南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Advanced Natural Language Processing

Lecture 9: Introduction to Large Language Models

陈冠华 CHEN Guanhua

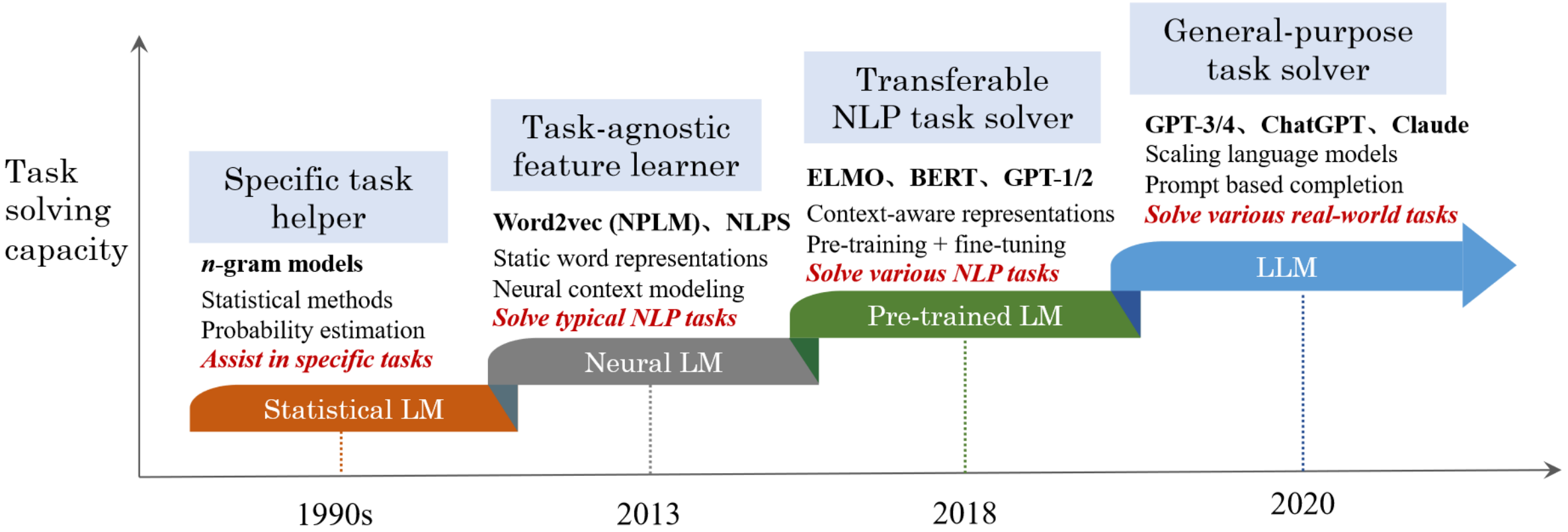
Department of Statistics and Data Science

Content

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Introduction
- Pretraining
- Instruction Tuning
- Alignment
- Future

Introduction



Introduction

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

[Overview Leaderboard | LMArena]

Rank (UB) ↑	Model ↑↓	Score ↑↓
1	gemini-2.5-pro	1452
1	claude-sonnet-4-5-20250929-thinking-32k	1448
1	claude-opus-4-1-20250805-thinking-16k	1448
2	chatgpt-4o-latest-20250326	1441
2	gpt-4.5-preview-2025-02-27	1441
2	gpt-5-high	1440
2	o3-2025-04-16	1440
2	claude-opus-4-1-20250805	1438
2	claude-sonnet-4-5-20250929	1437
3	qwen3-max-preview	1434
3	qwen3-max-preview	1434
8	qwen3-max-2025-09-23	1425
9	gpt-5-chat	1426
9	glm-4.6	1421
9	grok-4-fast	1420
9	deepseek-v3.2-exp-thinking	1418
11	claude-opus-4-20250514-thinking-16k	1420
11	qwen3-235b-a22b-instruct-2507	1418
11	qwen3-vl-235b-a22b-instruct	1418
11	deepseek-r1-0528	1417
11	kimi-k2-0905-preview	1416
11	deepseek-v3.2-exp	1416
11	deepseek-v3.1-terminus	1415
11	deepseek-v3.1-thinking	1415
11	deepseek-v3.1	1415
11	deepseek-v3.1-terminus-thinking	1414
12	kimi-k2-0711-preview	1416

