

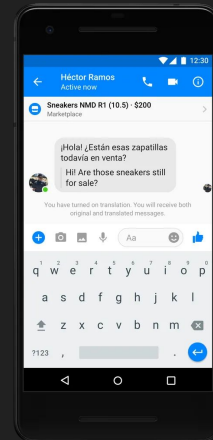
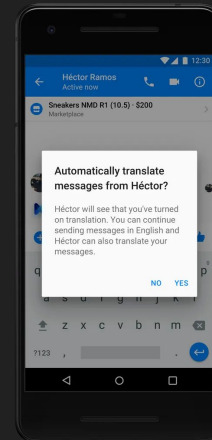
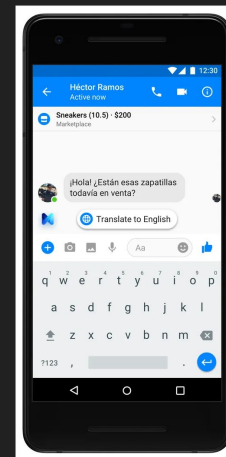
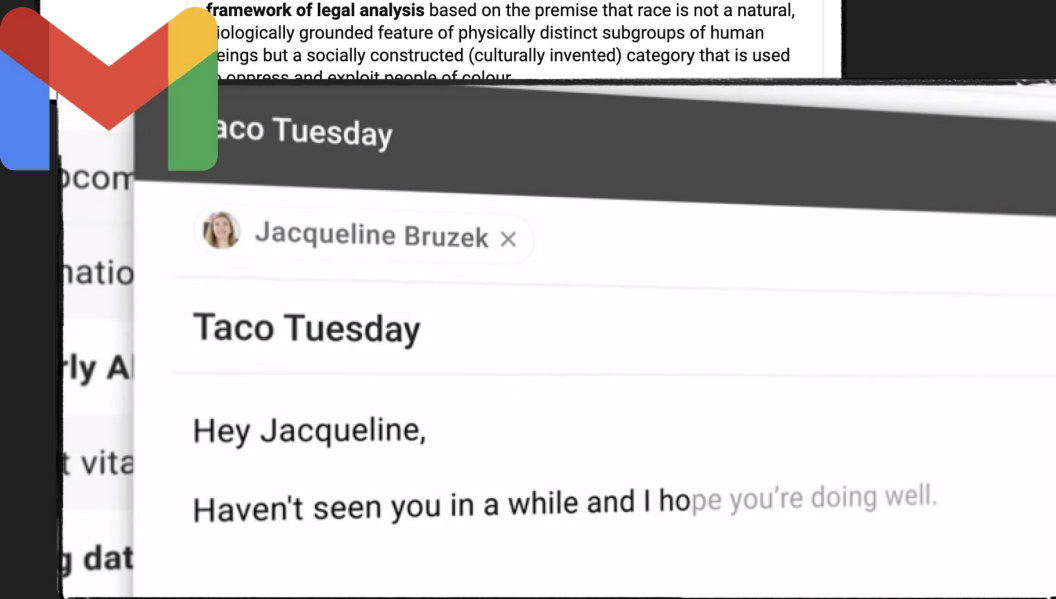
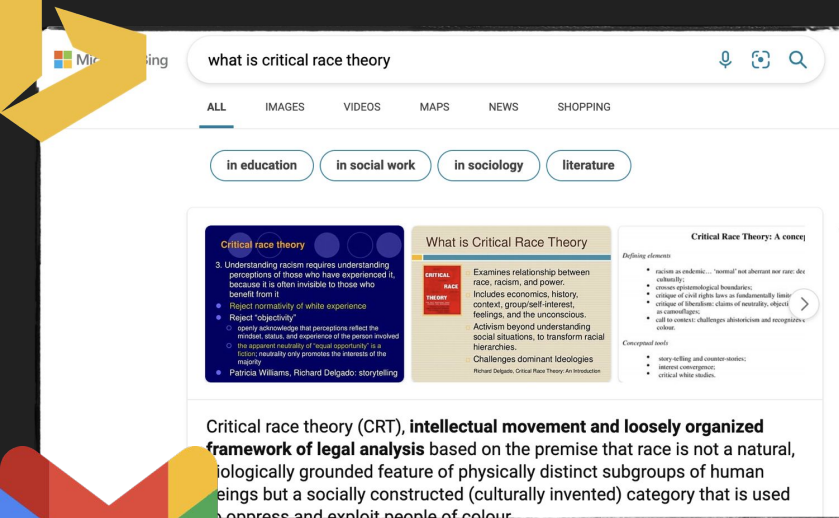
Building AI Applications in Minutes

Rajiv Shah

raj@hf.com
@rajistics

Colab Notebook: https://bit.ly/raj_food





Anyone Can use these Technologies!

Let me show you how!!



Build AI App!



What kind of
food is this
image?



Translate the
kind of food into
Arabic





Data Scientist

Roadmap

Mathematics

- Linear Algebra
- Analytics Geometry
- Matrix
- Vector Calculus
- Optimization
- Regression
- Dimensionality Reduction
- Density Estimation
- Classification

Probability

- Discrete Distribution
 - Binomial
 - Bernoulli
 - Geometric etc
- Continuous Distribution
 - Uniform
 - Exponential
 - Gamma
- Normal Distribution
- Introduction to Probability
- 1D Random Variable
- Function of One Random Variable
- Joint Probability Distribution

Statistics

- Introduction to Statistics
- Data Description
- Random Samples
- Sampling Distribution
- Parameter Estimation
- Hypotheses Testing
- ANOVA
- Reliability Engineering
- Stochastic Process
- Computer Simulation
- Design of Experiments
- Simple Linear Regression
- Correlation
- Multiple Regression
- Nonparametric Statistics
 - Sign Test
 - The Wilcoxon Signed-Rank Test
 - The Wilcoxon Rank Sum Test
 - The Kruskal-Wallis Test
- Statistical Quality Control
- Basic of Graphs

Programming

- | Python | R |
|--|---|
| Python Basics <ul style="list-style-type: none">• List• Set• Tuples• Dictionary• Function, etc. | R Basic <ul style="list-style-type: none">• Vector• List• Data Frame• Matrix• Array, etc |
| NumPy | dplyr |
| Pandas | ggplot2 |
| Matplotlib/Seaborn, etc. | Tidyr |
| | Shiny, etc. |
| DataBase | Other |
| SQL | Data Structure <ul style="list-style-type: none">• Array, etc |
| MongoDB | Web Scrapping |
| | Linux |
| | Git |

Machine Learning

- | Introduction | Intermediate |
|---|--|
| <ul style="list-style-type: none">• How Model Works• Basic Data Exploration• First ML Model• Model Validation• Underfitting & Overfitting• Random Forests• scikit-learn | <ul style="list-style-type: none">• Handling Missing Values• Handling Categorical Variables• Pipelines• Cross-Validation• XGBoost• Data Leakage |

