Anchored Diffusion Language Model

NeurIPS 2025

Litu Rout, Constantine Caramanis, and Sanjay Shakkottai

The University of Texas at Austin

The University of Texas at Austin
Machine Learning Laboratory

The University of Texas at Austin
Chandra Department of Electrical and Computer Engineering
Cockrell School of Engineering

IFML

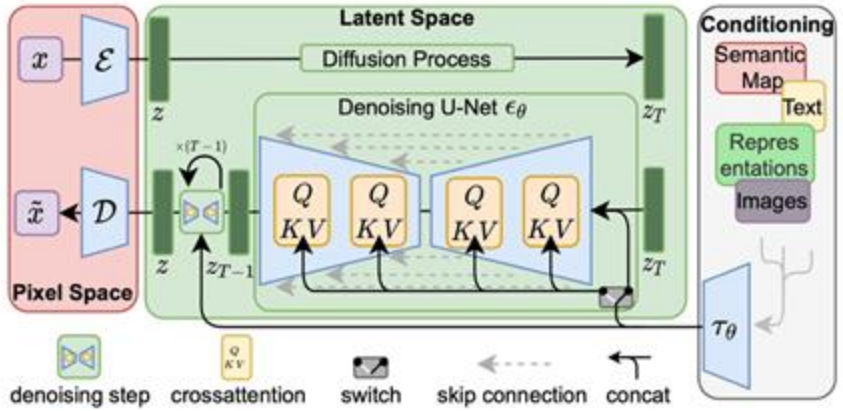
AI EDGE Institute

LDOS

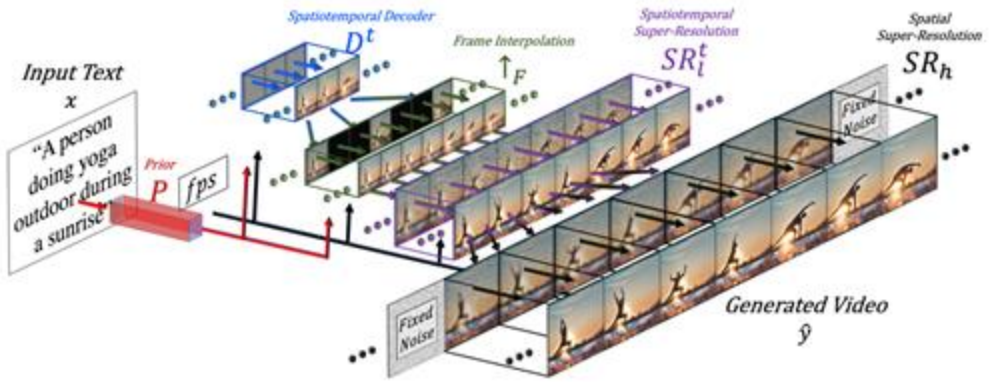
WNCG

Generative Modeling with Diffusion

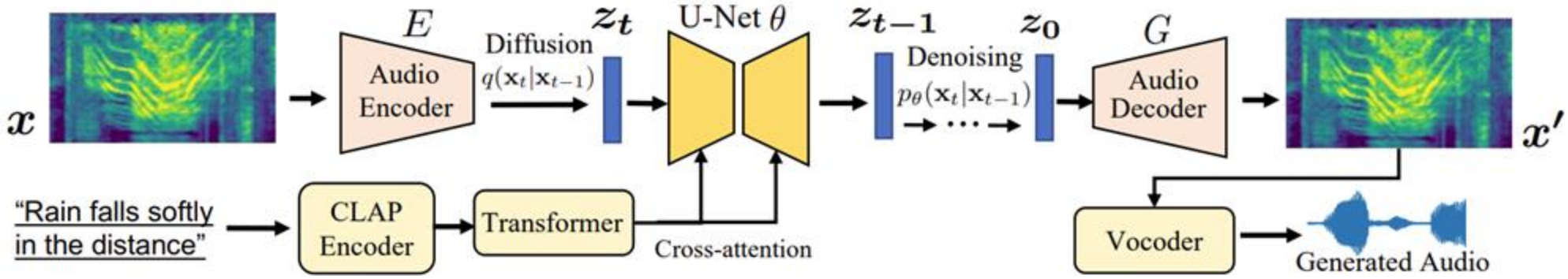
Diffusion Model for Images



Diffusion Model for Video



Diffusion Model for Audio



[1] Image: https://arxiv.org/pdf/2112.10752 [2] Audio: https://text-to-audio.github.io/ [3] Video: https://arxiv.org/pdf/2209.14792

