

TODO: Token limit problems

Priority: High

Date: October 19

Problem:

- long docs exceed 1M context
- [long-2-76 was 5K token limit]
- Current chunking: 500 chars, 100 overlap
- Summaries will too long

Actions:

- ① Benchmark chunk effectiveness
- ② Test overlap percentages
- ③ Measure token distribution
- ④ Consider dynamic chunking
- ⑤ Explore compression techniques

Bob is leading the optimization effort

Target: Oct 25

Effort: 4-6 hours