Deep Learning

- | | |
|--|--|
| <ul style="list-style-type: none">• Artificial Neural Network• Convolutional Neural Network• Recurrent Neural Network• Keras• PyTorch• TensorFlow | <ul style="list-style-type: none">• A Single Neuron• Deep Neural Network• Stochastic Gradient Descent• Overfitting and Underfitting• Dropout Batch Normalization• Binary Classification |
|--|--|

Feature Engineering

- Baseline Model
- Categorical Encodings
- Feature Generation
- Feature Selection

Natural language Processing

- Text Classification
- Word Vectors

Data Visualization Tools

- Excel VBA
- BI (Business Intelligence)
 - Tableau
 - Power BI
 - Qlik View
 - Qlik Sense

Deployment

- Microsoft Azure
- Heroku
- Google Cloud Platform
- Flask
- Django

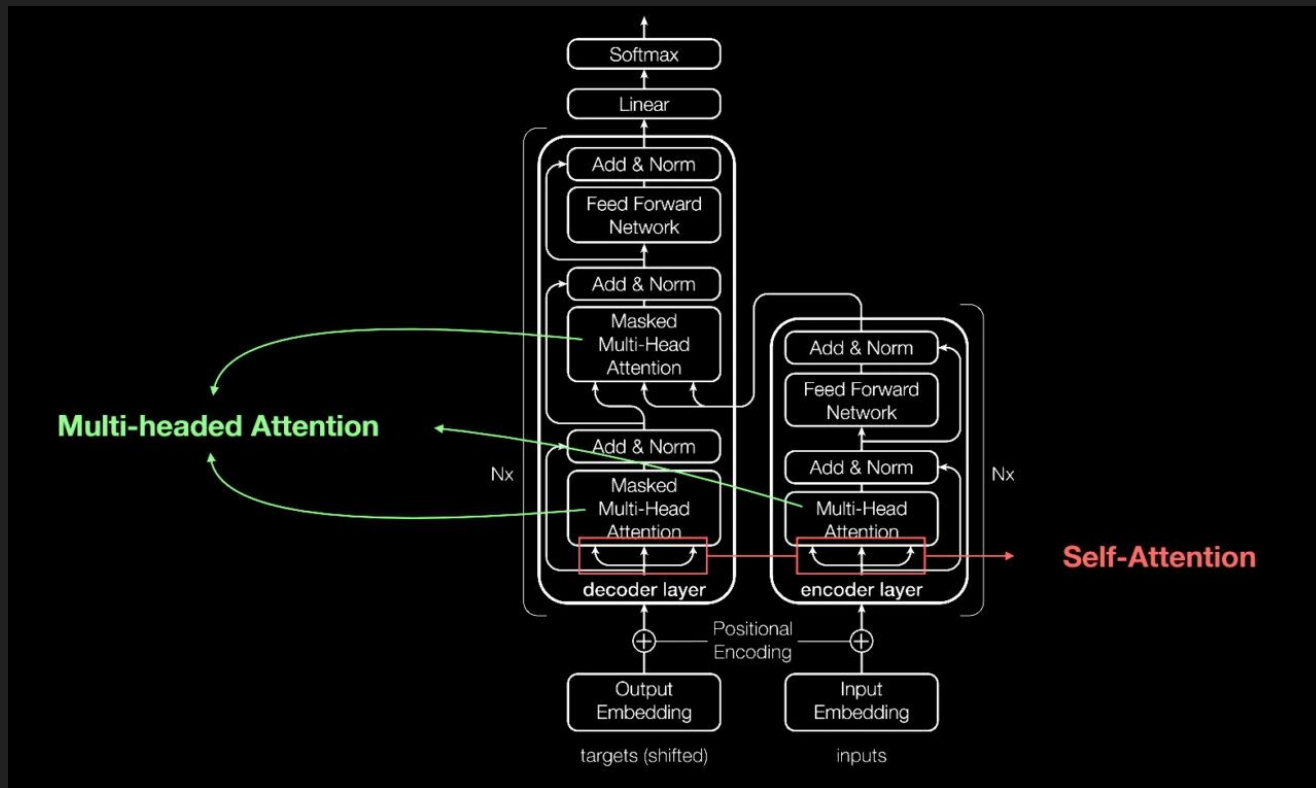
Other Points

- Domain Knowledge
- Communication Skill
- Reinforcement Learning
- Case Studies
 - Data Science at Netflix
 - Data Science at Flipkart
 - Project on Credit Card Fraud Detection
 - Project on Movie Recommendation , etc.

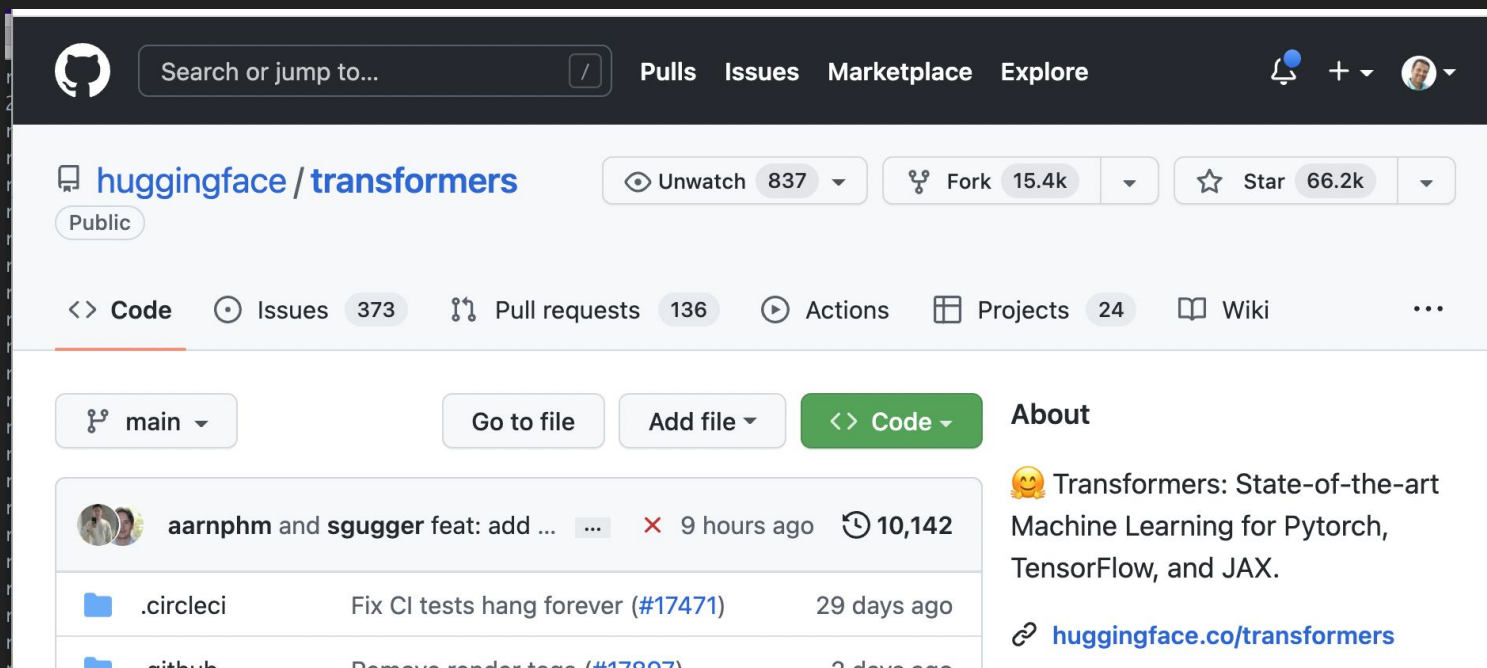
Keep Practicing



Transformer Models



Transformer Library



The screenshot shows the GitHub repository page for `huggingface/transformers`. The repository is public and has 66.2k stars, 15.4k forks, and 837 unwatchers. The repository is categorized as `Code` and has 373 issues, 136 pull requests, 24 projects, and a wiki. The repository is currently on the `main` branch. The repository description is: "Transformers: State-of-the-art Machine Learning for Pytorch, TensorFlow, and JAX." The repository is linked to huggingface.co/transformers.

