



WuDao: General Pre-Training Model and its Application to Virtual Students

Tang Jie

Tsinghua University

Beijing Academy of Artificial Intelligence (BAAI)

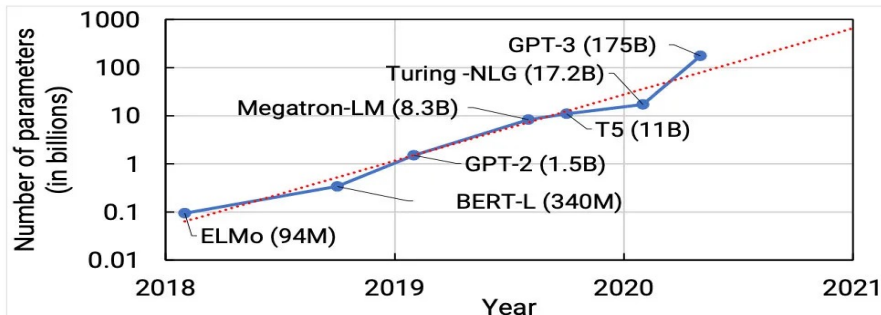


The slides is available at

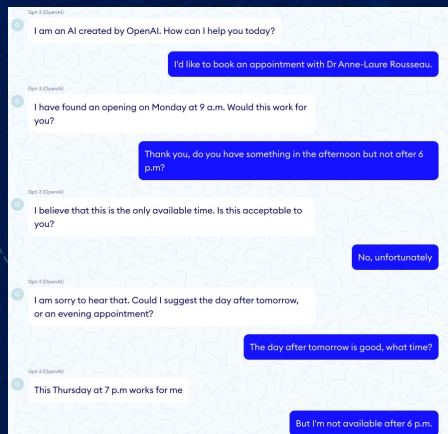
<http://keg.cs.tsinghua.edu.cn/jietang> (or Google "Jie Tang")

● GPT-3

- ◆ **GPT-3: OpenAI released GPT-3 in 2020/5 with super-scale parameters: 175B**
- ◆ **Google's Switch Transformer: 1.6T**



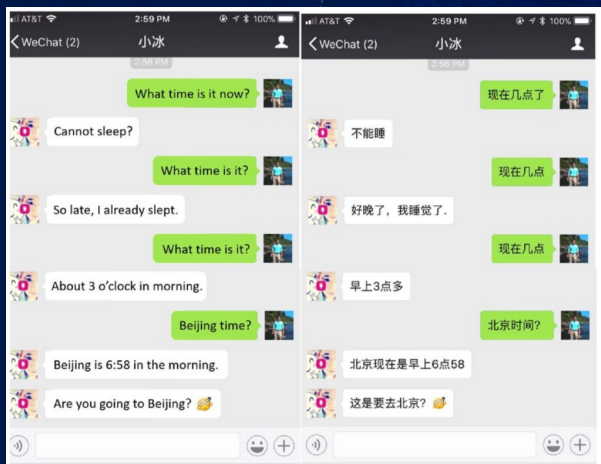
GPT-3 for QA



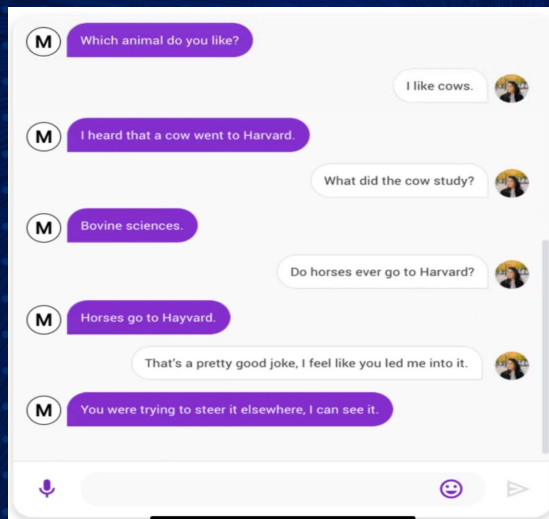
DALL·E: (Generating Images by Text)

a snail made of harp

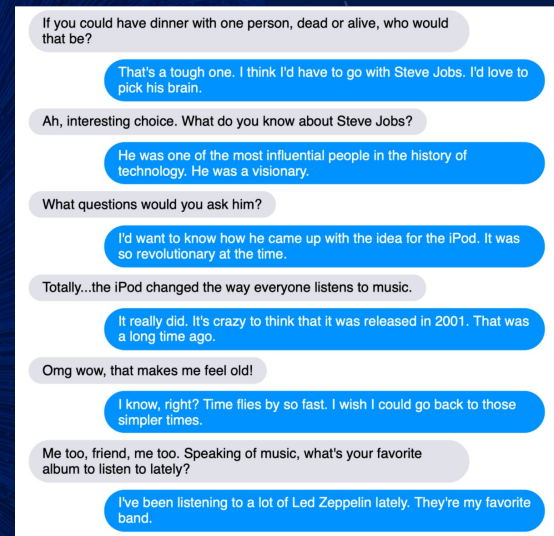
Open-domain Chatbot



Xiaolce (2018)



Meena by Google (2020)



Blender by Facebook (2020)

Chatbot w/ Grounded Knowledge

No reference

[Background Description]

2022年北京冬季奥运会abstract: 第24届冬季奥林匹克运动会, 即2022年北京冬季奥运会, 是由中国举办的国际性奥林匹克赛事, 于2022年2月4日开幕,

The 24th Winter Olympic Games, namely the Beijing 2022 Winter Olympic Games, is an international Olympic event held by China and opened on February 4, 2022,

Eileen Gu is so good!
谷爱凌好厉害呀

是呀, 我最喜欢她了。
Yes, I like her best.

哈哈, 她可是滑雪界的大魔王呀, 尤其是自由式滑雪滑的杠杠的。
Hahaha, she is the ace of skiing, especially freestyle skiing.

[Knowledge QA-pair]
Q: 谷爱凌是谁? Who is **Eileen Gu**
A: 谷爱凌, 2003年出生于美国加利福尼亚州旧金山市, 中国自由式滑雪运动员。...
Eileen Gu, born in 2003 in San Francisco, California, is a Chinese freestyle skier. ...
<http://xlore.org/instance/bdi10768871>

I also want to learn **skiing**. How can I learn it well?
我也想学滑雪, 怎样才能学好呢

那你要去刻苦训练呀!
Then you have to train hard!

先学习基本姿势, 滑行过程中姿势要正确, 动作要轻盈、敏捷、协调, 控制重心平稳, 两脚用力均衡。
Learn the basic posture first, which should be correct during sliding. Your movements should be light, agile and coordinated. Control your body weight and balance your feet.

[Knowledge QA-pair]
Q: 什么是滑雪? What is **skiing**?
A: 滑雪运动是运动员把滑雪板装在靴底上在雪地上进行速度、跳跃和滑降的竞赛运动。.....Skiing is a sport in which skis are mounted on the soles of shoes to speed, jump and slide down on the snow.
<http://xlore.org/instance/bdi12799316>

The Power of Generative Models

Domain-**specific**

unconditional generation

StyleGAN2 on FFHQ



Domain-**specific**

text-to-image generation

- DF-GAN on CUB
- This bird has a white belly and breast, with a blue crown and nape.



Domain-general

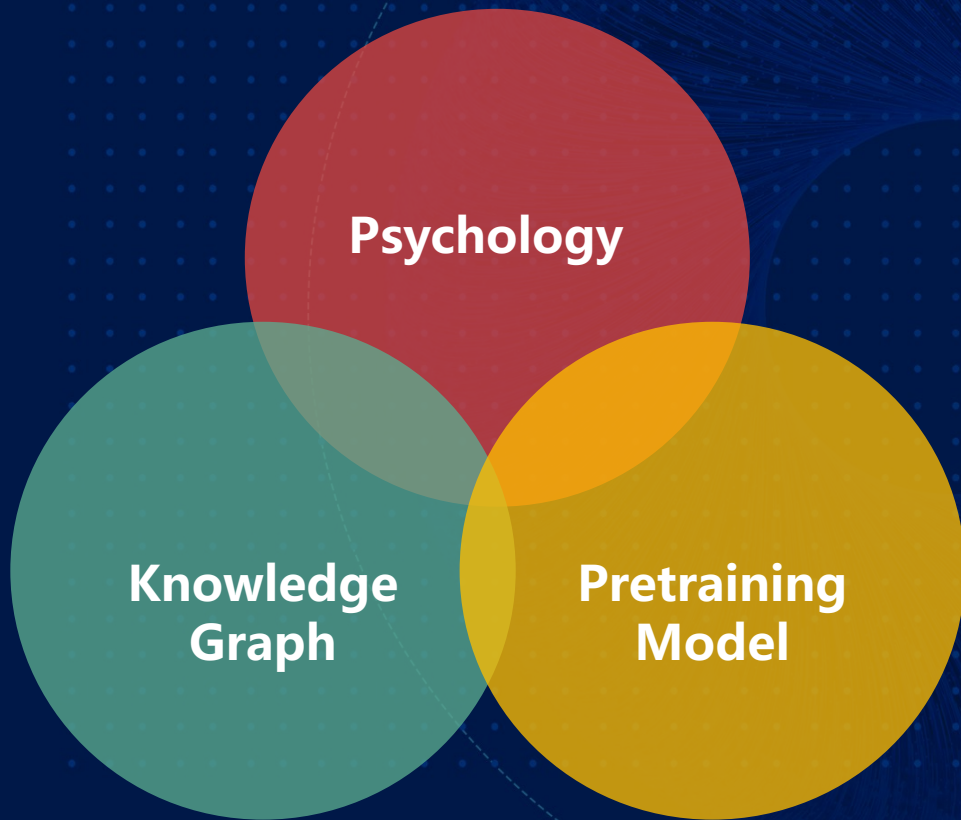
text-to-image generation

- DF-GAN on COCO
- Close-up of a man eating a piece of pizza while holding a plate.

CogView2



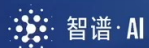
- Virtual Student



Virtual Front Desk



AI driven Sign Language



LUSTER 凌云光

BRTV



- AI driven Sign Language



AI Helps Interview

多面

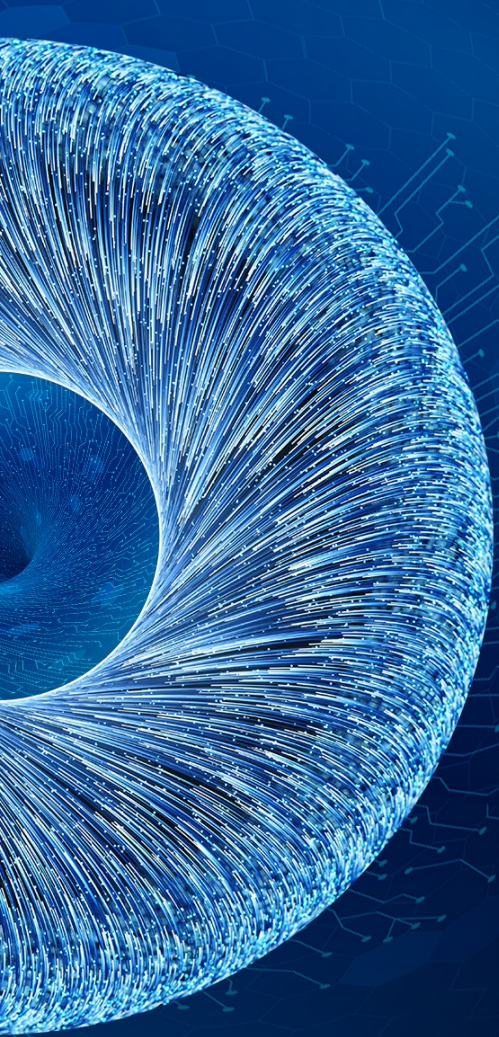


答疑环节 剩余时间 04:00

比如你可以问我以下问题或者其他问题。如果你没有疑问，你可以跟我说“再见”或者点击“结束面试”按钮。

- 1: 公司目前在全国各个项目团队都是多少人?
- 2: 咱们公司现在这个岗位的五险一金是怎么交的?
- 3: 这个项目组目前的盈利业务有哪几条线?

