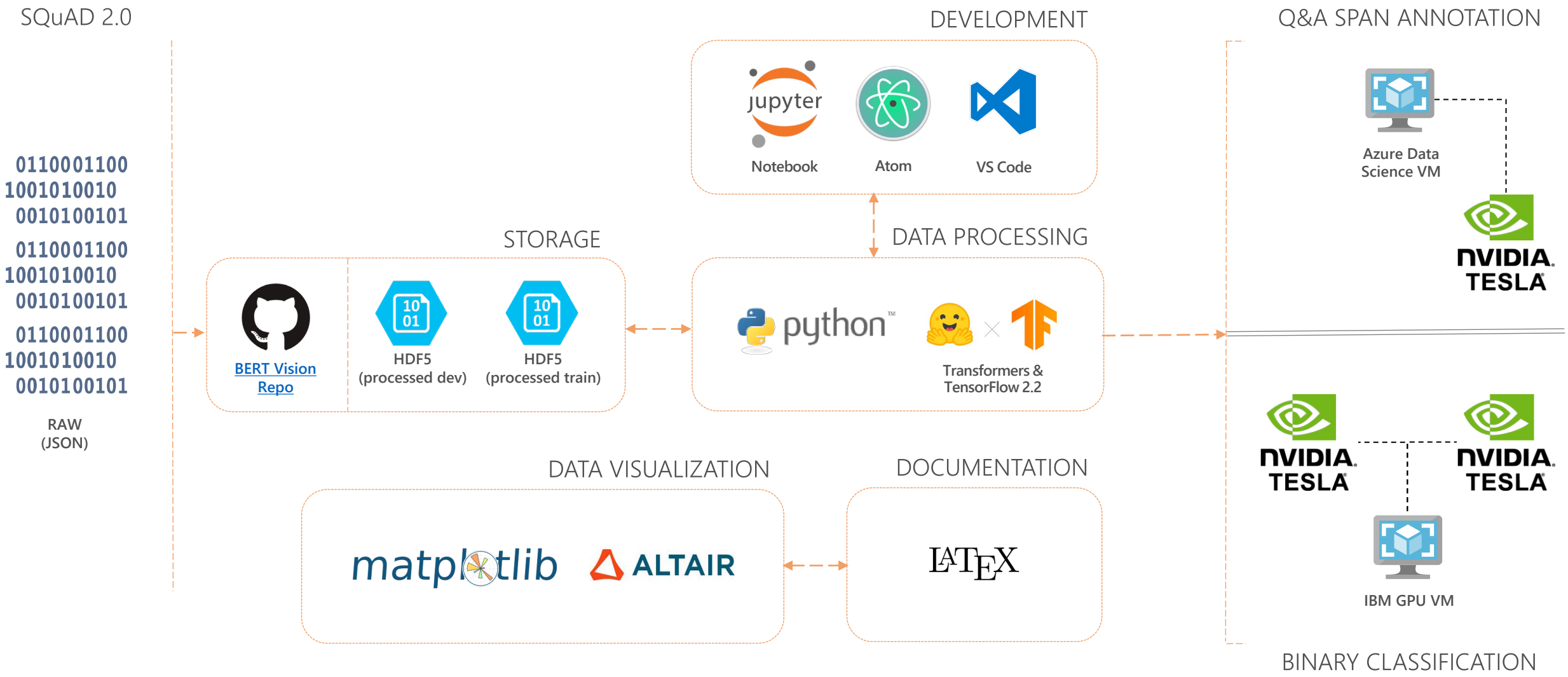
# BERT Vision

*Improving span annotation and classification task performance using parameter-efficient model architectures trained on BERT's hidden state activations.*

