🤗 Transformers

build passing | license Apache-2.0 | website online | release v4.21.2

Contributor Covenant v2.0 adopted | DOI 10.5281/zenodo.7019769

為 Jax、PyTorch 以及 TensorFlow 打造的先進自然語言處理函式庫

🤗 Transformers 提供了數以千計的預訓練模型，支援 100 多種語言的文本分類、資訊擷取、問答、摘要、翻譯、文本生成。它的宗旨是讓最先進的 NLP 技術人人易用。

🤗 Transformers 提供了便於快速下載和使用的API，讓你可以將預訓練模型用在給定文本、在你的資料集上微調然後經由 model hub 與社群共享。同時，每個定義的 Python 模組架構均完全獨立，方便修改和快速研究實驗。

🤗 Transformers 支援三個最熱門的深度學習函式庫： Jax, PyTorch 以及 TensorFlow — 並與之完美整合。你可以直接使用其中一個框架訓練你的模型，然後用另一個載入和推論。

線上Demo

你可以直接在 model hub 上測試大多數的模型。我們也提供了 私有模型託管、模型版本管理以及推論API。

這裡是一些範例：

- 用 BERT 做遮蓋填詞
- 用 Electra 做專有名詞辨識
- 用 GPT-2 做文本生成
- 用 RoBERTa 做自然語言推論
- 用 BART 做文本摘要
- 用 DistilBERT 做問答
- 用 T5 做翻譯

Write With Transformer，由 Hugging Face 團隊所打造，是一個文本生成的官方 demo。

如果你在尋找由 Hugging Face 團隊所提供的客製化支援服務

快速上手

我們為快速使用模型提供了 `pipeline` API。 Pipeline 包含了預訓練模型和對應的文本預處理。下面是一個快速使用 pipeline 去判斷正負面情緒的例子：

```
>>> from transformers import pipeline

# 使用情緒分析 pipeline
>>> classifier = pipeline('sentiment-analysis')
>>> classifier('We are very happy to introduce pipeline to the transformers repository.')
[{'label': 'POSITIVE', 'score': 0.9996980428695679}]
```

第二行程式碼下載並快取 pipeline 使用的預訓練模型，而第三行程式碼則在給定的文本上進行了評估。這裡的答案"正面" (positive) 具有 99.97% 的信賴度。

許多的 NLP 任務都有隨選即用的預訓練 `pipeline`。例如，我們可以輕鬆地從給定文本中擷取問題答案：

```
>>> from transformers import pipeline

# 使用問答 pipeline
>>> question_answerer = pipeline('question-answering')
>>> question_answerer({
...     'question': 'What is the name of the repository ?',
...     'context': 'Pipeline has been included in the huggingface/transformers repository'
... })
{'score': 0.30970096588134766, 'start': 34, 'end': 58, 'answer': 'huggingface/transformers'}
```

除了提供問題解答，預訓練模型還提供了對應的信賴度分數以及解答在 tokenized 後的文本中開始和結束的位置。你可以從這個教學了解更多 `pipeline` API支援的任務。

要在你的任務中下載和使用任何預訓練模型很簡單，只需三行程式碼。這裡是 PyTorch 版的範例：

```
>>> from transformers import AutoTokenizer, AutoModel

>>> tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")
>>> model = AutoModel.from_pretrained("bert-base-uncased")

>>> inputs = tokenizer("Hello world!", return_tensors="pt")
>>> outputs = model(**inputs)
```

這裡是對應的 TensorFlow 程式碼:

```
>>> from transformers import AutoTokenizer, TFAutoModel

>>> tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")
>>> model = TFAutoModel.from_pretrained("bert-base-uncased")

>>> inputs = tokenizer("Hello world!", return_tensors="tf")
>>> outputs = model(**inputs)
```

Tokenizer 為所有的預訓練模型提供了預處理, 並可以直接轉換單一字串 (比如上面的例子) 或串列 (list)。它會輸出一個的字典 (dict) 讓你可以在下游程式碼裡使用或直接藉由 `**` 運算式傳給模型。

模型本身是一個常規的 [Pytorch `nn.Module`] 或 [TensorFlow `tf.keras.Model`] (取決於你的後端), 可依常規方式使用。 [這個教學]解釋了如何將這樣的模型整合到一般的 PyTorch 或 TensorFlow 訓練迴圈中, 或是如何使用我們的 `Trainer` API 在一個新的資料集上快速進行微調。

為什麼要用 transformers?

