

ModernBERT

The [original BERT paper](#) came out in 2018, around 7 years ago at time of writing. However, it is still referred to and used as a strong baseline in a number of NLP tasks. [ModernBERT](#) was created by HuggingFace. ModernBERT is a drop-in replacement for use in problems where BERT may have previously been used and like the original, has a base and a large variant.

ModernBERT also outperforms [DEBERTA-v3-base](#) which has been a favourite of NLP practitioners for a few years thanks to its few-shot and zero-shot capabilities.

Comparison Table

	BERT Base	ModernBERT Base	BERT Large	ModernBERT Large
# Params	110M	149M	340M	395M
Context Size	512	8192	512	8192
BEIR	38.9	41.6	38.9	44.0
MLDR_{OOD}	23.9	27.4	23.3	34.3
MLDR _{ID}	32.2	44.0	31.7	48.6
BEIR (CoBERT)	49.0	51.3	49.5	52.4
MLDR _{OOD} (CoBERT)	28.1	80.2	28.5	80.4
GLUE	84.7	88.5	85.2	90.4
CSN	41.2	56.4	41.6	59.5
SQA	59.5	73.6	60.8	83.9

OOD=out of domain ID=In Domain

[mmBERT](#) is a modern multi-lingual encoder-only model based on ModernBERT.

Revision #5

Created 18 September 2025 07:04:32 by James

Updated 18 September 2025 09:21:56 by James