**Predictive Information for Future Attention**
**(Position provides MORE information than learned features)**

Position alone outperforms KV features!

Mutual Information (bits)

- Key vectors: 0.12
- Value vectors: 0.08
- Position: 0.31
- Prefill attention: 0.28
- Combined (SIP): 0.35

Feature Type