Wudao—Pretrain the world

Tang Jie

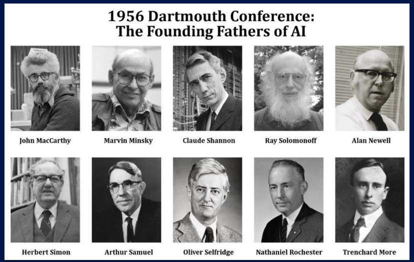
Tsinghua University

Beijing Academy of Artificial Intelligence (BAAI)

The slides will be available soon at http://keg.cs.tsinghua.edu.cn/jietang (or Google "Jie Tang")

AI History

Symbolic



1st

Symbolic model, perceptron

Recognition

2nd

Big data

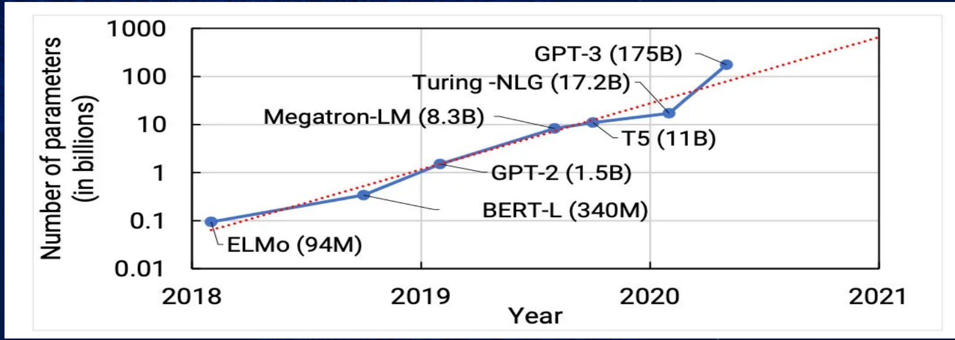
Cognition

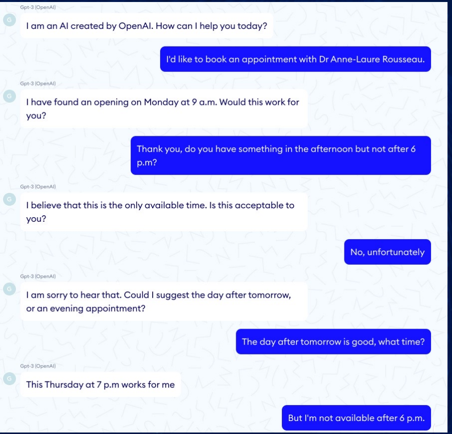
3rd

Data and Knowledge

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

QA

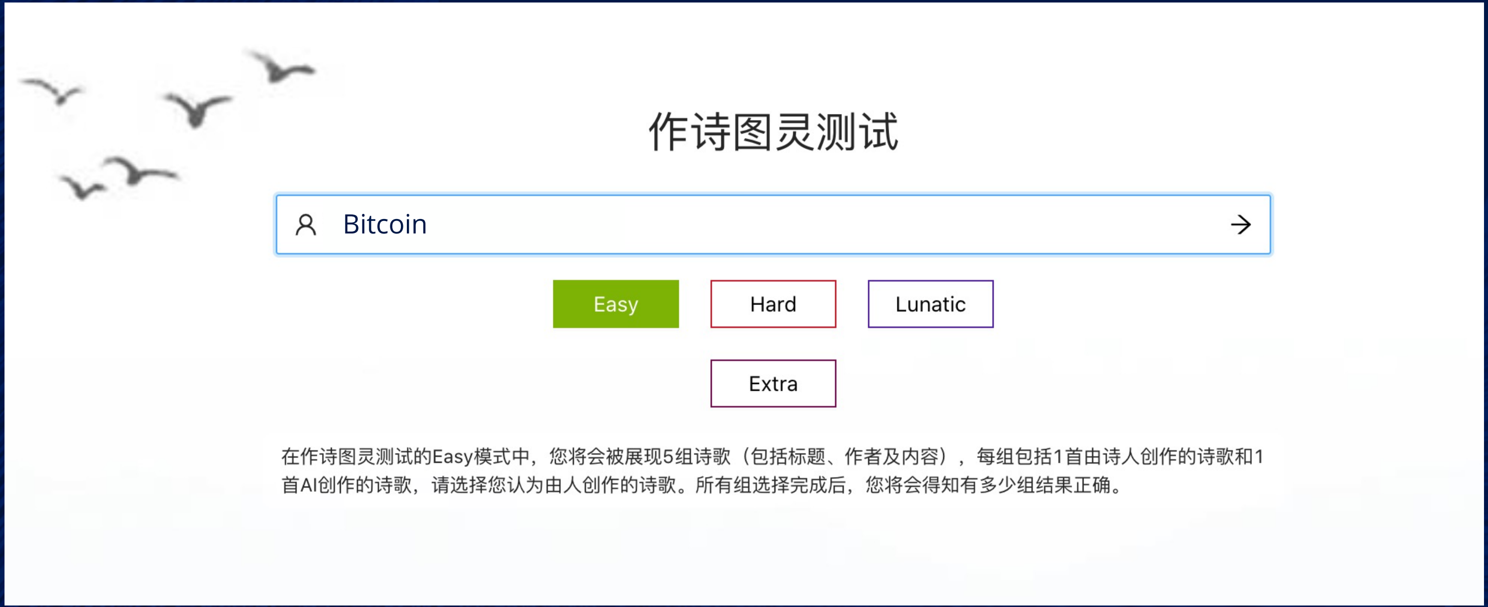
Question: Which moment did you want to live in forever?

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/



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

磻溪连灞水，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。

文渊

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.

