

The background is a solid dark blue. It is decorated with various hand-drawn, abstract elements in bright colors: yellow, red, and green. These include squares, rectangles, circles, and lines, some of which are partially cut off by the edges of the frame. The overall style is modern and artistic.

Curso Data Science

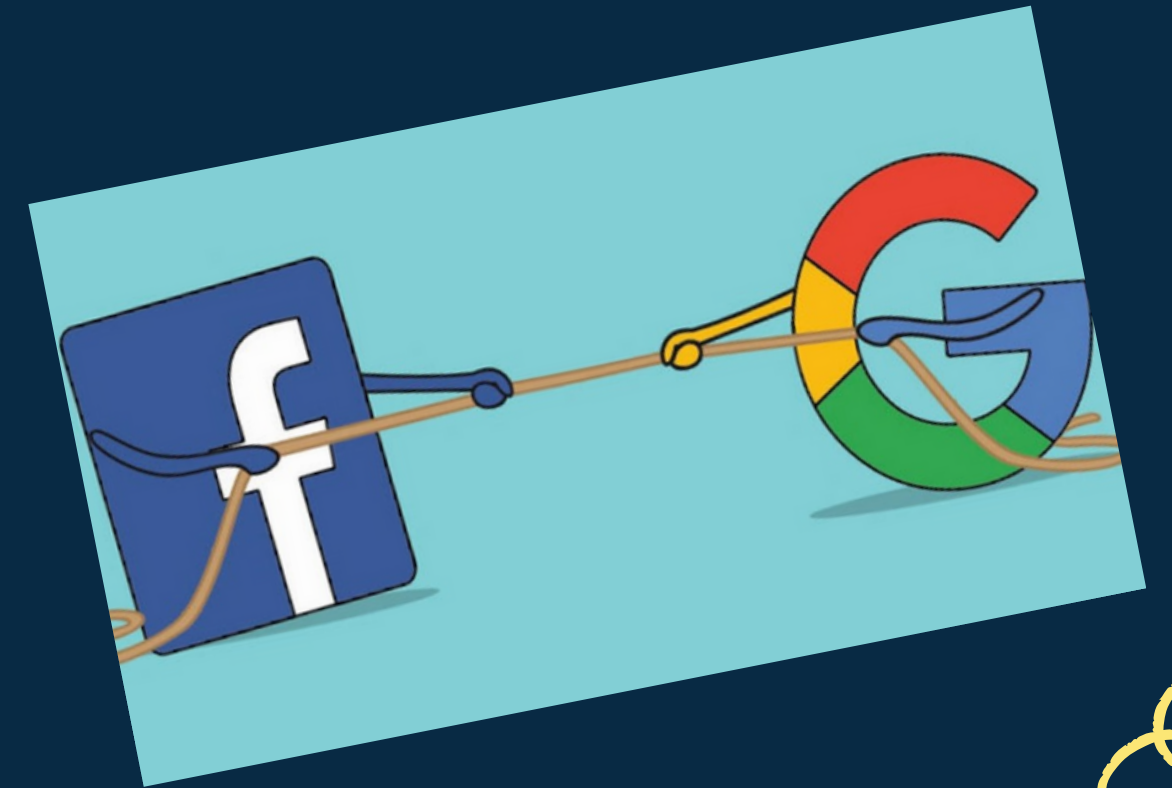
RoBERTa

NICOLL CRESPO RONCALLO

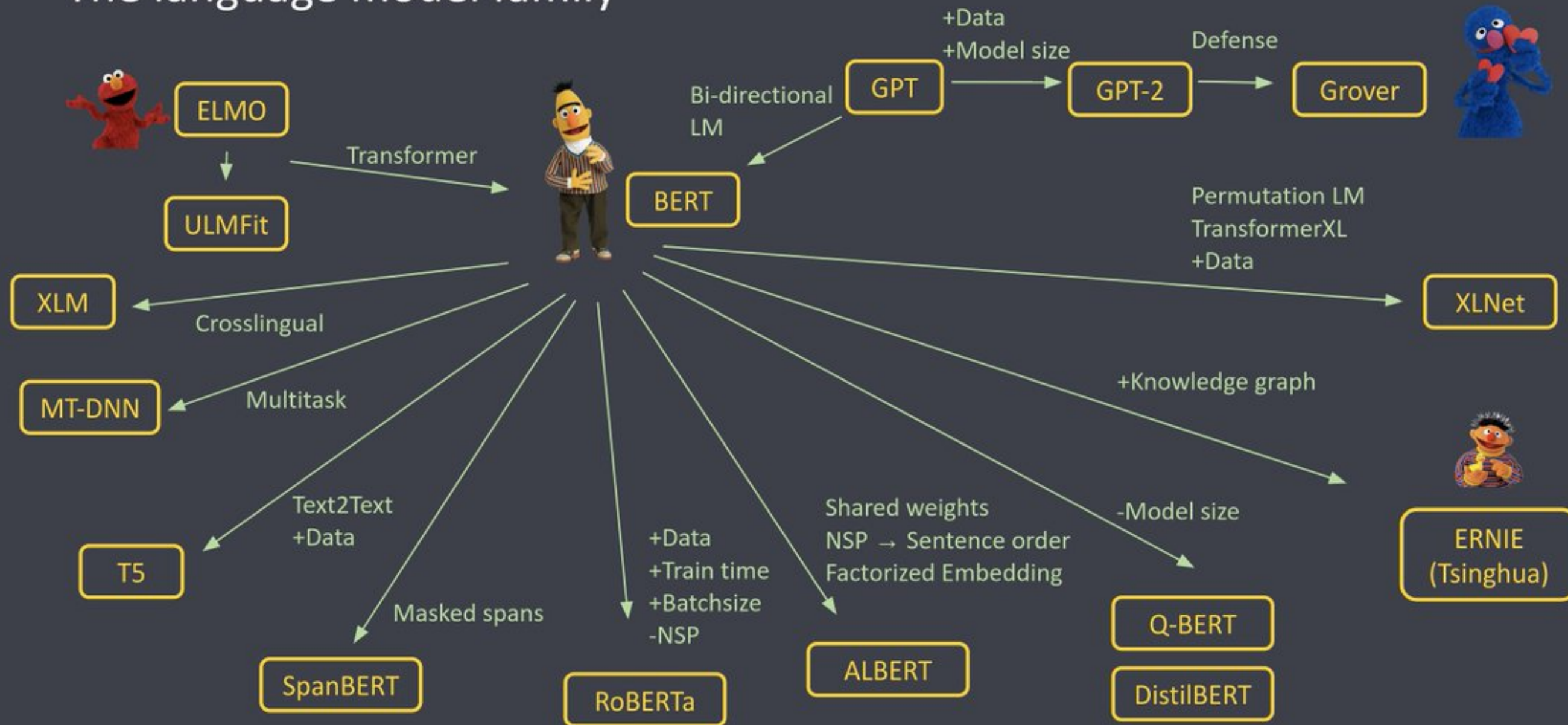
# DEFINICION

RoBERTa es un modelo de NLP  
desarrollado por Facebook

Está basado en el modelo BERT. Pero se re-entrenó con mejoras en su metodología, más datos y mayor poder computacional, obteniendo mejores resultados en GLUE (General Language Understanding Evaluation)



# The language model family



# METODOLOGIA



IMPLEMENTADO EN  
PYTORCH

vs. TensorFlow



ELIMINACION DEL  
ENTRENAMIENTO NSP

(Next Sentence Prediction)



USO DE LOTES MÁS  
GRANDES

(8000 VS. 256)



