MTEB Aggregation Masks Italian Performance

Models above line: Better on Italian than overall score suggests

Qwen3-8B
ml-e5-large
BGE-M3
ml-e5-base
Qwen3-4B
ml-e5-small
13.8 pts gap
ml-e5-small (finetuned)
e5-small-v2

X-axis: **Overall MTEB Score**
Y-axis: **Italian-Only NDCG@10**

Legend:
- Base models
- Fine-tuned models
- ml-e5-small (base)