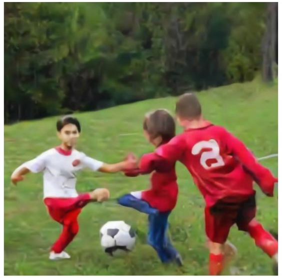
太空秋色凉，独鸟下微阳。
三径池塘静，六街车马忙。
渐能高酒户，始是入诗狂。
官冷且无事，追陪慎莫忘。

Draw Pictures

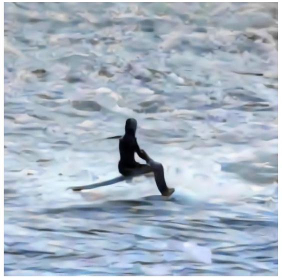
https://wudao.aminer.cn/CogView/index.html

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

Compare with Dalle

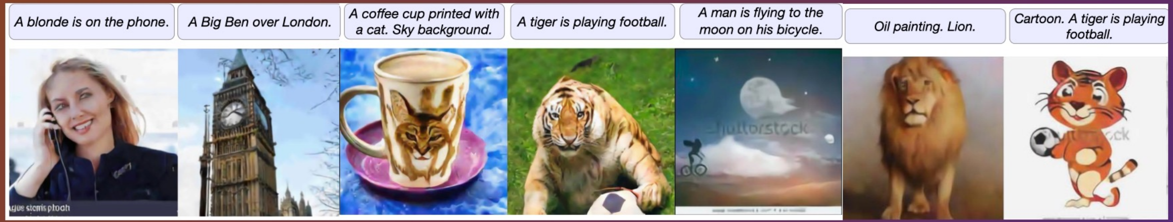
https://wudao.aminer.cn/CogView/index.html

Alibaba DAMO Academy

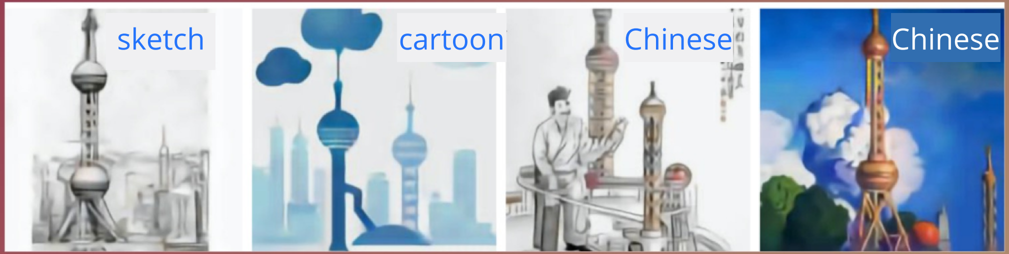
AI-aided Design

Z. Lin et al. M6: Multi-Modality-to-Multi-Modality Multitask Mega-transformer for Unified Pretraining. **KDD'21**.

Better than DALL.E on MS COCO



Different Styles



Turing test

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

Core Algorithms

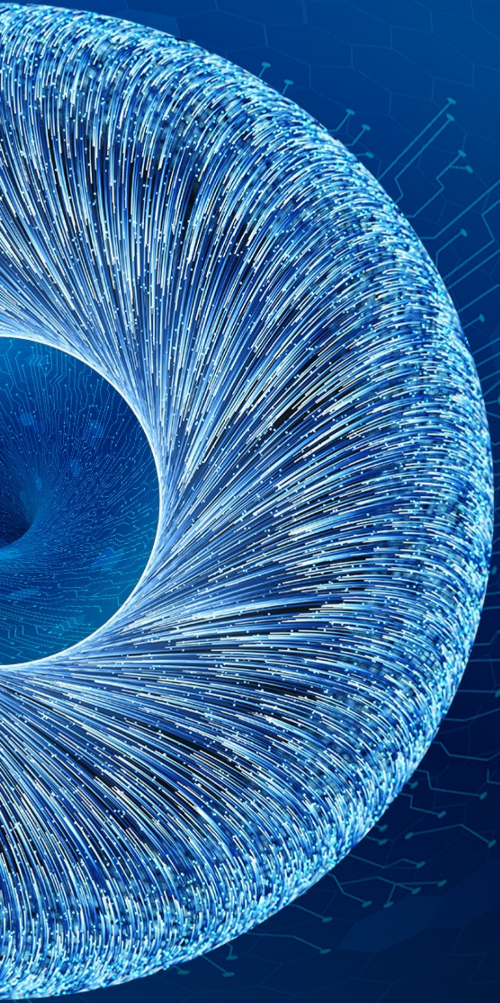
Controllable Generation via Inverse Prompting

CogView: Text-to-Image Generation

Prompt Tuning vs. Fine Tuning

GLM: General Language Model Pretraining

How to build a trillion-scale model



WUDAO 2.0 is the World's largest

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

Z. Ma et al. BAGUALU: Targeting Brain Scale Pretrained Models with over 37 Million Cores. **PPoPP'22.**

GLM: General Language Model Pretraining with Autoregressive Blank Infilling

Z. Du et al. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. ACL’22.

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

Z. Du et al. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. ACL'22.

All NLP Tasks are Generations

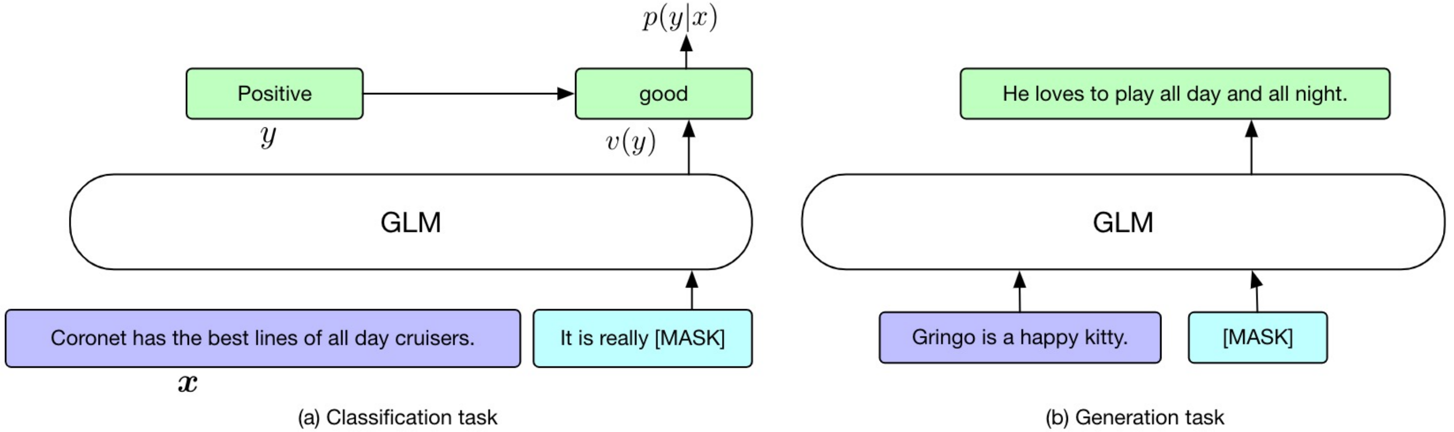


Figure 3. GLM finetune framework. (a) Formulation of the sentiment classification task as blank infilling with GLM. (b) GLM for text generation given the context. This can be the language modeling in the zero-shot setting, or seq2seq with fine-tuning.