1. 便於使用的先進模型:

   - NLU 和 NLG 上性能卓越
   - 對教學和實作友好且低門檻
   - 高度抽象, 使用者只須學習 3 個類別
   - 對所有模型使用的制式化API

2. 更低的運算成本, 更少的碳排放:

   - 研究人員可以分享預訓練的模型而非從頭開始訓練
   - 工程師可以減少計算時間以及生產成本
   - 數十種模型架構、兩千多個預訓練模型、100多種語言支援

3. 對於模型生命週期的每一個部分都面面俱到:

   - 訓練先進的模型, 只需 3 行程式碼
   - 模型可以在不同深度學習框架之間任意轉換
   - 為訓練、評估和生產選擇最適合的框架, 並完美銜接

4. 為你的需求輕鬆客製化專屬模型和範例:

   - 我們為每種模型架構提供了多個範例來重現原論文結果
   - 一致的模型內部架構
   - 模型檔案可單獨使用, 便於修改和快速實驗

什麼情況下我不該用 transformers?

- 本函式庫並不是模組化的神經網絡工具箱。模型文件中的程式碼並未做額外的抽象封裝, 以便研究人員快速地翻閱及修改程式碼, 而不會深陷複雜的類別包裝之中。

- `Trainer` API 並非相容任何模型，它只為本函式庫中的模型最佳化。對於一般的機器學習用途，請使用其他函式庫。
- 儘管我們已盡力而為，[examples 目錄](#)中的腳本也僅為範例而已。對於特定問題，它們並不一定隨選即用，可能需要修改幾行程式碼以符合需求。

安裝

使用 pip

這個 Repository 已在 Python 3.6+、Flax 0.3.2+、PyTorch 1.3.1+ 和 TensorFlow 2.3+ 下經過測試。

你可以在[虛擬環境](#)中安裝 😊 Transformers。如果你還不熟悉 Python 的虛擬環境，請閱此[使用者指引](#)。

首先，用你打算使用的版本的 Python 創建一個虛擬環境並進入。

然後，你需要安裝 Flax、PyTorch 或 TensorFlow 其中之一。對於該如何在你使用的平台上安裝這些框架，請參閱 [TensorFlow 安裝頁面](#)、[PyTorch 安裝頁面](#) 或 [Flax 安裝頁面](#)。

當其中一個後端安裝成功後，😊 Transformers 可依此安裝：

```
pip install transformers
```

如果你想要試試範例或者想在正式發布前使用最新開發中的程式碼，你必須[從原始碼安裝](#)。

使用 conda

自 Transformers 4.0.0 版始，我們有了一個 conda channel：`huggingface`。

😊 Transformers 可以藉由 conda 依此安裝：

```
conda install -c huggingface transformers
```

要藉由 conda 安裝 Flax、PyTorch 或 TensorFlow 其中之一，請參閱它們各自安裝頁面的說明。

模型架構

😊 Transformers 支援的[所有的模型檢查點](#)，由[使用者](#)和[組織](#)上傳，均與 huggingface.co [model hub](#) 完美結合。