huggingface / transformers

Public

Unwatch 837 Fork 15.4k Star 66.2k

Code Issues 373 Pull requests 136 Actions Projects 24 Wiki

main Go to file Add file Code

About

🤖 Transformers: State-of-the-art Machine Learning for Pytorch, TensorFlow, and JAX.

🔗 huggingface.co/transformers

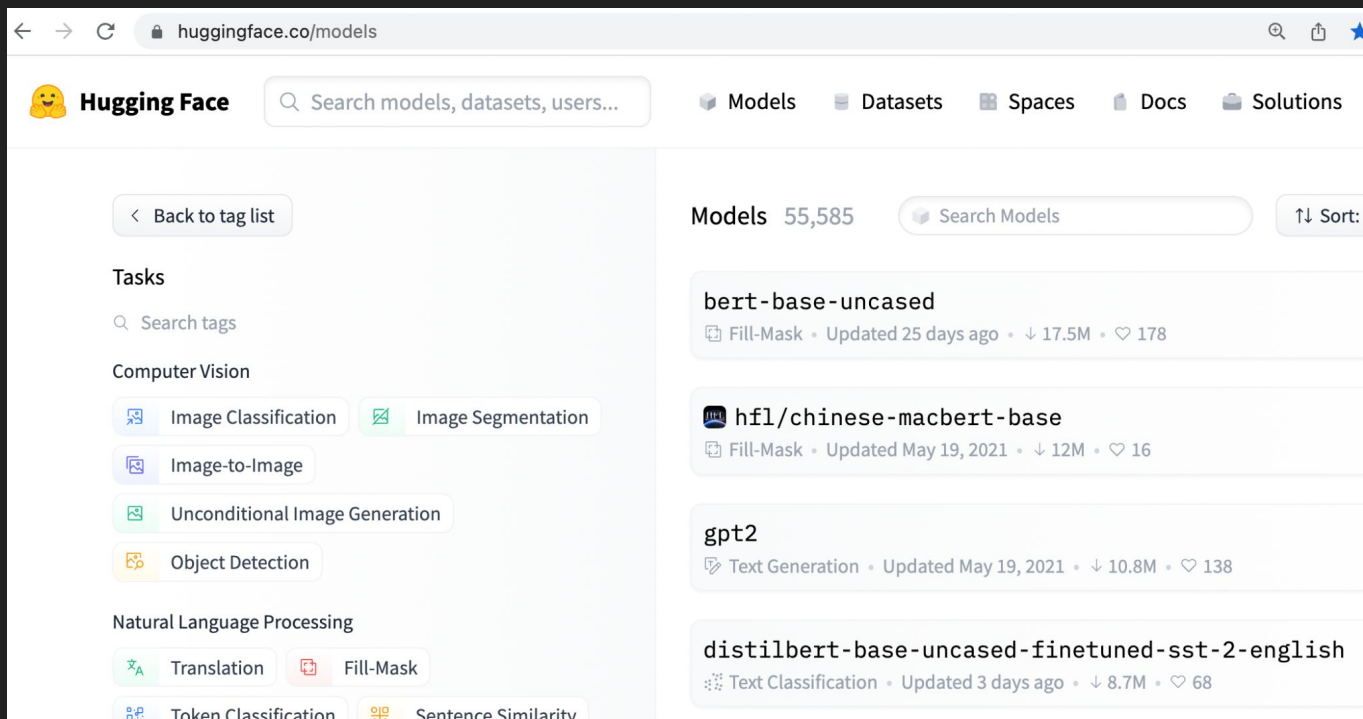
aarnphm and sgugger feat: add ... 9 hours ago 10,142

.circleci Fix CI tests hang forever (#17471) 29 days ago

github Remove render tags (#17897) 2 days ago



Hugging Face Public Hub



The screenshot shows the Hugging Face Public Hub website. The browser address bar displays `huggingface.co/models`. The page header includes the Hugging Face logo, a search bar for models, datasets, and users, and navigation links for Models, Datasets, Spaces, Docs, and Solutions.

The main content area is divided into two sections. On the left is a sidebar with a 'Back to tag list' button and a 'Tasks' section. The 'Tasks' section includes a search bar and a list of task categories: Computer Vision (Image Classification, Image Segmentation, Image-to-Image, Unconditional Image Generation, Object Detection) and Natural Language Processing (Translation, Fill-Mask, Token Classification, Sentence Similarity). On the right is the 'Models' section, which shows a total of 55,585 models. It includes a search bar for models and a 'Sort' button. Below this, a list of models is displayed, each with its name, task, update date, size, and number of likes.

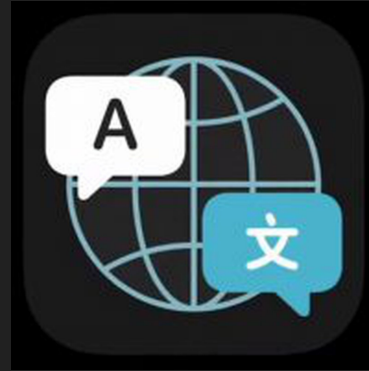
Model Name	Task	Updated	Size	Likes
<code>bert-base-uncased</code>	Fill-Mask	Updated 25 days ago	17.5M	178
<code>hfl/chinese-macbert-base</code>	Fill-Mask	Updated May 19, 2021	12M	16
<code>gpt2</code>	Text Generation	Updated May 19, 2021	10.8M	138
<code>distilbert-base-uncased-finetuned-sst-2-english</code>	Text Classification	Updated 3 days ago	8.7M	68



