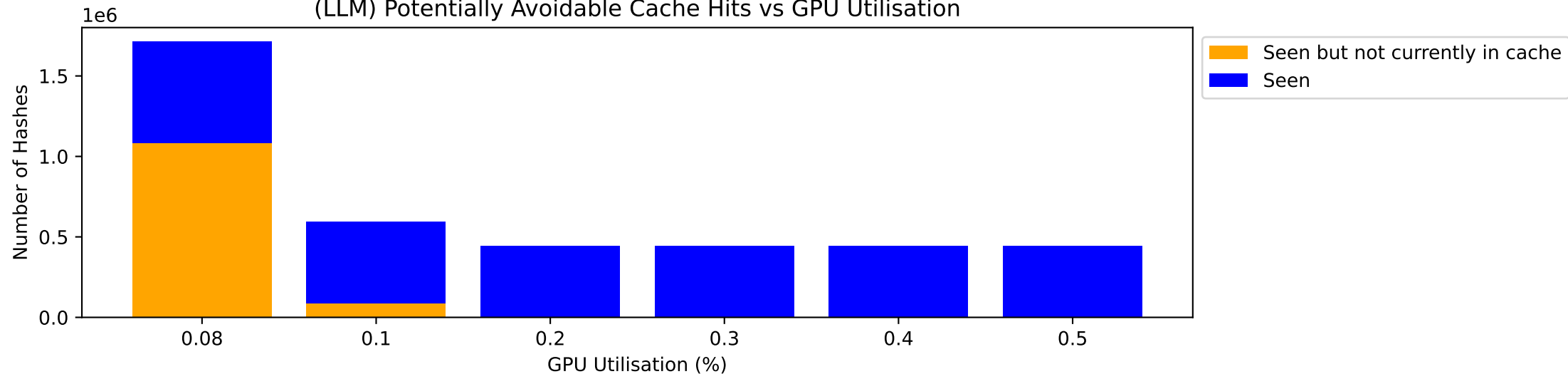


(LLM) Potentially Avoidable Cache Hits vs GPU Utilisation



(PRM) Potentially Avoidable Cache Hits vs GPU Utilisation

