

# Contradictory, My Dear Watson

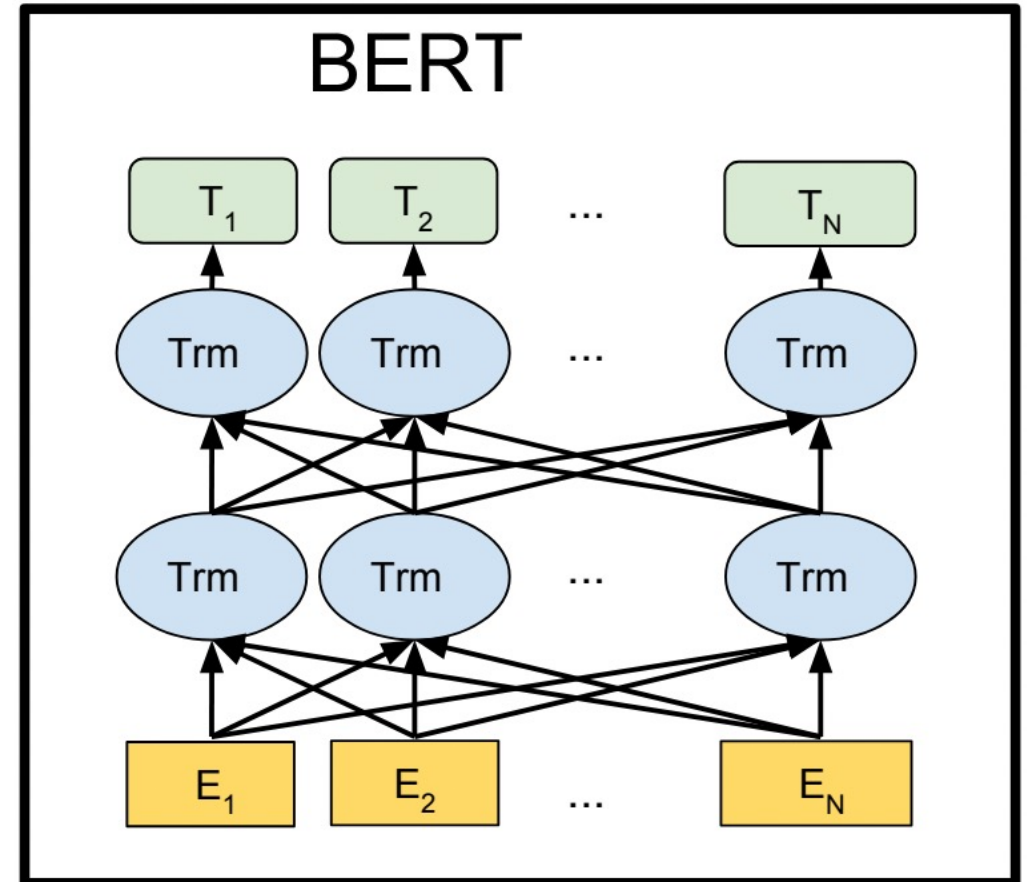
Kaggle Competition  
Final Report  
AI2 Seminar (NAIL052)

