simple way to train and use PyTorch models with multi-GPU, TPU, mixed-precision. Python, 506 stars, 14 forks.
- huggingface\_hub**: Client library to download and publish models and other files on the huggingface.co hub. Python, 65 stars, 7 forks.

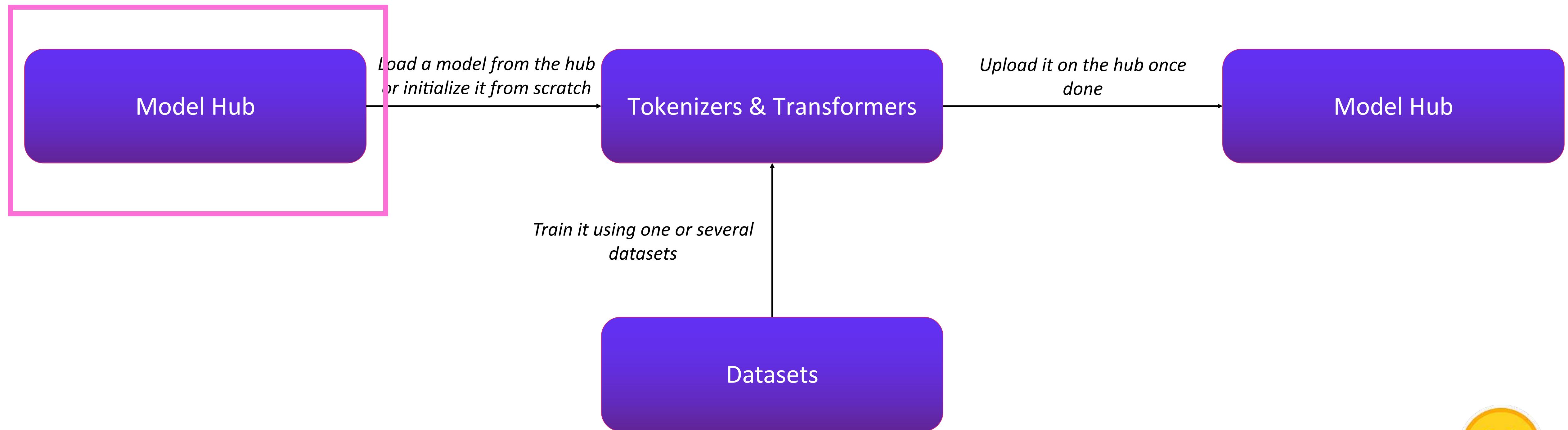


# The Hugging Face ecosystem





# The Hugging Face ecosystem



# The Model Hub



Search models, datasets, users...

Models Datasets Resources Solutions Pricing



## Tasks

- Fill-Mask Question Answering
- Summarization Table Question Answering
- Text Classification Text Generation
- Text2Text Generation Token Classification
- Translation Zero-Shot Classification
- Sentence Similarity +9

## Libraries

- PyTorch TensorFlow JAX +18

## Datasets

- wikiann common\_voice wikipedia indic\_glue
- dcep europarl jrc-acquis squad bookcorpus c4
- +354

## Languages

- en bn es fr de sv fi multilingual

Models 14,282

Search Models

Sort: Most Downloads

bert-base-uncased

Fill-Mask • Updated May 18 • 8,472k

xlm-roberta-base

Fill-Mask • Updated Dec 11, 2020 • 5,290k

distilbert-base-uncased

Fill-Mask • Updated Dec 11, 2020 • 2,863k

roberta-base

Fill-Mask • Updated Dec 11, 2020 • 2,227k

bert-base-cased

Fill-Mask • Updated May 18 • 1,982k

sentence-transformers/paraphrase-xlm-r...

Feature Extraction • Updated 11 days ago • 1,444k

bert-base-chinese

Fill-Mask • Updated May 18 • 1,361k

distilbert-base-uncased-finetuned-sst-2-e...

