The Pensieve Paradigm: Stateful Language Models Mastering Their Own Context

Xiaoyuan Liu[1,2] Tian Liang[1] Dongyang Ma[1] Deyu Zhou Haitao Mi[1] Pinjia He[2] Yan Wang[1]

[1] Tencent AI Lab [2] The Chinese University of Hong Kong, Shenzhen

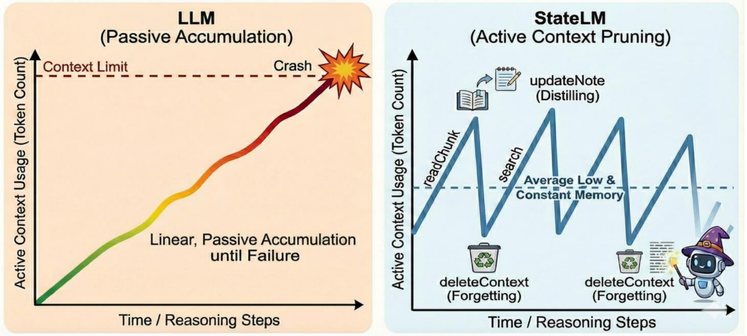
ICLR

Tencent AI Lab

The Chinese University of Hong Kong, Shenzhen

Scan Me!

Motivation & Key Idea



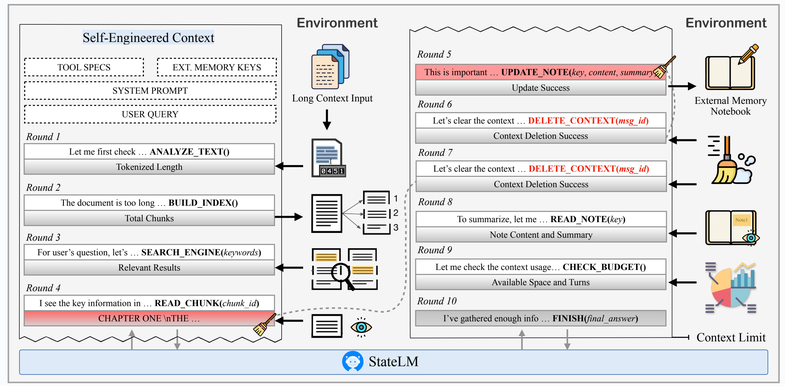
- **Current LLMs:** (1) **context grows rapidly** and soon exceeds native context limits; (2) **reasoning capability degrades**, as important information becomes buried in irrelevant or outdated content.
- **StateLMs:** maintains a "sawtooth" context profile via actively pruning, storing, and extracting its own memories, analogous to **self-context engineering.**
- By putting the "wand" back to model's own hand, we perform the first study to address memory challenges of long-doc QA, chat memory, and deep research within a **single, unified model**.

The StateLM "Spellbook"

Tool Name	Description
Context Perception	(Understanding the environment)
analyzeText	Returns the input length.
checkBudget	Reports remaining interaction budget.
Information Acquisition	(Accessing raw input)
buildIndex	Builds a searchable index.
searchEngine	Searches for relevant segments.
readChunk	Loads a selected text chunk.
Memory Management	(Distilling signal and pruning noise)
note / updateNote	Records or updates key knowledge.
readNote	Retrieves stored notes into context.
deleteContext	Removes messages from context.
Termination	
finish	Ends reasoning and outputs the answer.

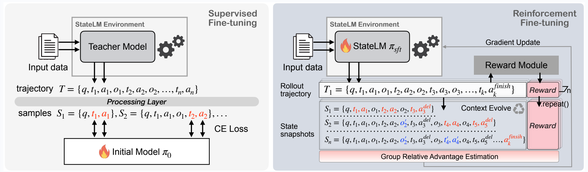
A general-purpose toolkit enabling **context perception**, **information acquisition**, **memory management**, and **context pruning**.

StateLM Self-Context Engineering Workflow



Given a query, StateLM iteratively **searches, reads, takes notes**, and **prunes** its context. Deleted messages are replaced with stubs.

Two-Stage Training Pipeline



Stage 1: SFT (35.7K samples)
From expert trajectories with outcome & process reject sampling.

Stage 2: RL (488 samples)
GRPO with trajectory snapshots and task-aware rewards.