Siduo "Stone" Jiang *and* William Casey King *and* Cristopher Benge

# BERT Vision

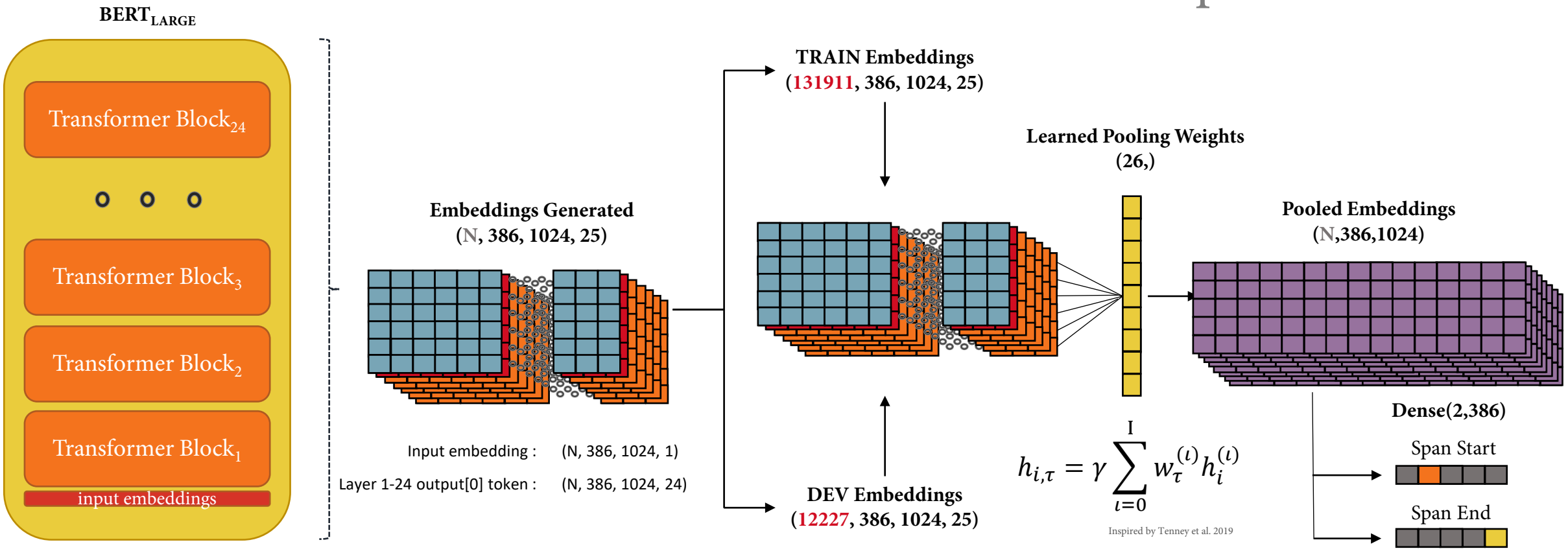
## Development Pipeline Components



Fine-tuned on SQuAD v2 for span annotation task (Q&A)
1/10th epoch – 9/10th epochs, 1 epoch – 6 epochs.