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

Z. Du et al. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. ACL'22.

A New Pretraining Framework

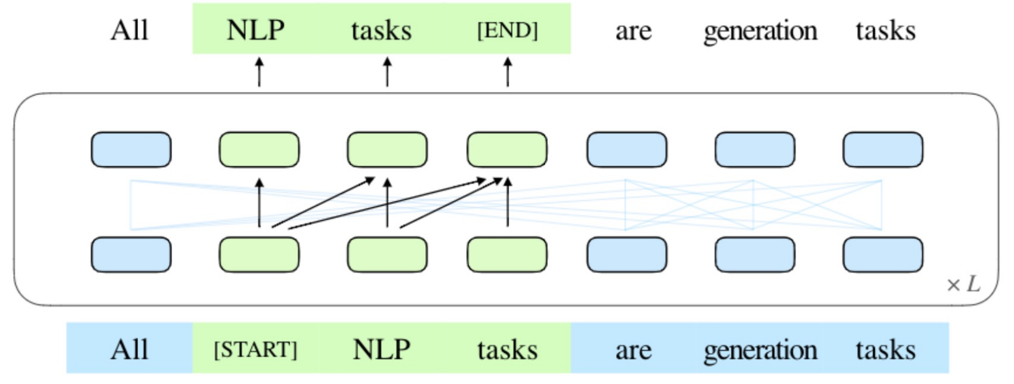
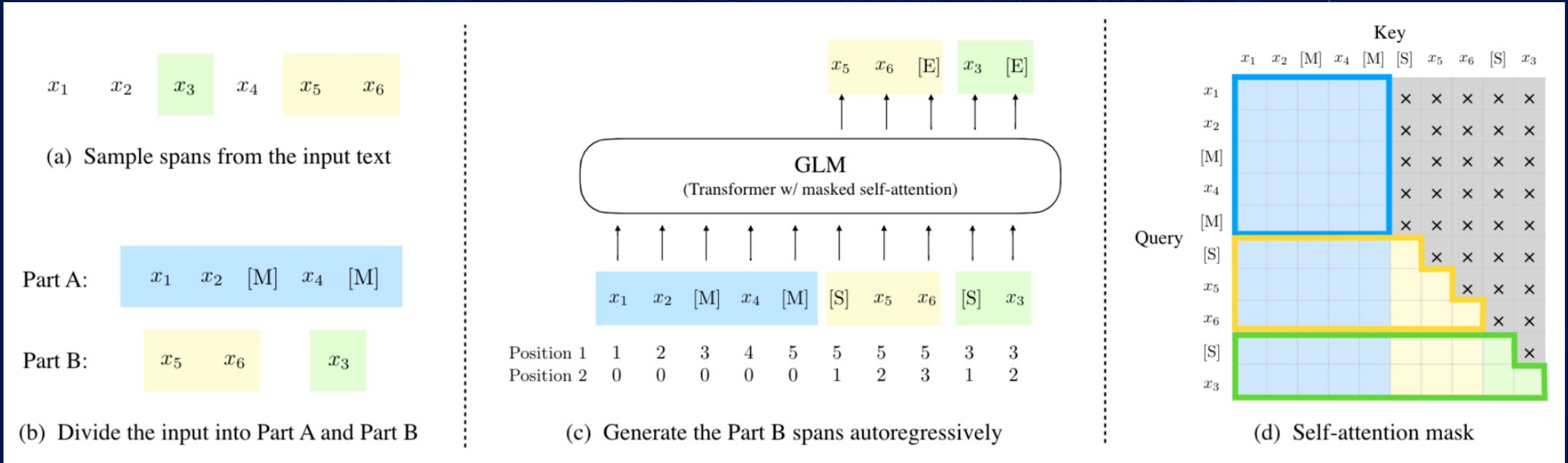


Figure 1. Illustration of GLM. We blank out text spans (green part) and GLM is trained to generate them in an autoregressive fashion.

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



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
$\text{BERT}_{\text{Base}}$	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**
$\text{BERT}_{\text{Large}}$	76.3/75.6	69.0	64.4	73.6	80.1	**71.0**	94.8/92.9	71.9/24.1	72.0
$\text{UniLM}_{\text{Large}}$	80.0/79.1	72.0	65.4	76.5	80.5	69.7	91.0/91.1	77.2/38.2	74.1
$\text{GLM}_{\text{Large}}$	81.7/81.1	76.0	**81.7**	74.0	**82.1**	68.5	96.1/94.6	**77.1/36.3**	77.0
$\text{GLM}_{\text{Large}}$ (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
$\text{BART}_{\text{Large}}$ *	88.3/87.8	60.0	65.4	84.5	84.3	69.0	90.5/92.9	81.8/48.0	76.0
$\text{RoBERTa}_{\text{Large}}$ *	89.0/88.4	**90.0**	63.5	87.0	**86.1**	**72.6**	96.1/94.6	**84.4/52.9**	81.5
$\text{GLM}_{\text{RoBERTa}}$	**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

Z. Du et al. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. ACL’22.

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
$\text{UniLM}_{\text{Large}}$	38.5	19.5	35.8
GLM $_{\text{Large}}$	38.6	19.7	36.0
GLM $_{\text{Large}}$ (multi-task)	38.5	19.4	35.8
GLM $_{\text{410M}}$ (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.

