Validation metrics over training

Validation loss ↓

Top-10% retrieval ↑

- Fine-tune (summaries)
- Fine-tune (abstract)
- Fine-tune head (summaries)
- Train from scratch (summaries)
- Shuffle within batch

Random expectation