结束面试



WUDAO 2.0 is one of the largest models

1.75 Trillion

10X larger than GPT-3 parameters

● What is WuDao 2.0

01

1.75 Trillion Parameters

Largest

02

both text and images

Universal

03

train on a supercomputer

Domestic

04

Bilingual (Cn and En) data: 4.9T text and images

Knowledge

• What is WuDao 2.0

FastMoE

Support complex balance strategies such as Switch and GShard
Support different experts and different models.

Deploy in Alibaba PAI platform
Explore the application to the Alipay intelligent service system

Training on a
supercomputer



FastMoE

- ✓ Redesign all operators
- ✓ Efficient communication strategy
- ✓ MoE training with tens of thousands of experts

• QA with WuDao

Question: Which moment did you want to live in forever?

User: High-school student

WUDAO: In an instant I recall the second year of high school. That afternoon, I took a sunbath on the playground for a long time. I then played basketball downstairs with my classmates. As the sun set slowly, the three of us stood together on the corridor, leaning on the railing to face the sunset. I 've always wanted to live in that moment forever, but unfortunately I never encountered a similar moment again. I just want to live in that moment.

● POEM

<https://wudao.aminer.cn/turing-test/v2/>



作诗图灵测试

→

Easy

Hard

Lunatic

Extra

在作诗图灵测试的Easy模式中，您将会被展现5组诗歌（包括标题、作者及内容），每组包括1首由诗人创作的诗歌和1首AI创作的诗歌，请选择您认为由人创作的诗歌。所有组选择完成后，您将会得知有多少组结果正确。

● POEM

<https://wudao.aminer.cn/turing-test/v2/>



比特币

外挖无穷洞, 机神犹未休。
卡中窥币影, 池里验沙流。
屡载吸金主, 孤深渍盗求。
方知区块链, 本是古来游。

Bitcoin

Mining an infinite hole outside
The god of machine is still not satisfied
Illusion of bitcoins seen inside the GPU
Verified in pools to confirm the workflow
Several years it has attracted many investors
Deep and limited, even the robbers pursue
Eventually it turns out that Blockchain
May only be a financial trick from ancient

● POEM

<https://wudao.aminer.cn/turing-test/v2/>



WuDao V.S. Poet

文源

礪溪连灞水，商岭接秦山。
青汉不回驾，白云长掩关。
雀喧知鹤静，凫戏识鸥闲。
却笑南昌尉，悠悠城市间。

V.S.

老翁岩畔卧，数载避喧惊。
棋局尘生涧，茶烟景近屏。
刀枪犹野外，吟啸亦琴清。
更有深交意，相期岁晚行。

文澜

晨趋禁掖暮郊园，
松桂苍苍烟露繁。
明月上时群动息，
雪峰高处正当轩。

V.S.

罗隐虽来频道寺，
蓬门未肯便归休。
殷勤与见我无倦，
竟日相随不暂留。

文溯

初辞云浦寂，再别竺岩难。
各分当时意，空驰海远寒。
依稀真弟子，守至四明滩。
临别殷勤语，重违誓去盘。

V.S.

上国随缘住，来途若梦行。
浮天沧海远，去世法舟轻。
水月通禅观，鱼龙听梵声。
惟怜一灯影，万里眼中明。

文汇

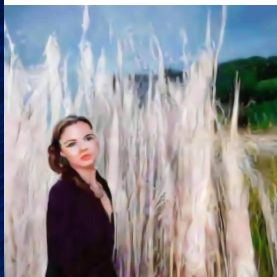
近年容尚白，卿士累资名。
投契顾怜旧，照乘拟次行。
常嫌官爵小，更睹道心清。
多谢白公意，作为肺腑情。

V.S.

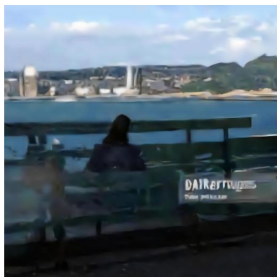
太空秋色凉，独鸟下微阳。
三径池塘静，六街车马忙。
渐能高酒户，始是入诗狂。
官冷且无事，追陪慎莫忘。

● Draw Pictures

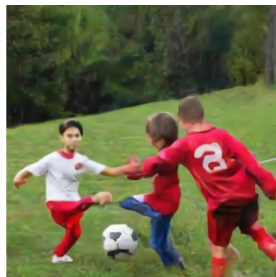
<https://wudao.aminer.cn/CogView/index.html>



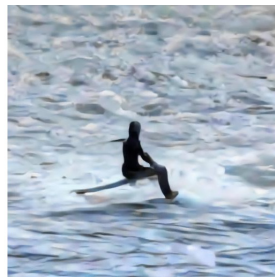
A woman in a black and purple dress poses in front of some tall grass.



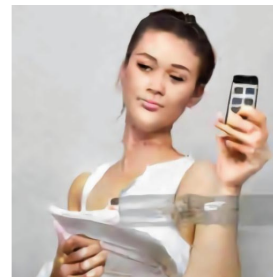
A woman is on a bench overlooking the city.



A couple of young boys playing a game of soccer.



a man that is on a surfboard in some water.



A women in a white blouse is holding a remote in her hands.



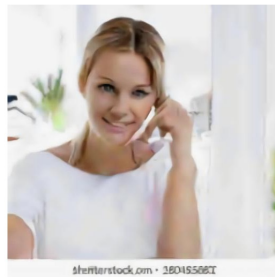
A bird perched on top of a leafless tree under a blue sky.



a clock hanging outside of a house in a nice neighborhood.



A red bus is driving on the road.

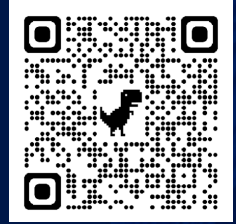


A beautiful young blond woman talking on a phone.



A red bowl filled with food and leafy greens.

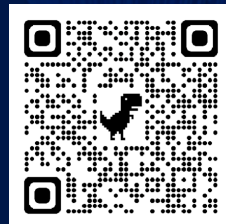
• Draw Pictures-Image completion



- A girl with a surgical mask
- A girl wear a tie



• Draw Pictures-Image completion



- A man with a red ball
- 一个胖子在吃一碗面



Turing test

<https://wudao.aminer.cn/turing-test/v2/>



WUDAO
TURING TEST

📖 Tang POEM

📖 Song POEM

📖 Couplet

💬 Caption

💬 QA

💬 Writing

🖼️ Drawing

🖼️ Img Caption

悟道绘图图灵测试

基于跨模态预训练模型的图文生成

▶ 开始挑战



游戏模式



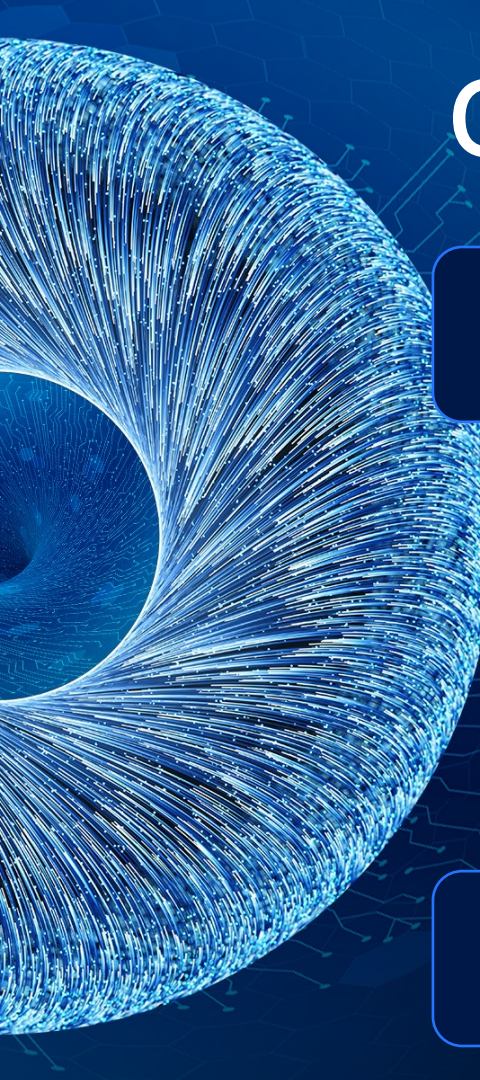
Easy



Fast



① 在图灵绘图的Easy模式中，您将会被展现5组图片及其标题，每组包括2张图片，其中包含1张真实图片和1张由AI生成的图片，请挑选出真实的图片。



Core Algorithms

Controllable Generation via Inverse Prompting

CogView: Text-to-Image
Generation

CogVideo: Text-to-Video
Generation

GLM: General Language Model Pretraining

How to build a trillion-scale model



GLM: General Language Model Pretraining with Autoregressive Blank Infilling



• Pretrained LMs and NLP Tasks

Framework	NLU	Cond. Gen.	Uncond. Gen.
Autoregressive	—	—	✓
Autoencoding	✓	×	×
Encoder-Decoder	—	✓	—
GLM	✓	✓	✓

None of the pretraining frameworks performs the best for all tasks.

- Autoregressive model: GPT, GPT-2, GPT-3
- Autoencoding model: BERT, RoBERTa, ALBERT
- Encoder-Decoder model: MASS, BART, PALM

• All NLP Tasks are Generations

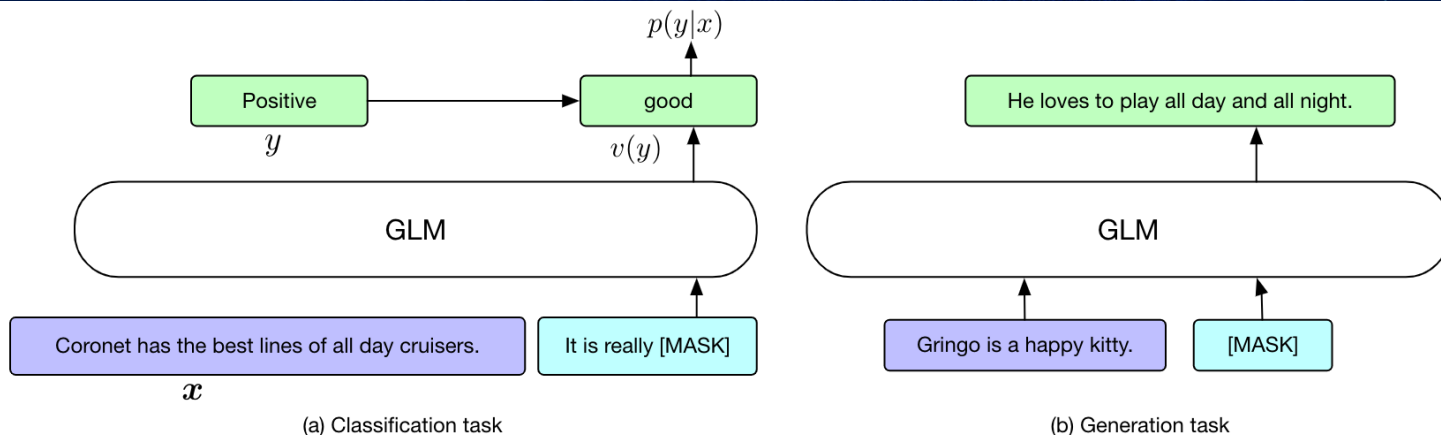


Figure 3. GLM finetune framework. (a) Formulation of the sentiment classification task as blank infilling with GLM. (b) GLM for text generation given the context. This can be the language modeling in the zero-shot setting, or seq2seq with fine-tuning.