Text Classification • Updated Feb 9 • 1,246k

distilbert-base-cased

Updated Dec 11, 2020 • 1,154k

roberta-large

Fill-Mask • Updated May 21 • 939k

gpt2

Text Generation • Updated May 19 • 724k

bert-base-multilingual-cased

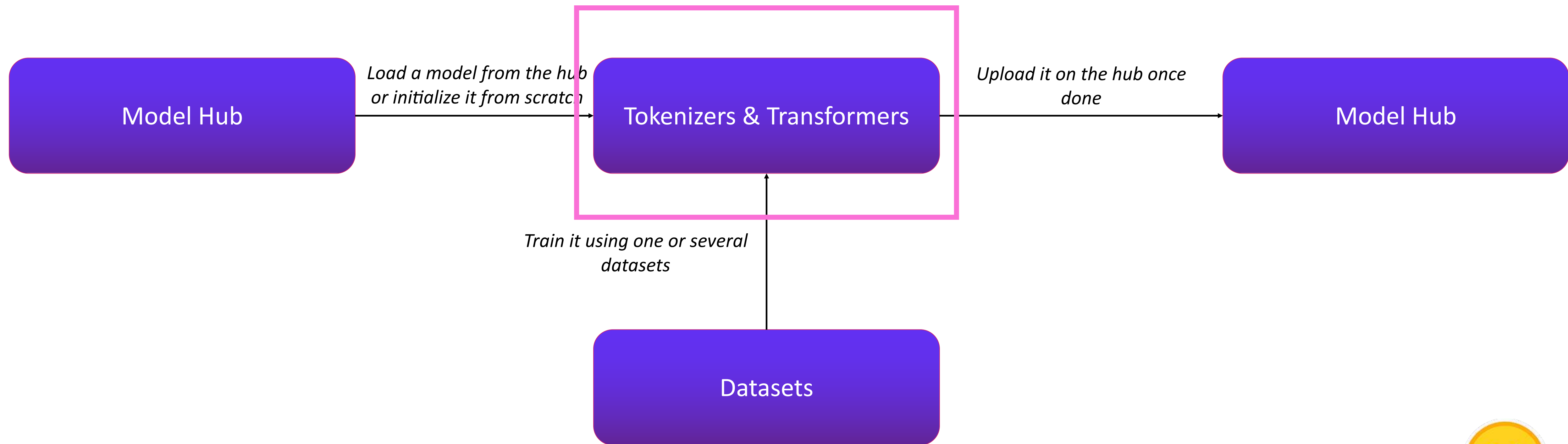
Fill-Mask • Updated May 18 • 629k

finiteautomata/hate-sentiment-analysis

distilroberta-base



# The Hugging Face ecosystem



# Tokenizers and Transformers

```
huggingface@transformers:~  
from transformers import AutoTokenizer, AutoModelForMaskedLM  
tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")  
model = AutoModelForMaskedLM.from_pretrained("bert-base-uncased")
```