Martina Ciklamíniová

May 21 2024

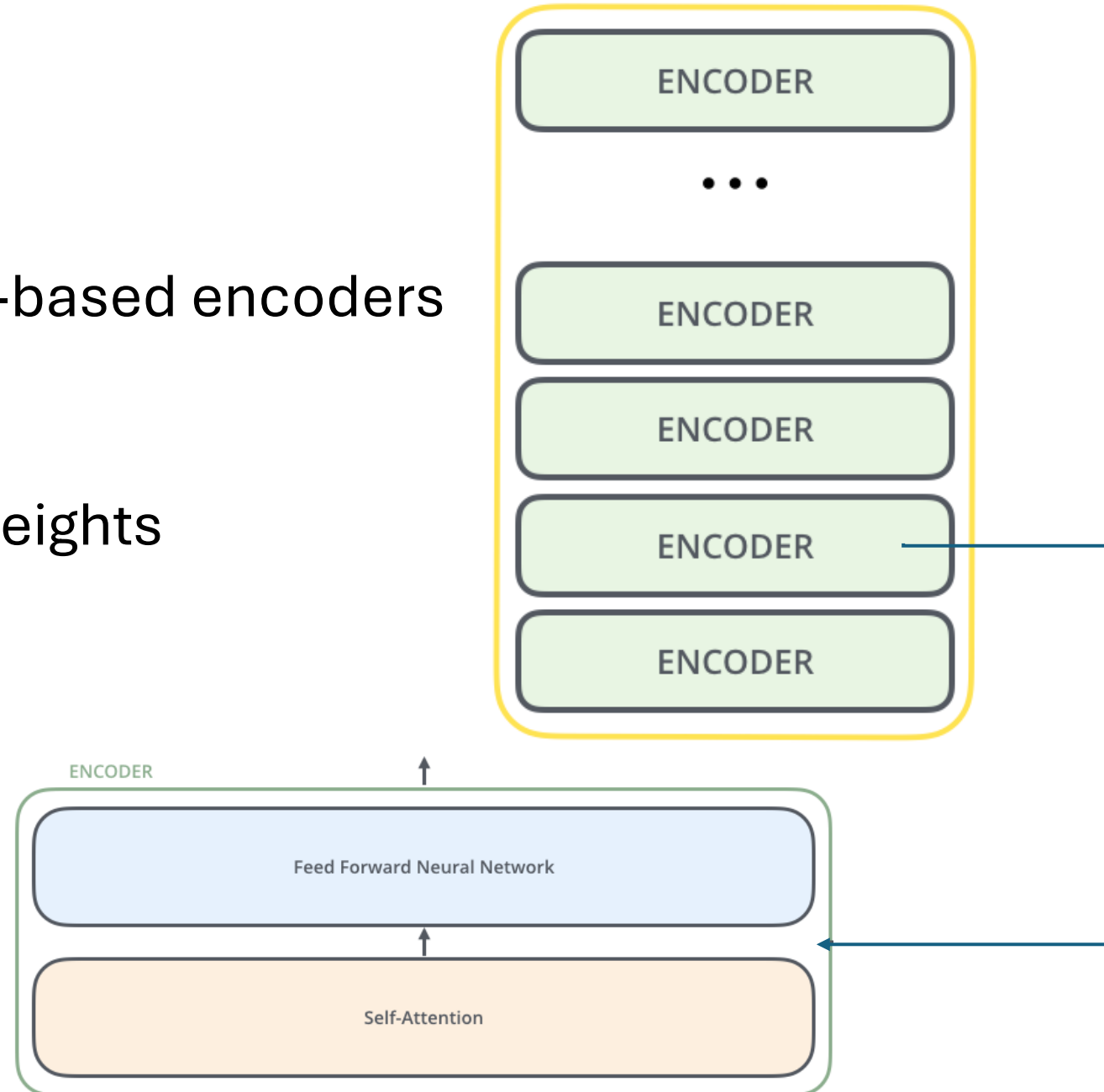
# BERT

- Bidirectional Encoder Representations for Transformers
- released by Google in 2018
- pre-trained using Wikipedia content with a shared vocabulary across 104 languages
  - Masked Language Model (MLM)
  - Next Sentence Prediction (NSP) tasks



# BERT

- stack of identical Transformer-based encoders
  - self-attention sub-layer
  - feed-forward sub-layer
- encoder layers don't share weights



