

TODO: Token limit problems

Priority: High

Date: October 19

Problem:

- long docs exceed LM context
- llm-2-7b has 4K token limit
- Current chunking: 500 Chrs, 100 overlap
- Summaries still too long

Actions:

- Benchmark chunk effectiveness
- Test overlap percentages
- Measure token distribution
- Consider dynamic chunking
- Explore compression techniques

Owner: Bob

Target: Oct 25

Effort: 4-6 hours