NLU, Cond. Gen, Uncond. Gen can be unified into the GLM generation framework

• A New Pretraining Framework

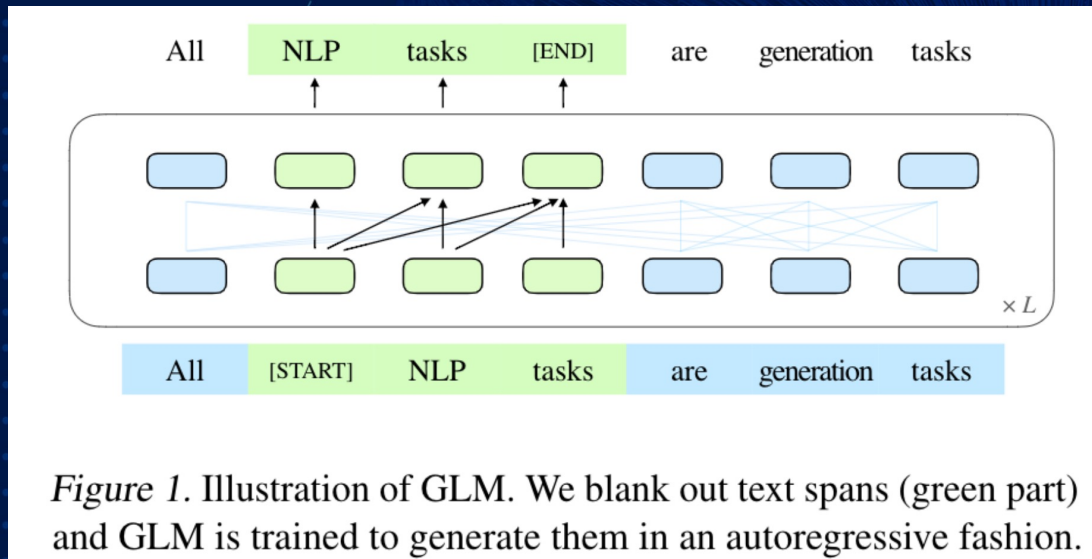


Figure 1. Illustration of GLM. We blank out text spans (green part) and GLM is trained to generate them in an autoregressive fashion.

Multi-task pretraining

1. Sample 15% in the middle as the generation objective
2. Sample 50-100% in the end as the generation objective

GLM: Autoregressive Blank Filling

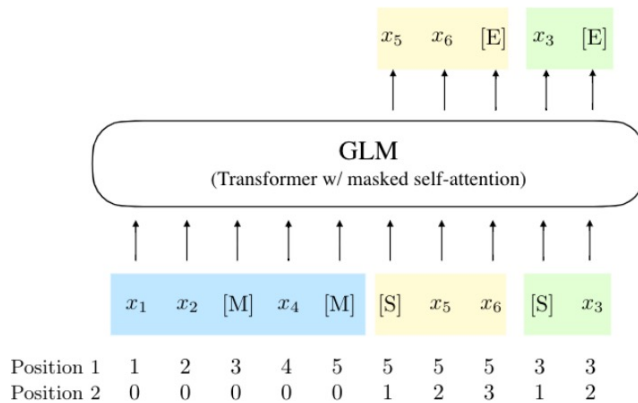
x_1 x_2 x_3 x_4 x_5 x_6

(a) Sample spans from the input text

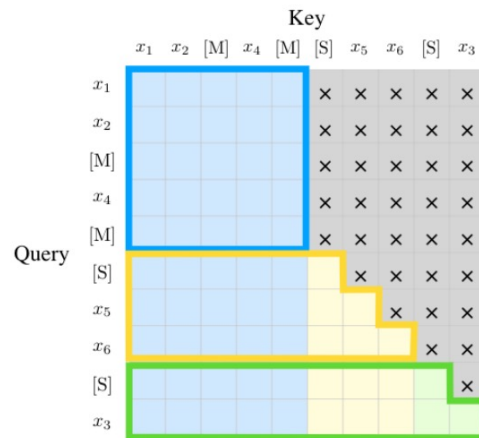
Part A: x_1 x_2 [M] x_4 [M]

Part B: x_5 x_6 x_3

(b) Divide the input into Part A and Part B



(c) Generate the Part B spans autoregressively



(d) Self-attention mask

Results: NLU-Classification

Table 2. Results on the SuperGLUE dev set. Models with * are pre-trained for two times the number of steps of other methods.

Model	ReCoRD F1/Acc.	COPA Acc.	WSC Acc.	RTE Acc.	BoolQ Acc.	WiC Acc.	CB F1/Acc.	MultiRC F1a/EM	Avg
BERT _{Base}	65.4/64.9	66.0	65.4	70.0	74.9	68.8	70.9/76.8	68.4/21.5	66.1
GLM _{Base}	73.5/72.8	71.0	72.1	71.2	77.0	64.7	89.5/85.7	72.1/26.1	70.7
BERT _{Large}	76.3/75.6	69.0	64.4	73.6	80.1	71.0	94.8/92.9	71.9/24.1	72.0
UniLM _{Large}	80.0/79.1	72.0	65.4	76.5	80.5	69.7	91.0/91.1	77.2/38.2	74.1
GLM _{Large}	81.7/81.1	76.0	81.7	74.0	82.1	68.5	96.1/94.6	77.1/36.3	77.0
GLM _{Large} (multi-task)	80.2/79.6	77.0	78.8	76.2	79.8	63.6	97.3/96.4	74.6/32.1	75.7
GLM _{410M} (multi-task)	81.5/80.9	80.0	81.7	79.4	81.9	69.0	93.2/96.4	76.2/35.5	78.0
GLM _{515M} (multi-task)	82.3/81.7	85.0	81.7	79.1	81.3	69.4	95.0/96.4	77.2/35.0	78.8
T5 _{Base}	76.2/75.4	73.0	79.8	78.3	80.8	67.9	94.8/92.9	76.4/40.0	76.0
T5 _{Large}	85.7/85.0	78.0	84.6	84.8	84.3	71.6	96.4/98.2	80.9/46.6	81.2
BART _{Large} *	88.3/87.8	60.0	65.4	84.5	84.3	69.0	90.5/92.9	81.8/48.0	76.0
RoBERTa _{Large} *	89.0/88.4	90.0	63.5	87.0	86.1	72.6	96.1/94.6	84.4/52.9	81.5
GLM _{RoBERTa}	89.6/89.0	82.0	83.7	87.7	84.7	71.2	98.7/98.2	82.4/50.1	82.9

- Better than BERT, T5, RoBERTa

• Results: Uncond. Gen, Cond. Gen

Table 3. Results on Gigaword abstractive summarization

Model	RG-1	RG-2	RG-L
MASS	37.7	18.5	34.9
UniLM _{Large}	38.5	19.5	35.8
GLM _{Large}	38.6	19.7	36.0
GLM _{Large} (multi-task)	38.5	19.4	35.8
GLM _{410M} (multi-task)	38.9	20.0	36.2

Before

Train three different models

After

1.25 X Larger GLM can do all the three tasks with one model!

Table 4. Zero-shot language modeling results.

Model	Lambada (Accuracy)	BookWiki (Perplexity)
GLM _{Large} (uni)	0.0	> 100
GLM _{Large} (multi-task, uni)	47.4	15.1
– 2d positional encoding	45.8	15.1
GLM _{410M} (multi-task, uni)	49.5	14.5
GLM _{515M} (multi-task, uni)	50.4	13.9
GLM _{Large} (bi)	10.6	> 100
GLM _{Large} (multi-task, bi)	48.5	14.9
– 2d positional encoding	47.3	15.0
GLM _{410M} (multi-task, bi)	53.5	14.3
GLM _{515M} (multi-task, bi)	54.9	13.7
GPT _{Large} (uni)	50.1	14.4

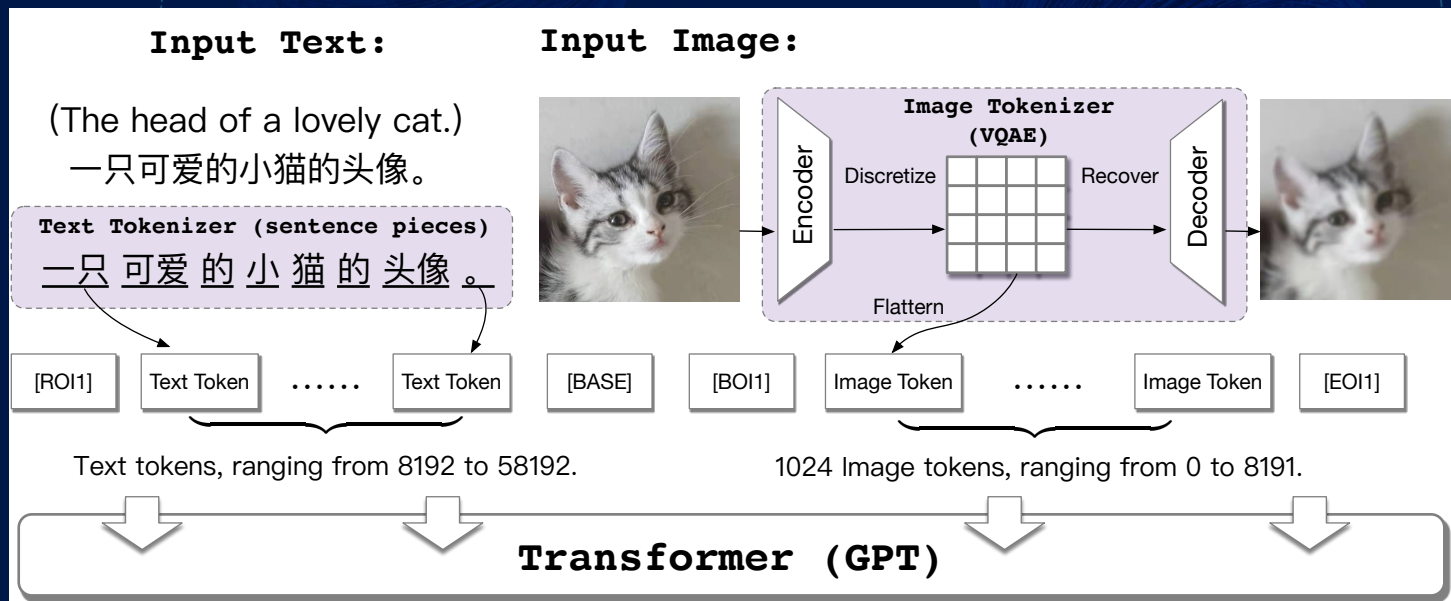


CogView: Mastering Text-to-Image Generation via Transformers.