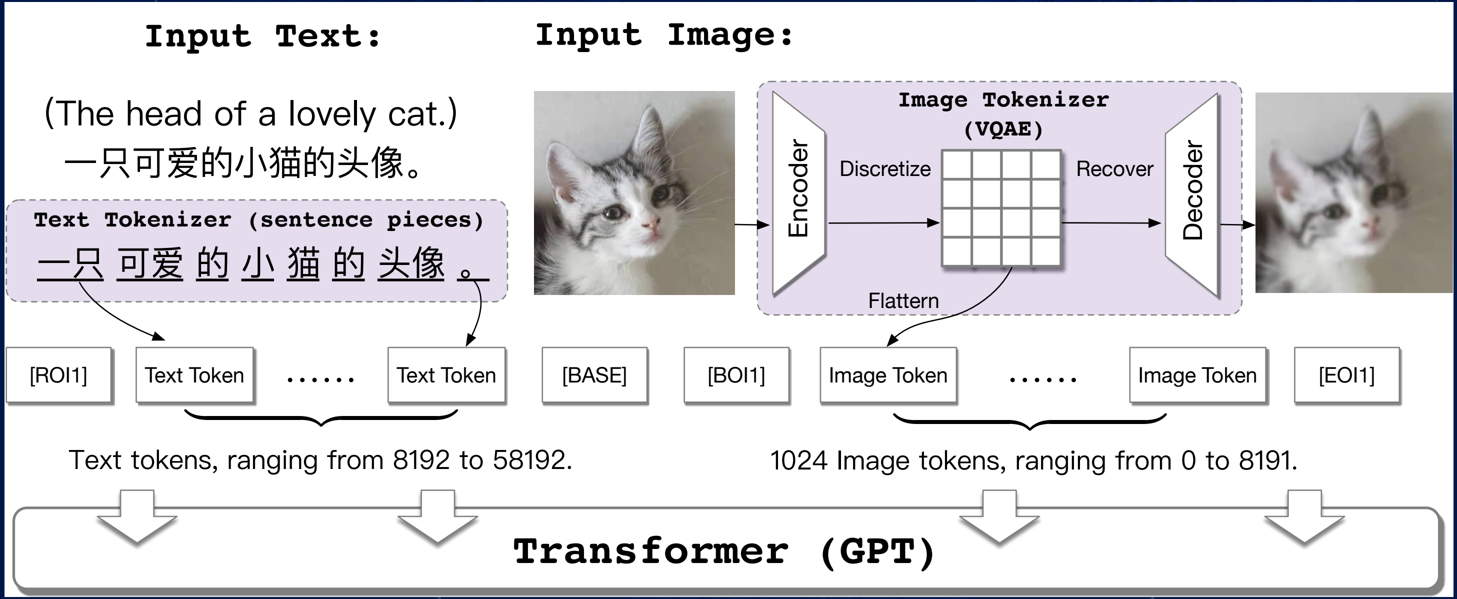
Model	Lambada (Accuracy)	BookWiki (Perplexity)
GLM $_{\text{Large}}$ (uni)	0.0	> 100
GLM $_{\text{Large}}$ (multi-task,uni)	47.4	15.1
– 2d positional encoding	45.8	15.1
GLM $_{\text{410M}}$ (multi-task,uni)	49.5	14.5
GLM $_{\text{515M}}$ (multi-task,uni)	**50.4**	**13.9**
GLM $_{\text{Large}}$ (bi)	10.6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

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

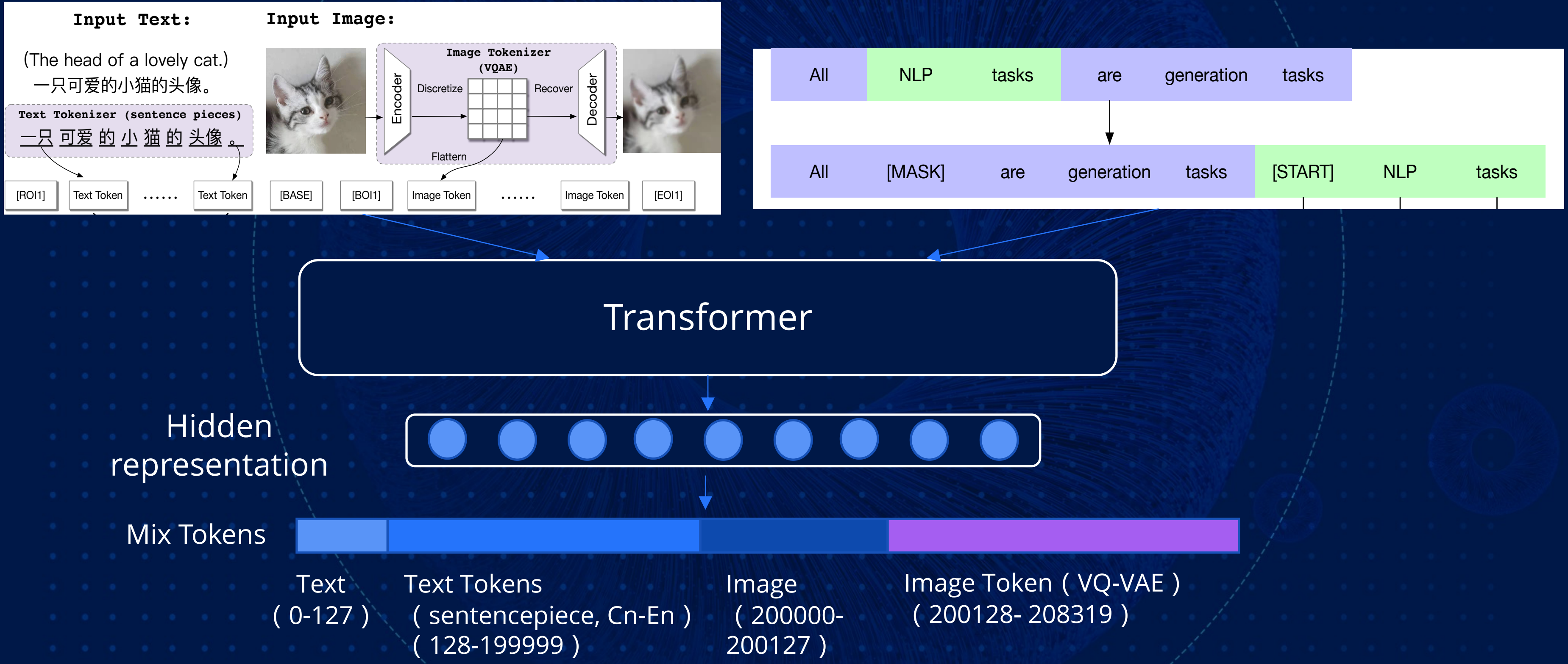
M. Ding et al. CogView: Mastering Text-to-Image Generation via Transformers. NeurIPS'21.

