Summarization Method Selection Guide

Document Count	Quality Need	Budget	Recommended Method
Single	High	High	GPT-4 + Few-Shot
Single	High	Low	FLAN-T5-XL Local
Single	Medium	Any	GPT-3.5 + Zero-Shot
Multiple	High	High	GPT-4 + Map-Reduce
Multiple	Medium	Low	FLAN-T5 + Chunking
Many (>10)	Any	Low	Extractive + LLM Polish

Cost Level:

High Cost

Medium Cost

Low Cost

Key Factors:

• Token limits constrain input size

• Quality vs Cost tradeoff

• Local models for sensitive data