Read-Disturb Accumulation in LLM Inference
Hypothesis 1: Unbounded Linear Growth

Legend:
- Measured
- Linear fit (R²=0.999)
- Projected

100K tokens
→ 63,572 reads
(18.4× baseline)

Reads = 0.636 × Tokens − 3218

X-axis: Tokens Generated (×1000)
Y-axis: Cumulative Flash Reads