目前的檢查點數量：`models 68,379`

😊 Transformers 目前支援以下的架構（模型概覽請參閱[這裡](#)）：

1. **[ALBERT](#)** (from Google Research and the Toyota Technological Institute at Chicago) released with the paper [ALBERT: A Lite BERT for Self-supervised Learning of Language Representations](#), by Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut.
2. **[BART](#)** (from Facebook) released with the paper [BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension](#) by Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov and Luke Zettlemoyer.
3. **[BARThez](#)** (from École polytechnique) released with the paper [BARThez: a Skilled Pretrained French Sequence-to-Sequence Model](#) by Moussa Kamal Eddine, Antoine J.-P. Tixier, Michalis Vazirgiannis.
4. **[BARTpho](#)** (from VinAI Research) released with the paper [BARTpho: Pre-trained Sequence-to-Sequence Models for Vietnamese](#) by Nguyen Luong Tran, Duong Minh Le and Dat Quoc Nguyen.
5. **[BEiT](#)** (from Microsoft) released with the paper [BEiT: BERT Pre-Training of Image Transformers](#) by Hangbo Bao, Li Dong, Furu Wei.

6. **BERT** (from Google) released with the paper [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#) by Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova.
7. **BERT For Sequence Generation** (from Google) released with the paper [Leveraging Pre-trained Checkpoints for Sequence Generation Tasks](#) by Sascha Rothe, Shashi Narayan, Aliaksei Severyn.
8. **BERTweet** (from VinAI Research) released with the paper [BERTweet: A pre-trained language model for English Tweets](#) by Dat Quoc Nguyen, Thanh Vu and Anh Tuan Nguyen.
9. **BigBird-Pegasus** (from Google Research) released with the paper [Big Bird: Transformers for Longer Sequences](#) by Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, Amr Ahmed.
10. **BigBird-RoBERTa** (from Google Research) released with the paper [Big Bird: Transformers for Longer Sequences](#) by Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, Amr Ahmed.
11. **Blenderbot** (from Facebook) released with the paper [Recipes for building an open-domain chatbot](#) by Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, Jason Weston.
12. **BlenderbotSmall** (from Facebook) released with the paper [Recipes for building an open-domain chatbot](#) by Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, Jason Weston.
13. **BORT** (from Alexa) released with the paper [Optimal Subarchitecture Extraction For BERT](#) by Adrian de Wynter and Daniel J. Perry.
14. **ByT5** (from Google Research) released with the paper [ByT5: Towards a token-free future with pre-trained byte-to-byte models](#) by Linting Xue, Aditya Barua, Noah Constant, Rami Al-Rfou, Sharan Narang, Mihir Kale, Adam Roberts, Colin Raffel.
15. **CamemBERT** (from Inria/Facebook/Sorbonne) released with the paper [CamemBERT: a Tasty French Language Model](#) by Louis Martin*, Benjamin Muller*, Pedro Javier Ortiz Suárez*, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah and Benoît Sagot.
16. **CANINE** (from Google Research) released with the paper [CANINE: Pre-training an Efficient Tokenization-Free Encoder for Language Representation](#) by Jonathan H. Clark, Dan Garrette, Iulia Turc, John Wieting.
17. **CLIP** (from OpenAI) released with the paper [Learning Transferable Visual Models From Natural Language Supervision](#) by Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever.
18. **ConvBERT** (from YituTech) released with the paper [ConvBERT: Improving BERT with Span-based Dynamic Convolution](#) by Zihang Jiang, Weihao Yu, Daquan Zhou, Yunpeng Chen, Jiashi Feng, Shuicheng Yan.
19. **ConvNeXT** (from Facebook AI) released with the paper [A ConvNet for the 2020s](#) by Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie.
20. **CPM** (from Tsinghua University) released with the paper [CPM: A Large-scale Generative Chinese Pre-trained Language Model](#) by Zhengyan Zhang, Xu Han, Hao Zhou, Pei Ke, Yuxian Gu, Deming Ye, Yujia Qin, Yusheng Su, Haozhe Ji, Jian Guan, Fanchao Qi, Xiaozhi Wang, Yanan Zheng, Guoyang Zeng, Huanqi Cao, Shengqi Chen, Daixuan Li, Zhenbo Sun, Zhiyuan Liu, Minlie Huang, Wentao Han, Jie Tang, Juanzi Li, Xiaoyan Zhu, Maosong Sun.
21. **CTRL** (from Salesforce) released with the paper [CTRL: A Conditional Transformer Language Model for Controllable Generation](#) by Nitish Shirish Keskar*, Bryan McCann*, Lav R. Varshney, Caiming Xiong and Richard Socher.
22. **Data2Vec** (from Facebook) released with the paper [Data2Vec: A General Framework for Self-supervised Learning in Speech, Vision and Language](#) by Alexei Baevski, Wei-Ning Hsu, Qiantong Xu, Arun Babu, Jiatao Gu, Michael Auli.
23. **DeBERTa** (from Microsoft) released with the paper [DeBERTa: Decoding-enhanced BERT with Disentangled Attention](#) by Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen.
24. **DeBERTa-v2** (from Microsoft) released with the paper [DeBERTa: Decoding-enhanced BERT with Disentangled Attention](#) by Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen.
25. **Decision Transformer** (from Berkeley/Facebook/Google) released with the paper [Decision Transformer: Reinforcement Learning via Sequence Modeling](#) by Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, Igor Mordatch.
26. **DeiT** (from Facebook) released with the paper [Training data-efficient image transformers & distillation through attention](#) by Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou.