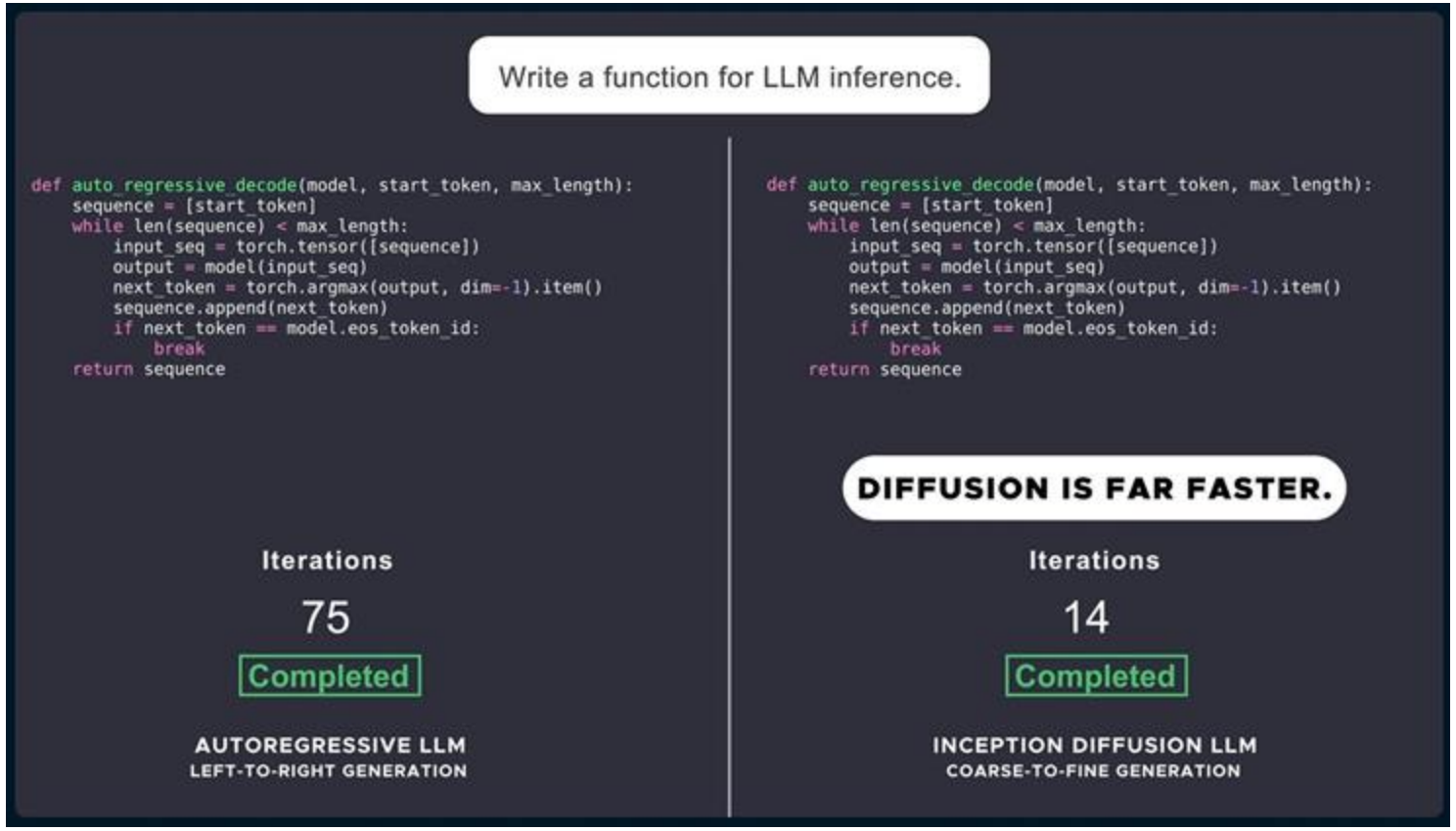
Large-Scale Diffusion Language Models

Gemini Diffusion

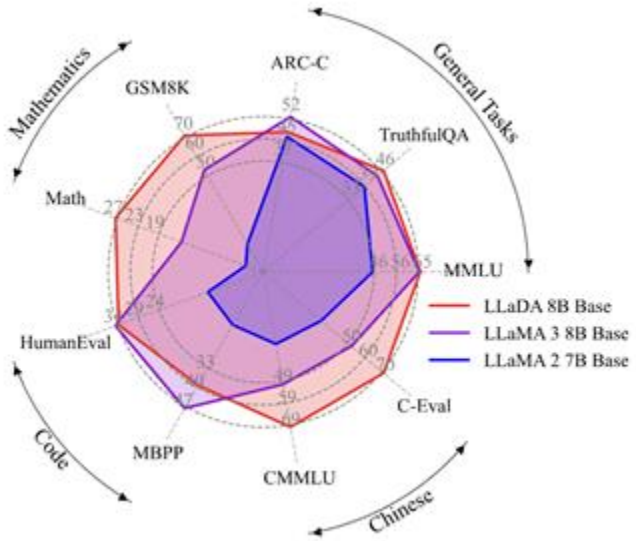
Mercury

Can be more accurate

Significantly faster