Indian Food Detector



Image Classifier → Text



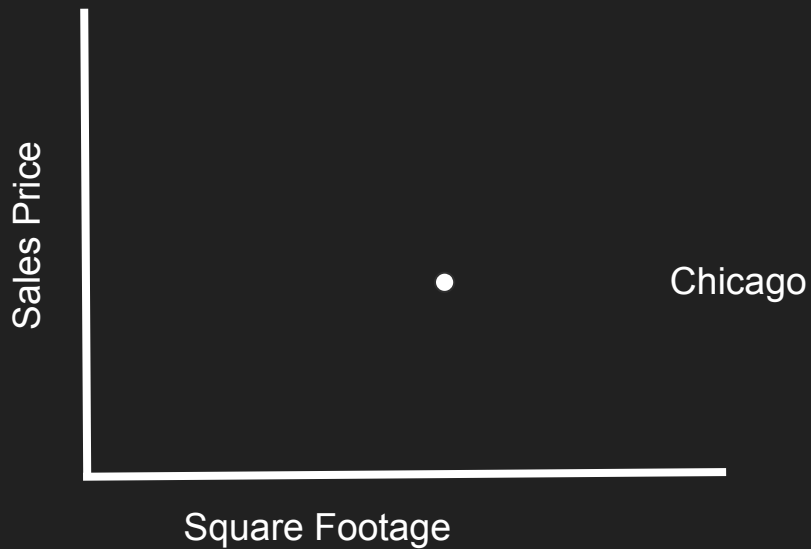
Text → Translator



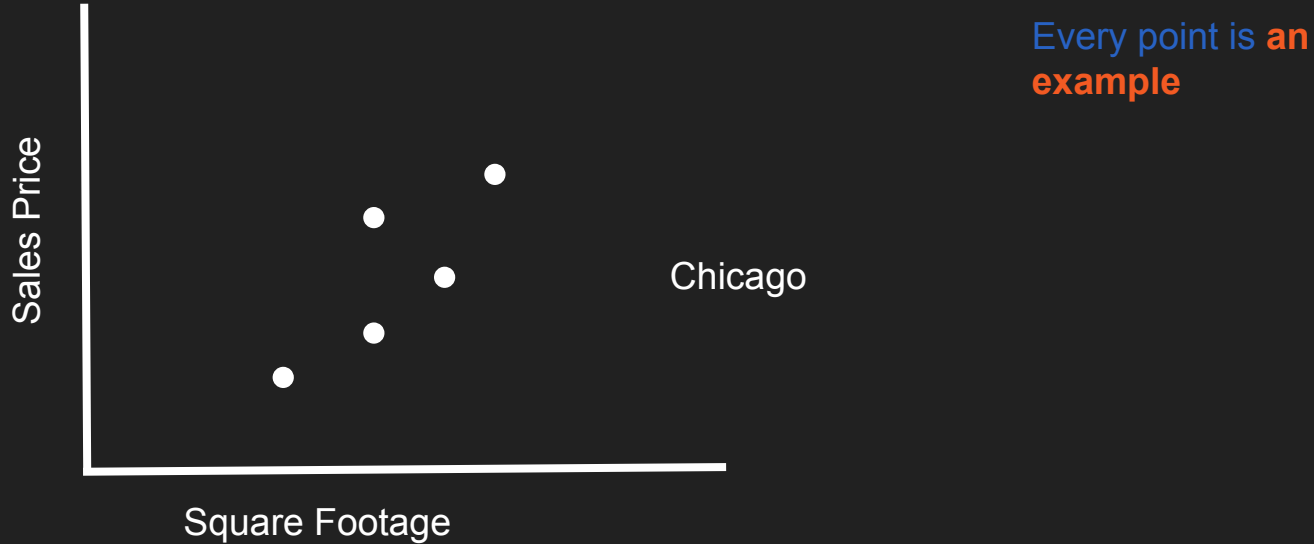
The conventional path . . .



