Large Language Diffusion Models

상태	완료
Journal	arXiv(China)
Year	2025.02
Summary	랜덤 마스킹과 복원 학습, Reversal Reasoning 에 강점을 갖는 diffusion 기반 언어모델
Limitations	기존 파이프라인을 사용하여 최적화는 진행하지 않음, 최적화 및 멀티모달 학습 필요
Link	https://arxiv.org/pdf/2502.09992
category	Fine-tuning LLM Pre-training

TWITTER BANNER TITLE META TAG
TWITTER BANNER DESCRIPTION META TAG
https://ml-gsai.github.io/LLaDA-demo/

Introduction

- Autoregressive model(ARM)
  - next-token prediction paradigm (ex. RNN, Transformers, BERT, GPT)
  - autoregressive 모델이 LLM이 보이는 지능을 달성하기 위한 유일한 실행 가능한 경로인가?
    - 저자들은 여기에 의문을 가짐 why? → autoregressive 모델은 한계가 있기 때문
  - LLM 필수 속성을 뒷받침하는 근간은 autoregressive 자체가 아니라 생성 모델링 원칙임
  - but, LLM의 내재적 한계는 autoregressive 특성에서 직접적으로 기인함
    - 순차적인 생성 → 높은 계산 비용, 역방향 추론의 한계 등등...
- ARM의 한계
  - 확장성
    - transformer, model, data size, fisher consistency 간 상호작용의 결과임
    - 즉, ARM 고유의 결과가 아님
    - 근거 : 시각 데이터에서의 diffusion transformer 성공
  - instruction-following / in-context learning
    - 구조적으로 일관된 언어적 과제에 대한 모든 적절한 조건부 생성 모델의 내재적 속성
    - 즉, ARM의 독점적 장점이 아님
- **LLaDA : Large Language Diffusion with mAsking**
  - 전통적인 ARM과 달리, LLaDA는 **이산적 무작위 마스킹 프로세스**를 포함하고 mask predictor를 훈련시켜 **reverse 프로세스를 근사**하는 masked diffusion model(MDM)을 활용함
  - 데이터 준비, pre-training, SFT, evaluation의 표준 파이프라인 채택
  - **양방향 의존성(bidirectional dependencies)**
    - 기존의 많은 LLM들은 자기회귀 모델처럼 왼쪽에서 오른쪽 또는 오른쪽에서 왼쪽으로 한 방향으로만 텍스트를 예측
    - but, LLaDA는 텍스트의 앞과 뒤를 동시에 고려할 수 있는 양방향 의존성을 가질 수 있음
    - 즉, 텍스트의 전체 문맥을 더 잘 이해하고 예측할 수 있음

- **모델 분포와 로그 우도(log-likelihood)**
  - LLaDA는 학습 과정에서 모델 분포를 만들고, 이 분포가 얼마나 잘 데이터를 설명하는지를 평가하는 로그 우도를 최적화함
  - 여기서 하한(lower bound)을 최적화한다는 것은 실제 모델의 우도를 직접적으로 최적화하는 대신, 이를 근사하는 방식으로 최적화하여 더 효율적으로 학습한다는 의미

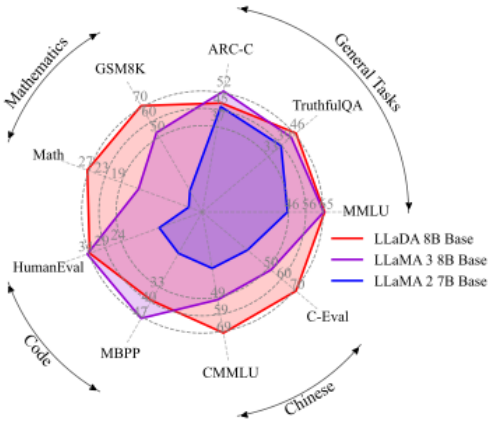

Figure 1. **Zero/Few-Shot Benchmarks.** We scale LLaDA to an unprecedented size of 8B parameters from scratch, achieving competitive performance with strong LLMs (Dubey et al., 2024).