CogView: Text–to–Image Generation

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



CogView Model

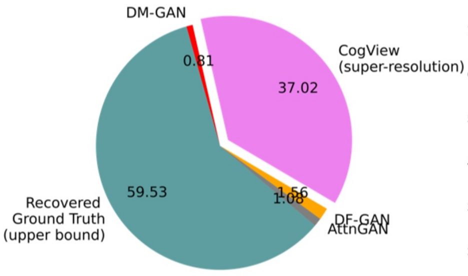


Results

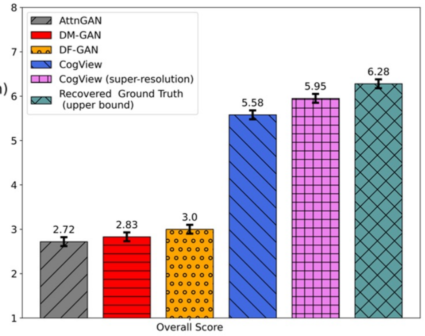
- Codes and Models: https://github.com/THUDM/CogView
- Demo website: http://wudao.aminer.cn/CogView/index.html

37% of the users favor the generated images by CogView

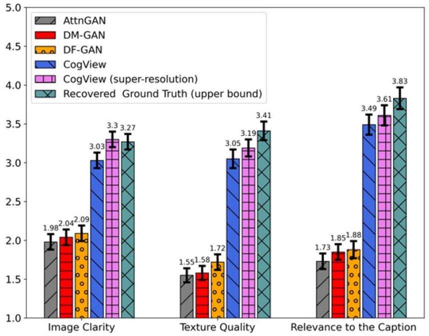
Better than DALL.E



(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.

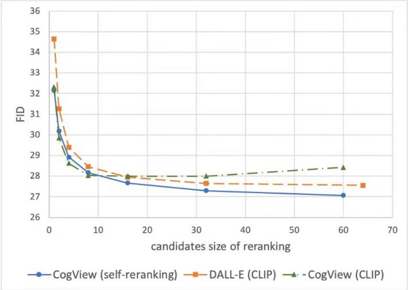
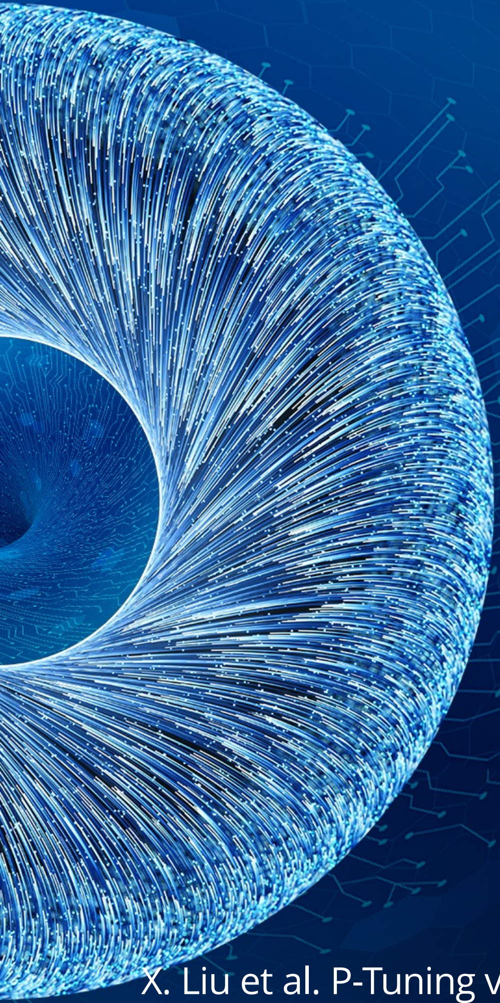


Table 1: Metrics for machine evaluation. Statistics about DALL-E are extracted from their figures. FID-k means that all the images are blurred by a Gaussian Filter with radius k.

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**



Prompt Tuning vs. Fine Tuning

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

Prompt

- Handwriting Prompt（Brown et al 2020）
- Discrete Prompt（Jiang et al 2020, Shin et al 2020, Gao et al 2020）
- Continuous Prompt（P-Tuning）

Prompt	P@1
[X] is located in [Y]. *(original)*	31.29
[X] is located in which country or state? [Y].	19.78
[X] is located in which country? [Y].	31.40
[X] is located in which country? In [Y].	51.08

Table 1. Case study on LAMA-TREx P17 with bert-base-cased. A single-word change in prompts could yield a drastic difference.

- Discrete prompt is very sensitive to the noise
- Is easy to overfit dev/test sets

P–Tuning

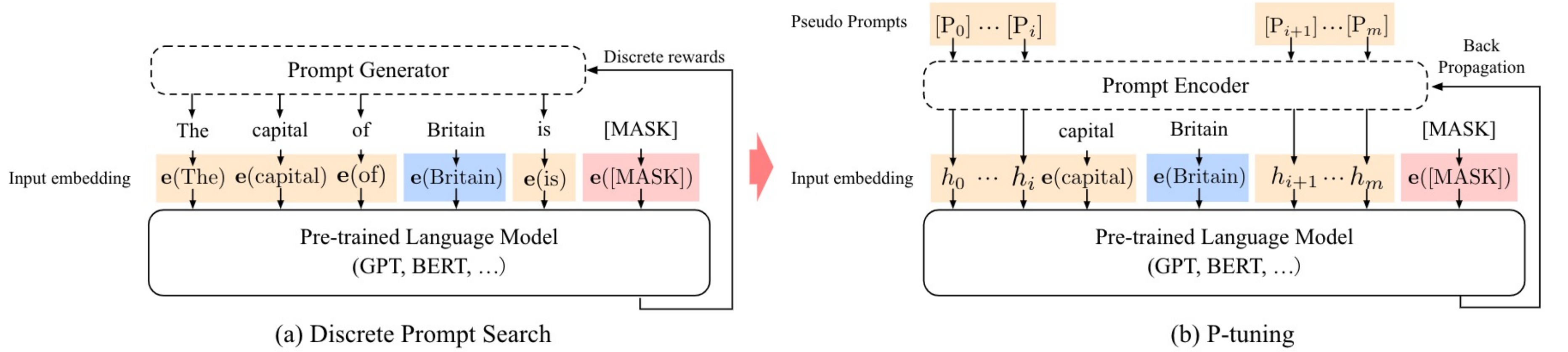


Figure 2. An example of prompt search for "The capital of Britain is [MASK]". Given the context (blue zone, "Britain") and target (red zone, "[MASK]"), the orange zone refer to the prompt tokens. In (a), the prompt generator only receives discrete rewards; on the contrary, in (b) the pseudo prompts and prompt encoder can be optimized in a differentiable way. Sometimes, adding few task-related anchor tokens (such as "capital" in (b)) will bring further improvement.