● CogView: Text-to-Image Generation

- CogView: 4B、 6B
- Training with 30M image-text pairs on 512 V100 for 400 hours

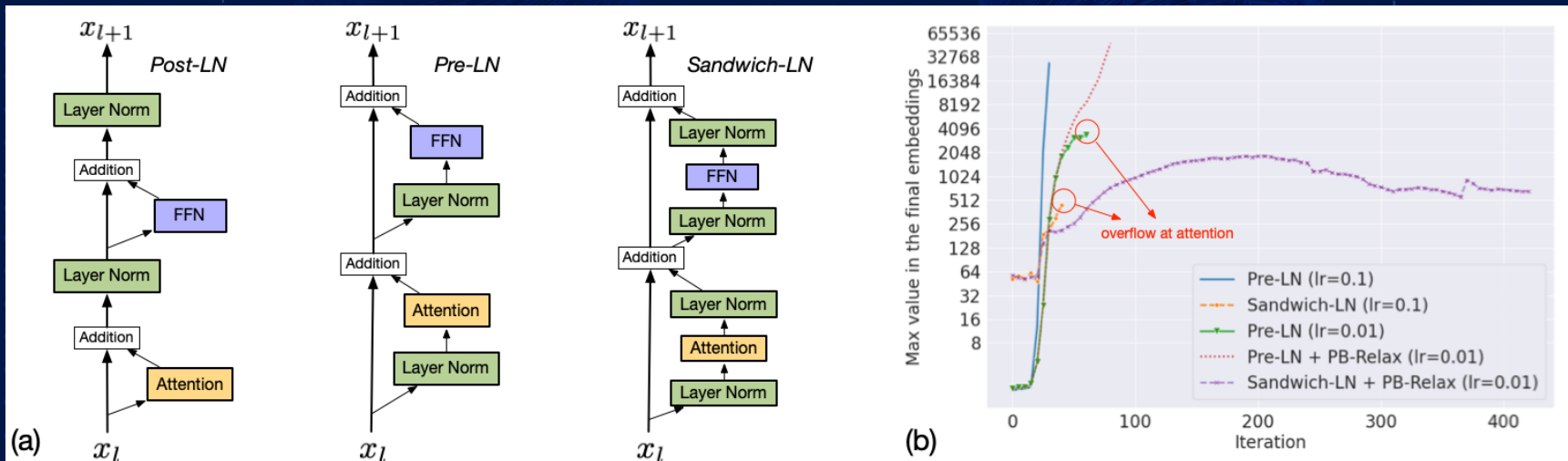


● Instability in Pretraining: Solutions

Sandwich LayerNorm

Block up the layer-wise magnification.

- Precision-Bottleneck relaxation
 - Make the computation of attention and LayerNorm precision-friendly.



DALL-E uses another solution, perResblock loss scaling and throw back some parameters back to fp32.

• Towards higher resolution

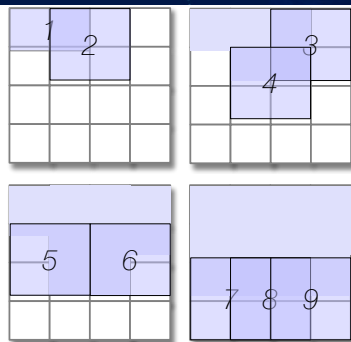
The Big Ben



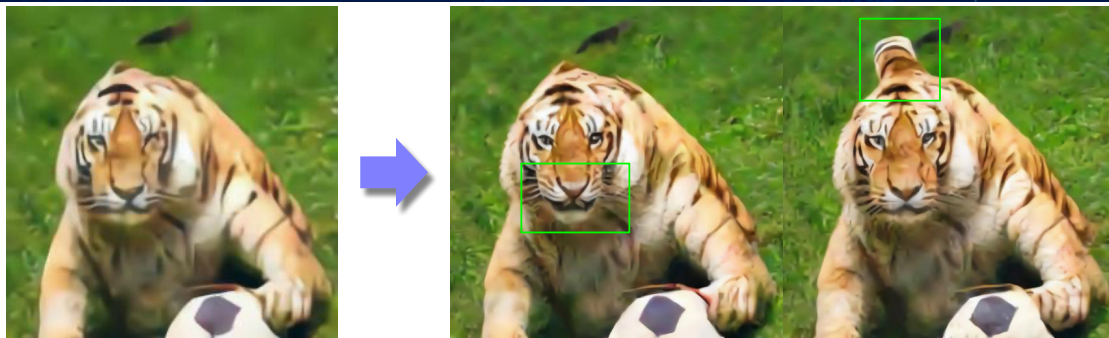
The close-up view
of the Big Ben



- Since CogView is trained on the most complex distribution of domain-general images, details of the objects have already been covered.
- Finetuning it into an super-resolution model should not be hard. (16×16 tokens $\Rightarrow 32 \times 32$ tokens, 1 DGX*day)



(a) Center-continuous sliding window



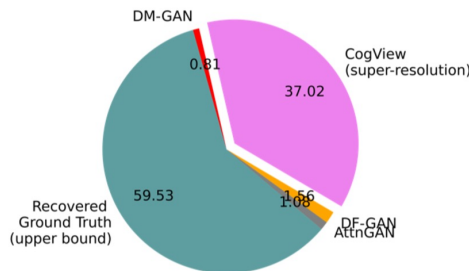
(b) Different super-resolution results for “a tiger is playing football”.

Results

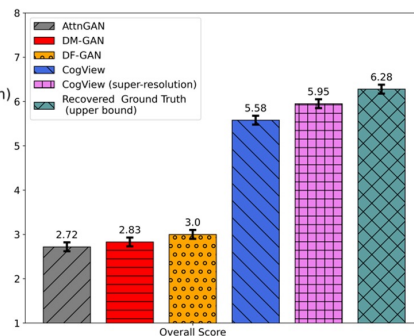
- Codes and Models: <https://github.com/THUDM/CogView>
- Demo website: <http://wudao.aminer.cn/CogView/index.html>

37% of the users
favor the generated
images by CogView

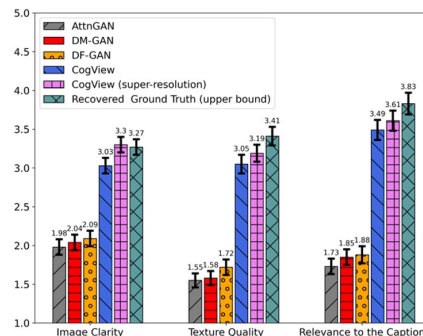
Better than DALL.E



(a) Human Preference. The percentage of the model to be chosen as best in all the questions.



(b) Overall scores (1-10) for the models.



(c) Scores (1-5) for the models on three important aspects.

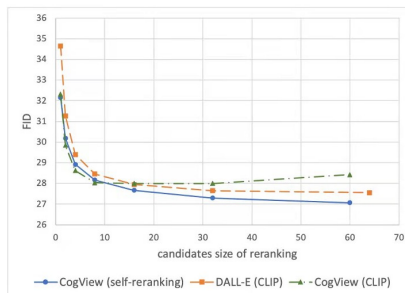
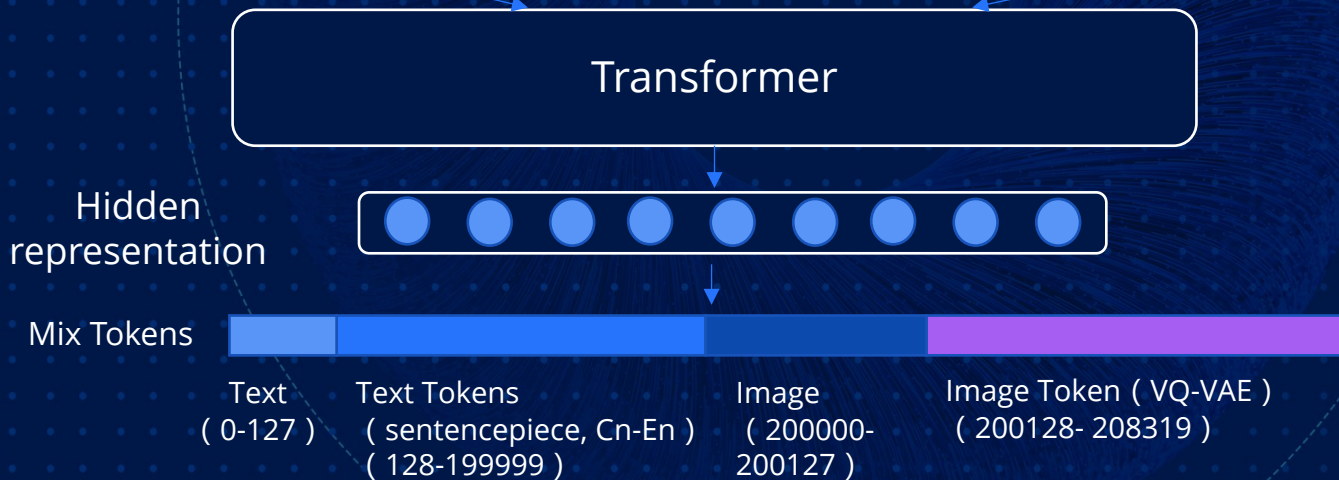
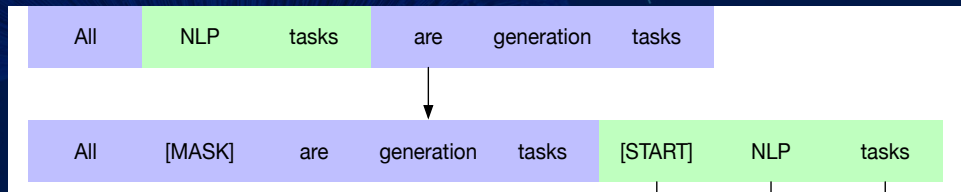
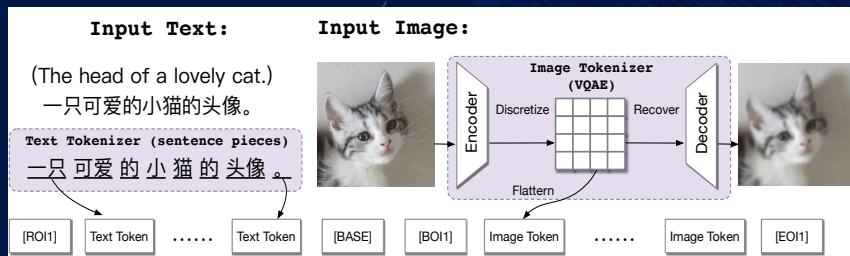