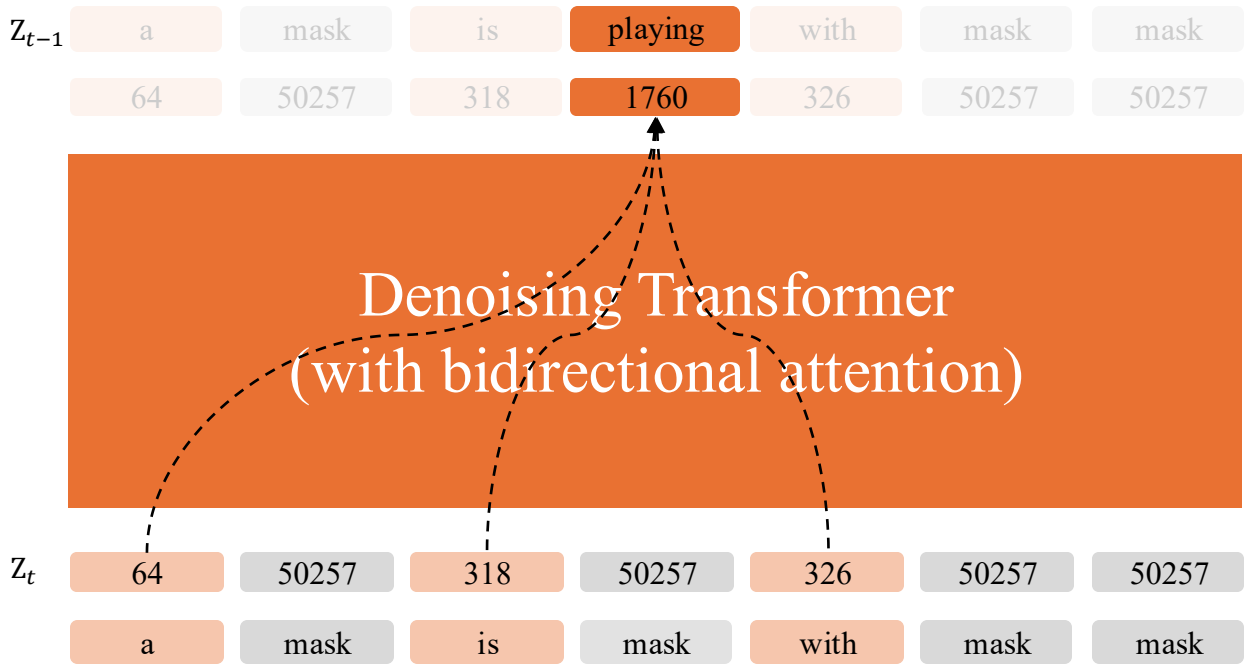


LLaDA



Gemini Diffusion: https://deepmind.google/models/gemini-diffusion/
LLaDA: https://arxiv.org/abs/2502.09992
Mercury: https://www.inceptionlabs.ai/