Introduction

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Vision Arena | LMArena

Rank (UB) ↑	Model ↑↓	Score ↑↓
1	gemini-2.5-pro	1241
1	chatgpt-4o-latest-20250326	1234
2	gpt-4.5-preview-2025-02-27	1220
3	o3-2025-04-16	1213
3	gpt-5-chat	1212
3	gpt-4.1-2025-04-14	1207
3	gemini-2.5-flash	1203
3	claude-opus-4-20250514-thinking-16k	1203
3	claude-opus-4-20250514-thinking-16k	1203
3	gemini-2.5-flash-preview-09-2025	1201
3	hunyuan-vision-1.5-thinking	1200 Preliminary
3	claude-sonnet-4-20250514-thinking-32k	1198
3	qwen3-vl-235b-a22b-instruct	1197
5	gpt-5-high	1194
5	qwen3-vl-235b-a22b-thinking	1189
5	claude-3-7-sonnet-20250219-thinking-32k	1189
6	o4-mini-2025-04-16	1194
6	gpt-4.1-mini-2025-04-14	1194

Introduction

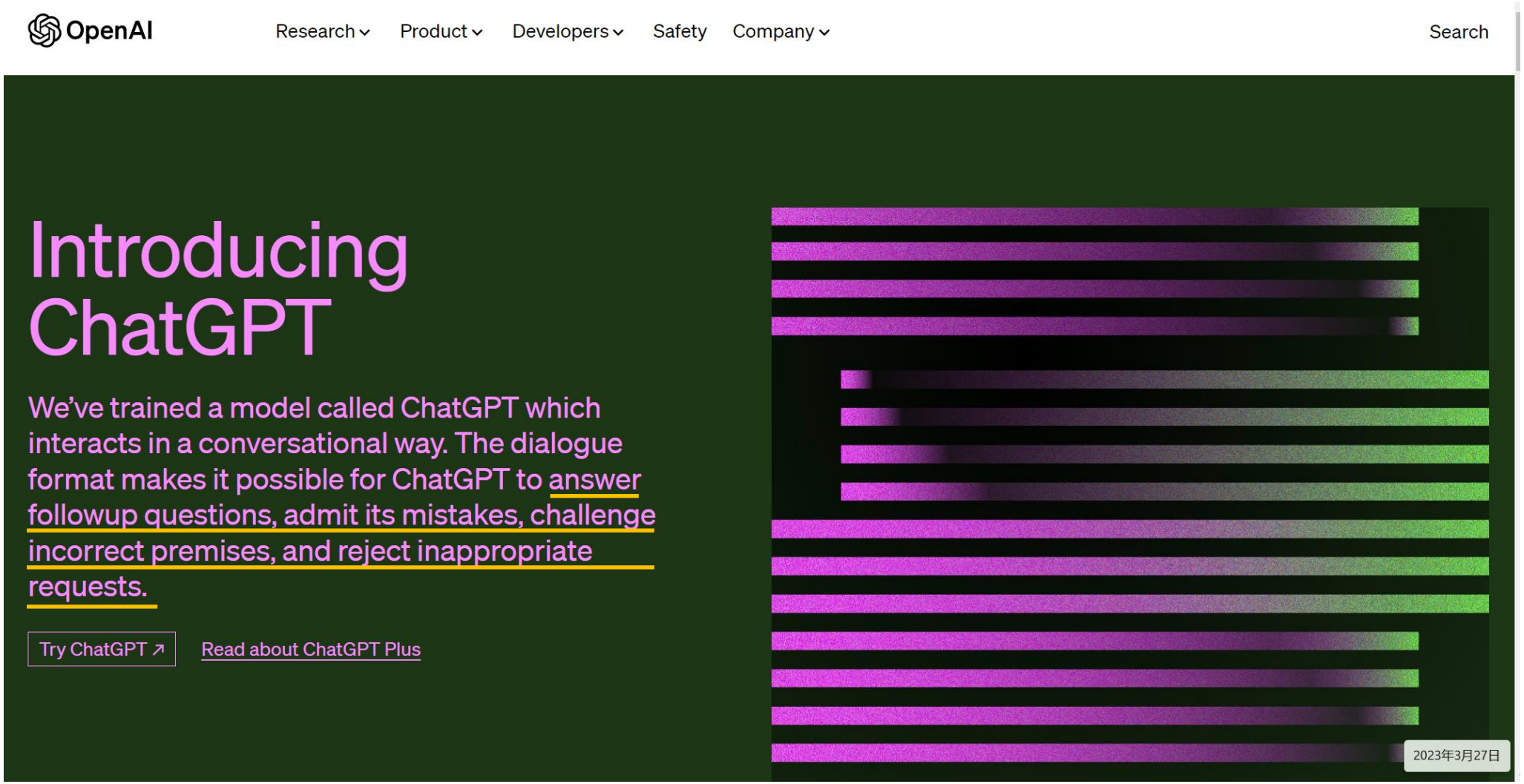
Sam Altman's blog post "The Intelligence Age"

- Visions
  - AI can help solve hard problems
  - Drastically improve quality of life
  - Create significant societal progress.
  - Personal AI teams aiding individuals
  - Personalized virtual tutors for education
  - Breakthroughs in healthcare and software development
- Access to abundant compute resources and stresses the importance of navigating the challenges and risks of this new era, including its impact on labor markets.