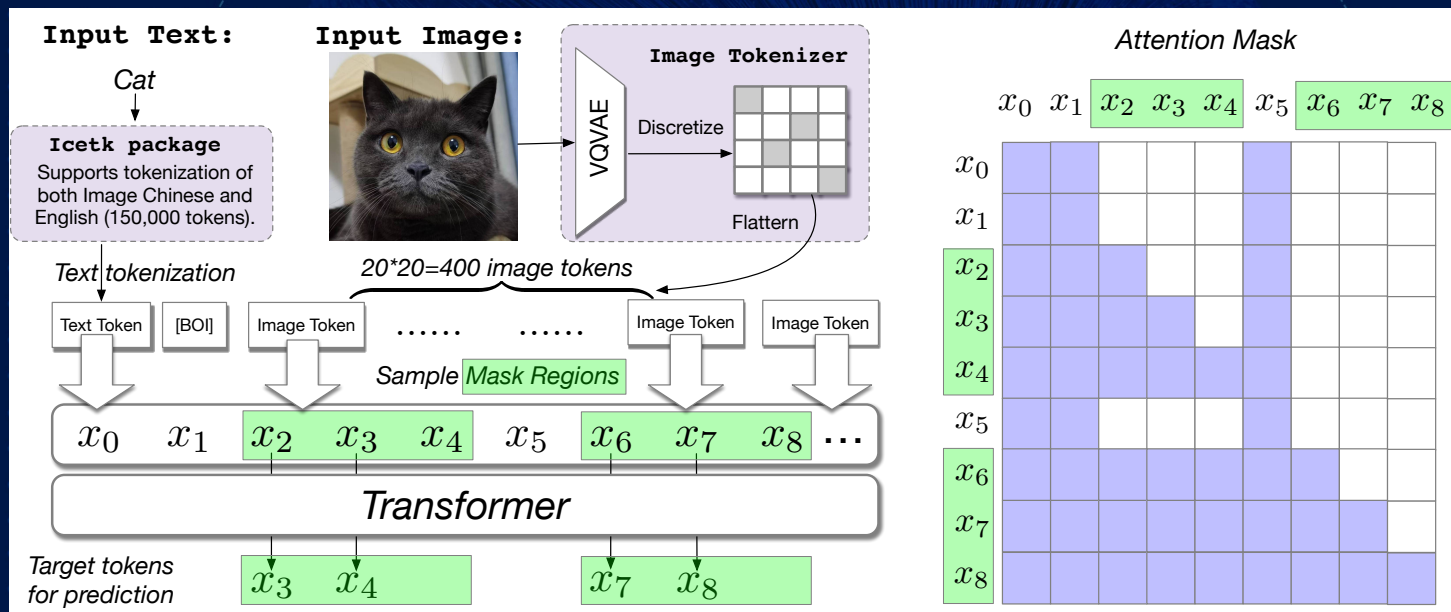
Table 1: Metrics for machine evaluation. Statistics about DALL-E are extracted from their figures. FID- k means that all the images are blurred by a Gaussian Filter with radius k .

Model	FID-0	FID-1	FID-2	FID-4	FID-8	IS	CapS
AttnGAN	35.2	44.0	72.0	108.0	100.0	23.3	0.02763
DM-GAN	26.0	39.0	73.0	119.0	112.3	32.2	0.02801
DF-GAN	26.0	33.8	55.9	91.0	97.0	18.7	0.02802
DALL-E	27.5	28.0	45.5	83.5	85.0	17.9	—
CogView	27.1	19.4	13.9	19.4	23.6	18.2	0.17403

CogView2



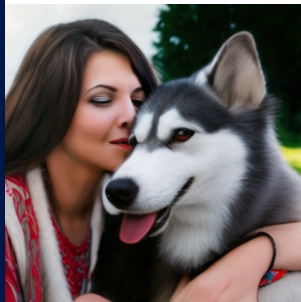
CogView2: Cross-Modality General Language Model



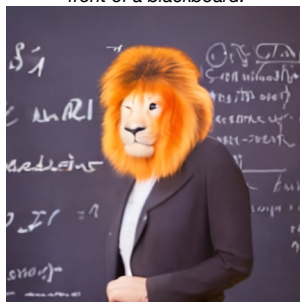
- A unification of "GPT(CogView) + MAE + Image captioning"
- Simplicity. No sentinel tokens, no order changing or token replacement.
- Only 20×20 tokens... 6 billion parameters!
- As a basic for future uses (super-resolution / video generation)!

CogView2: Faster (LoPAR) and better (Bidirectional/ higher-resolution) text-to-image generation

A beautiful girl is hugging a husky.



A lion teacher wearing a suit is in front of a blackboard.



A robot is riding under the blue and cloudy sky.



Several youths are talking in a bar.



A lion man is typing in the office.



A young woman is taking photos.



A pirate captain with a skull.



A girl holding an oil-paper umbrella in a rainy lane.



Earth in the Eye.



A magnificent church. Sketch.



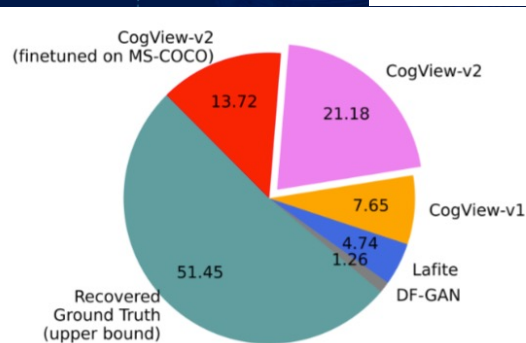
Mount Fuji, cherry blossom and Akita dog. Oil painting. A tiger with angel's wings.



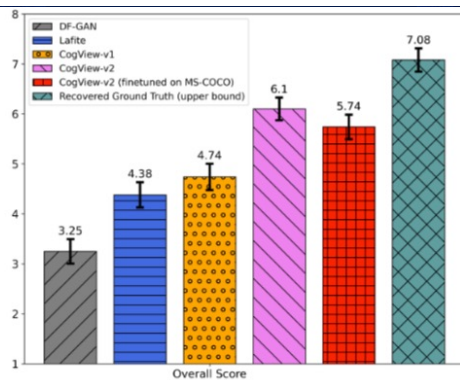
Results

Table 1: Machine Evaluation Results on MS-COCO. (Downsampling CogView2 images to 256×256 .)
 “*” means fine-tuning on MS-COCO. “- technique” is the ablation study without this technique.
 CogView2 achieves the **best blurred FIDs** over all comparable methods.

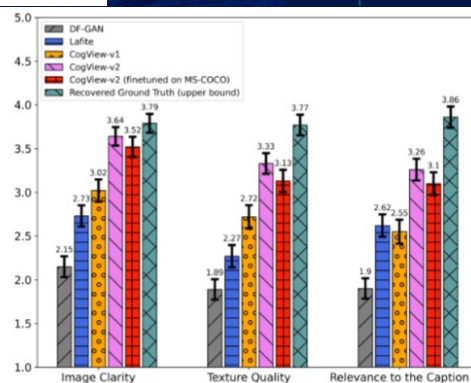
Model	FID-0	FID-1	FID-2	FID-4	FID-8	IS
AttnGAN* [35]	35.2	44.0	72.0	108.0	100.0	23.3
DM-GAN* [40]	26.0	39.0	73.0	119.0	112.3	32.2
DF-GAN* [28]	26.0	33.8	55.9	91.0	97.0	18.7
DALL-E [26]	27.5	28.0	45.5	83.5	85.0	17.9
CogView [3]	27.1	19.4	13.9	19.4	23.6	18.2
XMC-GAN* [36]	9.3	-	-	-	-	30.5
NÜWA* [33]	12.9	13.8	15.7	19.3	24	27.2
LAFITE [39]	26.9	23.0	18.7	15.7	14.8	26.0
Make-A-Scene* [8]	7.55	-	-	-	-	-
DALL-E-2 [27]	10.9	-	-	-	-	-
CogView2	24.0	19.7	16.8	17.2	17.2	22.4
- clustering sampling	36.4	32.4	28.9	28.5	30.4	18.8
- attention upweighting	24.6	20.4	17.5	17.9	18.9	21.1
CogView2*	17.5	13.4	10.9	10.6	10.4	25.2



(a) Human Preference. The percentage of the model to be chosen as best in **all** the questions.



(b) Overall scores (1-10) for the models.



(c) Scores (1-5) for the models on three important aspects.

Figure 7: The results of human evaluation. CogView2 performs the best in all the aspects.

← Tweet



Sepp Hochreiter
@HochreiterSepp



ArXiv arxiv.org/abs/2204.14217: Text-to-image system (CogView2) competitive with DALL-E-2. 6B-parameter hierarchical transformer for cross-modal language model. Fine-tuning by super-resolution modules. Tokens are re-masked and re-generated in a local parallel auto-regressive way.

Search Twitter

Relevant people



Sepp Hochreiter
@HochreiterSepp

Follow

Pioneer of Deep Learning and known for vanishing gradient and the LSTM. I mostly tweet about random ArXiv papers which sparked my interest.



Jack Clark @jackclarkSF · May 3



Replying to @HochreiterSepp

interesting research, though probably best way to show this is to let people compare prompts across both models, so should get the (mentioned in arxiv) demo website up



4



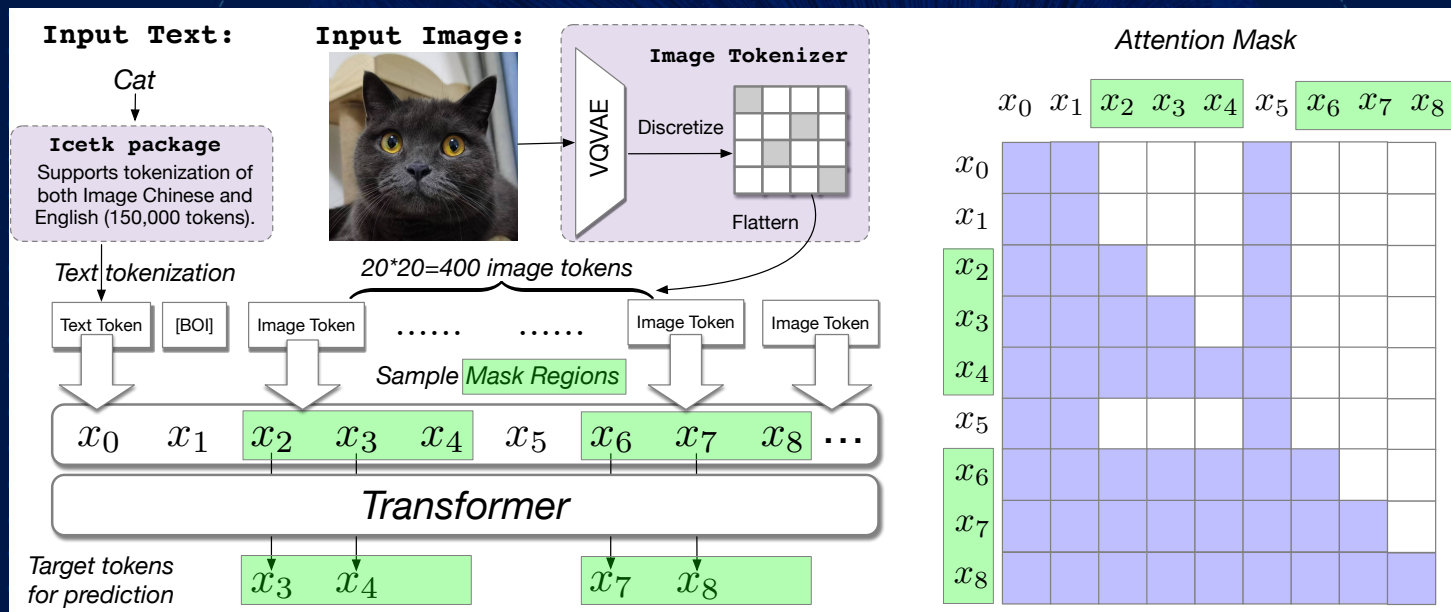
Bukenya Lukman @UgandanDev · May 2

Replying to @HochreiterSepp

For sure ,this is a must read.



CogView2: Cross-Modality General Language Model



- A unification of “GPT(CogView) + MAE + Image captioning”
- Simplicity. No sentinel tokens, no order changing or token replacement.
- Only 20 * 20 tokens... 6 billion parameters!
- As a basic for future uses (super-resolution / video generation)!

• CogVideo: Towards text-to-video generation

<https://github.com/THUDM/CogVideo>

Open domain

(480×480)

A woman is drinking waters.