# Tokenizers & Transformers

- **50+** Model architectures

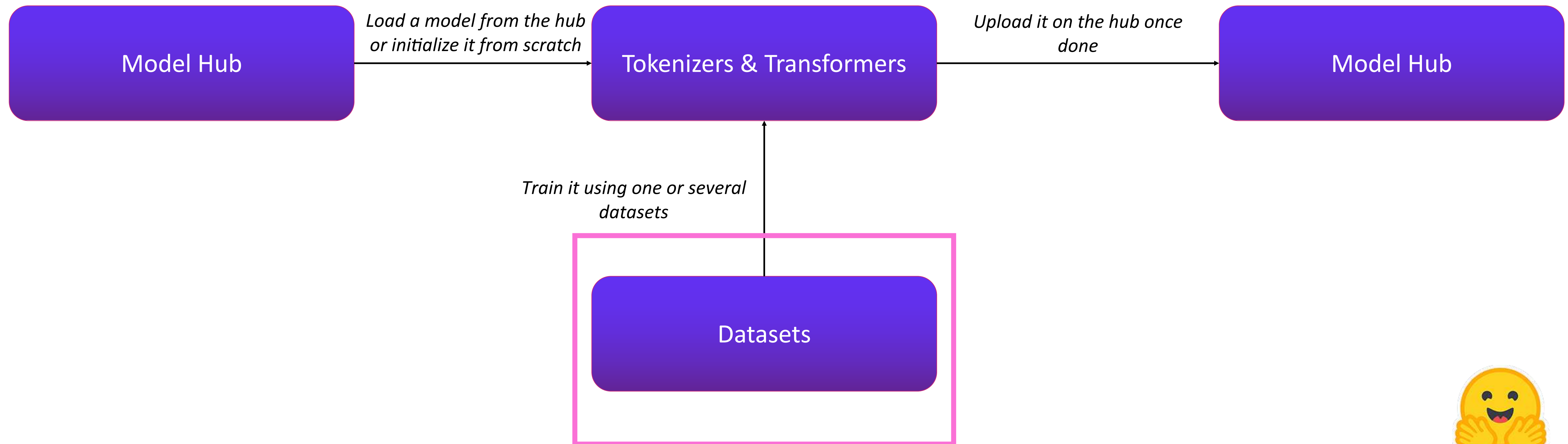
- **Used by more than 5,000** companies

- **Simple API** across all architectures

```
huggingface@transformers:~$  
from transformers import AutoTokenizer, AutoModelForMaskedLM  
tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")  
model = AutoModelForMaskedLM.from_pretrained("bert-base-uncased")
```



# The Hugging Face ecosystem





# The Datasets Hub



- Models
- Datasets
- Resources
- Solutions
- Pricing

## Task Category

- text-classification
- conditional-text-generation
- question-answering
- structure-prediction
- sequence-modeling
- other + 8

## Task

- machine-translation
- named-entity-recognition
- sentiment-classification
- language-modeling
- extractive-qa
- open-domain-qa + 159

## Language

- en
- de
- es
- fr
- pt
- pl + 197

## Multilinguality

- monolingual
- multilingual
- translation
- other-programming-languages
- other-language-learner
- fa + 2

## Size

Datasets 1537

Sort: Alphabetical

### acronym\_identification

Acronym identification training and development sets for the acronym identification task at SDU@AAAI-21.

### ade\_corpus\_v2

ADE-Corpus-V2 Dataset: Adverse Drug Reaction Data. This is a dataset for Classification if a sentence is ADE-related (True) or not (False) an...

### adversarial\_qa

AdversarialQA is a Reading Comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles usi...

### aeslc

A collection of email messages of employees in the Enron Corporation. There are two features: - email\_body: email body text....

### afrikaans\_ner\_corpus

Named entity annotated data from the NCHLT Text Resource Development: Phase II Project, annotated with PERSON, LOCATION...

### ag\_news

AG is a collection of more than 1 million news articles. News articles have been gathered from more than 2000 news sources by...

### ai2\_arc

A new dataset of 7,787 genuine grade-school level, multiple-choice science questions, assembled to encourage research in advanced...

### air\_dialogue

AirDialogue, is a large dataset that contains 402,038 goal-oriented conversations. To collect this dataset, we create a contextgenerator...

### ajgt\_twitter\_ar

Arabic Jordanian General Tweets (AJGT) Corpus consisted of 1,800 tweets annotated as positive and negative. Modern Standard Arabic...

### allegro\_reviews

Allegro Reviews is a sentiment analysis dataset, consisting of 11,588 product reviews written in Polish and extracted from Allegro.pl - a...







# Datasets



Search models, datasets, users...

Models

Datasets

Pricing

Resources

We're hiring!

Log In

Sign Up

Task Category

text-classification conditional-text-generation  
structure-prediction question-answering  
sequence-modeling other +8

Datasets 838

Search Datasets

Sort: Alphabetical

## The Hub:

Task

machine-translation named-entity-recognition  
sentiment-classification language-modeling

extractive-qa multi-class-classification +

Language

en es fr de pl pt +197

Multilinguality

monolingual multilingual translation

other-language-learner fa en +1

Size

10K<n<100K 1K<n<10K 100K<n<1M n<1K

1M<n<10M 10M<n<100M +4

acronym\_identification

Acronym identification training and development sets for the acronym identification task at SDU@AAAI-21.

ade\_corpus\_v2

ADE-Corpus-V2 Dataset: Adverse Drug Reaction Data. This is a dataset for Classification if a sentence is ADE-related (True) or not (False) an...

adversarial\_qa

AdversarialQA is a Reading Comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles usi...

enron\_emails

A collection of email messages of employees in the Enron Corporation. There are two features: - email\_body: email body text...

eflora\_ner\_corpus

Named entity annotated data from the NCHLT Text Resource Development: Phase II Project, annotated with PERSON, LOCATION...

ag\_news

AG is a collection of more than 1 million news articles. News articles have been gathered from more than 2000 news sources by...

gsm8k

A new dataset of 7,787 genuine grade-school level, multiple-choice science questions, assembled to encourage research in advanced...

air\_dialogue

AirDialogue, is a large dataset that contains 402,038 goal-oriented conversations. To collect this dataset, we create a contextgenerator...

ajgt\_twitter\_ar

Arabic Jordanian General Tweets (AJGT) Corpus consisted of 1,800 tweets annotated as positive and negative. Modern Standard Arabic...

allegro\_reviews

Allegro Reviews is a sentiment analysis dataset, consisting of 11,588 product reviews written in Polish and extracted from Allegro.pl - a...