拆解近1个月来全球AI领导者的6次访谈和2篇万字长文-36氪

Introduction

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

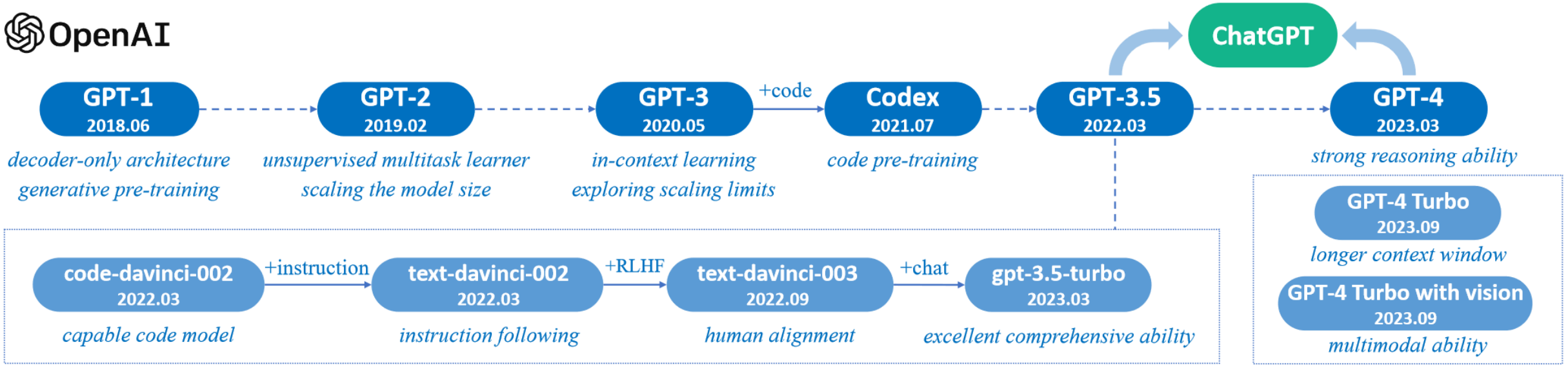


https://openai.com/blog/chatgpt