PRE-ENTRENAMIENTO  
CON 160 GB DE TEXTO

vs. 16 GB



USO DE  
ENMARCAMIENTO  
DINAMICO

VS. Estatico



MAYOR TIEMPO DE  
ENTRENAMIENTO

4-5 veces mas



# RESULTADOS



SUPERÓ A BERT Y XLNET EN  
RACE Y EN GLUE



PUNTUACIÓN PROMEDIO MÁS ALTA EN  
LA TABLA DE CLASIFICACION DE GLUE

( 88,5 )



NUEVO PUNTAJE MAS ALTO EN LA TABLA  
DE CLASIFICACIÓN PÚBLICA DE SQUAD

	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
<i>Single-task single models on dev</i>										
BERT <sub>LARGE</sub>	86.6/-	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-	-
XLNet <sub>LARGE</sub>	89.8/-	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa	<b>90.2/90.2</b>	<b>94.7</b>	<b>92.2</b>	<b>86.6</b>	<b>96.4</b>	<b>90.9</b>	<b>68.0</b>	<b>92.4</b>	<b>91.3</b>	-
<i>Ensembles on test (from leaderboard as of July 25, 2019)</i>										
ALICE	88.2/87.9	95.7	<b>90.7</b>	83.5	95.2	92.6	<b>68.6</b>	91.1	80.8	86.3
MT-DNN	87.9/87.4	96.0	89.9	86.3	96.5	92.7	68.4	91.1	89.0	87.6
XLNet	90.2/89.8	98.6	90.3	86.3	<b>96.8</b>	<b>93.0</b>	67.8	91.6	<b>90.4</b>	88.4
RoBERTa	<b>90.8/90.2</b>	<b>98.9</b>	90.2	<b>88.2</b>	96.7	92.3	67.8	<b>92.2</b>	89.0	<b>88.5</b>



Model	SQuAD 1.1		SQuAD 2.0	
	EM	F1	EM	F1
<i>Single models on dev, w/o data augmentation</i>				
BERT <sub>LARGE</sub>	84.1	90.9	79.0	81.8
XLNet <sub>LARGE</sub>	<b>89.0</b>	94.5	86.1	88.8
RoBERTa	88.9	<b>94.6</b>	<b>86.5</b>	<b>89.4</b>
<i>Single models on test (as of July 25, 2019)</i>				
XLNet <sub>LARGE</sub>			86.3 <sup>†</sup>	89.1 <sup>†</sup>
RoBERTa			86.8	89.8
XLNet + SG-Net Verifier			<b>87.0<sup>†</sup></b>	<b>89.9<sup>†</sup></b>

Model	Accuracy	Middle	High
<i>Single models on test (as of July 25, 2019)</i>			
BERT <sub>LARGE</sub>	72.0	76.6	70.1
XLNet <sub>LARGE</sub>	81.7	85.4	80.2
RoBERTa	<b>83.2</b>	<b>86.5</b>	<b>81.3</b>

# APLICACION

Wang, Y., Sun, Y., Ma, Z., Gao, L., Xu, Y., & Sun, T. (2020, August). Application of pre-training models in named entity recognition.

## CONSIDERACIONES

- Problema de NER
- Uso del MSRA-2006 dataset
- Uso de 2 epocas
- Razon de aprendizaje de  $5e-5$
- Tamaño de lote de 16

## MODELOS

- BERT
- ERNIE
- ERNIE2.0-tiny
- RoBERT

## RESULTADOS

THE RESULTS OF NER USING DIFFERENT PRE-TRAINING MODELS

Models	Precision/%	Recall/%	F1/%
Baseline	92.54	88.20	90.32
BERT-base	92.68	94.18	93.30
ERNIE	92.92	94.07	93.37
ERNIE-tiny	83.89	89.88	86.52
RoBERTa	<b>93.64</b>	<b>94.93</b>	<b>94.17</b>



# REFERENCIAS

1. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
2. Wang, Y., Sun, Y., Ma, Z., Gao, L., Xu, Y., & Sun, T. (2020, August). Application of pre-training models in named entity recognition. In 2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) (Vol. 1, pp. 23-26). IEEE.