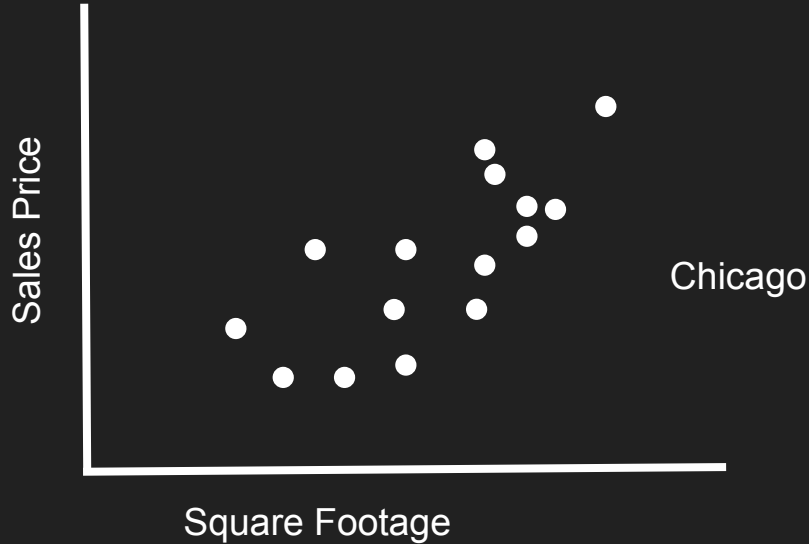
Machine Learning



Machine Learning



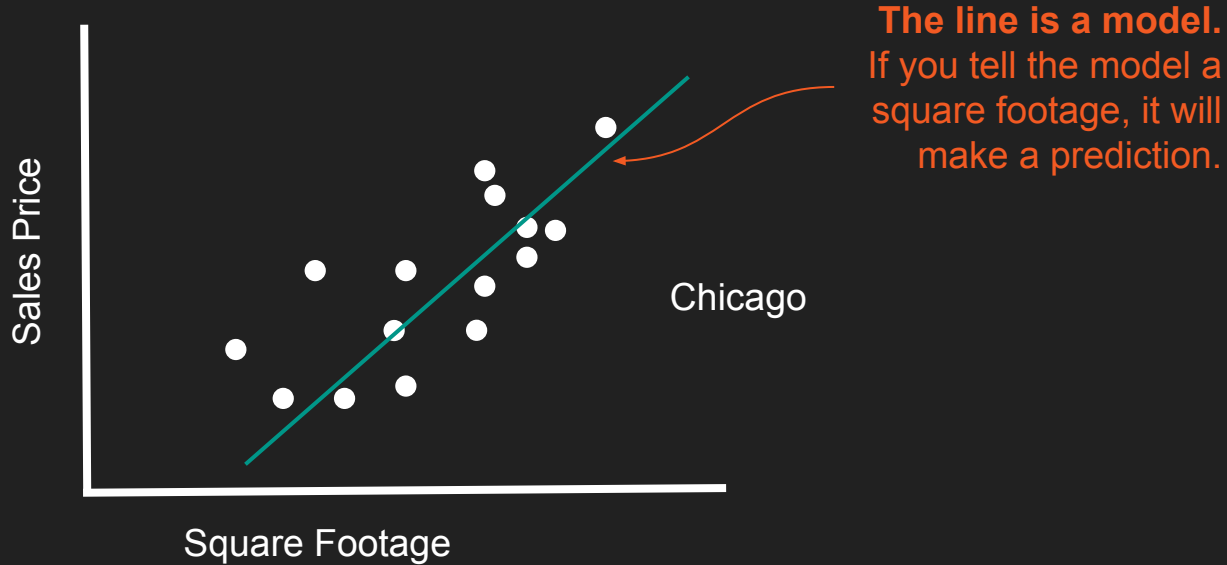
Machine Learning



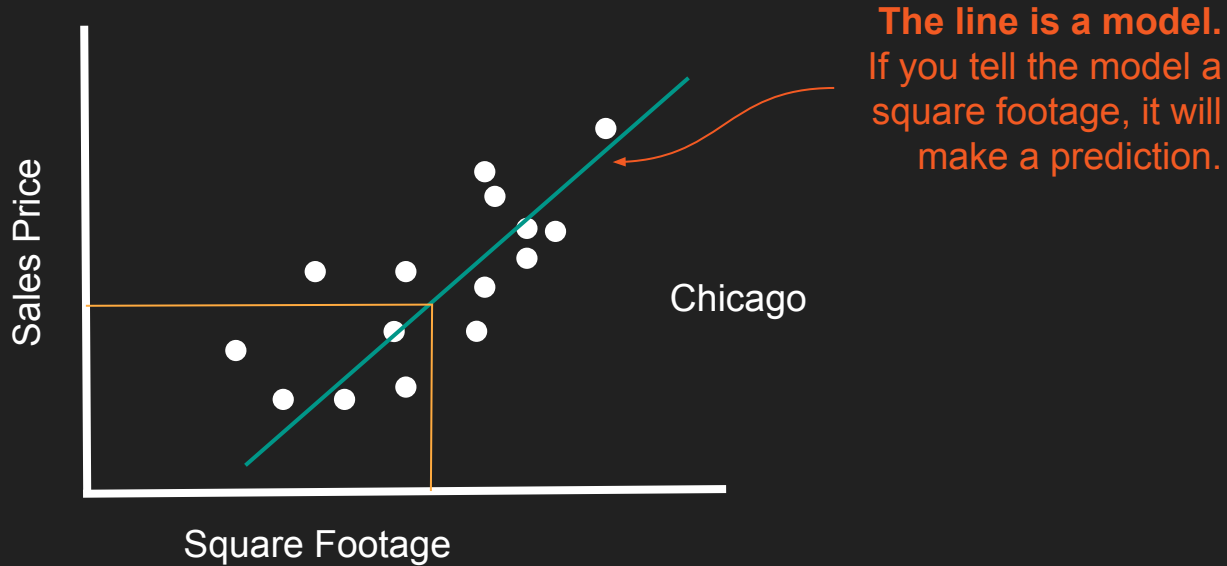
Every point is **an example** that the machine learns from



Machine Learning



Machine Learning



Data
Prep

Labeling

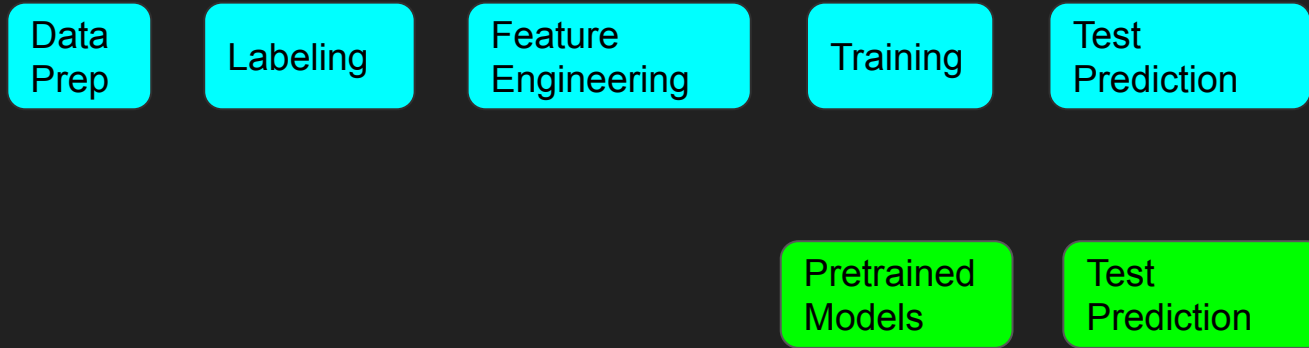
Feature
Engineering

Training

Test
Prediction



Pre-trained Models

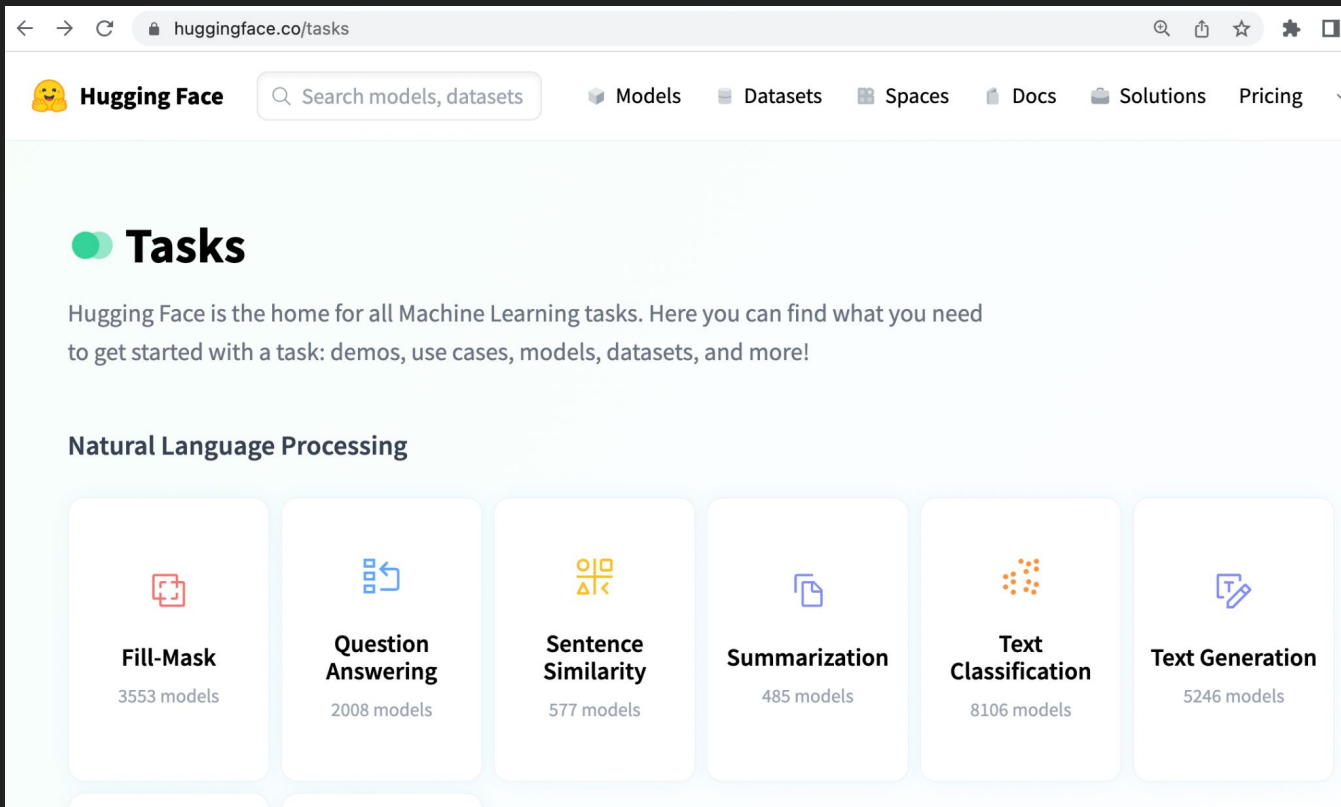


Pre-trained Models

- Trained on enormous datasets
- Contain million to billions of parameters
- Even take months to train