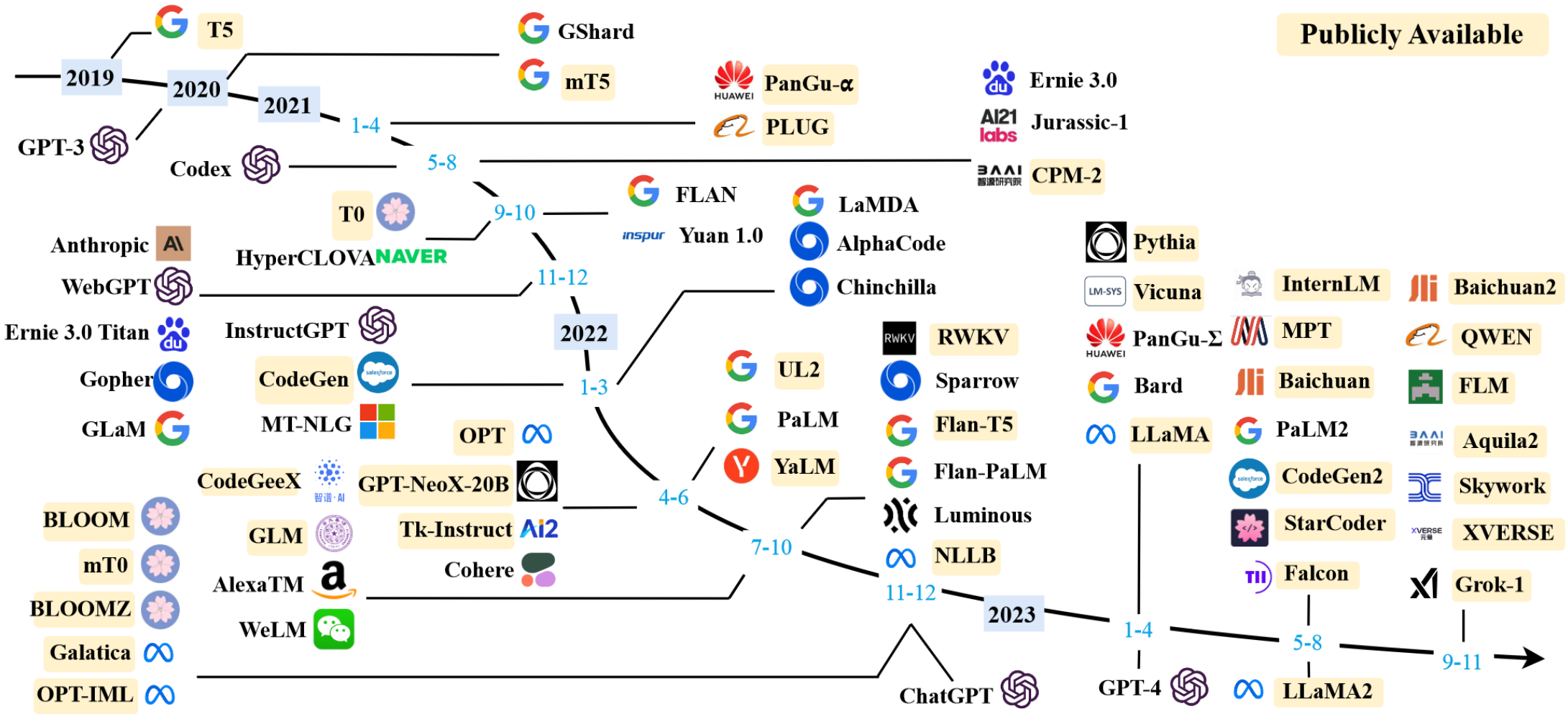
History of ChatGPT

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

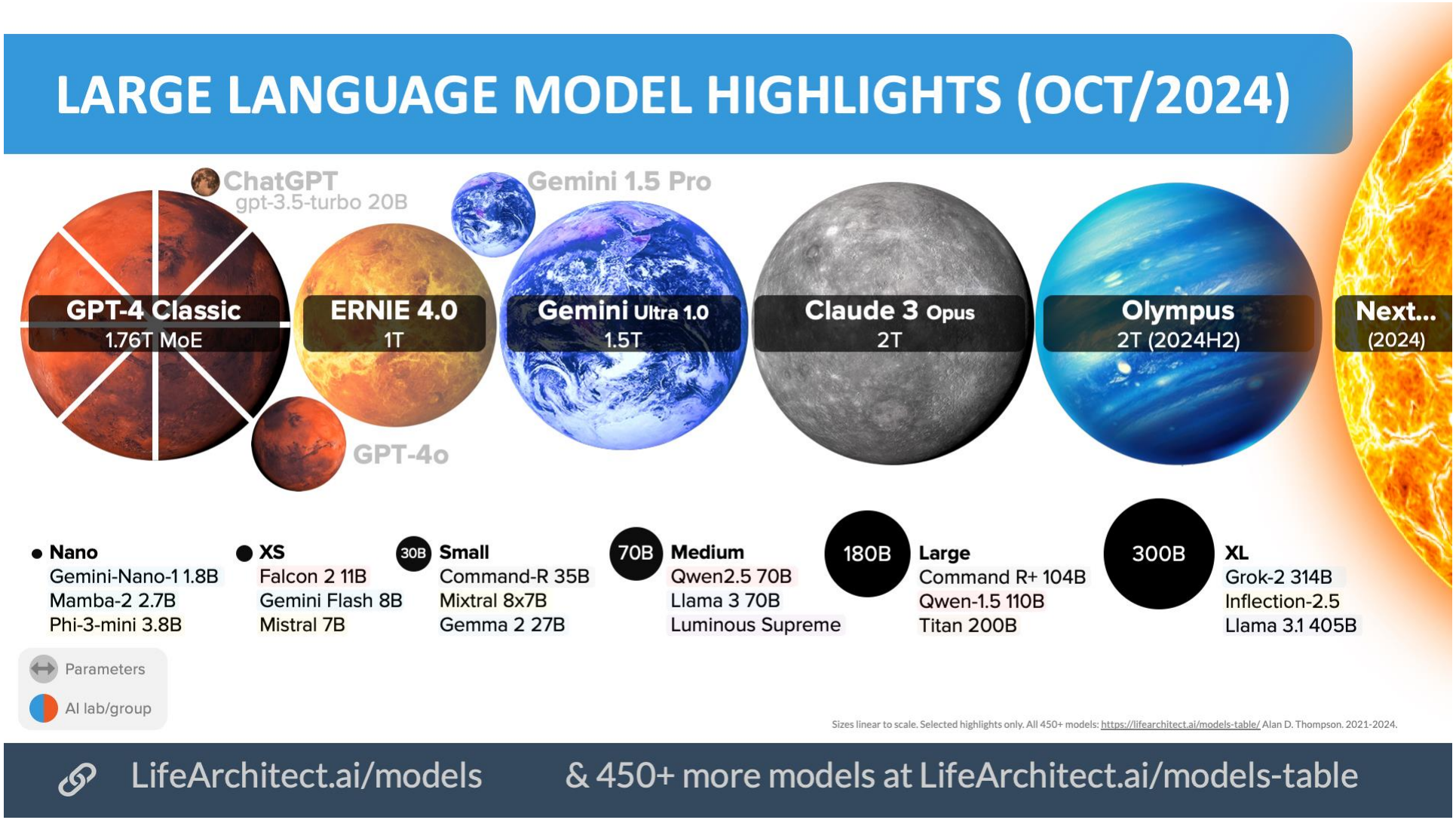
- From GPT-1 to ChatGPT