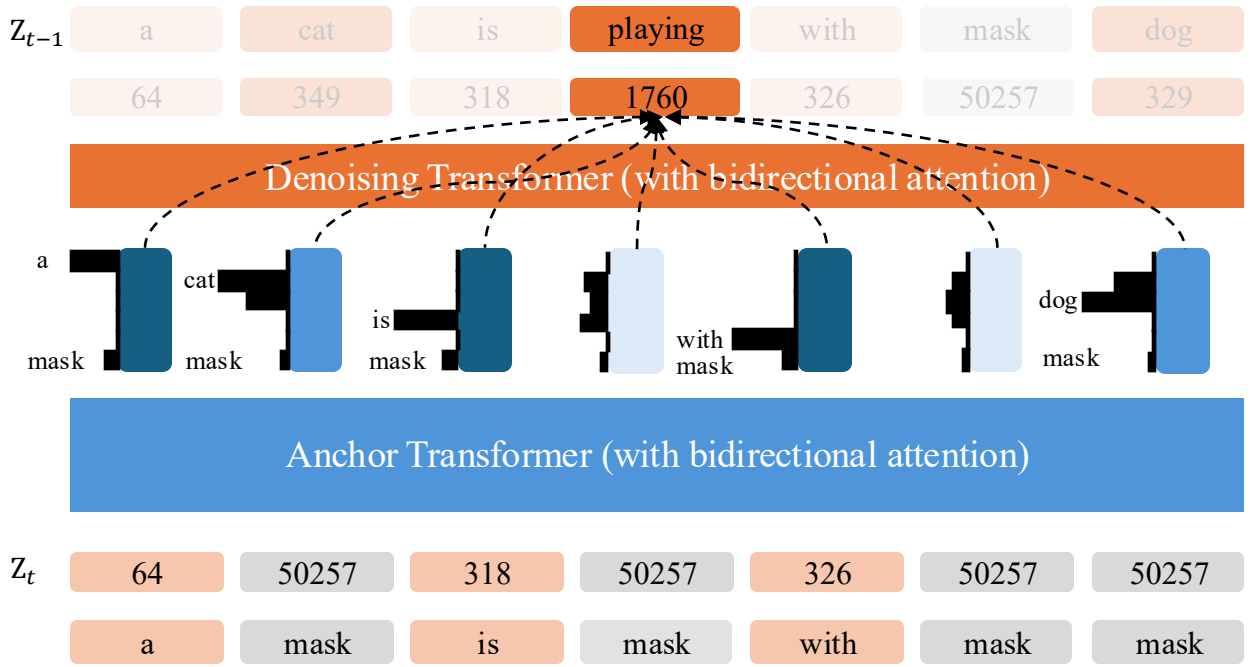
Standard Approach: Masked Diffusion Language Model



- **Denoising Transformer**: Unmasks using already unmasked tokens
- **Our approach:** Improve context using semantically important tokens

D3PM: https://arxiv.org/abs/2107.03006 MDLM: https://s-sahoo.com/mdlm/ MD4: https://arxiv.org/abs/2406.04329

Our Approach: Anchored Diffusion Language Model



- **Anchor Transformer**: Outputs a sequence of anchor predictions
  - Mixture of important tokens interpreted as soft samples
- **Denoising Transformer**: Unmasks tokens using anchored predictions

Our Key Idea: Anchoring

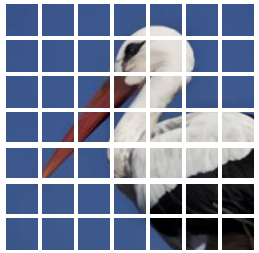
Anchors: Tokens whose inclusion as conditioning variables yields a substantial reduction in the conditional entropy of the remaining tokens

Examples:

1. To decode the sentence: "a **cat** is playing with a **dog**"
   - Tokens '**cat**' and '**dog**' are more useful than 'a' or 'is'

2. To solve the math question: "Janet's ducks lay **16** eggs per day. She eats **three** for breakfast every morning and bakes muffins for her friends every day with **four**. She sells the remainder at the farmers' market daily for $**2** per fresh duck egg. How much in dollars does she make every day at the farmers' market?"
   - Numbers **'16', 'three', 'four',** and **'2'** are more useful than 'breakfast' or 'muffins'