Let's explore some pretrained models at [hf.co/tasks](https://huggingface.co/tasks)



The screenshot shows the Hugging Face website's 'Tasks' page. The browser address bar displays 'huggingface.co/tasks'. The navigation bar includes the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, Docs, Solutions, and Pricing. The main heading is 'Tasks', accompanied by a green toggle switch. Below this, a paragraph states: 'Hugging Face is the home for all Machine Learning tasks. Here you can find what you need to get started with a task: demos, use cases, models, datasets, and more!'. A section titled 'Natural Language Processing' features six task cards: Fill-Mask (3553 models), Question Answering (2008 models), Sentence Similarity (577 models), Summarization (485 models), Text Classification (8106 models), and Text Generation (5246 models). Each card contains an icon representing the task.

Hugging Face Search models, datasets Models Datasets Spaces Docs Solutions Pricing

Tasks

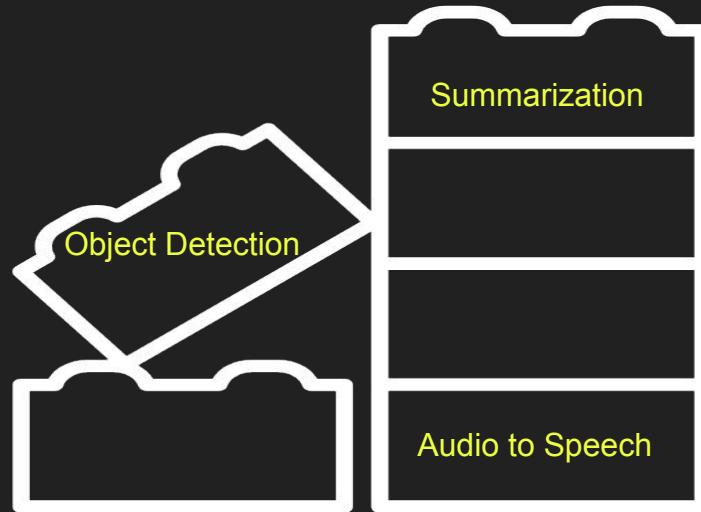
Hugging Face is the home for all Machine Learning tasks. Here you can find what you need to get started with a task: demos, use cases, models, datasets, and more!

Natural Language Processing

Task	Number of Models
Fill-Mask	3553 models
Question Answering	2008 models
Sentence Similarity	577 models
Summarization	485 models
Text Classification	8106 models
Text Generation	5246 models



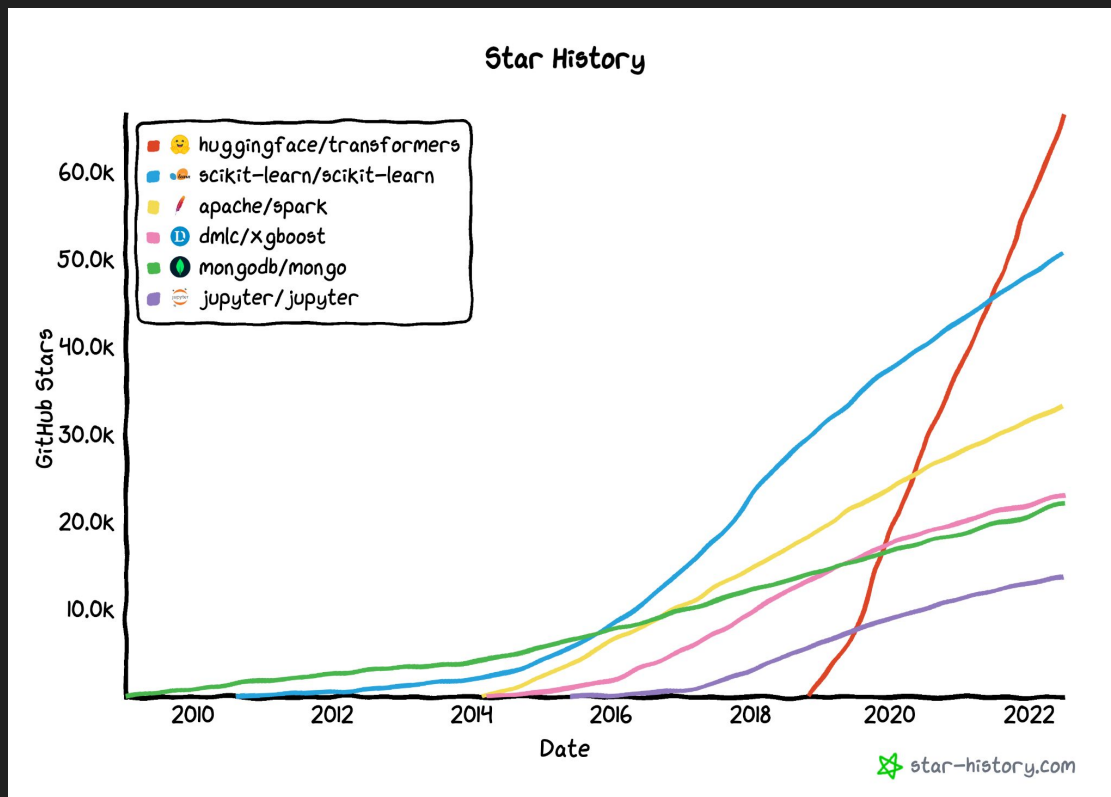
Tasks - Modern Building Blocks of AI



Let's Code with Pre-Trained Models



Rise of Transformers



<https://star-history.com/>



[Pulls](#)[Issues](#)[Marketplace](#)[Explore](#)

huggingface / transformers

[Unwatch](#) 837[Fork](#) 15.4k[Star](#) 66.2k[Public](#)[Code](#)[Issues](#) 373[Pull requests](#) 136[Actions](#)[Projects](#) 24[Wiki](#)[main](#)[Go to file](#)[Add file](#)[Code](#)

About



aarnphm and sgugger feat: add ... 9 hours ago 10,142

.circleci

Fix CI tests hang forever (#17471)

29 days ago

github

Remove render tags (#17897)

2 days ago

🤗 Transformers: State-of-the-art Machine Learning for Pytorch, TensorFlow, and JAX.

huggingface.co/transformers





Hugging Face

🔍 Search models, datasets, u

📁 Models

📁 Datasets

📁 Spaces

📁 Docs

📁 Solutions

Pricing

• Transformers

🔍 Search documentation



V4.20.1



EN



66,859

GET STARTED

🤗 Transformers

Quick tour

Installation

TUTORIALS

Pipelines for inference

Load pretrained instances with
an AutoClass

Preprocess

Fine-tune a pretrained model

Distributed training with 🤗
Accelerate



Transformers

State-of-the-art Machine Learning for PyTorch, TensorFlow and JAX.