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

Results on LAMA

Prompt type	Model	P@1
Original (MP)	BERT-base	31.1
	BERT-large	32.3
	E-BERT	36.2
Discrete	LPAQA (BERT-base)	34.1
	LPAQA (BERT-large)	39.4
	AutoPrompt (BERT-base)	43.3
P-tuning	BERT-base	48.3
	BERT-large	**50.6**

Model	MP	FT	MP+FT	P-tuning
BERT-base (109M)	31.7	51.6	52.1	52.3 (+20.6)
-AutoPrompt (Shin et al., 2020)	-	-	-	45.2
BERT-large (335M)	33.5	54.0	55.0	54.6 (+21.1)
RoBERTa-base (125M)	18.4	49.2	50.0	49.3 (+30.9)
-AutoPrompt (Shin et al., 2020)	-	-	-	40.0
RoBERTa-large (355M)	22.1	52.3	52.4	53.5 (+31.4)
GPT2-medium (345M)	20.3	41.9	38.2	46.5 (+26.2)
GPT2-xl (1.5B)	22.8	44.9	46.5	54.4 (+31.6)
MegatronLM (11B)	23.1	OOM*	OOM*	**64.2** (+41.1)

* MegatronLM (11B) is too large for effective fine-tuning.

Table 2. Knowledge probing Precision@1 on LAMA-34k (left) and LAMA-29k (right). P-tuning outperforms all the discrete prompt searching baselines. And interestingly, despite fixed pre-trained model parameters, P-tuning overwhelms the fine-tuning GPTs in LAMA-29k. (MP: Manual prompt; FT: Fine-tuning; MP+FT: Manual prompt augmented fine-tuning; PT: P-tuning).

- **Significantly boost performance on LAMA**
- **Pre-trained models have learned more knowledge than we thought.**

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

Results on Few-shot NLU

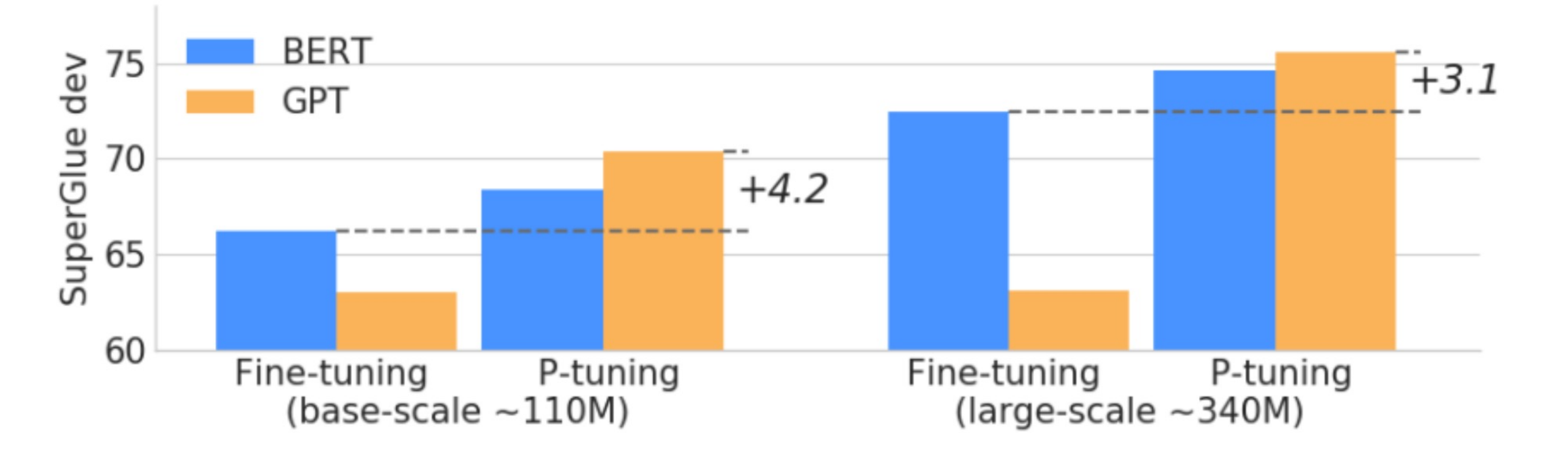
Dev size	Method	BoolQ (Acc.)	CB (Acc.)	CB (F1)	WiC (Acc.)	RTE (Acc.)	MultiRC (EM)	MultiRC (F1a)	WSC (Acc.)	COPA (Acc.)
32	PET*	73.2±3.1	82.9±4.3	74.8±9.2	51.8±2.7	62.1±5.3	33.6±3.2	74.5±1.2	79.8±3.5	85.3±5.1
	PET best†	75.1	86.9	83.5	52.6	65.7	35.2	75.0	80.4	83.3
	P-tuning	77.8 (+4.6)	92.9 (+10.0)	92.3 (+17.5)	56.3 (+4.5)	76.5 (+14.4)	36.1 (+2.5)	75.0 (+0.5)	84.6 (+4.8)	87.0 (+1.7)
Full	GPT-3	77.5	82.1	57.2	55.3	72.9	32.5	74.8	75.0	92.0
	PET‡	79.4	85.1	59.4	52.4	69.8	37.9	77.3	80.1	95.0
	iPET§	80.6	92.9	92.4	52.2	74.0	33.0	74.0	-	-

* We report the average and standard deviation of each candidate prompt's average performance.
† We report the best performed prompt selected on *full* dev dataset among all candidate prompts.
‡ With additional ensemble and distillation.
§ With additional data augmentation, ensemble, distillation and self-training.

- **Few-shot SOTA on SuperGLUE**
- **A more reasonable few-shot setting: small train and SMALL DEV**

