

Runtime comparison of kv-caching and no-cache sampling