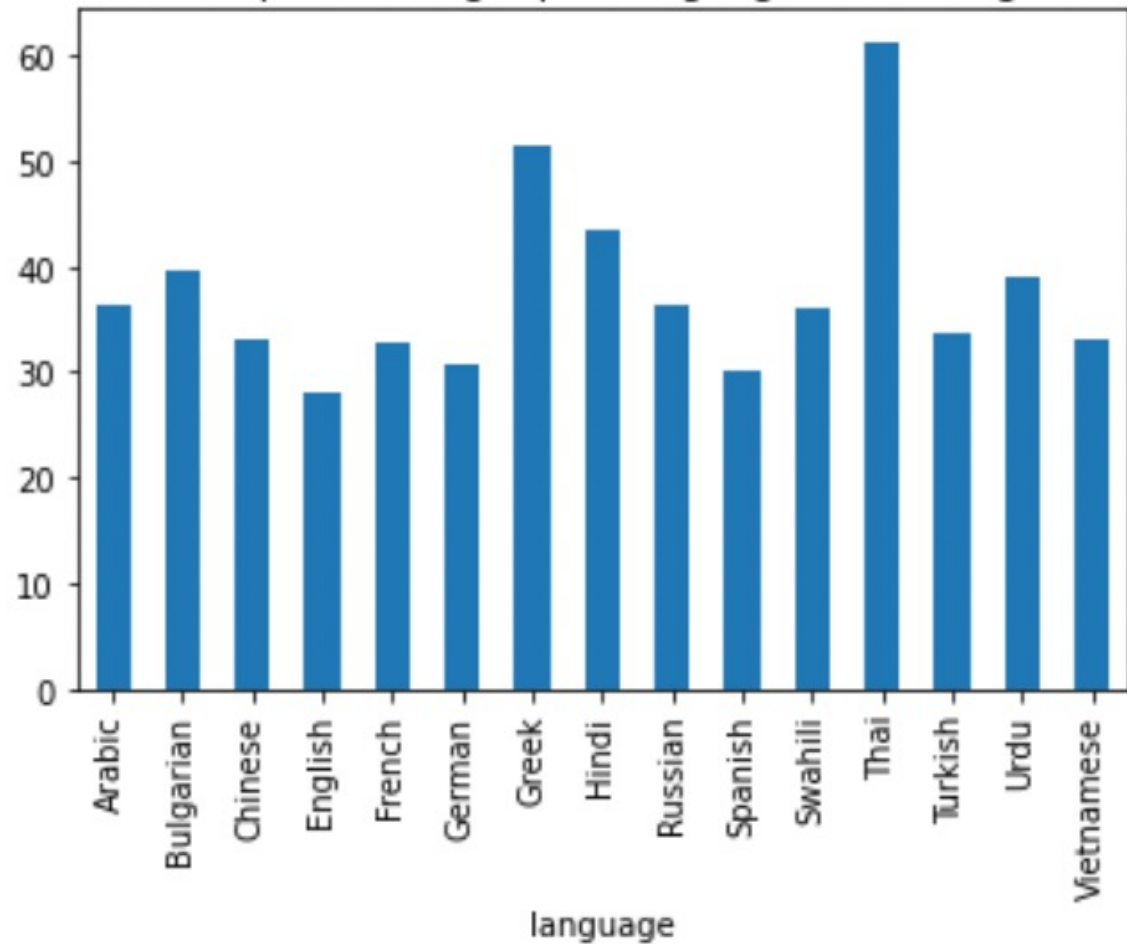
# XLM-RoBERTa

- extension of RoBERTa (an improvement over BERT)
- employs a similar architecture to BERT but benefits from larger-scale pre-training and more aggressive training strategies
- multilingual model aiming to provide robust performance across different languages without the need for language-specific models.