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

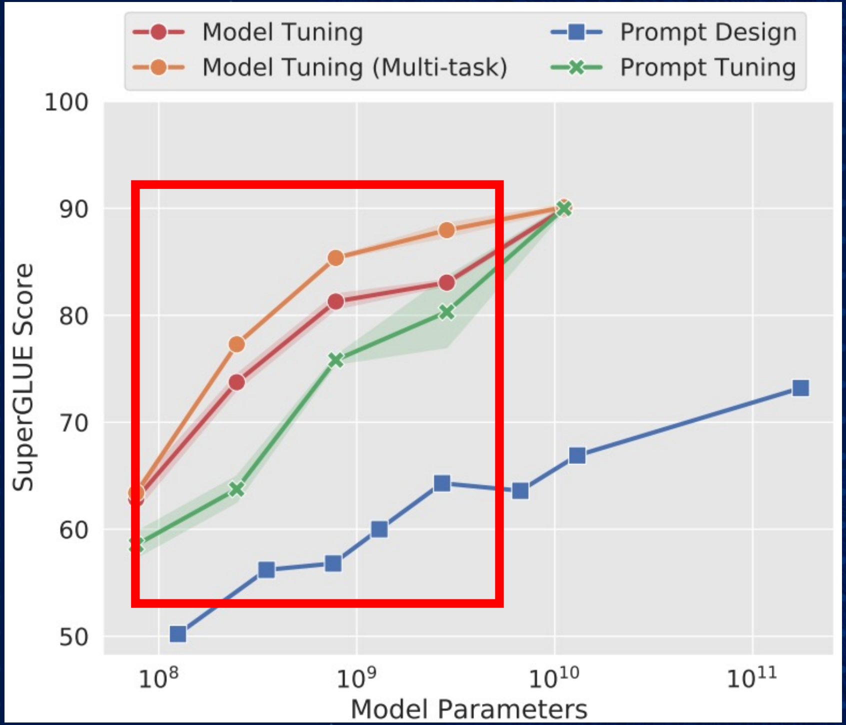
P–Tuning for GPT



- Boost GPT on NLU
- Improve BERT on NLU

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

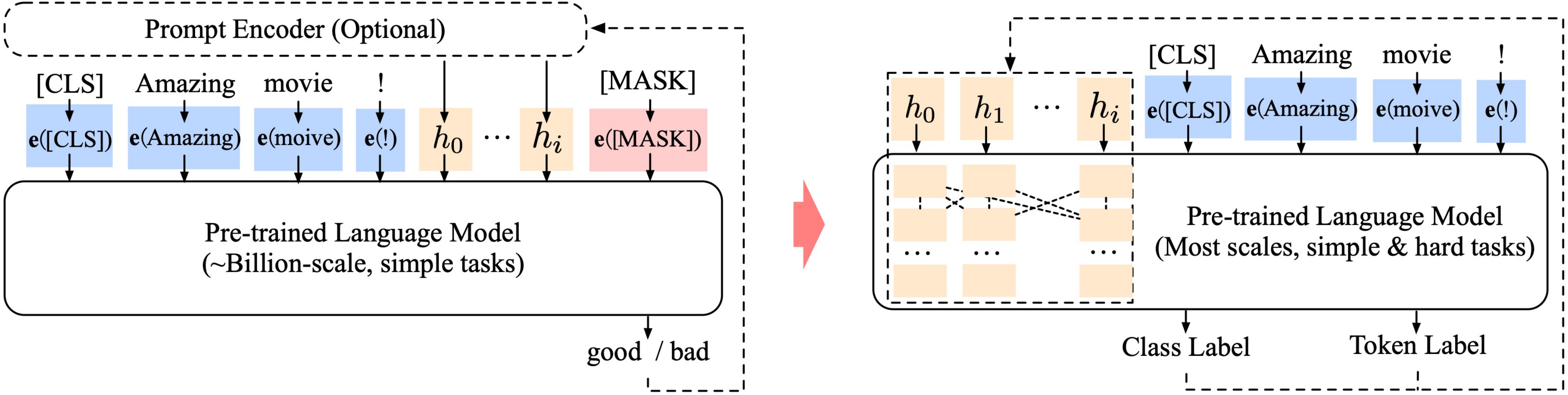
One more thing



P-tuning outperforms fine-tuning only when the # of parameters **>10B!**

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

One more thing



(a) Prompt tuning & P-tuning

(b) P-tuning v2

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

Finally, P-tuning >= fine-tuning

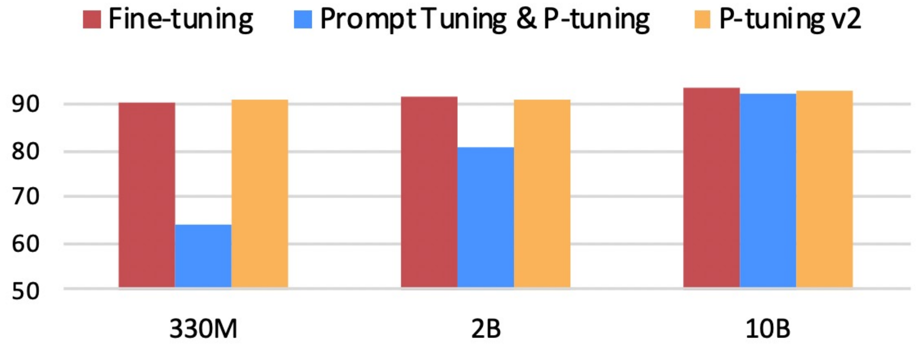


Figure 1: Average scores on RTE, BoolQ and CB of SuperGlue dev. With 0.1% task-specific parameters, P-tuning v2 can be comparable to fine-tuning across different scales of pre-trained models, while prompt tuning & P-tuning can only make it at over 10B scales.

P-tuning achieves comparable performance with only 0.1% task-specific parameters, as fine-tuning even with small models (<10B).

X. Liu et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.

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

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 relativeness is weak!

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

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.