27. **DETR** (from Facebook) released with the paper [End-to-End Object Detection with Transformers](#) by Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko.
28. **DialoGPT** (from Microsoft Research) released with the paper [DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation](#) by Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, Bill Dolan.
29. **DistilBERT** (from HuggingFace), released together with the paper [DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter](#) by Victor Sanh, Lysandre Debut and Thomas Wolf. The same method has been applied to compress GPT2 into [DistilGPT2](#), RoBERTa into [DistilRoBERTa](#), Multilingual BERT into [DistilmBERT](#) and a German version of DistilBERT.
30. **DiT** (from Microsoft Research) released with the paper [DiT: Self-supervised Pre-training for Document Image Transformer](#) by Junlong Li, Yiheng Xu, Tengchao Lv, Lei Cui, Cha Zhang, Furu Wei.
31. **DPR** (from Facebook) released with the paper [Dense Passage Retrieval for Open-Domain Question Answering](#) by Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih.
32. **DPT** (from Intel Labs) released with the paper [Vision Transformers for Dense Prediction](#) by René Ranftl, Alexey Bochkovskiy, Vladlen Koltun.
33. **ELECTRA** (from Google Research/Stanford University) released with the paper [ELECTRA: Pre-training text encoders as discriminators rather than generators](#) by Kevin Clark, Minh-Thang Luong, Quoc V. Le, Christopher D. Manning.
34. **EncoderDecoder** (from Google Research) released with the paper [Leveraging Pre-trained Checkpoints for Sequence Generation Tasks](#) by Sascha Rothe, Shashi Narayan, Aliaksei Severyn.
35. **FlauBERT** (from CNRS) released with the paper [FlauBERT: Unsupervised Language Model Pre-training for French](#) by Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoît Crabbé, Laurent Besacier, Didier Schwab.
36. **FNet** (from Google Research) released with the paper [FNet: Mixing Tokens with Fourier Transforms](#) by James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, Santiago Ontanon.
37. **Funnel Transformer** (from CMU/Google Brain) released with the paper [Funnel-Transformer: Filtering out Sequential Redundancy for Efficient Language Processing](#) by Zihang Dai, Guokun Lai, Yiming Yang, Quoc V. Le.
38. **GLPN** (from KAIST) released with the paper [Global-Local Path Networks for Monocular Depth Estimation with Vertical CutDepth](#) by Doyeon Kim, Woonghyun Ga, Pyungwhan Ahn, Donggyu Joo, Sehwan Chun, Junmo Kim.
39. **GPT** (from OpenAI) released with the paper [Improving Language Understanding by Generative Pre-Training](#) by Alec Radford, Karthik Narasimhan, Tim Salimans and Ilya Sutskever.
40. **GPT Neo** (from EleutherAI) released in the repository [EleutherAI/gpt-neo](#) by Sid Black, Stella Biderman, Leo Gao, Phil Wang and Connor Leahy.
41. **GPT-2** (from OpenAI) released with the paper [Language Models are Unsupervised Multitask Learners](#) by Alec Radford*, Jeffrey Wu*, Rewon Child, David Luan, Dario Amodei** and Ilya Sutskever**.
42. **GPT-J** (from EleutherAI) released with the paper [kingoflolz/mesh-transformer-jax](#) by Ben Wang and Aran Komatsuzaki.
43. **Hubert** (from Facebook) released with the paper [HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units](#) by Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, Abdelrahman Mohamed.
44. **I-BERT** (from Berkeley) released with the paper [I-BERT: Integer-only BERT Quantization](#) by Sehoon Kim, Amir Gholami, Zhewei Yao, Michael W. Mahoney, Kurt Keutzer.
45. **ImageGPT** (from OpenAI) released with the paper [Generative Pretraining from Pixels](#) by Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, Ilya Sutskever.
46. **LayoutLM** (from Microsoft Research Asia) released with the paper [LayoutLM: Pre-training of Text and Layout for Document Image Understanding](#) by Yiheng Xu, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming Zhou.
47. **LayoutLMv2** (from Microsoft Research Asia) released with the paper [LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding](#) by Yang Xu, Yiheng Xu, Tengchao Lv, Lei Cui, Furu Wei, Guoxin Wang, Yijuan Lu, Dinei Florencio, Cha Zhang, Wanxiang Che, Min Zhang, Lidong Zhou.
48. **LayoutXLM** (from Microsoft Research Asia) released with the paper [LayoutXLM: Multimodal Pre-training for Multilingual Visually-rich Document Understanding](#) by Yiheng Xu, Tengchao Lv, Lei Cui, Guoxin Wang, Yijuan Lu, Dinei Florencio, Cha Zhang, Furu Wei.