# Tokenization

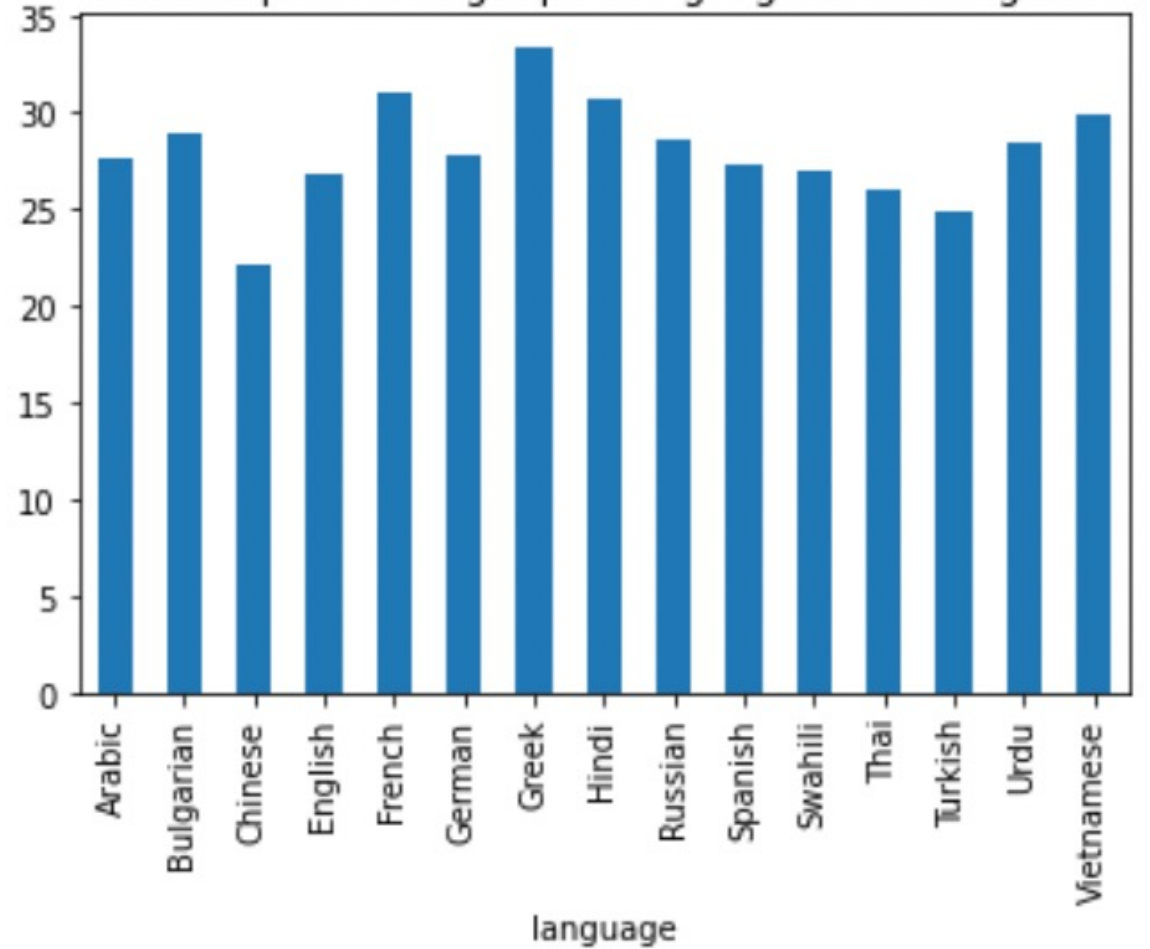
- BERT
  - no normalization on the input
  - WordPiece tokenizer a vocabulary size of 119,547
  - input sentences are tokenized into subwords (elements of the dictionary)
- XLM-RoBERTa
  - SentencePiece tokenizer with a vocabulary size of 250,002
  - converts the inputs into unicode characters (includes spaces)
  - then uses multiple subword algorithms to tokenize the words.
- converting token strings to ids

Mean Sequence Length per Language on Training Data









Using WordPiece tokenizer

Mean Sequence Length per Language on Training Data



Using SentencePiece tokenizer

# Leaderboard

#	Team	Members	Score	Entries	Last
1	Marek Nurzynski		0.99037	1	15d
2	Marc Rosales		0.92916	9	2mo
3	Muneeb Ahméd		0.89470	2	2mo
4	joey123i		0.89143	1	17d
5	Matteo Malosetti		0.88931	1	2mo
21	<b>Martina Ciklamíniová</b>		0.61347	10	19m



Your Best Entry!

Your submission scored 0.34071, which is not an improvement of your previous score. Keep trying!

# Other results

- <https://github.com/wchowdhu/textual-entailment-recognition/tree/main>

<b>Model</b>	<b>Pooling</b>	<b>Accuracy</b>
Random Choice	-	0.333
M-BERT	<i>[CLS]</i>	0.650
XLM-R	<i>[CLS]</i>	0.780
XLM-R + Back-Translation	<i>[CLS]</i>	0.803
XLM-R + Back-Translation	<i>[CLS]</i> + Fully Connected Layer	0.808
XLM-R + Back-Translation + XNLI-val	<i>[CLS]</i>	0.821
XLM-R + Back-Translation + XNLI-val	<i>[CLS]</i> + Fully Connected Layer	0.830
XLM-R + Back-Translation + XNLI-all	Average	0.867
XLM-R + Back-Translation + XNLI-all	<i>[CLS]</i>	0.874
XLM-R + Back-Translation + XNLI-all	<i>[CLS]</i> + Fully Connected Layer	0.874



# Difficulties

- Improper model download in Kaggle
- Version mismatches
- Unclear Documentation
- Poor timemanagement