- 기여점
  - 확장성 : 10^{23} FLOPs 계산 효율성 ↔ ARM baseline과 비교되는 성능
  - in-context learning : zero/few-shot learning task에서 LLaMA2 7B 이김 ↔ LLaMA3 8B랑 비슷
  - instruction-following : SFT 이후 지침을 따르는 능력 향상 → multi-turn dialogue에서 확인 가능
  - reversal reasoning : 역추론의 저주를 극복함 ↔ GPT-4o보다 뛰어난 성과

Approach

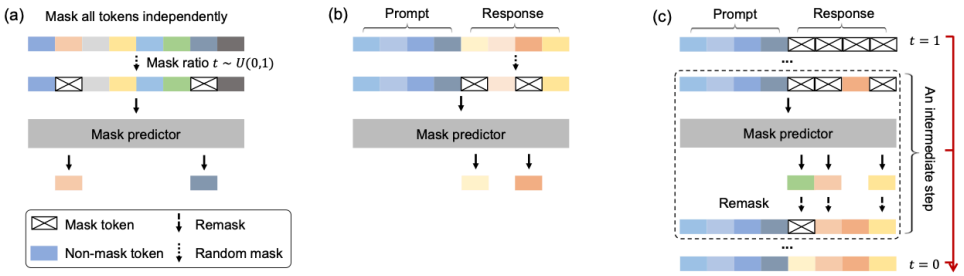

Figure 2. **A Conceptual Overview of LLaDA.** (a) Pre-training. LLaDA is trained on text with random masks applied independently to all tokens at the same ratio $t \sim U[0, 1]$. (b) SFT. Only response tokens are possibly masked. (c) Sampling. LLaDA simulates a diffusion process from $t = 1$ (fully masked) to $t = 0$ (unmasked), predicting all masks simultaneously at each step with flexible remask strategies.

Probabilistic Formulation

- LLaDA는 데이터를 마스크하는 전방 과정(forward process) 과 이를 역으로 복원하는 역방향 과정(reverse process) 을 통해 학습
- 이 과정은 자기회귀 방식과 달리 양방향 의존성을 가지며, 매우 유연한 학습 방식을 제공

Pre-training

- forward process
  - LLaDA는 random masking 기법을 사용하여 모든 토큰을 독립적으로 마스킹함

    - 마스킹 비율 t는 0과 1 사이의 균등 분포(Uniform Distribution: U[0,1])에서 샘플링됨
    - 이 과정에서 마스킹된 토큰과 마스킹되지 않은 토큰이 랜덤하게 선택
        - 기존의 BERT 같은 Masked Language Model(MLM)과 달리, **고정된 비율이 아닌 동적인 마스킹 비율 사용**
- reverse process
    - Mask predictor(transformer)는 주어진 마스킹된 데이터를 보고 **원래의 문장을 복원하도록 학습**
    - cross-entropy loss를 사용하여 마스킹된 토큰에 대한 예측 성능 평가

$$\mathcal{L}(\theta) \triangleq -\mathbb{E}_{t,x_0,x_t}\left[\frac{1}{t}\sum_{i=1}^{L}\mathbf{1}[x_t^i=\mathbf{M}]\log p_\theta(x_0^i \mid x_t)\right]$$

🔥 LLaDA는 문장의 일부를 **랜덤하게 마스킹한 후, 이를 복원**하는 방식으로 학습

Supervised Fine-Tuning

- SFT 단계에서는 Prompt와 Response 형식의 데이터 사용
- p_0인 것을 보면 prompt는 항상 전체를 주고, **response만 masking 적용**
- Mask predictor는 응답을 복원하는 방식으로 학습
- <EOS> 토큰을 통해 모든 데이터의 길이를 동일하게 유지 → 450만 pair datasets

$$-\mathbb{E}_{t,p_0,r_0,r_t}\left[\frac{1}{t}\sum_{i=1}^{L'}\mathbf{1}[r_t^i=\mathbf{M}]\log p_\theta(r_0^i \mid p_0, r_t)\right]$$