🤗 Transformers provides APIs to easily download and train state-of-the-art pretrained models. Using pretrained models can reduce your compute costs, carbon footprint, and save you time from training a model from scratch. The models can be used across different modalities such as:

- 📄 Text: text classification, information extraction, question answering, summarization, translation, and text generation in over 100 languages.
- 🖼️ Images: image classification, object detection, and segmentation.

🤗 Transformers

If you are looking for
support from the Hugging
team

Contents

Supported models

Supported frameworks

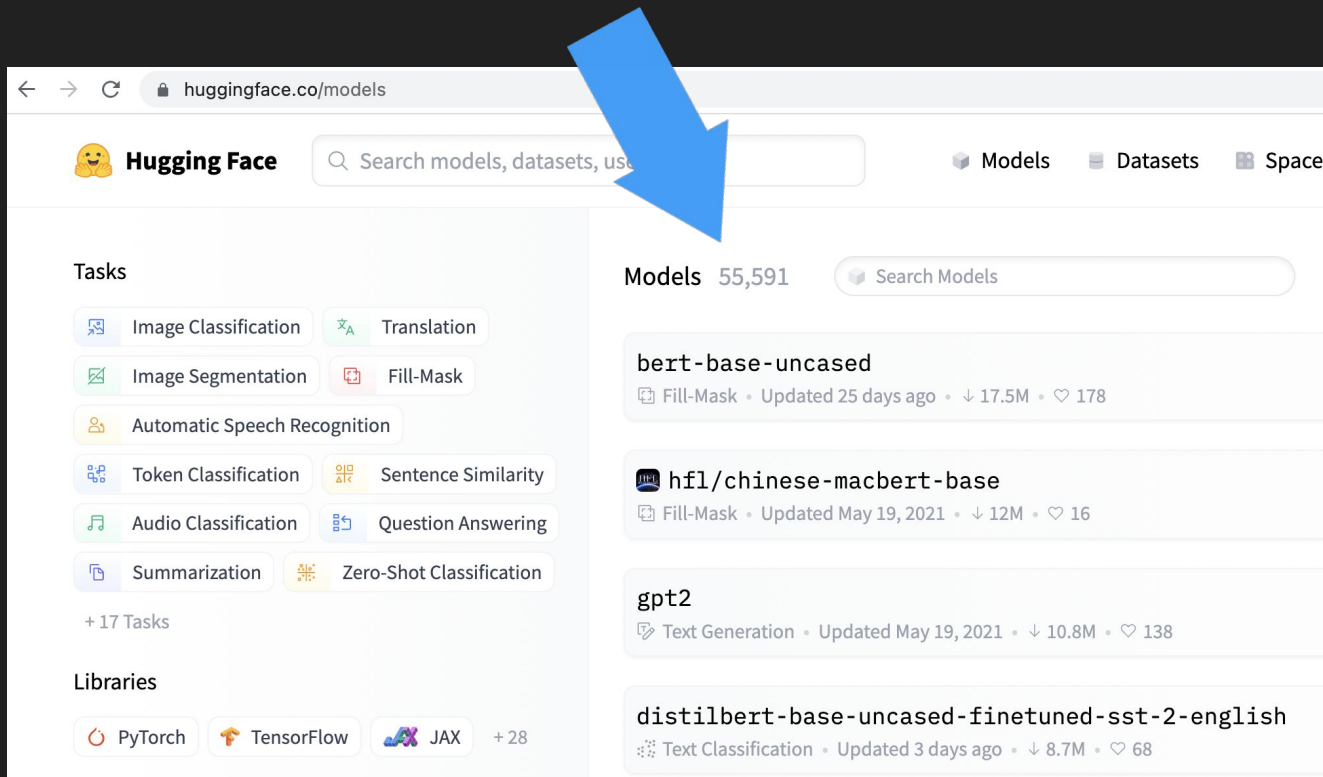


Let's use the transformer's library in the Colab Notebook

Colab Notebook: https://bit.ly/raj_food



Why so many models?



The screenshot shows the Hugging Face website interface. At the top, the URL is `huggingface.co/models`. The Hugging Face logo is on the left, and a search bar is in the center. A large blue arrow points to the search bar. Below the search bar, there are tabs for "Models", "Datasets", and "Spaces". On the left side, there are sections for "Tasks" and "Libraries". The "Tasks" section lists various tasks like Image Classification, Translation, Image Segmentation, Fill-Mask, Automatic Speech Recognition, Token Classification, Sentence Similarity, Audio Classification, Question Answering, Summarization, and Zero-Shot Classification. The "Libraries" section lists PyTorch, TensorFlow, and JAX. On the right side, there is a list of models. The first model is `bert-base-uncased`, which is a Fill-Mask model updated 25 days ago, with 17.5M downloads and 178 likes. The second model is `hfl/chinese-macbert-base`, which is a Fill-Mask model updated May 19, 2021, with 12M downloads and 16 likes. The third model is `gpt2`, which is a Text Generation model updated May 19, 2021, with 10.8M downloads and 138 likes. The fourth model is `distilbert-base-uncased-finetuned-sst-2-english`, which is a Text Classification model updated 3 days ago, with 8.7M downloads and 68 likes.

huggingface.co/models

Hugging Face

Search models, datasets, use

Models Datasets Spaces

Tasks

- Image Classification
- Translation
- Image Segmentation
- Fill-Mask
- Automatic Speech Recognition
- Token Classification
- Sentence Similarity
- Audio Classification
- Question Answering
- Summarization
- Zero-Shot Classification