- Largest hub of ready-to-use datasets

- 1000+ Datasets available

- 450+ languages and dialects supported





# Datasets

The python library:

- Load any dataset in one line

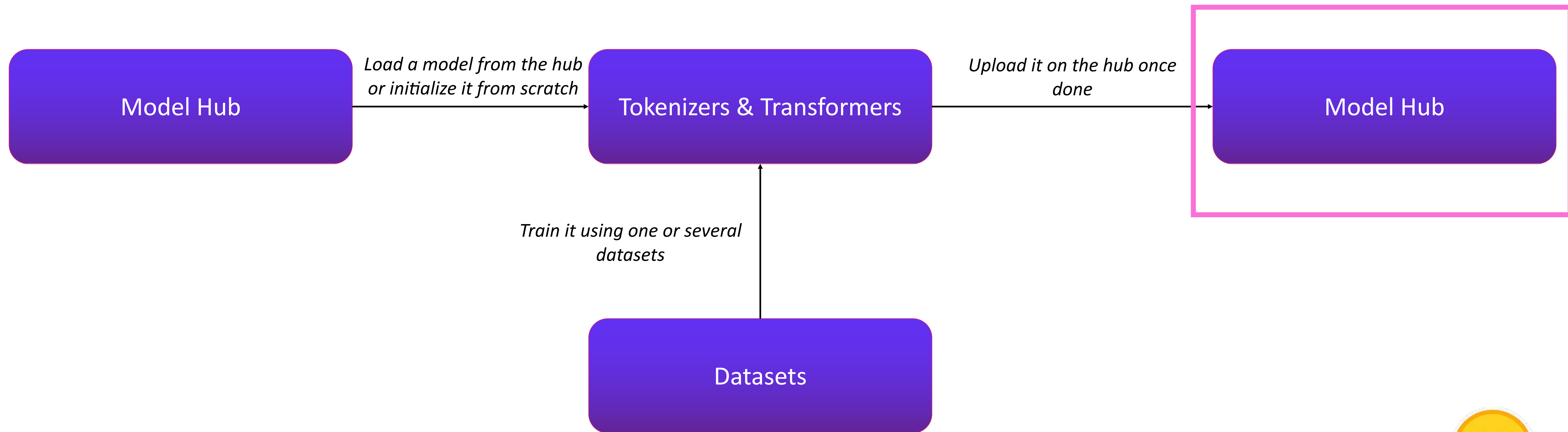
- Supports huge datasets without RAM limitations

- Fast iterations and querying





# The Hugging Face ecosystem





# Hugging Face Hub

```
● ● ●  
from huggingface_hub import HfApi()  
api = HfApi()
```



# Hugging Face Hub

```
class DummyModel(nn.Module, ModelHubMixin):

    def __init__(self, **kwargs):
        super().__init__()
        self.config = kwargs.pop("config", None)
        self.l1 = nn.Linear(2, 2)

    def forward(self, x):
        return self.l1(x)

model = DummyModel()
model.save_pretrained("my-dummy-model")
model.push_to_hub("my-dummy-model", organization="huggingface")

# Reload it from any device!

model = DummyModel.from_pretrained("huggingface/my-dummy-model")
```





# HuggingFace Hub

- Python API

- Simple mixin for your PyTorch Module

```
from huggingface_hub import HfApi()
```

