

# Pre-trained Models

We provide a large collection of baselines and checkpoints for NLP pre-trained models.

## How to Load Pretrained Models

### How to Initialize from Checkpoint

**Note:** TF-HUB/Savedmodel is the preferred way to distribute models as it is self-contained. Please consider using TF-HUB for finetuning tasks first.

If you use the [NLP training library](#), you can specify the checkpoint path link directly when launching your job. For example, to initialize the model from the checkpoint, you can specify `--`

`params_override=task.init_checkpoint=PATH_TO_INIT_CKPT` as:

```
python3 train.py \  
--params_override=task.init_checkpoint=PATH_TO_INIT_CKPT
```

### How to load TF-HUB SavedModel

Finetuning tasks such as question answering (SQuAD) and sentence prediction (GLUE) support loading a model from TF-HUB. These built-in tasks support a specific `task.hub_module_url` parameter. To set this parameter, replace

`--params_override=task.init_checkpoint=...` with `--`

`params_override=task.hub_module_url=TF_HUB_URL`, like below:

```
python3 train.py \  
--params_override=task.hub_module_url=https://tfhub.dev/tensorflow/bert_en_uncased_L-  
12_H-768_A-12/3
```

## BERT

Public BERT pre-trained models released by the BERT authors.

We released both checkpoints and tf.hub modules as the pretrained models for fine-tuning. They are TF 2.x compatible and are converted from the checkpoints released in TF 1.x official BERT repository [google-research/bert](#) in order to keep consistent with BERT paper.

### Checkpoints

Model	Configuration	Training Data	Checkpoint & Vocabulary	TF-HUB SavedModels
BERT-base uncased English	uncased_L-12_H-768_A-12	Wiki + Books	<a href="#">uncased_L-12_H-768_A-12</a>	<a href="#">BERT-Base, Uncased</a>
BERT-base cased English	cased_L-12_H-768_A-12	Wiki + Books	<a href="#">cased_L-12_H-768_A-12</a>	<a href="#">BERT-Base, Cased</a>
BERT-large uncased English	uncased_L-24_H-1024_A-16	Wiki + Books	<a href="#">uncased_L-24_H-1024_A-16</a>	<a href="#">BERT-Large, Uncased</a>
BERT-large cased English	cased_L-24_H-1024_A-16	Wiki + Books	<a href="#">cased_L-24_H-1024_A-16</a>	<a href="#">BERT-Large, Cased</a>

BERT-large, Uncased (Whole Word Masking)	wwm_uncased_L-24_H-1024_A-16	Wiki + Books	<a href="#">wwm_uncased_L-24_H-1024_A-16</a>	<a href="#">BERT-Large, Uncased (Whole Word Masking)</a>
BERT-large, Cased (Whole Word Masking)	wwm_cased_L-24_H-1024_A-16	Wiki + Books	<a href="#">wwm_cased_L-24_H-1024_A-16</a>	<a href="#">BERT-Large, Cased (Whole Word Masking)</a>
BERT-base Multilingual	multi_cased_L-12_H-768_A-12	Wiki + Books	<a href="#">multi_cased_L-12_H-768_A-12</a>	<a href="#">BERT-Base, Multilingual Cased</a>
BERT-base Chinese	chinese_L-12_H-768_A-12	Wiki + Books	<a href="#">chinese_L-12_H-768_A-12</a>	<a href="#">BERT-Base, Chinese</a>

You may explore more in the TF-Hub BERT collection: <https://tfhub.dev/google/collections/bert/1>

## BERT variants

We also have pretrained BERT models with variants in both network architecture and training methodologies. These models achieve higher downstream accuracy scores.

Model	Configuration	Training Data	TF-HUB SavedModels	Comment
BERT-base talking heads + ggelu	uncased_L-12_H-768_A-12	Wiki + Books	<a href="#">talkheads_ggelu_base</a>	BERT-base trained with <a href="#">talking heads attention</a> and <a href="#">gated GeLU</a> .
BERT-large talking heads + ggelu	uncased_L-24_H-1024_A-16	Wiki + Books	<a href="#">talkheads_ggelu_large</a>	BERT-large trained with <a href="#">talking heads attention</a> and <a href="#">gated GeLU</a> .
LAMBERT-large uncased English	uncased_L-24_H-1024_A-16	Wiki + Books	<a href="#">lambert</a>	BERT trained with LAMB and techniques from RoBERTa.

## ALBERT

The academic paper that describes ALBERT in detail and provides full results on a number of tasks can be found here: <https://arxiv.org/abs/1909.11942>.

We released both checkpoints and tf.hub modules as the pretrained models for fine-tuning. They are TF 2.x compatible and are converted from the ALBERT v2 checkpoints released in the TF 1.x official ALBERT repository [google-research/albert](https://github.com/google-research/albert) in order to be consistent with the ALBERT paper.

Our current released checkpoints are exactly the same as the TF 1.x official ALBERT repository.

## Checkpoints

Model	Training Data	Checkpoint & Vocabulary	TF-HUB SavedModels
ALBERT-base English	Wiki + Books	<a href="#">ALBERT Base</a>	<a href="https://tfhub.dev/tensorflow/albert_en_base/3">https://tfhub.dev/tensorflow/albert_en_base/3</a>
ALBERT-large	Wiki +	<a href="#">ALBERT Large</a>	<a href="https://tfhub.dev/tensorflow/albert_en_large/3">https://tfhub.dev/tensorflow/albert_en_large/3</a>

English	Books		
ALBERT-xlarge English	Wiki + Books	<a href="#">ALBERT_XLarge</a>	<a href="https://tfhub.dev/tensorflow/albert_en_xlarge/3">https://tfhub.dev/tensorflow/albert_en_xlarge/3</a>
ALBERT-xxlarge English	Wiki + Books	<a href="#">ALBERT_XXLarge</a>	<a href="https://tfhub.dev/tensorflow/albert_en_xxlarge/3">https://tfhub.dev/tensorflow/albert_en_xxlarge/3</a>

## ELECTRA

[ELECTRA](#), which stands for "Efficiently Learning an Encoder that Classifies Token Replacements Accurately", is an efficient language pretraining method. In a nutshell, ELECTRA contains two transformer models, one called "generator" and the other called "discriminator". Given a masked sequence, the generator replaces words in masked positions with randomly generated words. The discriminator then takes the corrupted sentence as input and predicts whether each word is replaced by the generator or not. During the pretraining stage, ELECTRA jointly learns two models (i.e., trains the generator using masked language modeling (MLM) task, and trains the discriminator using replaced token detection (RTD) task). At the fine-tuning stage, the generator is discarded and the discriminator is used for downstream tasks (e.g., GLUE and SQuAD tasks).

## Checkpoints

The checkpoints are re-trained with the Electra code in this repository.

Model	Training Data	Checkpoint & Vocabulary
ELECTRA-small English	Wiki + Books	<a href="#">ELECTRA_Small</a> : the vocabulary is the same as BERT uncased English.
ELECTRA-base English	Wiki + Books	<a href="#">ELECTRA_Base</a> : the vocabulary is the same as BERT uncased English.