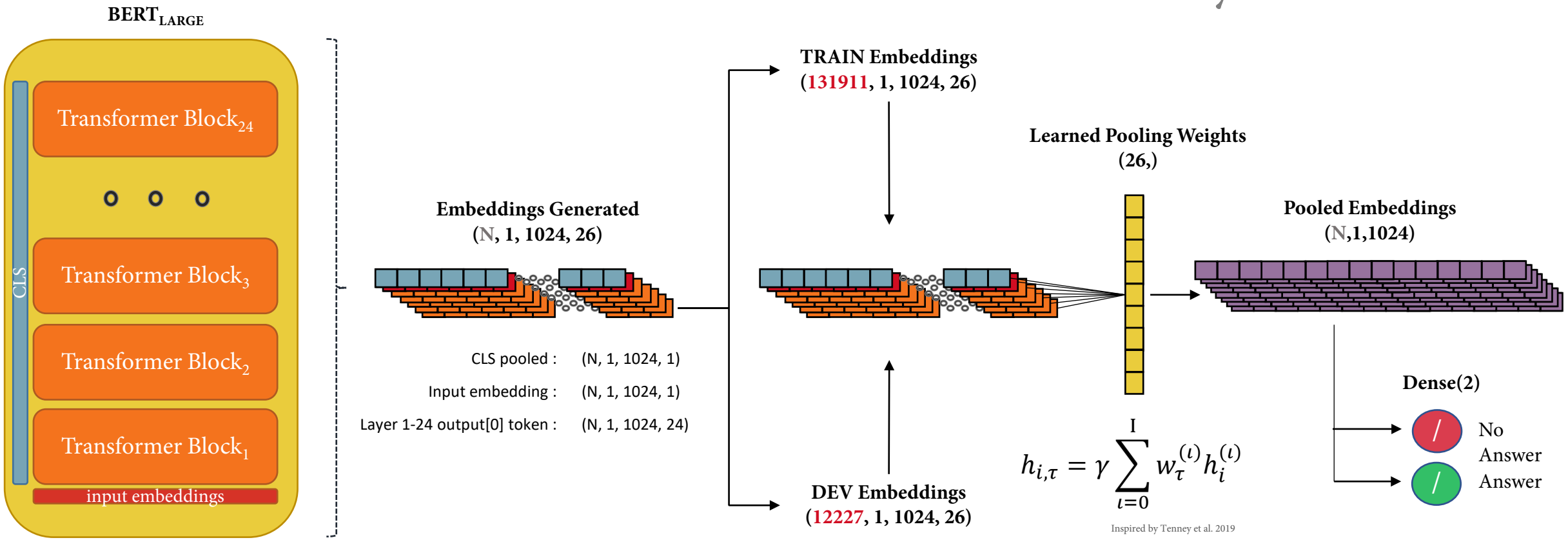
# BERTVision Data Pipeline

## Span Annotation



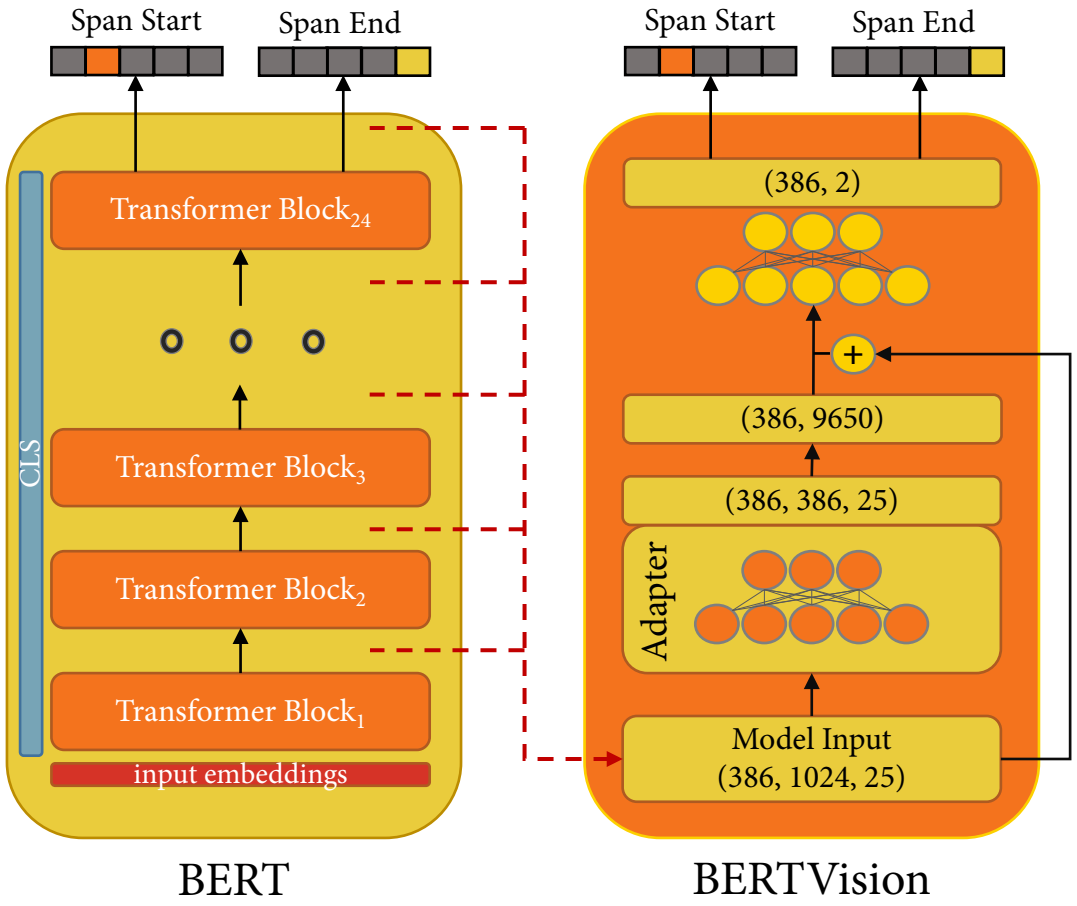
Fine-tuned on SQuAD v2 for binary classification task (Answer / No Answer)
1/10th epoch – 9/10th epochs, 1 epoch – 6 epochs.