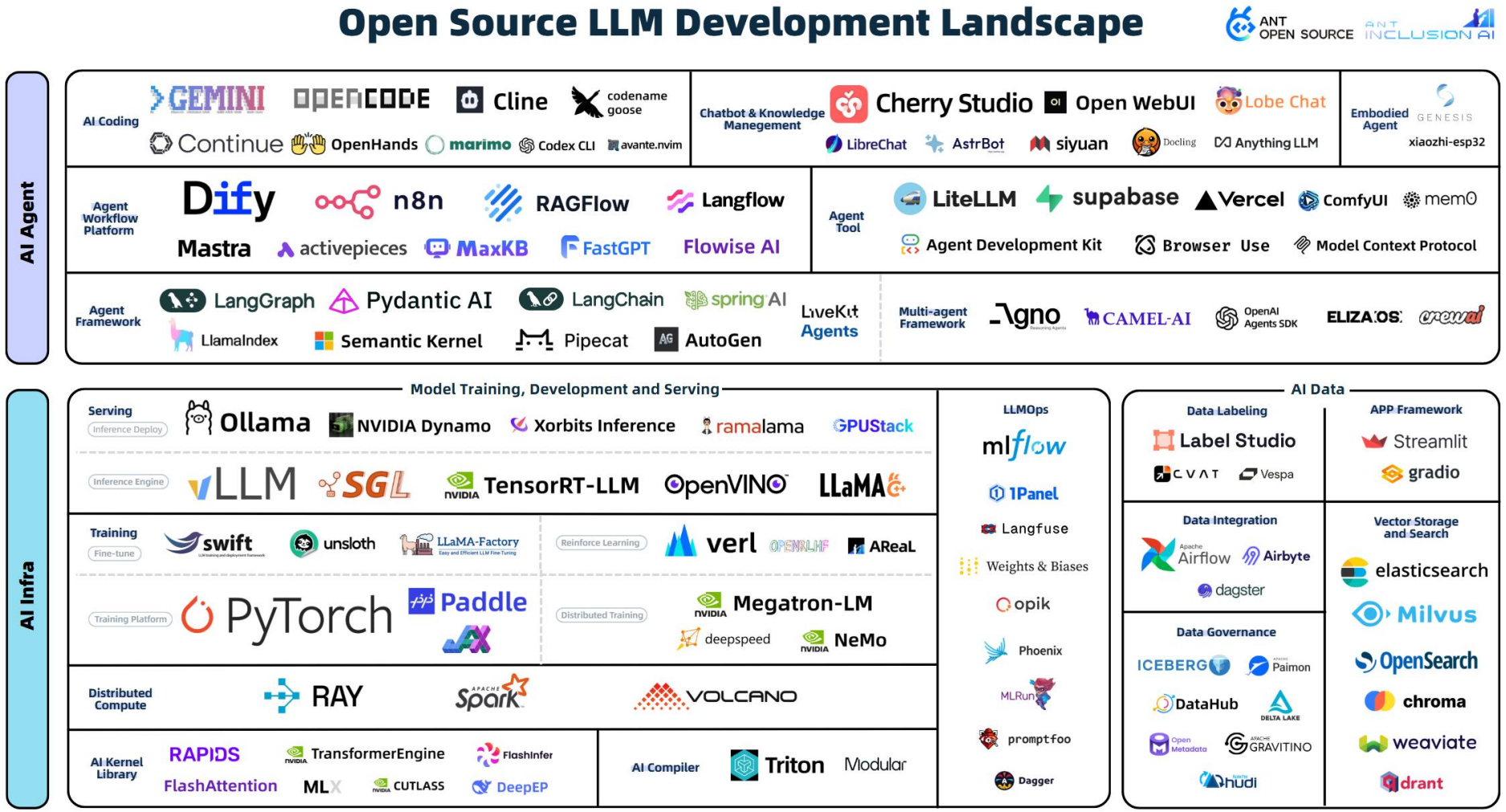
Introduction



Introduction



Introduction



https://github.com/antgroup/llm-oss-landscape

[Open Source LLM Development Landscape 2025]