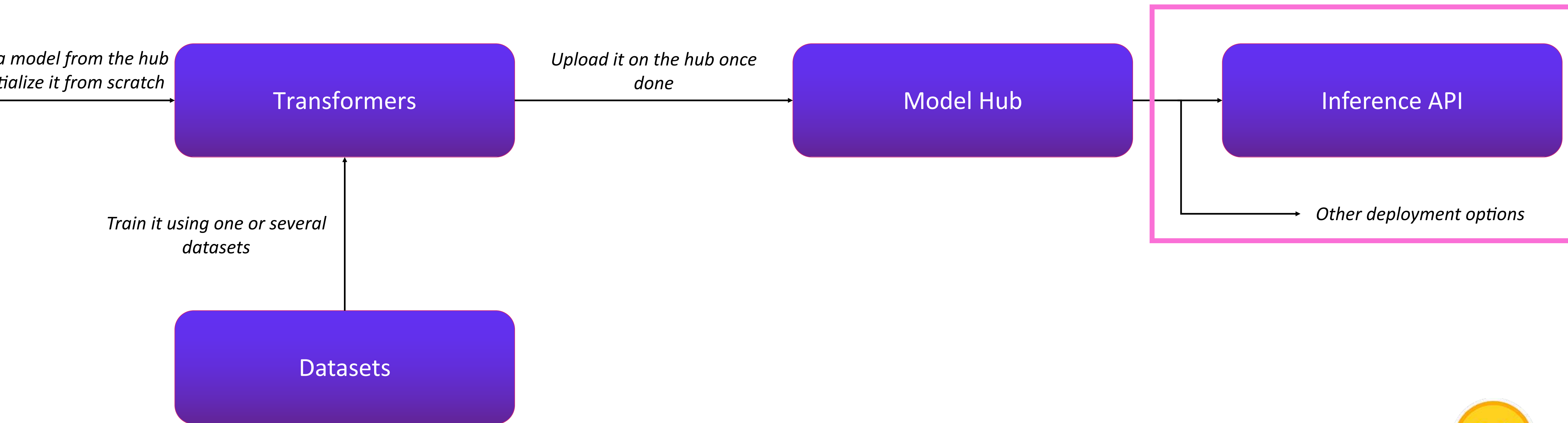
- Supports Transformers, AllenNLP, Asteroid, Spacy,

Timm ...





# The Hugging Face ecosystem



# Inference API

Fill-Mask distilbert-base-uncased

The goal of life is [MASK]. Compute

This model can be loaded on the Inference API on-demand.  
Computation time on cpu: cached.

happiness	0.036
survival	0.031
salvation	0.017
freedom	0.017
unity	0.015

</> JSON Output Maximize

Token Classification dbmdz/bert-finetuned-conll03

My name is Clara and I live in Berkeley, Californi Compute

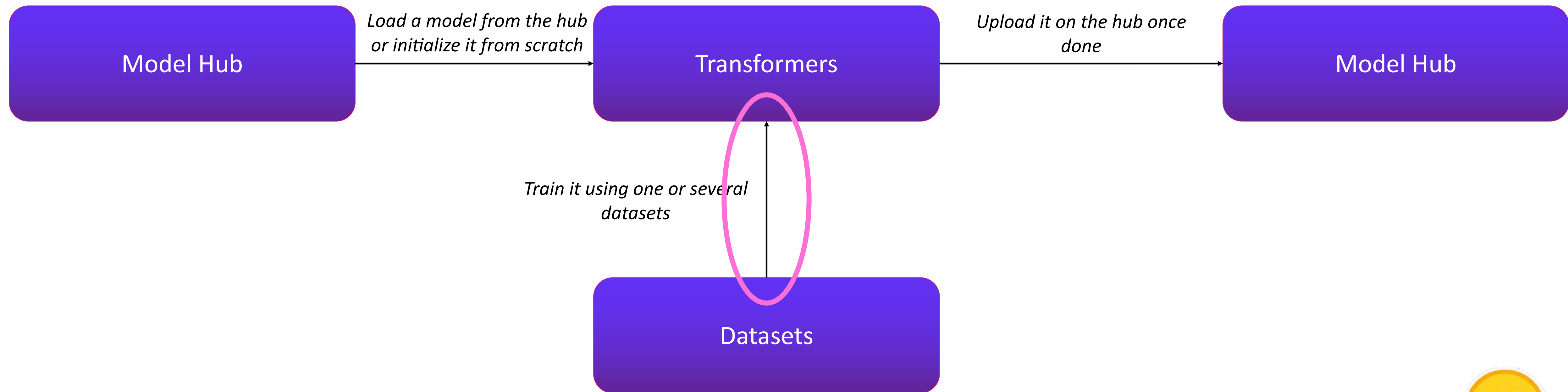
This model can be loaded on the Inference API on-demand.  
Computation time on cpu: cached.

