

Parámetros *	Transformer from scratch	Fine-tuned GPT2
learning_rate = 0.001 weight_decay = 0.001 batch_size = 64 context_size = 32 (configuración inicial)		
learning_rate = 0.001 weight_decay = 0.01 batch_size = 64 context_size = 32		
learning_rate = 0.01 weight_decay = 0.001 batch_size = 64 context_size = 32		
learning_rate = 0.001 weight_decay = 0.001 batch_size = 64 context_size = 24		
learning_rate = 0.001 weight_decay = 0.01 batch_size = 64 context_size = 24		
learning_rate = 0.01 weight_decay = 0.001 batch_size = 64 context_size = 24		

* Las variaciones de cada experimento con respecto a la configuración inicial se resaltan en rojo.