🔥 SFT 단계에서는 모델이 **프롬프트를 기반으로 올바른 응답을 예측하는 능력**을 학습

Inference

- LLaDA는 diffusion process를 통해 텍스트를 생성
- 처음에는 모든 토큰이 마스킹된(t=1) 상태에서 시작
- 이후 여러 단계에 걸쳐 **점진적으로 마스크를 해제하며 원래 텍스트를 복원**
- 각 단계에서 Mask predictor가 마스킹된 토큰을 예측하며, 리마스킹(remask) 전략을 사용하여 일부 토큰을 다시 마스킹하면서 점진적으로 정확도를 높임
- 샘플링 단계와 샘플링 길이를 하이퍼파라미터로 설정하여 다양한 텍스트를 생성

🔥 샘플링 과정에서는 모델이 **완전히 마스킹된 상태에서 점진적으로 원래 문장을 복원하는 방식**으로 텍스트를 생성

Experiments

Scalability of LLaDA on Language Tasks

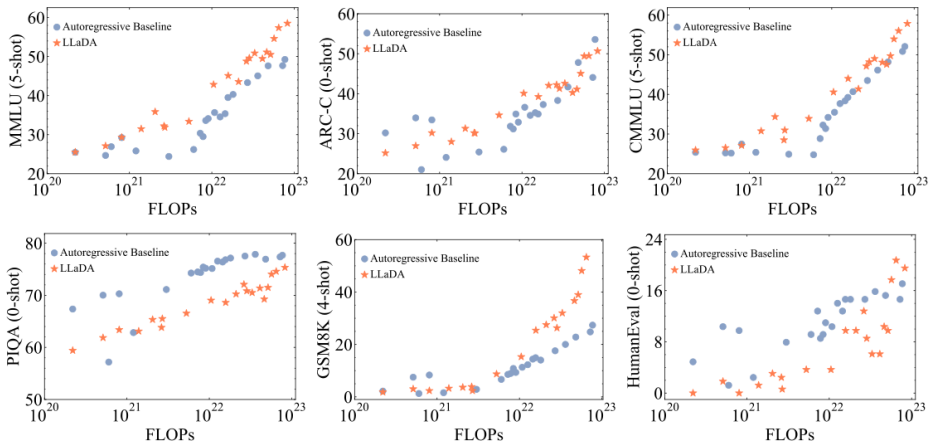

Figure 3. **Scalability of LLaDA.** We evaluate the performance of LLaDA and our ARM baselines trained on the same data across increasing computational FLOPs. LLaDA exhibits strong scalability, matching the overall performance of ARMs on six tasks.

- **크기가 커질수록 확장성**이 뛰어난 성능
- **MMLU**나 **GSM8K**와 같은 벤치마크에서 **ARM 모델**과 비교하여 우수한 확장성

Benchmark Results

Table 1. **Benchmark Results of Pre-trained LLMs.** * indicates that LLaDA 8B Base, LLaMA2 7B Base, and LLaMA3 8B Base are evaluated under the same protocol, detailed in Appendix B.5. Results indicated by † and ¶ are sourced from Chu et al. (2024); Yang et al. (2024) and Bi et al. (2024) respectively. The numbers in parentheses represent the number of shots used for evaluation. "-" indicates unknown data.