49. **LED** (from AllenAI) released with the paper [Longformer: The Long-Document Transformer](#) by Iz Beltagy, Matthew E. Peters, Arman Cohan.
50. **Longformer** (from AllenAI) released with the paper [Longformer: The Long-Document Transformer](#) by Iz Beltagy, Matthew E. Peters, Arman Cohan.
51. **LUKE** (from Studio Ousia) released with the paper [LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention](#) by Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto.
52. **LXMERT** (from UNC Chapel Hill) released with the paper [LXMERT: Learning Cross-Modality Encoder Representations from Transformers for Open-Domain Question Answering](#) by Hao Tan and Mohit Bansal.
53. **M2M100** (from Facebook) released with the paper [Beyond English-Centric Multilingual Machine Translation](#) by Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, Armand Joulin.
54. **MarianMT** Machine translation models trained using [OPUS](#) data by Jörg Tiedemann. The [Marian Framework](#) is being developed by the Microsoft Translator Team.
55. **MaskFormer** (from Meta and UIUC) released with the paper [Per-Pixel Classification is Not All You Need for Semantic Segmentation](#) by Bowen Cheng, Alexander G. Schwing, Alexander Kirillov
56. **MBart** (from Facebook) released with the paper [Multilingual Denoising Pre-training for Neural Machine Translation](#) by Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, Luke Zettlemoyer.
57. **MBart-50** (from Facebook) released with the paper [Multilingual Translation with Extensible Multilingual Pretraining and Finetuning](#) by Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, Angela Fan.
58. **Megatron-BERT** (from NVIDIA) released with the paper [Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism](#) by Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper and Bryan Catanzaro.
59. **Megatron-GPT2** (from NVIDIA) released with the paper [Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism](#) by Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper and Bryan Catanzaro.
60. **mLUKE** (from Studio Ousia) released with the paper [mLUKE: The Power of Entity Representations in Multilingual Pretrained Language Models](#) by Ryokan Ri, Ikuya Yamada, and Yoshimasa Tsuruoka.
61. **MPNet** (from Microsoft Research) released with the paper [MPNet: Masked and Permuted Pre-training for Language Understanding](#) by Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, Tie-Yan Liu.
62. **MT5** (from Google AI) released with the paper [mT5: A massively multilingual pre-trained text-to-text transformer](#) by Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, Colin Raffel.
63. **Nyströmformer** (from the University of Wisconsin - Madison) released with the paper [Nyströmformer: A Nyström-Based Algorithm for Approximating Self-Attention](#) by Yunyang Xiong, Zhanpeng Zeng, Rudrasis Chakraborty, Mingxing Tan, Glenn Fung, Yin Li, Vikas Singh.
64. **Pegasus** (from Google) released with the paper [PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization](#) by Jingqing Zhang, Yao Zhao, Mohammad Saleh and Peter J. Liu.
65. **Perceiver IO** (from Deepmind) released with the paper [Perceiver IO: A General Architecture for Structured Inputs & Outputs](#) by Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, João Carreira.
66. **PhoBERT** (from VinAI Research) released with the paper [PhoBERT: Pre-trained language models for Vietnamese](#) by Dat Quoc Nguyen and Anh Tuan Nguyen.
67. **PLBart** (from UCLA NLP) released with the paper [Unified Pre-training for Program Understanding and Generation](#) by Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray, Kai-Wei Chang.
68. **PoolFormer** (from Sea AI Labs) released with the paper [MetaFormer is Actually What You Need for Vision](#) by Yu, Weihao and Luo, Mi and Zhou, Pan and Si, Chenyang and Zhou, Yichen and Wang, Xinchao and Feng, Jiashi and Yan, Shuicheng.
69. **ProphetNet** (from Microsoft Research) released with the paper [ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training](#) by Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen,

