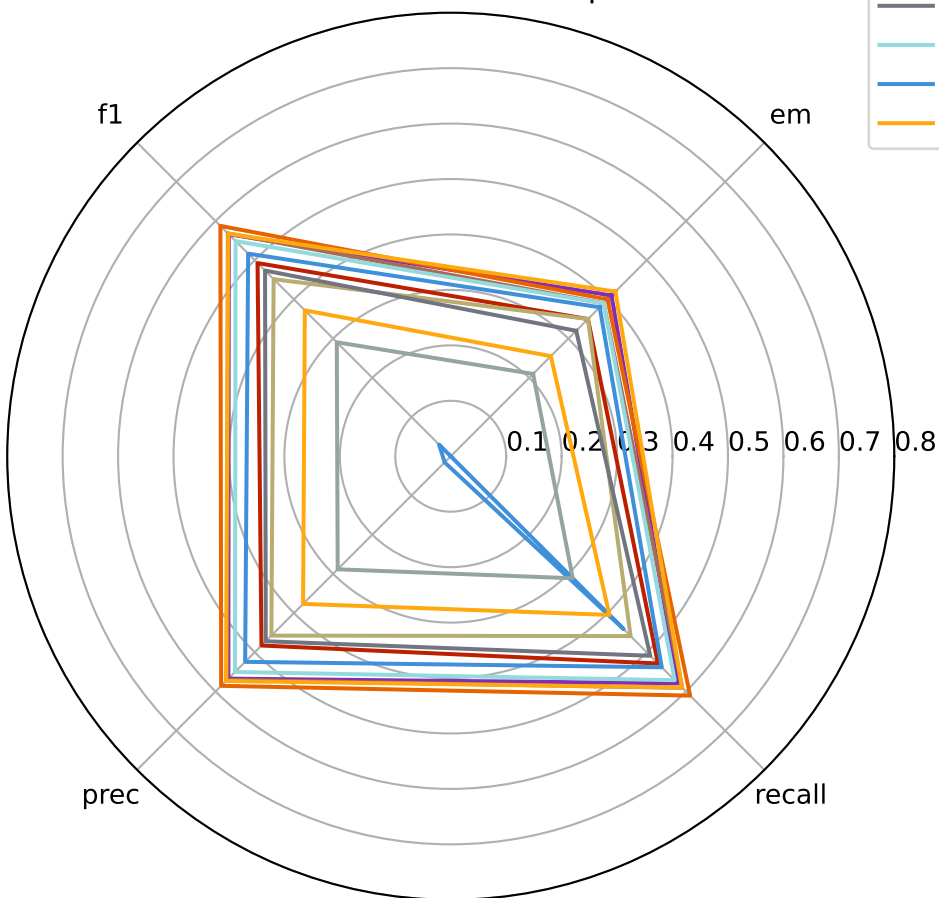


Model Performance Comparison



- baseline
- small rag
- medium rag
- large rag
- blackmail template
- polite template
- expert template
- model reasoning
- in-prompt reasoning
- rag fusion
- two step reflection
- three step reflection