the special tokens used by the tokenizer:

* [CLS]: Classification token
* [MASK]: Mask token for masked language modeling
* [PAD]: Padding token
* [SEP]: Separator token (for sentence pairs)
* [UNK]: Unknown token (for out-of-vocabulary words)

 the full configuration and vocabulary for a BERT-like tokenizer:

* **Version**: 1.0
* **Truncation**: Max length 128, truncates on the right, uses "LongestFirst" strategy
* **Padding**: Fixed length 128, pads on the right, pad token is [PAD] with ID 0
* **Special tokens**: [PAD] (0), [UNK] (100), [CLS] (101), [SEP] (102), [MASK] (103)
* **Normalizer**: Uses BertNormalizer (lowercases, handles Chinese chars, cleans text)
* **Pre-tokenizer**: BertPreTokenizer
* **Post-processor**: Adds [CLS] at start and [SEP] at end for single sequences, and appropriate tokens for pairs
* **Decoder**: WordPiece with ## as subword prefix
* **Model**: WordPiece with a large vocabulary (IDs and tokens shown), including many [unusedX] tokens for expansion

additional tokenizer configuration:

* **added\_tokens\_decoder**: Maps special token IDs to their properties (content, special status, etc.)
* **clean\_up\_tokenization\_spaces**: false
* **cls\_token**: [CLS]
* **do\_lower\_case**: true
* **mask\_token**: [MASK]
* **model\_max\_length**: 512
* **pad\_token**: [PAD]
* **sep\_token**: [SEP]
* **tokenizer\_class**: DistilBertTokenizer
* **unk\_token**: [UNK]

tracking the state and progress of model training:

* **Training progress**: 1 epoch completed, 213 global steps
* **Logging**: Contains a history of loss, learning rate, grad norm, and step at regular intervals
* **Evaluation**: Last eval loss was 0.1273, with runtime and throughput stats
* **Training parameters**:
  + Max steps: 639
  + Number of epochs: 3
  + Train batch size: 16
  + Logging every 10 steps
  + Save every 500 steps
  + Total FLOPs: 112,862,223,654,912
* **Callbacks and control flags**: Indicate current training state (not stopped, not saving, etc.)