Three Training Stages of LLMs

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Stages	Goal	Gained Abilities	Challenges
Pretraining	Strong foundation model	Text understanding/ generation, in-context learning, world knowledge, code understanding/generation, reasoning and CoT	Data scale/quality/ratio, Large-scale distributed training, Training stability, Code pretraining
Instruct tuning	Activating abilities	Following instructions, Generalizing to novel instructions, Complex reasoning/CoT	Large-scale diverse instruction data, Improve the code and reasoning abilities
Alignment	Align with human values	Unbiased/informative response Reject instructions that violate human values/beyond its ability	Alignment tax Align with SFT/RLHF

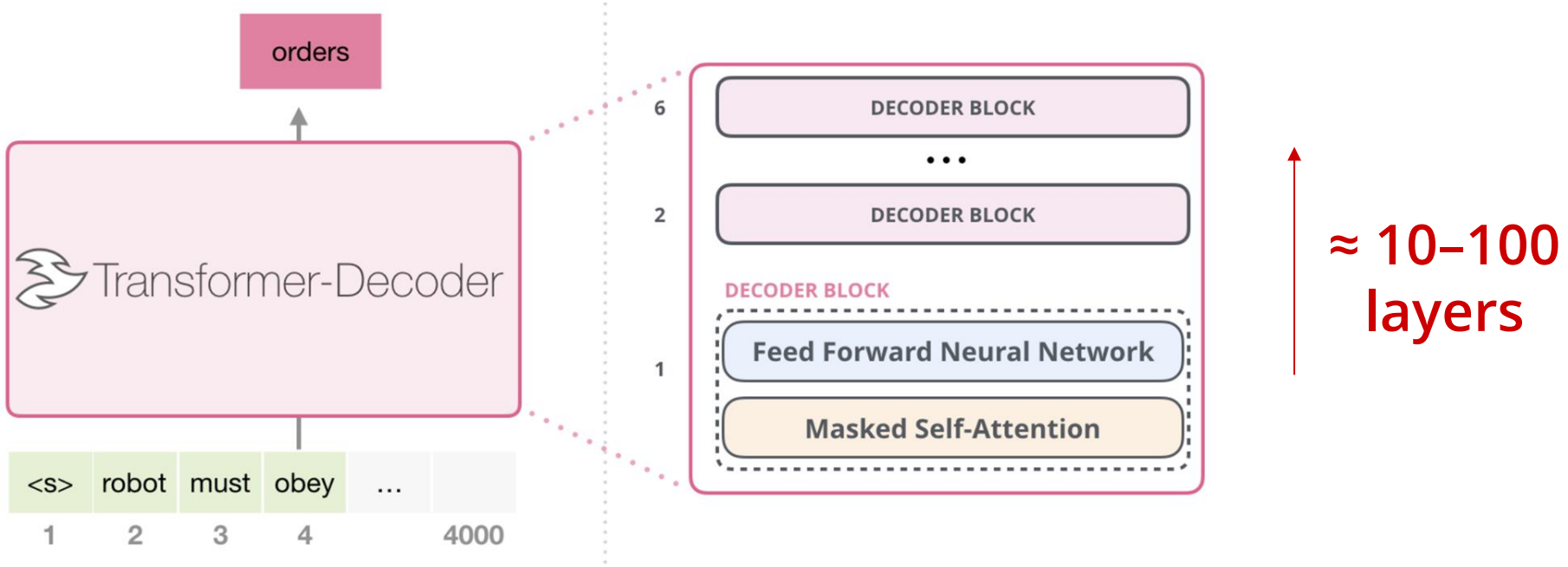
Content

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Introduction
- Pretraining
- Instruction Tuning
- Alignment
- Future

Pretrained Language Model

- Learn to predict the next token, $p(\text{next token} \mid \text{previous tokens})$
- Model structure

