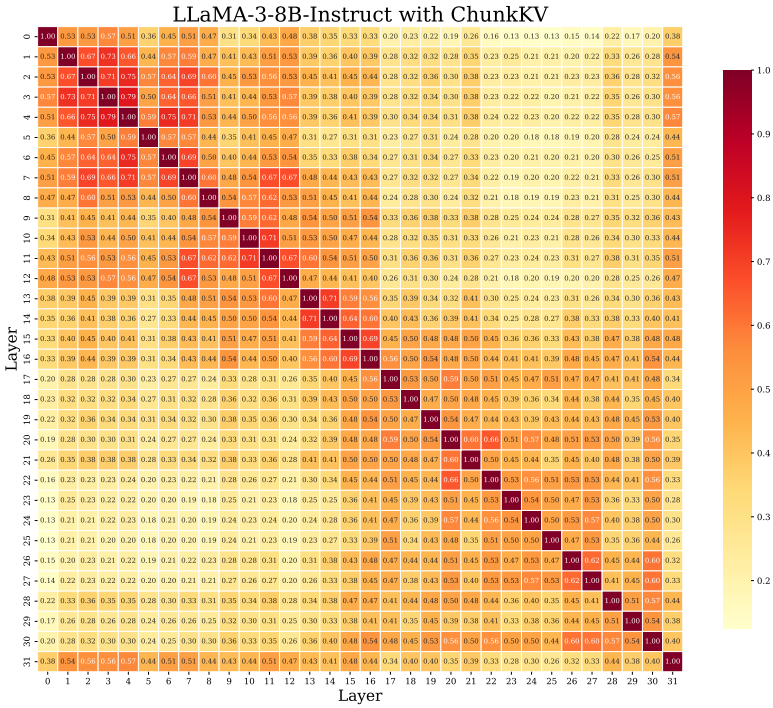

Figure 9: Layer-wise similarity heatmaps of the preserved KV cache indices by ChunkKV on LLaMA-3-8B-Instruct

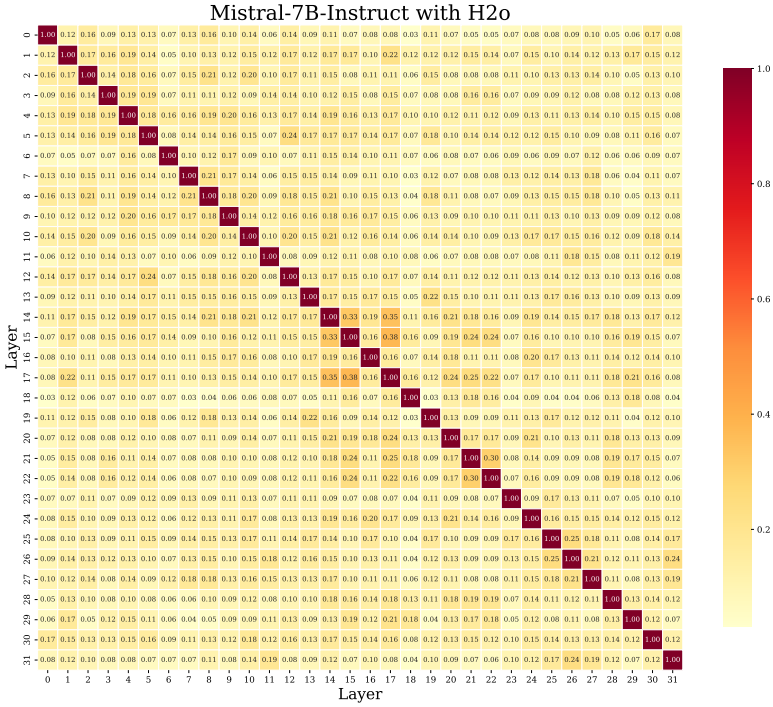

Figure 10: Layer-wise similarity heatmaps of the preserved KV cache indices by H2O on Mistral-7B-Instruct