

ALoDLM: Adaptively Looped Diffusion Language Models

Original summary — NOT the original paper or model card.

paper identified preprint · 2026-10-03 [10]

Authors/source: Liancheng Fang, Zhuowei Li, Youngeun Kim, Tianchen Zhao, Rajat Koner, Jiaye Wu et al. (full author list in manifest)

Coverage: Video topic

What the source reports

Adaptively looped diffusion language models at 1.7B and 8B scales. Shared-layer recurrence preserves unresolved token states, commits confident tokens, and learns token-specific compute allocation; authors evaluate eleven benchmarks. [10]

Managing expectations

Longer time to first token can hurt short or latency-sensitive responses. Throughput depends on input difficulty and domain; underrepresented domains can erase or reverse speed advantages over optimized autoregressive baselines. [10]

Original document and sharing status

The original paper has a noncommercial license. Because the site's qualifying noncommercial status is not established, this package links to the official original instead of rehosting it.

Recorded rights: CC BY-NC-SA 4.0 [3]

Sources and downloads

[Primary source / manuscript details \[10\]](#)

[Official original document / PDF download](#)

[Complete document library](#) · [Read the research roundup](#)

Source briefs are Managing Expectations editorial summaries, not independent replications, peer review, or medical advice. Source claims and release details may change.