Ruofei Zhang and Ming Zhou.

70. **QDQBert** (from NVIDIA) released with the paper [Integer Quantization for Deep Learning Inference: Principles and Empirical Evaluation](#) by Hao Wu, Patrick Judd, Xiaojie Zhang, Mikhail Isaev and Paulius Micikevicius.

71. **REALM** (from Google Research) released with the paper [REALM: Retrieval-Augmented Language Model Pre-Training](#) by Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat and Ming-Wei Chang.

72. **Reformer** (from Google Research) released with the paper [Reformer: The Efficient Transformer](#) by Nikita Kitaev, Łukasz Kaiser, Anselm Levskaya.

73. **RegNet** (from META Research) released with the paper [Designing Network Design Space](#) by Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, Piotr Dollár.

74. **RemBERT** (from Google Research) released with the paper [Rethinking embedding coupling in pre-trained language models](#) by Hyung Won Chung, Thibault Févry, Henry Tsai, M. Johnson, Sebastian Ruder.

75. **ResNet** (from Microsoft Research) released with the paper [Deep Residual Learning for Image Recognition](#) by Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.

76. **RoBERTa** (from Facebook), released together with the paper a [Robustly Optimized BERT Pretraining Approach](#) by Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov.

77. **RoFormer** (from ZhuiyiTechnology), released together with the paper a [RoFormer: Enhanced Transformer with Rotary Position Embedding](#) by Jianlin Su and Yu Lu and Shengfeng Pan and Bo Wen and Yunfeng Liu.

78. **SegFormer** (from NVIDIA) released with the paper [SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers](#) by Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Ping Luo.

79. **SEW** (from ASAPP) released with the paper [Performance-Efficiency Trade-offs in Unsupervised Pre-training for Speech Recognition](#) by Felix Wu, Kwangyoun Kim, Jing Pan, Kyu Han, Kilian Q. Weinberger, Yoav Artzi.

80. **SEW-D** (from ASAPP) released with the paper [Performance-Efficiency Trade-offs in Unsupervised Pre-training for Speech Recognition](#) by Felix Wu, Kwangyoun Kim, Jing Pan, Kyu Han, Kilian Q. Weinberger, Yoav Artzi.

81. **SpeechToTextTransformer** (from Facebook), released together with the paper [fairseq S2T: Fast Speech-to-Text Modeling with fairseq](#) by Changhan Wang, Yun Tang, Xutai Ma, Anne Wu, Dmytro Okhonko, Juan Pino.