My name is Clara **PER** and I live in Berkeley **LOC**, California  
**LOC**. I work at this cool company called Hugging Face **ORG**.

</> JSON Output Maximize



# The Hugging Face ecosystem





# Three ways to train a model



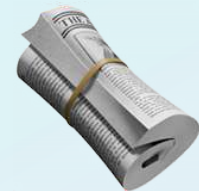
# Trainer



```
from transformers import AutoModelForSequenceClassification, Trainer, TrainingArguments
model = AutoModelForSequenceClassification.from_pretrained("bert-base-uncased")
training_args = TrainingArguments(
    output_dir="./my-finetuned-model", # output directory
    num_train_epochs=3, # total number of training epochs
    per_device_train_batch_size=16, # batch size per device during training
    per_device_eval_batch_size=64, # batch size for evaluation
    evaluation_strategy="epoch", # evaluate every epoch
    weight_decay = 0.01, # strength of weight decay
)
trainer = Trainer(
    model=model, # model to be trained
    args=training_args, # training arguments, defined above
    train_dataset=train_dataset, # the training set
    eval_dataset=test_dataset, # the evaluation set
    tokenizer=tokenizer, # the tokenizer used for preprocessing
)
train.train()
train.push_to_hub()
```







# Accelerate

```
import torch
import torch.nn.functional as F
from datasets import load_dataset
+ from accelerate import Accelerator

+ accelerator = Accelerator()
- device = 'cpu'
+ device = accelerator.device

model = torch.nn.Transformer().to(device)
optim = torch.optim.Adam(model.parameters())

dataset = load_dataset('my_dataset')
data = torch.utils.data.DataLoader(dataset, shuffle=True)

+ model, optim, data = accelerator.prepare(model, optim, data)
```

```
model.train()
for epoch in range(10):
    for source, targets in data:
        source = source.to(device)
        targets = targets.to(device)

        optimizer.zero_grad()

        output = model(source)
        loss = F.cross_entropy(output, targets)

- loss.backward()
+ accelerator.backward(loss)

        optimizer.step()
```







# Accelerate: handling devices

```
import torch
import torch.nn.functional as F
from datasets import load_dataset
+ from accelerate import Accelerator

+ accelerator = Accelerator()
- device = 'cpu'

- model = torch.nn.Transformer().to(device)
+ model = torch.nn.Transformer()
optim = torch.optim.Adam(model.parameters())

dataset = load_dataset('my_dataset')
data = torch.utils.data.DataLoader(dataset, shuffle=True)

+ model, optim, data = accelerator.prepare(model, optim, data)
```

```
model.train()
for epoch in range(10):
    for source, targets in data:
        - source = source.to(device)
        - targets = targets.to(device)

        optimizer.zero_grad()

        output = model(source)
        loss = F.cross_entropy(output, targets)

        - loss.backward()
        + accelerator.backward(loss)

        optimizer.step()
```





# Accelerate

## Currently supported:

- Training

- Evaluation

- CPU, GPU, Multi GPU, TPU, Mixed

precision

- AWS Sagemaker

```
import torch
import torch.nn as nn
from datasets import load_dataset
+ from accelerate import Accelerator

+ accelerator = Accelerator()
- device = 'cpu'
+ device = accelerator.device

model = torch.nn.Transformer().to(device)
optim = torch.optim.Adam(model.parameters())
dataset = load_dataset('my_dataset')
data = torch.utils.data.DataLoader(dataset, shuffle=True)

+ model, optim, data = accelerator.prepare(model, optim, data)
```

## Upcoming:

- Horovod

- FairScale

- DeepSpeed

- Others? 🙌

```
model.train()
for epoch in range(10):
    for source, targets in data:
        source = source.to(device)
        targets = targets.to(device)

        optimizer.zero_grad()

        output = model(source)
        loss = F.cross_entropy(output, targets)

        - loss.backward()
        + accelerator.backward(loss)

        optimizer.step()
```







# auto



*# Upload your model data*

```
autonlp upload --project sentiment_detection --split train  
              --col_mapping review:text,sentiment:target  
              --files ~/datasets/train.csv
```

*# Train your model*

```
autonlp train --project sentiment_detection
```

*# Use your model*

```
curl -X POST  
     -H "Authorization: Bearer API api_jeZrkpoqfjziaoRaerjlbRQeKykrop"  
     -H "Content-Type: application/json"  
     -d '{"inputs": "The goal of life is [MASK]"}'  
     https://api-inference.huggingface.co/models/sentiment_detection
```







