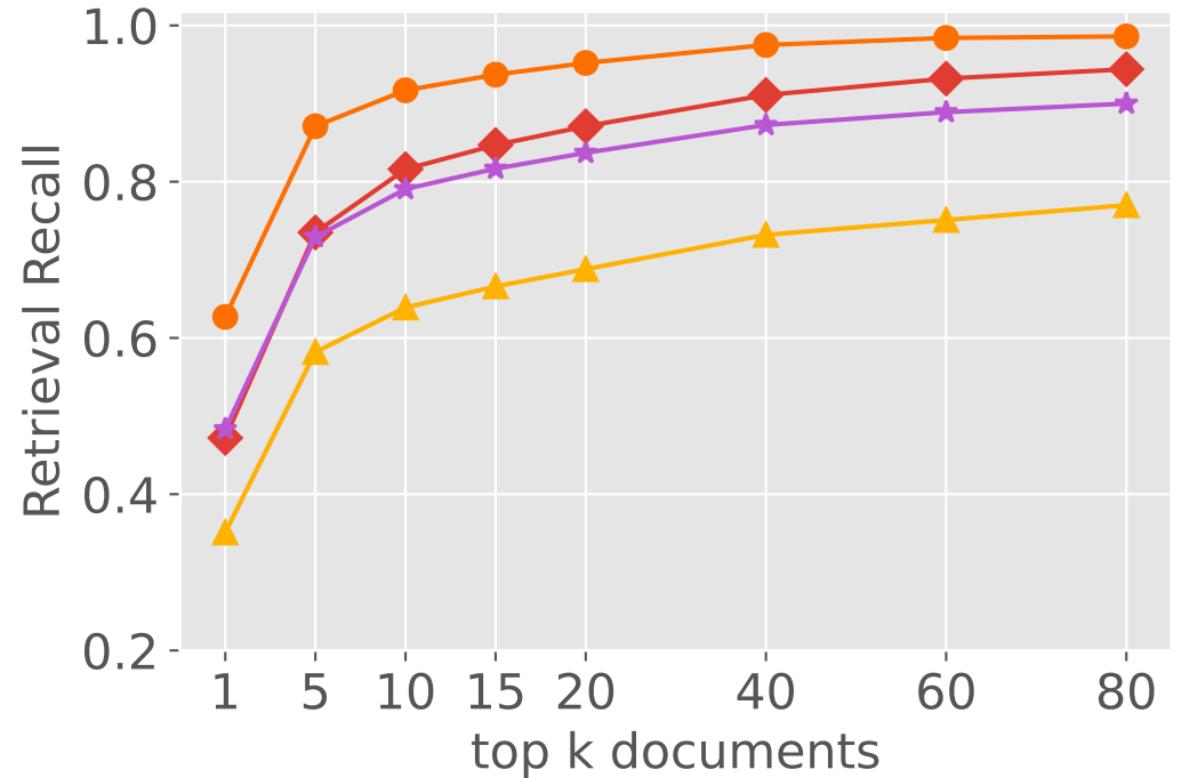
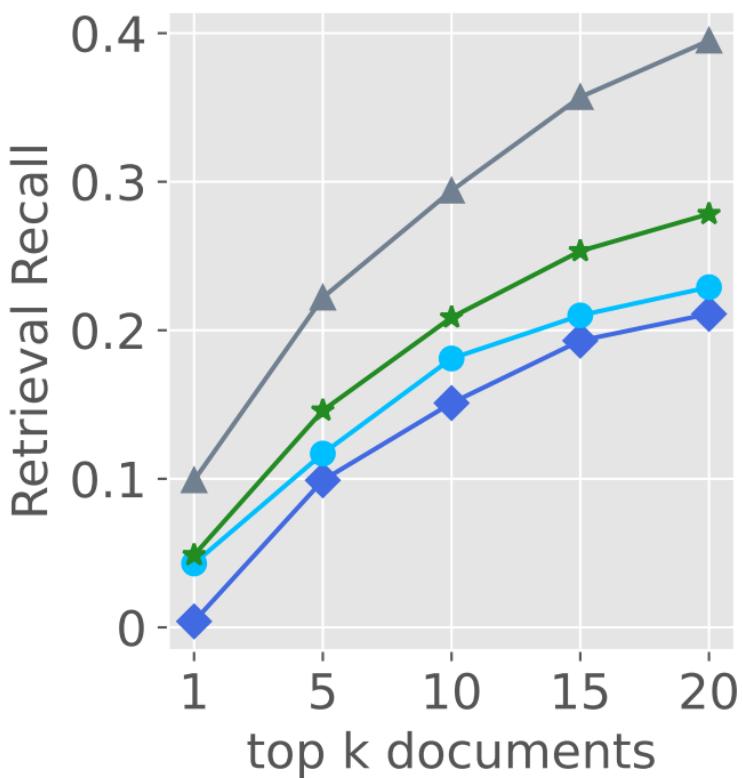


QA Datasets



Code Generation Datasets



Legend:

- NQ (Red Diamond)
- TriviaQA (Orange Circle)
- avg of QA tasks (Purple Star)
- HotpotQA (Yellow Triangle)
- CoNaLa (Blue Diamond)
- DS1000 (Cyan Circle)
- PNE (Grey Triangle)
- avg of code tasks (Green Star)