82. **SpeechToTextTransformer2** (from Facebook) released with the paper [Large-Scale Self- and Semi-Supervised Learning for Speech Translation](#) by Changhan Wang, Anne Wu, Juan Pino, Alexei Baevski, Michael Auli, Alexis Conneau.

83. **Splinter** (from Tel Aviv University) released with the paper [Few-Shot Question Answering by Pretraining Span Selection](#) by Ori Ram, Yuval Kirstain, Jonathan Berant, Amir Globerson, Omer Levy.

84. **SqueezeBert** (from Berkeley) released with the paper [SqueezeBERT: What can computer vision teach NLP about efficient neural networks?](#) by Forrest N. Iandola, Albert E. Shaw, Ravi Krishna, and Kurt W. Keutzer.

85. **Swin Transformer** (from Microsoft) released with the paper [Swin Transformer: Hierarchical Vision Transformer using Shifted Windows](#) by Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo.

86. **T5** (from Google AI) released with the paper [Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer](#) by Colin Raffel and Noam Shazeer and Adam Roberts and Katherine Lee and Sharan Narang and Michael Matena and Yanqi Zhou and Wei Li and Peter J. Liu.

87. **T5v1.1** (from Google AI) released with the paper [google-research/text-to-text-transfer-transformer](#) by Colin Raffel and Noam Shazeer and Adam Roberts and Katherine Lee and Sharan Narang and Michael Matena and Yanqi Zhou and Wei Li and Peter J. Liu.

88. **TAPAS** (from Google AI) released with the paper [TAPAS: Weakly Supervised Table Parsing via Pre-training](#) by Jonathan Herzig, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno and Julian Martin Eisenschlos.

89. **Transformer-XL** (from Google/CMU) released with the paper [Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context](#) by Zihang Dai*, Zhilin Yang*, Yiming Yang, Jaime Carbonell, Quoc V. Le, Ruslan Salakhutdinov.

90. **TrOCR** (from Microsoft) released with the paper [TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models](#) by Minghao Li, Tengchao Lv, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei.

91. **UniSpeech** (from Microsoft Research) released with the paper [UniSpeech: Unified Speech Representation Learning with Labeled and Unlabeled Data](#) by Chengyi Wang, Yu Wu, Yao Qian, Kenichi Kumatani, Shujie Liu, Furu Wei, Michael Zeng, Xuedong Huang.

92. **UniSpeechSat** (from Microsoft Research) released with the paper [UNISPEECH-SAT: UNIVERSAL SPEECH REPRESENTATION LEARNING WITH SPEAKER AWARE PRE-TRAINING](#) by Sanyuan Chen, Yu Wu, Chengyi Wang, Zhengyang Chen, Zhuo Chen, Shujie Liu, Jian Wu, Yao Qian, Furu Wei, Jinyu Li, Xiangzhan Yu.

93. **VAN** (from Tsinghua University and Nankai University) released with the paper [Visual Attention Network](#) by Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, Shi-Min Hu.

94. **ViLT** (from NAVER AI Lab/Kakao Enterprise/Kakao Brain) released with the paper [ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision](#) by Wonjae Kim, Bokyung Son, Ildoo Kim.

95. **Vision Transformer (ViT)** (from Google AI) released with the paper [An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale](#) by Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby.

96. **VisualBERT** (from UCLA NLP) released with the paper [VisualBERT: A Simple and Performant Baseline for Vision and Language](#) by Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, Kai-Wei Chang.

97. **ViTMAE** (from Meta AI) released with the paper [Masked Autoencoders Are Scalable Vision Learners](#) by Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick.

98. **Wav2Vec2** (from Facebook AI) released with the paper [wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations](#) by Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, Michael Auli.

99. **Wav2Vec2Phoneme** (from Facebook AI) released with the paper [Simple and Effective Zero-shot Cross-lingual Phoneme Recognition](#) by Qiantong Xu, Alexei Baevski, Michael Auli.