	LLaDA 8B*	LLaMA3 8B*	LLaMA2 7B*	Qwen2 7B†	Qwen2.5 7B†	Mistral 7B†	Deepseek 7B¶
Model	Diffusion	AR	AR	AR	AR	AR	AR
Training tokens	2.3T	15T	2T	7T	18T	-	2T
General Tasks							
MMLU	**65.9** (5)	65.4 (5)	45.9 (5)	70.3 (5)	74.2 (5)	64.2 (5)	48.2 (5)
BBH	49.8 (3)	**57.6** (3)	37.3 (3)	62.3 (3)	70.4 (3)	56.1 (3)	39.5 (3)
ARC-C	47.9 (0)	**53.1** (0)	46.3 (0)	60.6 (25)	63.7 (25)	60.0 (25)	48.1 (0)
Hellaswag	72.5 (0)	**79.1** (0)	76.0 (0)	80.7 (10)	80.2 (10)	83.3 (10)	75.4 (0)
TruthfulQA	**46.4** (0)	44.0 (0)	39.0 (0)	54.2 (0)	56.4 (0)	42.2 (0)	-
WinoGrande	74.8 (5)	**77.3** (5)	72.5 (5)	77.0 (5)	75.9 (5)	78.4 (5)	70.5 (0)
PIQA	74.4 (0)	**80.6** (0)	79.1 (0)	-	-	-	79.2 (0)
Mathematics & Science							
GSM8K	**70.7** (4)	53.1 (4)	14.3 (4)	80.2 (4)	85.4 (4)	36.2 (4)	17.4 (8)
Math	**27.3** (4)	15.1 (4)	3.2 (4)	43.5 (4)	49.8 (4)	10.2 (4)	6.0 (4)
GPQA	**26.1** (5)	25.9 (5)	25.7 (5)	30.8 (5)	36.4 (5)	24.7 (5)	-
Code							
HumanEval	33.5 (0)	**34.2** (0)	12.8 (0)	51.2 (0)	57.9 (0)	29.3 (0)	26.2 (0)
HumanEval-FIM	**73.8** (2)	73.3 (2)	26.9 (2)	-	-	-	-
MBPP	38.2 (4)	**47.4** (4)	18.4 (4)	64.2 (0)	74.9 (0)	51.1 (0)	39.0 (3)
Chinese							
CMMLU	**69.9** (5)	50.7 (5)	32.5 (5)	83.9 (5)	-	-	47.2 (5)
C-Eval	**70.5** (5)	51.7 (5)	34.0 (5)	83.2 (5)	-	-	45.0 (5)

Table 2. **Benchmark Results of Post-trained LLMs.** LLaDA only employs an SFT procedure while other models have extra reinforcement learning (RL) alignment. * indicates that LLaDA 8B Instruct, LLaMA2 7B Instruct, and LLaMA3 8B Instruct are evaluated under the same protocol, detailed in Appendix B.5. Results indicated by † and ¶ are sourced from Yang et al. (2024) and Bi et al. (2024) respectively. The numbers in parentheses represent the number of shots used for in-context learning. "-" indicates unknown data.

	LLaDA 8B*	LLaMA3 8B*	LLaMA2 7B*	Qwen2 7B†	Qwen2.5 7B†	Gemma2 9B†	Deepseek 7B¶
Model	Diffusion	AR	AR	AR	AR	AR	AR
Training tokens	2.3T	15T	2T	7T	18T	8T	2T
Post-training	SFT	SFT+RL	SFT+RL	SFT+RL	SFT+RL	SFT+RL	SFT+RL
Alignment pairs	4.5M	-	-	0.5M + -	1M + 0.15M	-	1.5M + -
General Tasks							
MMLU	65.5 (5)	**68.4** (5)	44.1 (5)	-	-	-	49.4 (0)
MMLU-pro	37.0 (0)	**41.9** (0)	4.6 (0)	44.1 (5)	56.3 (5)	52.1 (5)	-
Hellaswag	74.6 (0)	**75.5** (0)	51.5 (0)	-	-	-	68.5 (-)
ARC-C	**88.5** (0)	82.4 (0)	57.3 (0)	-	-	-	49.4 (-)
Mathematics & Science							
GSM8K	**78.6** (4)	78.3 (4)	29.0 (4)	85.7 (0)	91.6 (0)	76.7 (0)	63.0 (0)
Math	26.6 (0)	**29.6** (0)	3.8 (0)	52.9 (0)	75.5 (0)	44.3 (0)	15.8 (0)
GPQA	31.8 (5)	**31.9** (5)	28.4 (5)	34.3 (0)	36.4 (0)	32.8 (0)	-
Code							
HumanEval	47.6 (0)	**59.8** (0)	16.5 (0)	79.9 (0)	84.8 (0)	68.9 (0)	48.2 (-)
MBPP	34.2 (4)	**57.6** (4)	20.6 (4)	67.2 (0)	79.2 (0)	74.9 (0)	35.2 (-)