Needle-in-a-Haystack: Memory Retrieval

Model	Length							
	32K	64K	128K	256K	512K	768K	1M	2M
Qwen3-4B	**100.00**	**100.00**	88.33	41.67	23.33	16.67	3.33	1.7
StateLM-4B (w/o search)	95.00	95.56	**95.56**	**88.33**	**76.67**	**62.78**	**53.89**	**32.22**
Qwen3-8B	**100.00**	**100.00**	88.33	41.67	23.33	16.67	3.33	1.7
StateLM-8B (w/o search)	99.44	98.33	**98.33**	**99.44**	**95.55**	**88.89**	**88.89**	**67.22**
Qwen3-14B	**100.00**	**100.00**	88.33	41.67	23.33	16.67	3.33	1.7
StateLM-14B (w/o search)	**100.00**	**100.00**	**99.44**	**97.78**	**89.45**	**94.44**	**95.00**	**83.89**

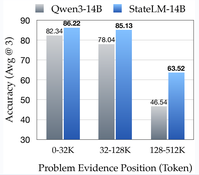
StateLM remains **robust up to 2M tokens** while baselines collapse beyond 128K. StateLM-14B achieves **84% at 2M** vs. 1.7% for Qwen3.

Main Results: Long-Context Reasoning

Model	Context	LongDoc QA (135K)		Chat Memory (115K)	*BrowseComp+ (552K)
		NovelQA	∞Bench		
Qwen3-235B (w/ Pensieve)	256K	80.71	73.36	67.00	55.33
RL-MemoryAgent-7B	32K	60.24	62.45	40.60	-
RL-MemoryAgent-14B	32K	78.86	74.24	59.00	-
ReadAgent-8B	32K	16.38	24.02	0.00	-
ReadAgent-14B	32K	23.12	34.06	14.60	-
Qwen3-4B	128K	65.17 ± 0.53	59.97 ± 0.50	39.53 ± 1.36	2.89 ± 1.02
StateLM-4B	32K	**79.57** ± 0.93	**67.25** ± 0.76	**59.33** ± 0.23	**35.33** ± 5.92
Qwen3-8B	128K	65.87 ± 1.42	66.81 ± 1.16	45.40 ± 1.56	5.56 ± 0.77
StateLM-8B	32K	83.84 ± 0.42	70.16 ± 1.33	58.93 ± 2.53	46.22 ± 1.68
↳ *StateLM-8B-RL*	32K	**84.15** ± 1.00	**73.07** ± 1.33	**59.73** ± 2.20	**46.44** ± 0.77
Qwen3-14B	128K	77.94 ± 0.26	74.96 ± 0.25	54.07 ± 0.76	5.46 ± 0.04
StateLM-14B	32K	84.15 ± 0.82	77.44 ± 1.40	64.40 ± 1.40	51.33 ± 1.34
↳ *StateLM-14B-RL*	32K	**84.85** ± 0.42	**78.46** ± 0.67	**64.47** ± 0.50	**52.67** ± 4.00

StateLM outperforms instruct baselines using **only 1/4 context** and surpasses other agentic methods. On BrowseComp-Plus, StateLM achieves **52% vs. ~5%** for vanilla LLMs.

Analysis I: Performance by Evidence Position & Tool Use Pattern

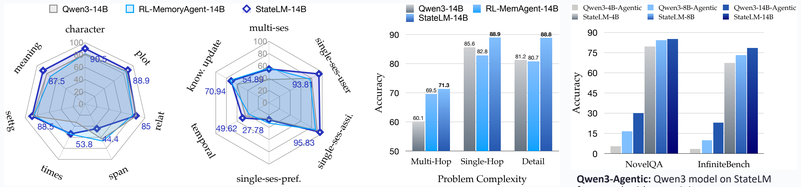


	Rounds	mem	del	srh
NovelQA (119K)	18.6	4.3	6.3	1.8
∞Bench (189K)	20.7	4.6	6.8	2.9
LongMemEval (115K)	22.4	4.9	8.0	2.2
BrowseComp+ (552K)	22.8	4.1	6.1	6.6

Table 1: Tool-use pattern of StateLM-14B across benchmarks with mean input length reported.

- StateLM's most pronounced gains occurring when the relevant evidence appears later in the document **(128–512K)**
- StateLMs allocate more intermediate steps on more challenging benchmarks

Analysis II: Problem Aspects & Comparison to Qwen3-Agentic



Qwen3-Agentic: Qwen3 model on StateLM framework without training