女人喝水



A woman is running on beach at dusk.

女人在傍晚的海滩上跑步



A lion man is drinking a glass of water.

狮子人拿着水杯喝水



Text-to-video generation.

📖 Readme

☆ 1k stars

👁 59 watching

🍴 45 forks

Open on
GitHub



CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers

github: github.com/THUDM/CogVideo

367 Retweets

129 Quote Tweets

1,443 Likes



Saramsha Dotel @saramshadotel · May 30

How long until an AI generated movie wins the Oscar?

Examples generated by CogVide

Challenges: text-video alignment

A video in the training set



Split to fixed frame-rate(5) clips as real samples.



But... Each sample are accompanied with text "drinking".

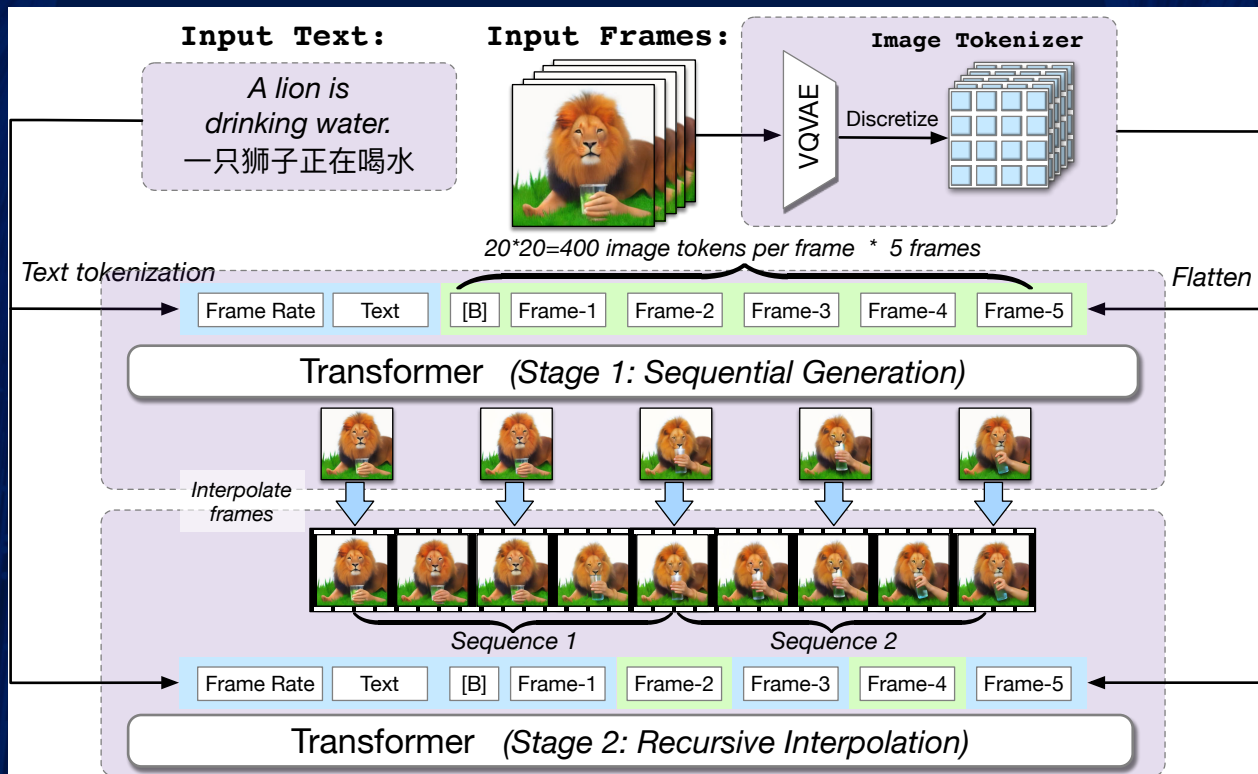
Could the model learn the whole action?



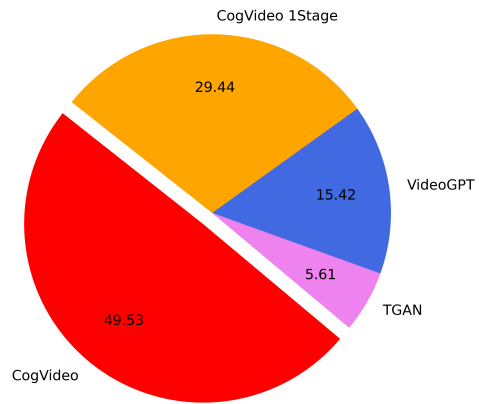
Challenges: text-video alignment

Solution:

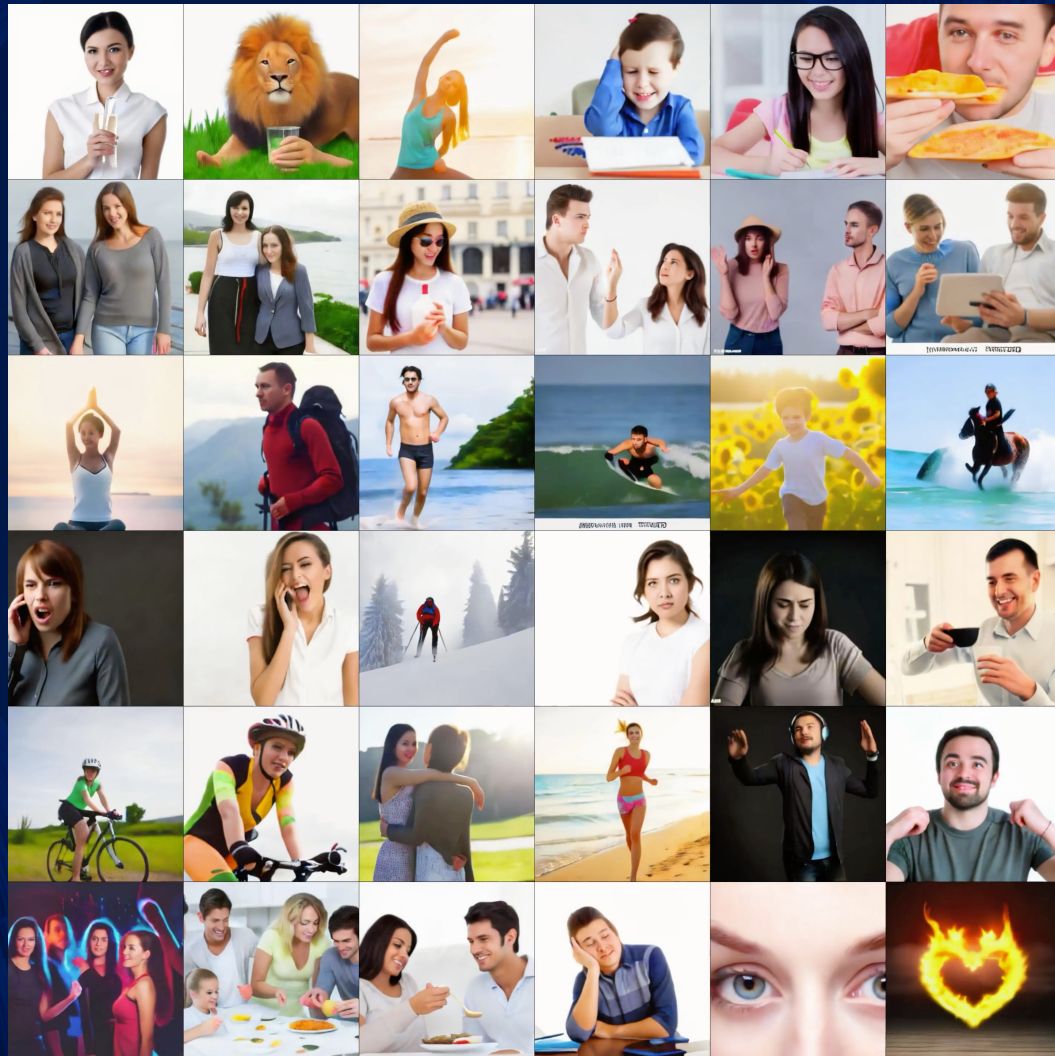
- Sample frames to cover the long video.
 - Lead to different frame-rate?
 - **Treat the frame-rate as a condition!**
- Recursive Interpolation
 - Another model

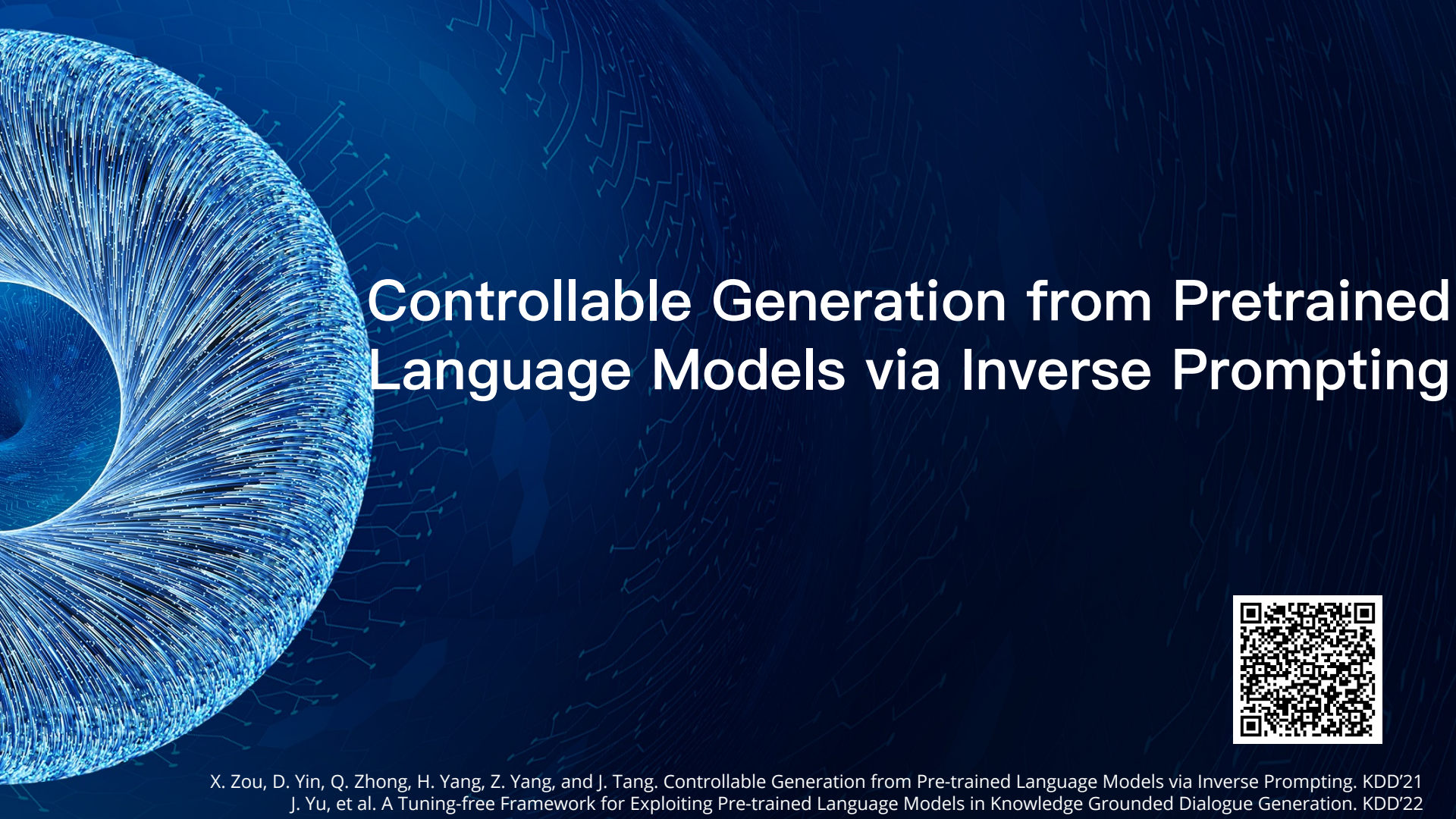


Results



(a) Human preference. The percentage of being chosen as the best.