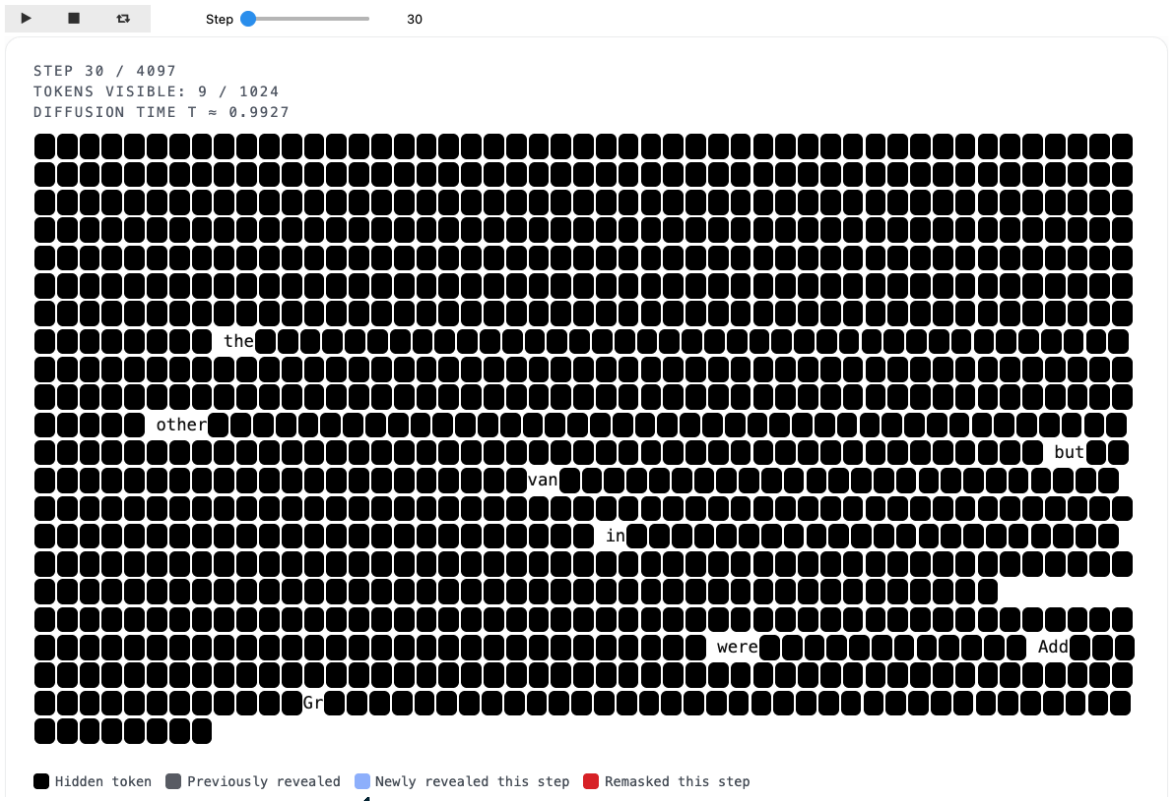
3. To reconstruct the image:



- Tokens in the beak or body (foreground) are more useful than tokens from blue background