# BERTVision Data Pipeline
## Binary Classification



# BERTVision Performance

## Span Annotation



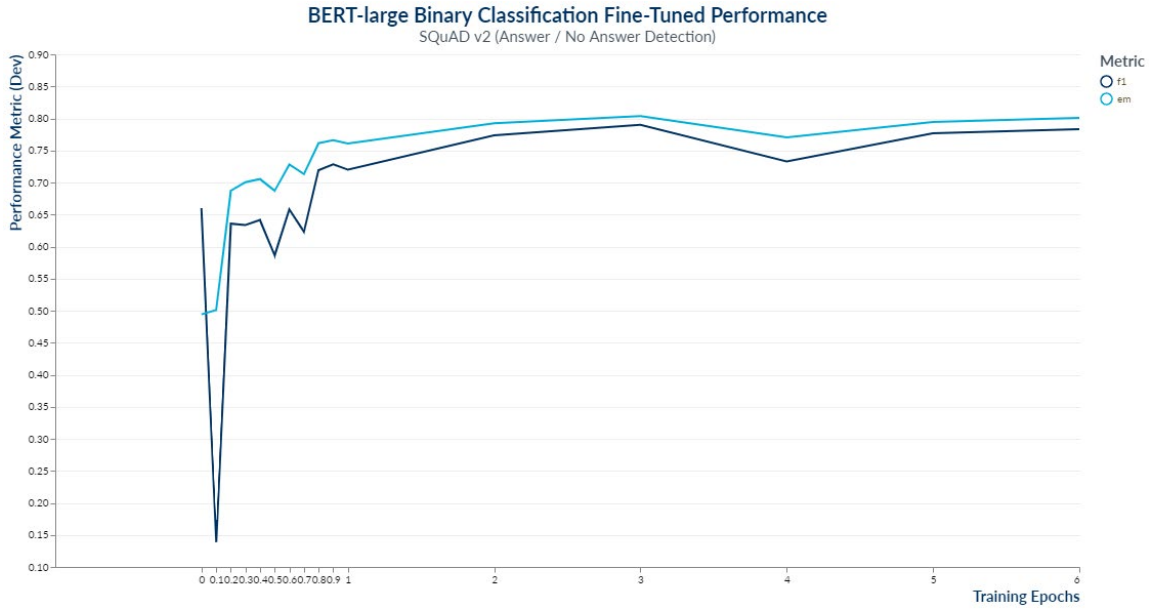

| Model | SQuAD2.0 EM | SQuAD2.0 F1 |
|---|---|---|
| BERT $\frac{3}{10}e$ | 0.654 | 0.702 |
| our model $\frac{3}{10}e$ | 0.699 | 0.740 |
| BERT $1e$ | 0.728 | 0.777 |
| our model $1e$ | 0.749 | 0.790 |
| ensemble BERT+our model $\frac{3}{10}e$ | 0.691 | 0.734 |
| **ensemble BERT+our model** $1e$ | 0.756 | 0.798 |