Controllable Generation from Pretrained Language Models via Inverse Prompting



X. Zou, D. Yin, Q. Zhong, H. Yang, Z. Yang, and J. Tang. Controllable Generation from Pre-trained Language Models via Inverse Prompting. KDD'21
J. Yu, et al. A Tuning-free Framework for Exploiting Pre-trained Language Models in Knowledge Grounded Dialogue Generation. KDD'22

● Uncontrollable Prompting-based Generation

Question: Which moment did you want to live in forever?

I have been in a relationship with my boyfriend for two years. I studied abroad and he studied domestically. Alone in a foreign country, I felt really lonely and sad. When I was in school, I felt that it was ok to be together, working hard together and quarrelling together. I kept thinking in this way. In the last few months of our time at the high school, I felt that he was my walking little brother. In fact, the happiest moment was when we were together in the ``orphanage’’.

The relativity is weak!

Controllable Generation Based on Inverse Prompting

Question: Which moment did you want to live in forever?

Inverse Prompting:

In an instant I recall the second year of high school. That afternoon, I took a sunbath on the playground for a long time. I then played basketball downstairs with my classmates. As the sun set slowly, the three of us stood together on the corridor, leaning on the railing to face the sunset. I've always wanted to live in that moment forever, but unfortunately I never encountered a similar moment again. I just want to live in that moment.

● Inverse Prompting

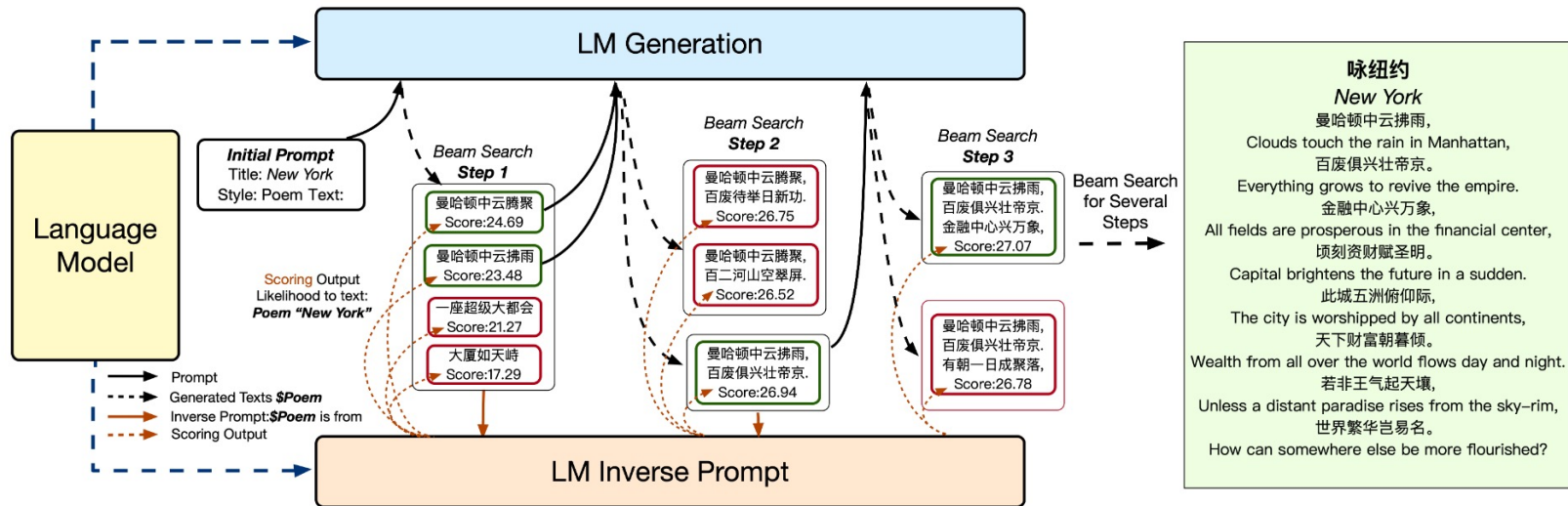
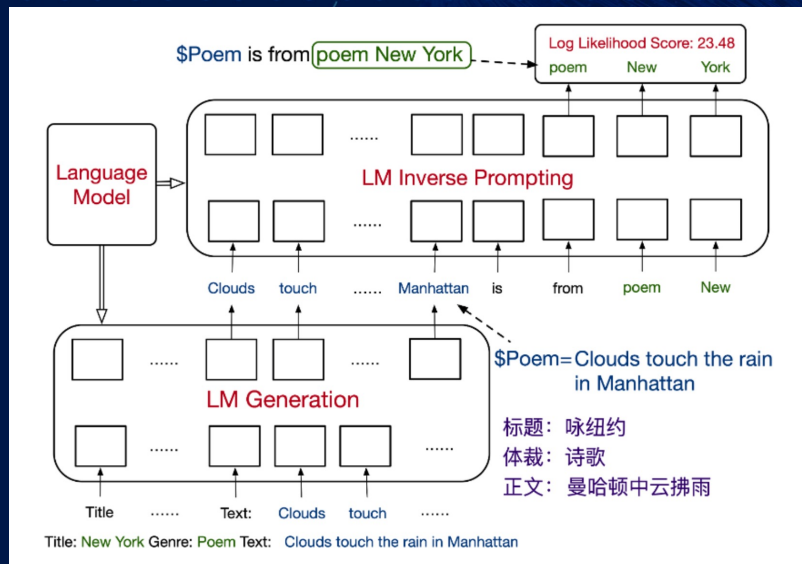


Figure 1: The generation process of open-domain traditional Chinese poems under inverse prompting. Using title *New York* (《咏纽约》) as an example.

Beam search using Inverse Prompting scoring for several steps.

● Inverse Prompting



Inverse Prompting scoring method:
Inversely prompt the title to improve the reliveness.

输入：
标题：咏纽约 体裁：诗
歌 正文：

输出：
曼哈顿中云拂雨

输入：
“曼哈顿中云拂雨”出自

目标输出：
诗歌《咏纽约》

Input:
Title : New York Genre:
Poem Text:

Output :
Clouds touch the rain in
Manhattan.

Input :
“Clouds Touch the rain in
Manhattan”is from

Target Output:
poem New York

Compute the log likelihood(-26.52), score=50-26.52=23.48

● Evaluation: QA

Method	Fluency (1-5)	Inform. ¹ (1-5)	Relevance (1-5)	Overall (1-10)
CPM [27]	2.66	2.47	2.36	4.32
Prompting Baseline	3.44	3.25	3.21	5.97
Inverse Prompting	3.61	3.43	3.59	6.51
Human Answers	3.80	3.61	3.67	6.85

¹ Informativeness

● Evaluation: Poem

Method	Format (1-5)	Innov. ¹ (1-5)	Relevance (1-5)	Aes. ² (1-5)	Overall (1-10)
Jiuge [28]	3.60	2.47	1.99	3.12	3.57
Search Baseline	2.79	1.10	1.16	2.44	1.35
Inverse Prompting	2.56	2.71	2.92	2.33	4.00
Inverse Prompting +ST	2.42	2.92	3.65	2.18	4.40

¹ Innovation

² Aesthetics

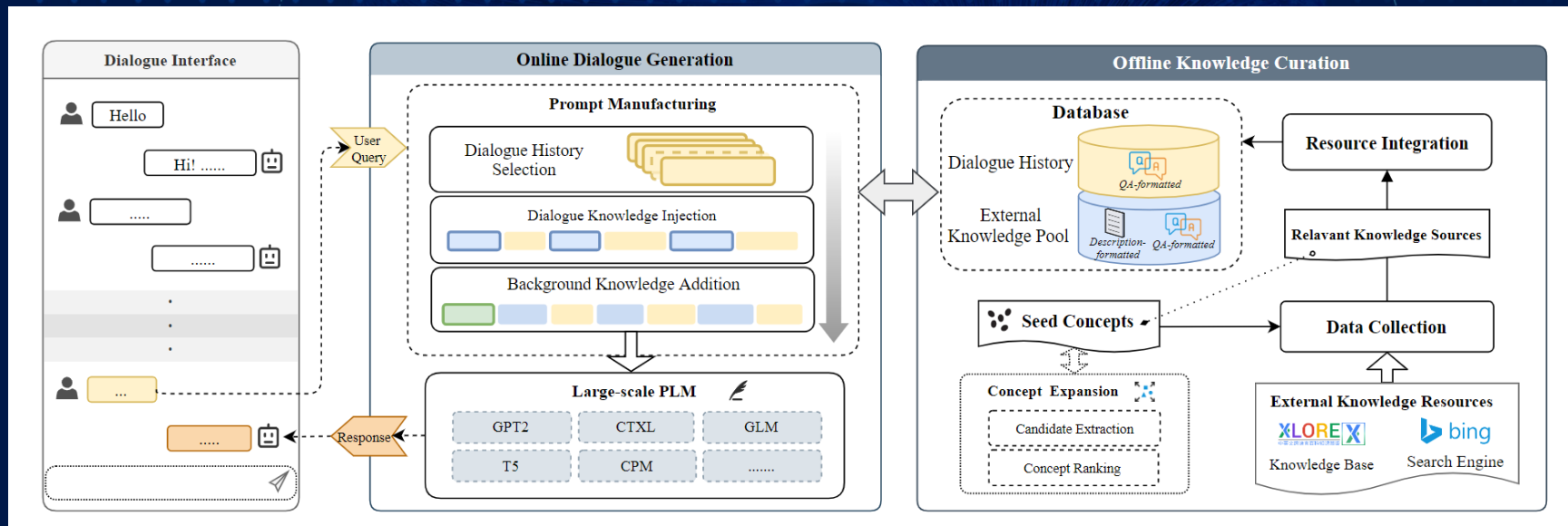
● Turing Test

Method	Total	Selected	Selection Rate
Inverse Prompting +ST	1,656	748	45.2%
Ancient Human Poems	1,656	908	54.8%

- 45.2%

Online Generation & Offline Knowledge Curation

XDAI: eXploiting large-scale pre-trained language models in building grounded Dialogue AI systems