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

Inverse Prompting

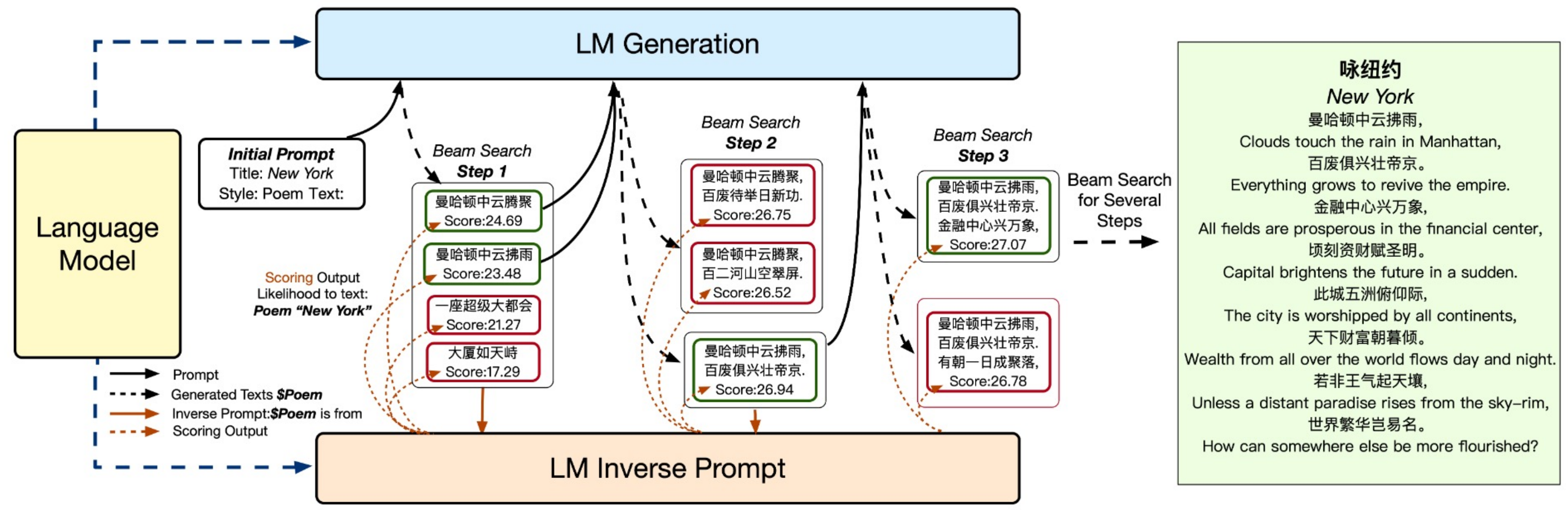
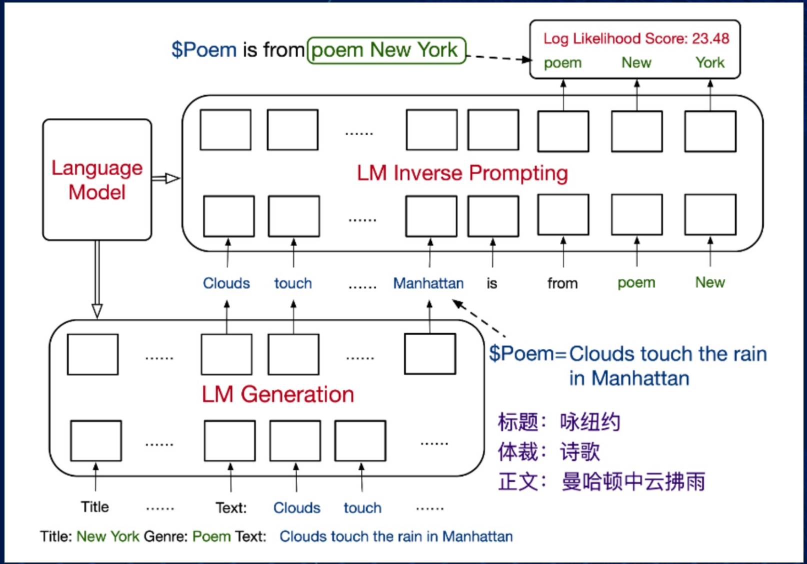


Figure 1: The generation process of open-domain traditional Chinese poems under inverse prompting. Using title *New York* (《咏纽约》) as an example.

Beam search using Inverse Prompting scoring for several steps.

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

Inverse Prompting



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

输入：
标题：咏纽约 体裁：诗歌 正文：

Input:
Title：New York Genre: Poem Text:

输出：
曼哈顿中云拂雨

Output：
Clouds touch the rain in Manhattan.

输入：
“曼哈顿中云拂雨”出自

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

目标输出：
诗歌《咏纽约》

Target Output:
poem New York

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

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

Final Generated Poem

咏纽约

曼哈顿中云拂雨，百废俱兴壮帝京。

金融中心兴万象，顷刻资财赋圣明。

此城五洲俯仰际，天下财富朝暮倾。

若非王气起天壤，世界繁华岂易名。

- *New York*
- Clouds touch the rain in Manhattan
- Everything grows to revive the empire
- All fields are prosperous in the financial center
- Capital brightens the future in a sudden
- The city is worshipped by all continents
- Wealth from all over the world flows day and ni
- Unless a distant paradise rises from the sky-rim
- How can somewhere else become more flouris

Combination of traditional Chinese poem & modern objects/images via Inverse Prompting!

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

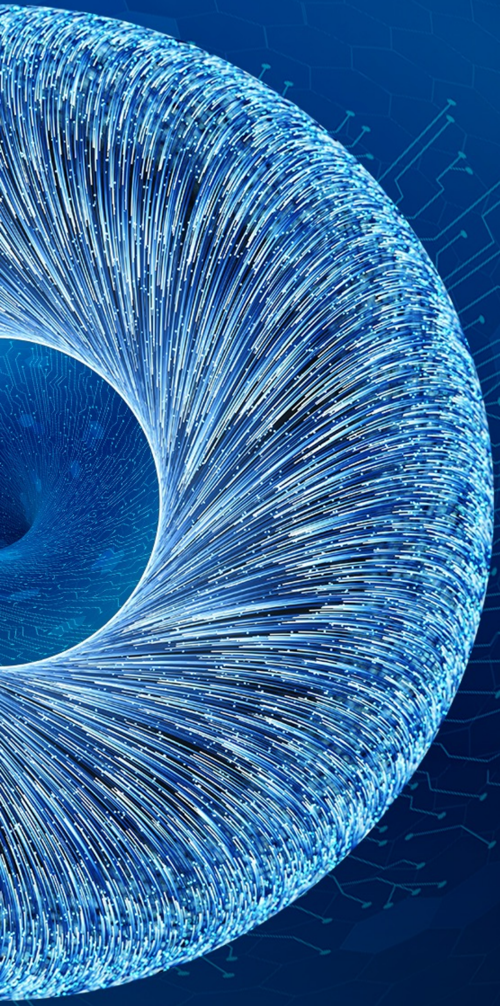
Method	Format (1-5)	Innov.[1] (1-5)	Relevance (1-5)	Aes.[2] (1-5)	Overall (1-10)
Jiuge [28]	**3.60**	2.47	1.99	**3.12**	3.57
Search Baseline	2.79	1.10	1.16	2.44	1.35
Inverse Prompting	2.56	2.71	2.92	2.33	4.00
Inverse Prompting +ST	2.42	**2.92**	**3.65**	2.18	**4.40**

[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%



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 paratmeters.

- 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

The slides will be available soon at http://keg.cs.tsinghua.edu.cn/jietang (or Google "Jie Tang")

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

Tang Jie, Lu Bai

Qiu Jiezhong, Xie Changyu, Xiao Yijia, Zeng Aohan, Li Ziang

FastMoE and trillion large model

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!