# Big Science

[bigscience.huggingface.co/](https://bigscience.huggingface.co/)



# What is Big Science?

*One-year research workshop on large multilingual datasets  
and large language models*

Analogy with the Large Hadron Collider at CERN:

- has involved **10,000+** researchers
- from **100+** countries
- lead to the discovery of **59 hadrons**
- publication of more than **2,800 papers** (🤯)

In many scientific fields (epidemiology, space, fusion...), **large-scale and worldwide research collaborations** create **tools** useful for the entire research community, like the LHC, ITER, ISS...

*Isn't it time to build similar **large, diverse, open research collaborations** in AI/NLP as well?*



# Why do this?

## Research

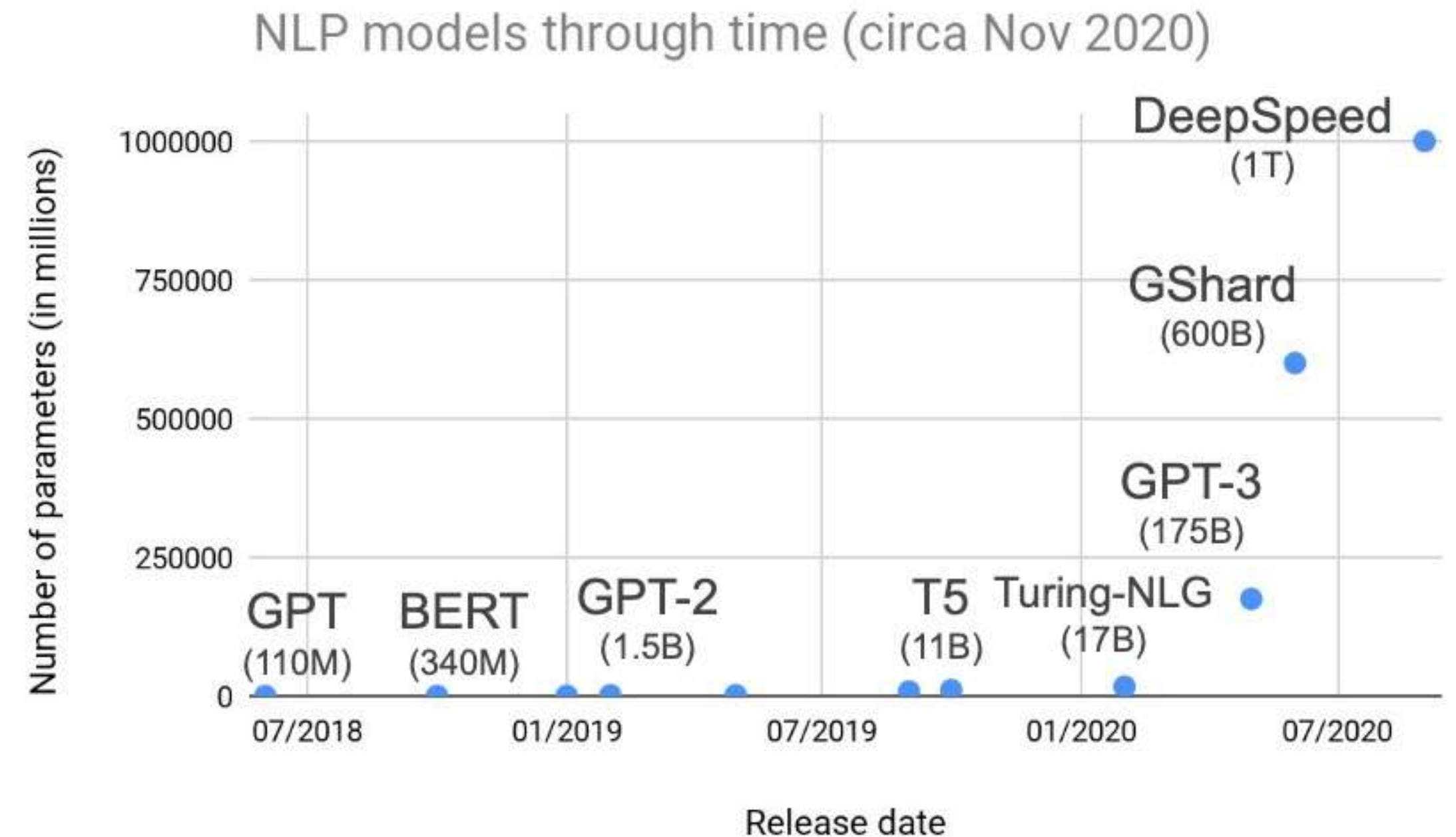
- Models not designed as general research tools
- Difficult involvement of academic researchers
- Lack of fields diversity of the research teams building them