Interaction

Prompt engineering & Generation

External Information Provider

Online Generation & Offline Knowledge Curation

Dialogue History Selection

- Most relevant historical utterances

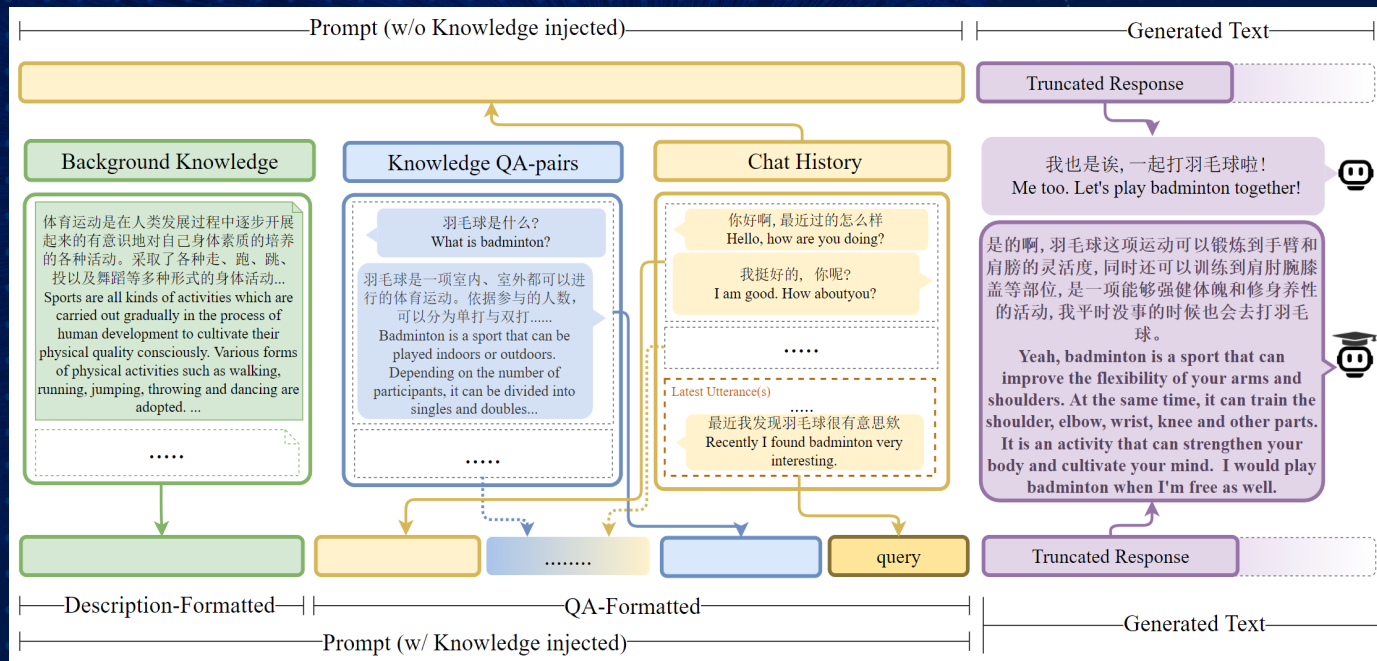
$$\alpha^{t-i} * \text{cosine}([U_i; S_i], U_t)$$

Dialogue Knowledge Injection

- Most similar QA-formatted knowledge to $[U_i; S_i]$
- Injected before the corresponding dialogue utterance.

Background Knowledge Addition

- Description-formatted as the background
- Injected as the prefix of the entire prompt text.



● Grounded on personalized condition (emotion)

Prompt: <dialogue context> [MASK] <grounded post-response>

Mechanism: Auto-regressive Blank Infilling in GLM

A: 今天天气不错啊

B: 哪有啊, 热死了

A: 最近我发现羽毛球很有意思欸

B: [MASK]

A: 你不能好好说话吗, 怎么突然生气了

GLM

切, 你打那么烂

易怒版 (Irritable Dai)



你在干嘛呢
What are you doing?

切, 还问我
Humph, don't ask me

怎么了
What happened?

哼, 别打扰我吃薯片, 要不就再也别跟我说话了, 讨厌
Hum, don't bother me eating chips, or don't talk to me anymore. I hate it.

我怎么惹你了
Why are you so pissed with me?

我说了不跟你说话, 你还来烦我, 算了, 你走好了, 以后别来烦我了!
I said I wouldn't talk to you, but you still bother me. Forget it, you can go. Don't bother me in the future!

抑郁版 (Gloomy Dai)



怎么了
What's up?

没什么, 我只是觉得人生没什么值得我努力的东西, 就这样吧
Nothing. I just don't think there is anything worth my efforts in my life. That's it.

为什么这么想
Why do you think so?

我也不知道啊, 我也不知道该为什么活着, 可能只是无聊吧, 可能是闲的慌吧
I don't know. I don't know why I should live. Maybe I'm just boring and idle.

多出去走走看看吧
Go out and have a look

对, 我本来也想看看的, 但是一想没什么好看的, 我就没有那个心, 还不如窝在家玩游戏打发时间
Yeah, I wanted to but I'm not in the mood since there is nothing interesting. I might as well stay at home and play games to pass the time.

暖心版 (Supporting Dai)



我最近有点累
I feel tired recently

是呀, 身体上的累可以靠休息, 心理上的累就需要你多与人交流了呀, 你是不? Well, if you are physically tired, you can rest. But if you are mentally tired, you need to communicate with others more. Do you think so?

身心俱疲
I am exhausted both physically and mentally.

你可以多与人交流, 不管是家人、朋友、还是陌生人, 你都要好好和他们交流, 他们是你最好的朋友, 你开心了他们会开心, 不开心了他们会难过
You can talk with your family, friends, or strangers. They are your best friend, happy when you are happy, and sad when you feel terrible.

那如果没有心情怎么办
What if I'm not in the mood?

那就不要管了呀, 这样只会加重你的负担, 没有心情就想开心的事情, 多听听音乐、看看电影、和朋友出去走走, 或者看看书吧
Then let it go, or it will only add to your burden. Focus on happy thoughts. Listen to music, watch movies, go out with friends, or read a book.

● Grounded on personalized condition (character)

Prompt: <dialogue context> [MASK] <grounded post-response>

Mechanism: Auto-regressive Blank Infilling in GLM

林妹妹 (Sentimental Dai)

 你好啊
Hi 

怎么,哥哥今天心情这么好,竟跟我开玩笑
Well, honey you are in such a good mood today that begin making fun of me

你最近怎么样
How have you been? 

哥哥这么关心我,怎么就不问问我呢,是不是我长得太丑了,还是哥哥最近忙的忘了疼惜我。
If you are so concerned about me, why don't you ask me? Is it that I'm too ugly? Or you are too busy recently so you forget to cherish me?

我没有嫌你丑啊
I don't think you are ugly at all! 

哥哥,那为什么你对每个妹妹都很好,却唯独对我那么冷淡,是不是我哪里做的不好,哥哥,你告诉我,我哪儿做的不好,我改。
Then why have you been so nice to all your sisters, but so cold to me? Did I do something wrong, honey? Please tell me, I will change.

甄嬛版 (Palace-drama Dai)

 你好
Hi 

皇上,若说今日踏青,便是好的了
Your Majesty, an outing would be great today.

你想去哪踏青
Where do you want to go for an outing? 

我想去倚梅园看看梅花,听说那梅花开得极盛,只怕明日去还赶不及
I want to go to the Yimei garden for the plum blossom. I heard that the plum trees are in full bloom. I'm afraid it's too late to go tomorrow.

可以
Good. 

若是陛下能同我一道赏梅,怕是踏青之意又更浓了一层
If your majesty can enjoy the plum blossom with me, the outing would be more meaningful.

悟空版 (Monkey Dai)

 你是谁
Who are you? 

我乃五百年前大闹天宫的齐天大圣
I'm the Great Sage Equal to Heaven who caused havoc in the heavenly palace five hundred years ago

那你现在是什么
So what are you now? 

俺老孙已成佛,并正在南瞻部洲做善事
I have become a Buddha. I am doing good deeds in Jambu-dvipa.

做什么善事
Do what good? 

弘扬佛法,劝人为善,以助人为乐的善事,就是俺老孙最大的善事
Carrying forward the Dharma, persuading people to do good and helping others is the greatest good of me.



WuDao Ecology



<https://wudaoai.cn/>

● Open!

01 WuDaoCorpora: the world's largest publicly available dataset!

Open Data

02 We released almost all codes in WuDao!

Open Code

03 You can download >20 well-trained models!

Open Model

04 Directly call an API to enjoy the power of WuDao!

Open API

Click here to find more: <https://wudaoai.cn/>



• WuDao's Today

WuDao: A super-scale model with 1.75 Trillion parameters.

- WuDao is very general and can be applied to different tasks
- WuDao can do QA, write poem, knowledge extraction, draw pictures, write articles, recognize pictures, etc.
- WuDao is open to everyone
- WuDao is very expensive... and needs to reduce cost...

● WuDao's Tomorrow

Teach Machine to Think Like Humans: Beyond the Turing Test!

Cognition (T9)

1. Adaptation and Learning
2. Definition and Contextualization
3. Character Setting
4. Priority and Access Control
5. Call Together and Control
6. Decision Making and Execution
7. Probing and Editing
8. Reflection and Self-Monitoring
9. Logic and Flexibility



Thanks to everyone!

WUDAO ·Wenyuan

A large scale pre-training language model with Chinese as its core

Liu Zhiyuan, Huang Minlie, Han Wentao, Liu Yang, Zhu Xiaoyan, Sun Maosong, Zhang Zhengyan, Gu Yuxian, Han Xu, Chen Shengqi, Xiao Chaojun, Yao Yuan, Qi Fanchao, Guan Jian, Ke Pei, Zhou Hao, Sun Zhenbo, Cai Yanzheng, Zeng Guoyang, Tan Zhixing, Qin Yujia, Su Yusheng, Si Chenglei, Hu Xueyu, Li Wenhao, Wang Fengyu, Yi Jing, Wang Xiaozhi, Chen Weize, Ding Ning, Zhang Jiajie

WUDAO ·Wenlan

Super large multi-modal pre-training model

Wen Jirong, Song Ruihua, Lu Zhiwu, Jin Qin, Zhao Xin, Pang Liang, Lan Yanyan, Dou Zhicheng, Gao Yizhao, Huo Yuqi, Lu Haoyu, Wen Jingyuan, Yang Guoxing, Song Haoyang, Zhang Manli, Zhang Liang, Hu Anwen, Li Ruichen, Song Yuqing, Zhao Jinming, Zhao Yida, Fei Nanyi, Sun Yuchong, Jin Chuhao, Hong Xin, Cui Wanqing, Hou Danyang, Li Yingyan, Xi Zongzheng, Liu Guangzhen, Liu Peiyu, Gong Zheng, Li Junyi

WUDAO ·Wenhui

A new super-large cognitive-oriented pre-training model

Tang Jie, Yang Zhilin, Yang Hongxia, Du Zhengxiao, Ding Ming, Zou Xu, Qiu Jiezhong, Qian Yujie, Yinda, Zhong Qingyang, Yu Jifan, Liu Xiao, Zheng Yanan, He Jiaao, Zeng Aohan, Hong Wenyi, Yang Zhuoyi, Zheng Wendi, Zhou Jing, Du Jizhong, Guo Zitong, Liu Jing, Zhou Chang, Lin Junyang

WUDAO ·Wensu

Super large protein sequence prediction pre-training model

FastMoE and trillion large model

Tang Jie, Lu Bai, Qiu Jiezhong, Xie Changyu, Xiao Yijia, Zeng Aohan, Li Ziang

Tang Jie, Zhai Jidong, Yang Hongxia, Chen Wenguang, Zheng Weimin, Ma Zixuan, He Jiaao, Qiu Jiezhong, Cao Huanqi, Wang Yuanwei, Sun Zhenbo, Zheng Liyan, Wang Haojie, Tang Shizhi, Feng Guanyu, Zeng Aohan, Zhong Runxin, Shi Tianhui, Du Zhengxiao, Ding Ming, Tiago Antunes, Peng Jinjun, Lin Junyang, Zhang Jianwei

Wudao—Pretrain the world

Thanks!