+ 17 Tasks

Libraries

- PyTorch
- TensorFlow
- JAX

+ 28

Models 55,591

Search Models

bert-base-uncased

Fill-Mask • Updated 25 days ago • ↓ 17.5M • ♥ 178

hfl/chinese-macbert-base

Fill-Mask • Updated May 19, 2021 • ↓ 12M • ♥ 16

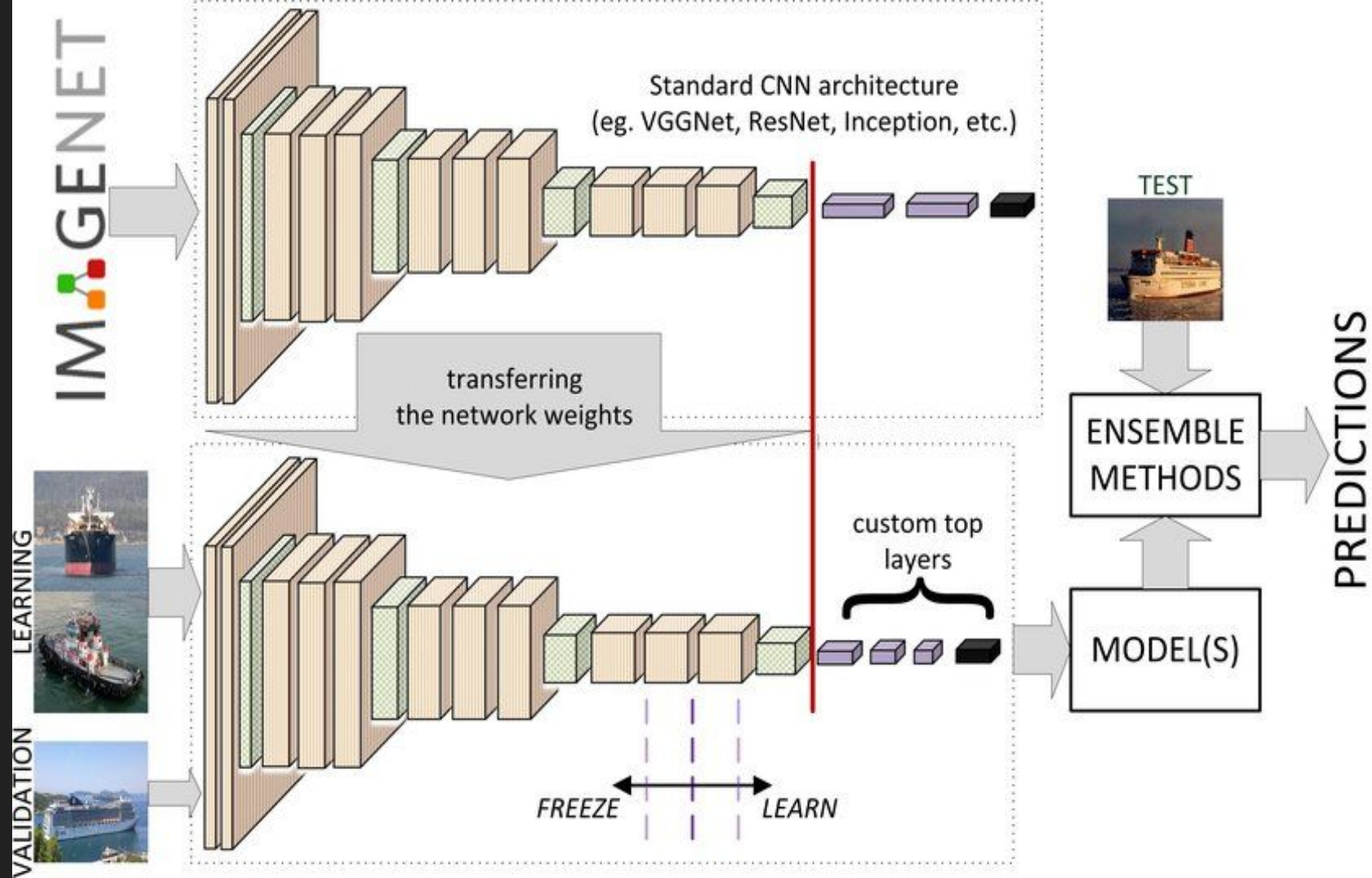
gpt2

Text Generation • Updated May 19, 2021 • ↓ 10.8M • ♥ 138

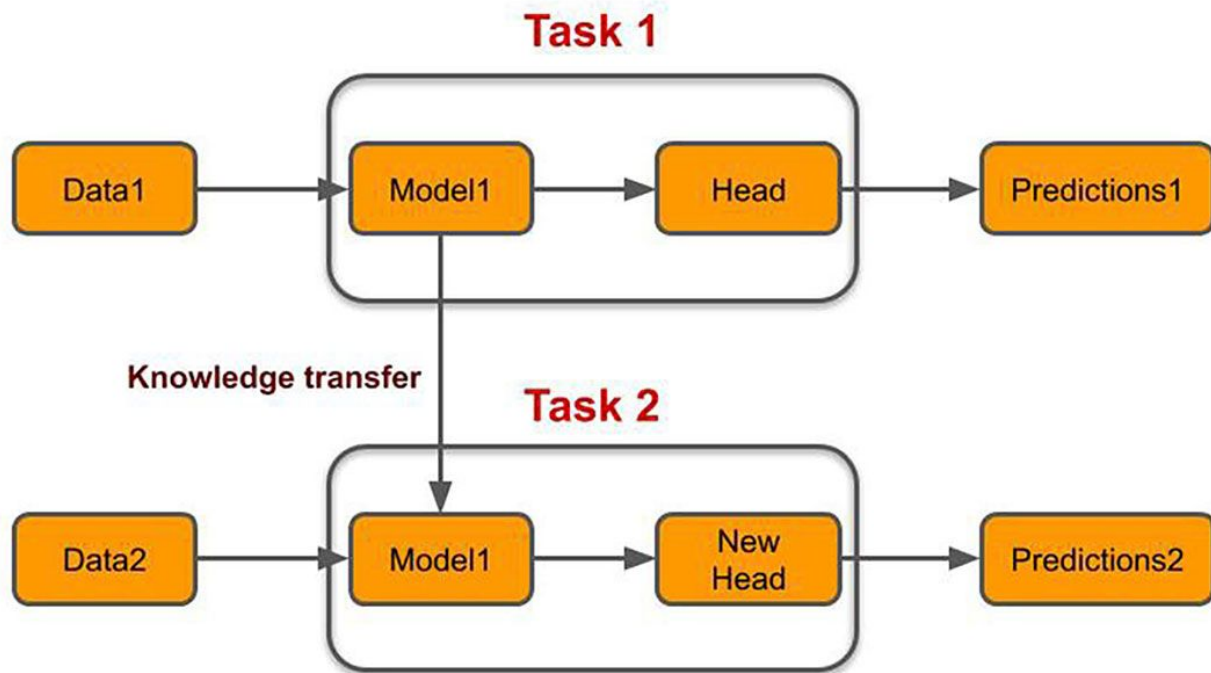
distilbert-base-uncased-finetuned-sst-2-english

Text Classification • Updated 3 days ago • ↓ 8.7M • ♥ 68





Transfer Learning



*"Transfer Learning
will be the next driver
of ML success!"*

- Andrew Ng



Show an example of a fine tuned model

Finbert Forward Looking Statements: <https://huggingface.co/yiyanghkust/finbert-fls>



You can search, test, code up, and
build a prediction pipeline for many
image, text, and audio tasks!!!



Web Apps for AI



Reasons You Should Build a ML Demo



1. Accessibility: get your models used

2. Understand real-world limitations of your model

3. It's easy! (👉 gradio.dev)



Let's build a web app in the colab notebook

Colab Notebook: https://bit.ly/raj_food

Spaces/App:

https://huggingface.co/spaces/rajistics/Indian_food_translator



Accomplished

- Use pretrained models in AI web apps
- Notebook is at: https://bit.ly/raj_food
- Please share feedback with me 🙏

Rajiv Shah

raj@hf.com

@rajistics (LinkedIn, Tik Tok, Twitter, Medium)