- LLaDA 8B는 **LLaMA2**와 **LLaMA3**와 같은 최신 모델과 **비교하여 유사한 성능**
- Qwen과 비교하면 그렇게 높은 성능은 아니지만, diffusion 모델로 이정도의 성능을 낼 수 있음을 보여줌
- 모델 학습에서 최적화를 하지 않고
- **수학**, **중국어** 등의 분야에서 **우수한 성능**

Reversal Reasoning and Analyses

Table 3. **Comparison in the Poem Completion Task.**

	Forward	Reversal
GPT-4o (2024-08-06)	**82.7**	34.3
Qwen2.5 7B Instruct	75.9	38.0
LLaDA 8B Instruct	48.8	**42.4**

- **시가의 역전 문제**에서 LLaDA는 **GPT-4o**보다 더 나은 성능
- LLaDA는 **자기회귀 모델**보다 **균형 잡힌 성능**을 보여, **역전 추론**에서도 뛰어난 성과

Case Studies

Table 4. **Visualization of the Sampling Process and a Generated Multi-round Dialogue.** In the response of LLaDA, darker colors indicate tokens predicted in the later stages of sampling, while lighter colors correspond to earlier predictions.

Sampling Process	
User	Lily can run 12 kilometers per hour for 4 hours. After that, she runs 6 kilometers per hour. How many kilometers can she run in 8 hours?
LLaDA	In 4 hours , Lily runs 12 * 4 = 48 kilometers . After 4 hours she runs 6 * 4 = 24 kilometers . In total she runs 48 + 24 = 72 kilometers in 8 hours . Result 72
Multi-round Dialogue Case	
User	Could you please share with me what the initial two lines of the renowned poem 'The Road Not Taken'?
LLaDA	Certainly! The initial two lines of the renowned poem "The Road Not Taken" by Robert Frost are: "Two roads diverged in a yellow wood, And sorry I could not travel both." These lines set the stage for the poem, introducing the idea of a speaker facing a choice between two paths.
User	Please help me translate into Chinese.
LLaDA	Sure! The first two lines of "The Road Not Taken" by Robert Frost can be translated into Chinese as: "两条路分岔在黄色的树林中，遗憾我不能同时走"
User	Please translate into German.
LLaDA	Sure! The first two lines of "The Road Not Taken" by Robert Frost can be translated into German as: "Zwei Wege trennten sich im gelben Wald, und leider konnte ich nicht beide tun."
User	Please also write a poem about life choices. I require 5 sentences, and the first word of each sentence starts with C as the first letter.
LLaDA	Certainly! Here's a poem about life choices, with each sentence starting with the letter "C": Choices cascade through life's journey, Creating connections with each decision, Careful consideration leads to better paths, Courageous choices shape our destiny, Choosing wisely guides us through the years.

- **다중 회전 대화**나 **응답 생성**에서 유연하게 처리하는 능력
- **명령어 따르기** 및 **문맥 학습**에서도 좋은 결과

Related Work

- 기존의 **확산 모델**(Diffusion Model)
- **대형 언어 모델**에서의 **자기회귀 모델**의 한계
- **LLaDA**가 기존 모델들과 비교할 때 어떤 점에서 차별화되는지 설명

Conclusion and Discussion

- LLaDA는 **자기회귀 모델**을 대체할 수 있는 **확산 모델**로서의 가능성
- **확장성**, **명령어 따르기**, **문맥 학습** 등에서 뛰어난 성능을 발휘
- 향후 연구 방향
  - **멀티모달 학습:** LLaDA를 비언어적 데이터(예: 이미지, 오디오)로 확장
  - **강화학습 기반 정렬(RLHF):** 모델의 **사용자 의도 이해력 향상**을 위한 후처리 탐색
  - **샘플링 최적화:** 샘플링 단계에서의 **리마스킹 전략 및 하이퍼파라미터 최적화** 연구
  - **모델 크기 확장:** 8B 이상의 모델로 실험하여 **궁극적인 성능 한계 탐색**
- 즉,
  - **ARMs:** "앞 단어"만 보고 "다음 단어"를 예측
  - **LLaDA:** "문장 전체"를 보고 "모든 마스킹된 단어"를 동시에 예측