## Environmental

- Training parallel models in private setting => duplication of energy requirements
- Carbon footprint not documented/taken into account

## Ethical and societal

- Shortcomings in the text corpora used to train these models
- Ethical/bias/usage question are usually asked a-posteriori





# But large-scale public compute exists

## Accelerated partition (or GPU partition)

- 261 four-GPU accelerated compute nodes with:
  - 2 Intel Cascade Lake 6248 processors (20 cores at 2.5 GHz), namely 40 cores per node
  - 192 GB of memory per node
  - 4 Nvidia Tesla V100 SXM2 GPUs (32 GB)
- 31 eight-GPU accelerated compute nodes, currently dedicated to the AI community with:
  - 2 Intel Cascade Lake 6226 processors (12 cores at 2.7 GHz), namely 24 cores per node
  - 20 nodes with 384 GB of memory and 11 nodes with 768 GB of memory
  - 8 Nvidia Tesla V100 SXM2 GPUs (32 GB)
- Extension in the summer of 2020, 351 four-GPU accelerated compute nodes with:
  - 2 Intel Cascade Lake 6248 processors (20 cores at 2.5 GHz), namely 40 cores per node
  - 192 GB of memory per node
  - 4 Nvidia Tesla V100 SXM2 GPUs (16 GB)



- Cumulated peak performance of 28 Pflop/s with a total of **2696 Nvidia V100 GPUs**

Jean Zay supercomputer in France





# A brief history

- 🐣 **Early 2021**: Discussions between Thomas Wolf (HuggingFace), Stéphane Requena (GENCI) and Pierre-François Lavallée (IDRIS)
- 👨‍👩‍👧‍👦 **Very quickly**: Science team of HF + many members of the French academic and industrial AI and NLP research communities joined the discussion to further develop the project leading to the grant application
- 📝 **February 2021**: Grant application for 5 million GPU hours
- 🌐 **Following the grant submission**: open/extend to international research community
- 🛠️ **When the project reached 200+ participants**: the organization of the project started to take shape and to adopt the structure of a research workshop
- 🚀 **19/04**: Grant accepted -- first half of the project
- 🌸 **28/04**: Kickoff event + project becomes public



# Core research questions

- **Large models + large datasets:** exhibit intriguing and quite surprising behaviors from a research point of view
- Raise **many research questions** across many fields/subfields of AI/NLP:
  - Fundamental:
    - limits of what can be done with purely statistical and text-based approaches?
    - notion of what is an NLP task and what is the relation between a task and a dataset
  - Bias/fairness:
    - notion of bias and its relation to the dataset and training objectives
    - representativeness and stereotypes
    - memorization versus generalization and personal information memorization
  - Environmental impact and carbon footprint
  - And so many others (interpretability, relation to cognitive processes, use in linguistics...)





# How can I participate?

- **General Advisor** (Steering Committee member):
  - *role* - give general scientific/organization advice - everyone here is in the SC by default
  - *time commitment* - **light** - reading a newsletter every 2 weeks - giving feedback/advice
- **Join a Working Group** (Organizing Committee member)
  - *role* - join one of the Working Groups to advise or participate (code, research...)
  - *time commitment* - **medium** - depends on the chosen WG
- **Chair/co-Chair a Working Group** (Organizing Committee member)
  - *role* - the chairs are responsible for providing the minimal amount of work necessary for having a barebone version of the task. If WG members are active, the chairs can mostly coordinate the effort and organise the decision process.
  - *time commitment* - **more significant** - depends on the chosen WG
- A **Workshop attendant** joining **live events** or some **community events** (tbd)
  - *role* - participating in the collaborative task following guidelines by the WG
  - *time commitment* - **free** - up to the attendant - open to anyone, beginners, people outside of the research fields, etc... - very accessible



