	OpenAI	Anthropic
Streaming Support	No	Yes
Pricing	50% cost reduction for cached tokens	90% cost reduction for cached tokens (reads), but 25% cost increase for cache writes
How to enable	Enabled automatically in the API	Enabled only after making specific changes to the request header or, if you are using the python SDK using **client.beta.prompt_caching. messages.create**
Cache control	Whatever you want to cache has to go in the prefix of the prompt	Fine grained control over what you cache using the **cache_control** block in the API.
		Sources
		https://openai.com/index/api-prompt-caching/
		https://docs.anthropic.com/en/docs/build-with-claude/prompt-caching
		https://www.anthropic.com/news/prompt-caching