100. **WavLM** (from Microsoft Research) released with the paper [WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing](#) by Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Furu Wei.

101. **XGLM** (From Facebook AI) released with the paper [Few-shot Learning with Multilingual Language Models](#) by Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, Xian Li.

102. **XLM** (from Facebook) released together with the paper [Cross-lingual Language Model Pretraining](#) by Guillaume Lample and Alexis Conneau.

103. **XLM-ProphetNet** (from Microsoft Research) released with the paper [ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training](#) by Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang and Ming Zhou.

104. **XLM-RoBERTa** (from Facebook AI), released together with the paper [Unsupervised Cross-lingual Representation Learning at Scale](#) by Alexis Conneau*, Kartikay Khandelwal*, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer and Veselin Stoyanov.

105. **XLM-RoBERTa-XL** (from Facebook AI) released with the paper [Larger-Scale Transformers for Multilingual Masked Language Modeling](#) by Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman, Alexis Conneau.

106. **XLNet** (from Google/CMU) released with the paper [XLNet: Generalized Autoregressive Pretraining for Language Understanding](#) by Zhilin Yang*, Zihang Dai*, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. Le.

107. **XLS-R** (from Facebook AI) released with the paper [XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale](#) by Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, Michael Auli.

108. **XLSR-Wav2Vec2** (from Facebook AI) released with the paper [Unsupervised Cross-Lingual Representation Learning For Speech Recognition](#) by Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli.

109. **YOSO** (from the University of Wisconsin - Madison) released with the paper [You Only Sample (Almost) by Zhanpeng Zeng, Yunyang Xiong, Sathya N. Ravi, Shailesh Acharya, Glenn Fung, Vikas Singh.

110. 想要貢獻新的模型？我們這裡有一份**詳細指引和模板**來引導你加入新的模型。你可以在 `templates` 目錄中找到它們。記得查看[貢獻指引](#)並在開始寫 PR 前聯繫維護人員或開一個新的 issue 來獲得 feedbacks。

要檢查某個模型是否已有 Flax、PyTorch 或 TensorFlow 的實作，或其是否在🤗 Tokenizers 函式庫中有對應的 tokenizer，敬請參閱[此表](#)。

這些實作均已於多個資料集測試（請參閱範例腳本）並應與原版實作表現相當。你可以在範例文件的[此節](#)中了解實作的細節。

了解更多

章節	描述
[文件](#)	完整的 API 文件和教學
[任務概覽](#)	🤗 Transformers 支援的任務
[預處理教學](#)	使用 `Tokenizer` 來為模型準備資料
[訓練和微調](#)	使用 PyTorch/TensorFlow 的內建的訓練方式或於 `Trainer` API 中使用 🤗 Transformers 提供的模型
[快速上手: 微調和範例腳本](#)	為各種任務提供的範例腳本
[模型分享和上傳](#)	上傳並與社群分享你微調的模型
[遷移](#)	從 `pytorch-transformers` 或 `pytorch-pretrained-bert` 遷移到 🤗 Transformers

引用

我們已將此函式庫的[論文](#)正式發表。如果你使用了 🤗 Transformers 函式庫，可以引用:

```
@inproceedings{wolf-etal-2020-transformers,
    title = "Transformers: State-of-the-Art Natural Language Processing",
    author = "Thomas Wolf and Lysandre Debut and Victor Sanh and Julien Chaumond and
Clement Delangue and Anthony Moi and Pierric Cistac and Tim Rault and Rémi Louf and
Morgan Funtowicz and Joe Davison and Sam Shleifer and Patrick von Platen and Clara Ma and
Yacine Jernite and Julien Plu and Canwen Xu and Teven Le Scao and Sylvain Gugger and
Mariama Drame and Quentin Lhoest and Alexander M. Rush",
    booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural
Language Processing: System Demonstrations",
    month = oct,
    year = "2020",
    address = "Online",
    publisher = "Association for Computational Linguistics",
    url = "https://www.aclweb.org/anthology/2020.emnlp-demos.6",
    pages = "38--45"
}
```