Credit: Math question from the GSM8K dataset

Masked DLM vs Anchored DLM – Inference Illustration

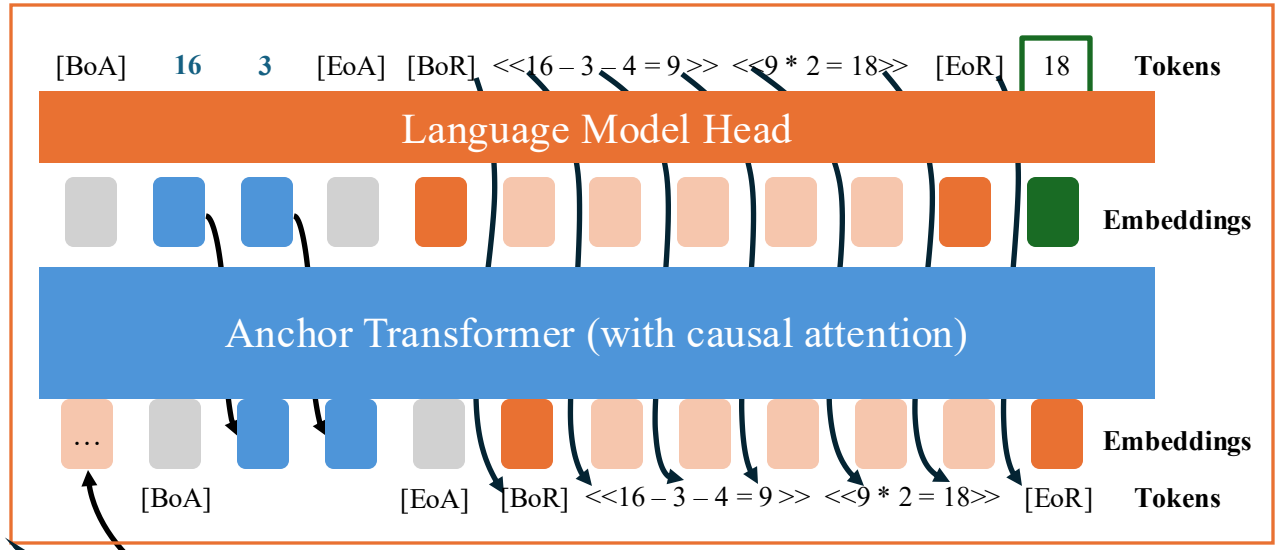


MDLM at time-step 30 / 4097

ADLM at time-step 30 / 4097

- Anchoring has 2 main advantages:
  - unmasks key words first
  - unmasks many tokens in parallel

Anchored Autoregressive Model: Training and Inference

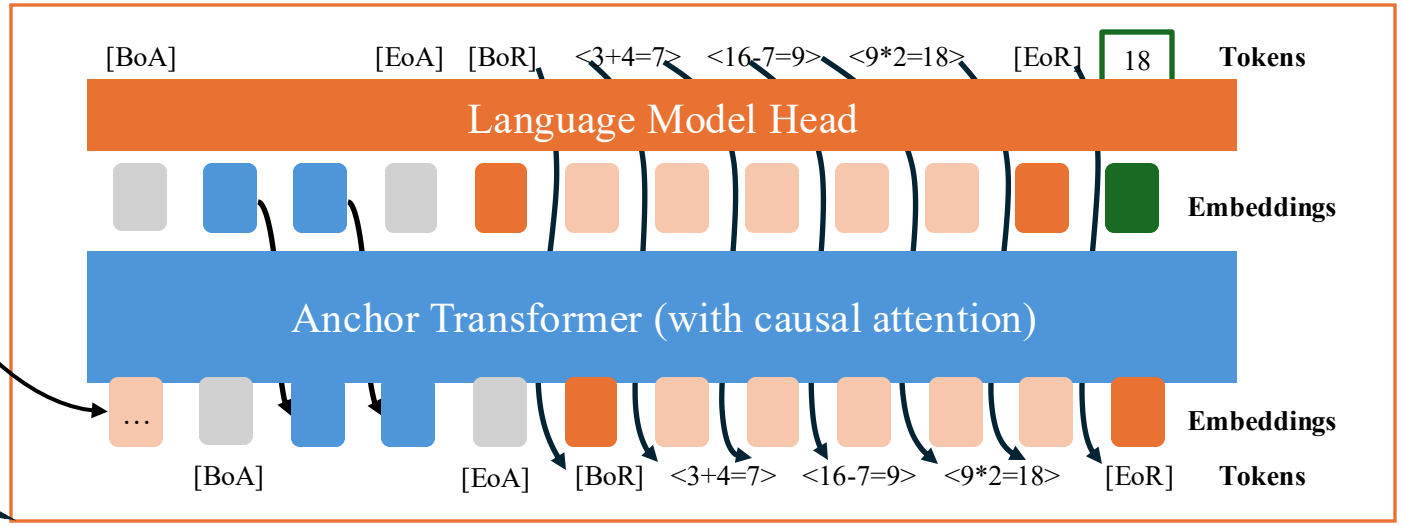


Anchoring enables "look-ahead" planning, unlike the standard left-to-right decoding of chain-of-thought (CoT) reasoning

Janet's ducks lay **16** eggs per day. She eats **three** for breakfast every morning and bakes muffins for her friends everyday with **four**. She sells the remainder at the farmers' market daily for $**2** per fresh duck egg. How much in dollars does she make every day at the farmers' market?

During Supervised Fine Tuning of LLM

During LLM Inference



📄 **Paper:**
openreview.net/pdf?id=E8adS5srd

🌐 **Project Page:**
anchored-diffusion-llm.github.io

💻 **Source Code:**
github.com/LituRout/ADLM