# BERT Fine-Tuned Performance

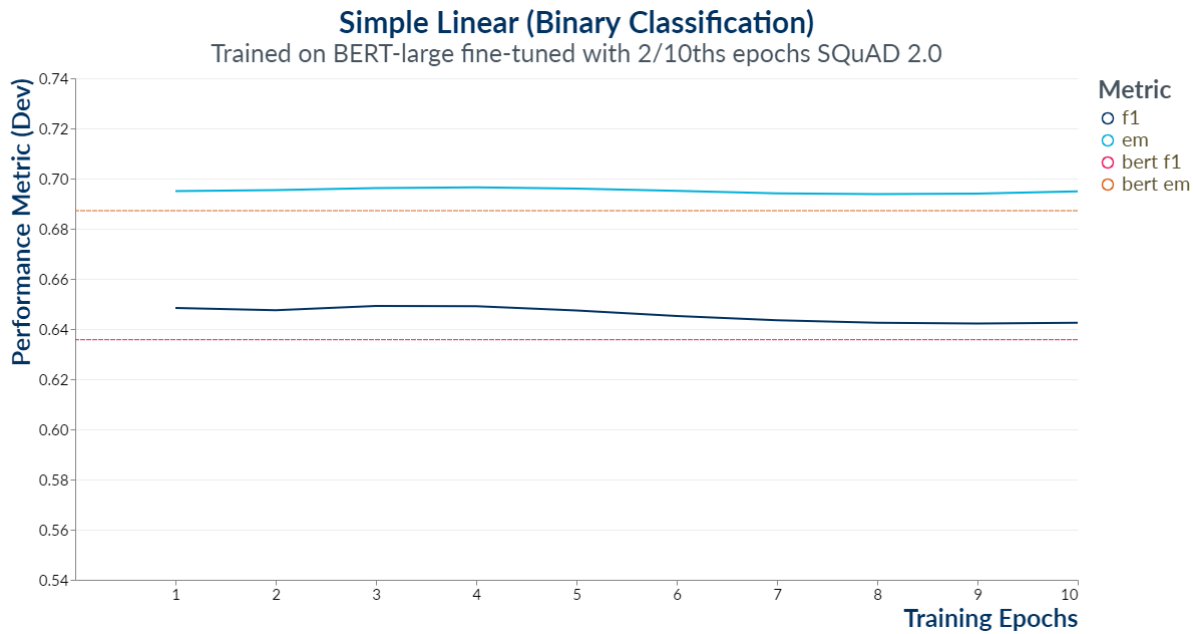
BERT-large Binary
Classification on SQuAD v2

| epoch | f1 | em |
|---|---|---|
| 0.1 | 13.891% | 50.088% |
| 0.2 | 63.574% | 68.719% |
| 0.3 | 63.378% | 70.050% |
| 0.4 | 64.146% | 70.530% |
| 0.5 | 58.607% | 68.711% |
| 0.6 | 65.798% | 72.795% |
| 0.7 | 62.309% | 71.296% |
| 0.8 | 71.928% | 76.122% |
| 0.9 | 72.822% | 76.594% |
| 1.0 | 72.017% | 76.097% |
| 2.0 | 77.364% | 79.255% |
| 3.0 | 79.010% | 80.359% |
| 4.0 | 73.276% | 77.049% |
| 5.0 | 77.688% | 79.449% |
| 6.0 | 78.310% | 80.064% |

```
Model: "BERT_SQuADv2_BinaryClassification"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
input_ids (InputLayer)       [(None, 386)]             0
_________________________________________________________________________________________________
input_masks (InputLayer)     [(None, 386)]             0
_________________________________________________________________________________________________
input_tokens (InputLayer)    [(None, 386)]             0
_________________________________________________________________________________________________
tf_bert_model (TFBertModel)  ((None, 386, 1024),  335141888   input_ids[0][0]
                                                               input_masks[0][0]
                                                               input_tokens[0][0]
_________________________________________________________________________________________________
dense_2 (Dense)              (None, 2)                 2050        tf_bert_model[0][1]
=================================================================================================
Total params: 335,143,938
Trainable params: 335,143,938
Non-trainable params: 0
_________________________________________________________________________________________________
```




# "TS" Model Performance (1 epoch)




Tenney Small on BERT 1
epochs fine-tuned

| epoch | f1 | em |
|---|---|---|
| 1.0 | 76.326% | 78.161% |
| 2.0 | 76.047% | 77.992% |
| 3.0 | 76.089% | 78.043% |
| 4.0 | 76.087% | 78.051% |
| 5.0 | 76.034% | 78.034% |
| 6.0 | 76.028% | 78.043% |
| 7.0 | 76.018% | 78.068% |
| 8.0 | 75.982% | 78.068% |
| 9.0 | 76.003% | 78.093% |
| 10.0 | 76.027% | 78.127% |

```
Model: "BinaryClassification_Adapter_Tenney"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_layer (InputLayer)     [(None, 1, 1024, 26)]     0
_________________________________________________________________
bert_concat_35 (BertConcat)  (None, 1, 1024)           27
_________________________________________________________________
tf_op_layer_Squeeze_30 (Tens [(None, 1024)]            0
_________________________________________________________________
dense_85 (Dense)             (None, 2)                 2050
=================================================================
Total params: 2,077
Trainable params: 2,077
Non-trainable params: 0
_________________________________________________________________
```