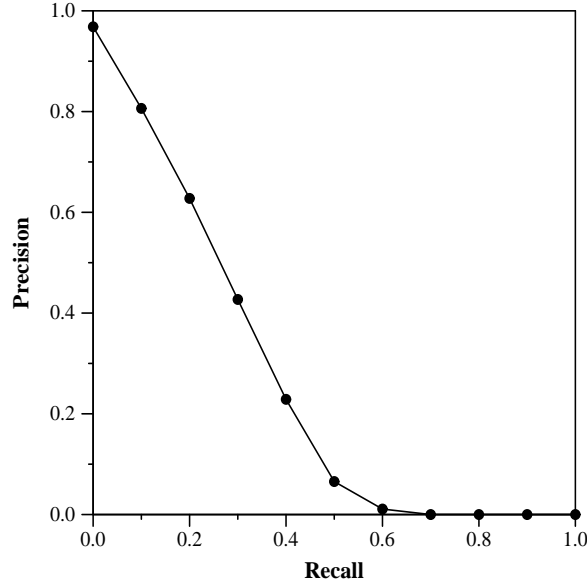


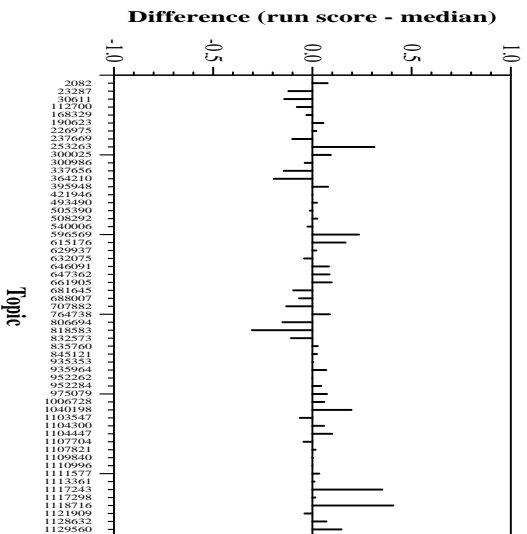
Deep Learning Track results — (bcai) Bosch Center for AI

Run Description	
Run ID:	bcai_bertmlens
Task:	Document Ranking
Subtask:	Full ranking from the collection
Topic type:	Automatic
Single-stage retrieval?	No
Dense retrieval?	Yes
Used deep nn model?	Pre-trained model
Type of training:	This year's MS MARCO training data; Previous year's MS MARCO training data
Pre-processing/indexing cost:	tokenization and indexing took several days on a 10-core server. document & query embedding took several days on a single RTX 3090.
Query processing cost:	First stage: 7 seconds per core (brute-force exact search) Second stage: 40 seconds per RTX 3090. Yes, this model ensemble *IS* ridiculously expensive. It can surely be distilled though into a single BERT model based, e.g., on MiniLM with a small loss of accuracy.

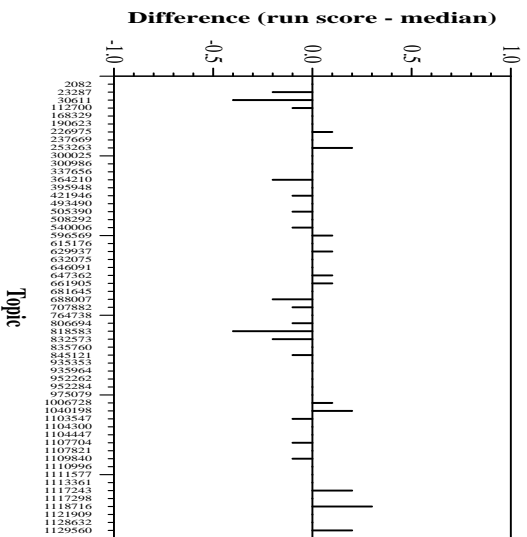
Overall measures	
Number of topics	57
Total number retrieved	5700
Total relevant	8203
Total relevant retrieved	2370
MAP	0.2559
Mean NDCG@10	0.6812
Mean Reciprocal Rank	0.9444

Document Level Averages	
	Precision
At 5 docs	0.8982
At 10 docs	0.8140
At 15 docs	0.7743
At 20 docs	0.7254
At 30 docs	0.6561
R-Precision	
Exact	0.3251

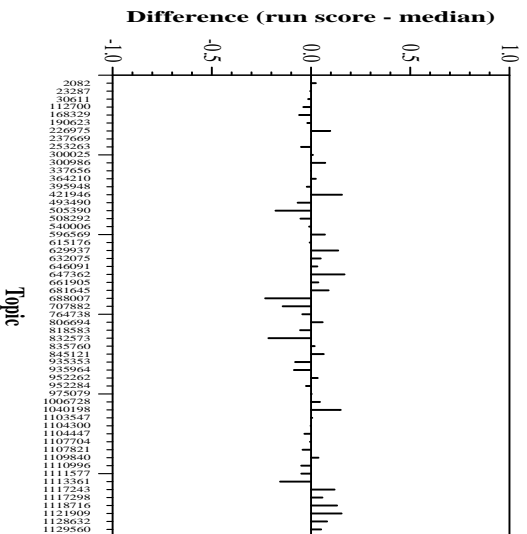




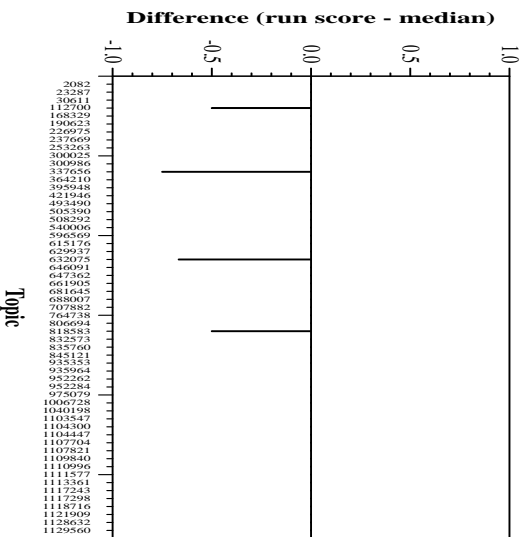
Per-topic difference from median for NDCG@10 for Document Ranking runs



Per-topic difference from median for Precision@10 for Document Ranking runs



Per-topic difference from median for Average Precision for Document Ranking runs



Per-topic difference from median for Reciprocal Rank for Document Ranking runs