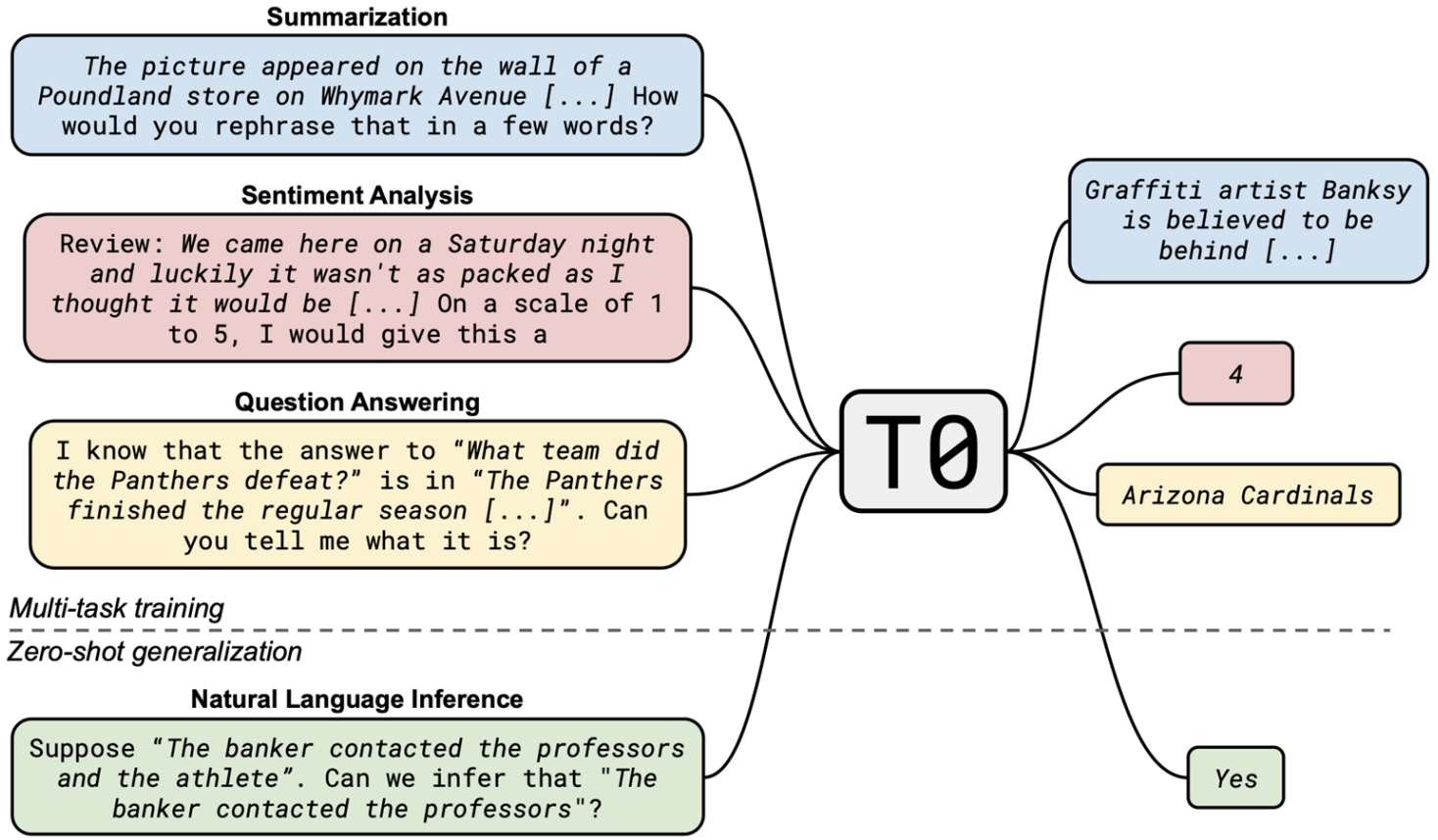


https://jalammar.github.io/illustrated-gpt2/

Diverse Tasks

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Diverse tasks can be modeled as next token prediction task



[Sanh et al., 2022: https://arxiv.org/pdf/2110.08207.pdf]

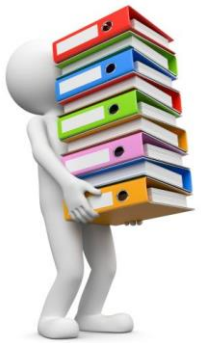
Code Pretraining

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Code with comments
  - Align with natural language
- Divide and conquer, solve problems by steps
- Long-distance dependance

```
112     def _set_current_step(self, epoch: int):
113         """Sets current step number.
114
115         Args:
116             epoch (int): Step number to be set.
117         """
118         self._cur_step = epoch * self._steps_per_epoch
119
120     def _call_timer(self, action: str, item: str, *args, **kwargs) -> None:
121         """Call timer funciton with a given timer name.
122
123         Args:
124             action (str): Function to be called on timer.
125             item (str): Name of the timer.
126             args (list): args used for action function.
127             kwargs (dict): kwargs used for action function.
128         """
129
130         if self._timer is not None:
131             getattr(self._timer, action)(item, *args, **kwargs)
132
133     def _reset_states(self) -> None:
134         """Clear trainer states"""
135         self.states = dict()
```

Pretrain-and-Finetune for Small LMs



Pre-train

PLM

- Next token prediction

Fine-tune

Model for Task 1

Model for Task 2

Model for Task 3

Downstream Tasks

Open-Sourced Pretrain Datasets

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- wikimedia/wikipedia · Datasets at Hugging Face
- HuggingFaceFW/fineweb-2 · Datasets at Hugging
- Facetogethercomputer/RedPajama-Data-1T · Datasets at Hugging Face
- allenai/dolma · Datasets at Hugging Face
- Skywork/SkyPile-150B · Datasets at Hugging Face
- CASIA-LM/ChineseWebText at main

Data Preprocessing Pipeline

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Source	Doc Type	UTF-8 bytes *(GB)*	Documents *(millions)*	Unicode words *(billions)*	Llama tokens *(billions)*
Common Crawl	web pages	9,812	3,734	1,928	2,479
GitHub	code	1,043	210	260	411
Reddit	social media	339	377	72	89
Semantic Scholar	papers	268	38.8	50	70
Project Gutenberg	books	20.4	0.056	4.0	6.0
Wikipedia, Wikibooks	encyclopedic	16.2	6.2	3.7	4.3
Total		**11,519**	**4,367**	**2,318**	**3,059**

Table 1: The Dolma corpus at-a-glance. It consists of three trillion tokens sampled from a diverse set of domains;

[2402.00159] Dolma: an Open Corpus of Three Trillion Tokens for Language Model Pretraining Research

Data Preprocessing Pipeline

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Acquisition & Language Filtering
- Quality Filtering
  - Model and heuristic filters
- Deduplication
- Content Filtering
  - Toxic, personal identifiable information
  - FastText classifiers, regular expressions

Data Preprocessing Pipeline

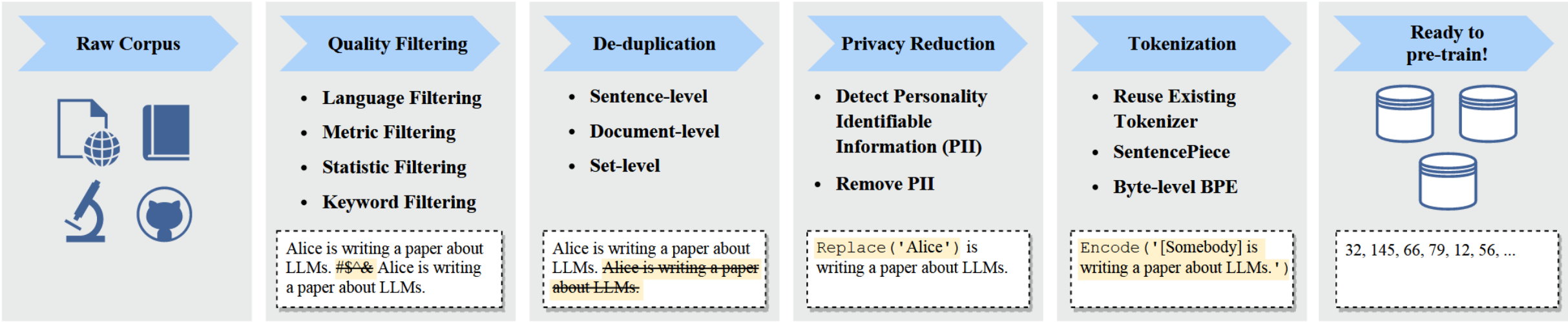


Fig. 7: An illustration of a typical data preprocessing pipeline for pre-training large language models.

What are the goals of pre-training?

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- The goals of pre-training are to get the language model to:
  - learn the structure of natural language
  - learn humans' understanding of the world (as encoded in the training data)
  - become incredibly good simulator of the training data
- Ideal data for pre-training:
  - Large!
  - Data that is high-quality, clean, and diverse.
  - Books, Wikipedia, news, scientific papers, etc.

Pre-training Data Reality

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- In practice, the web is the most viable option for data collection.
  - In the digital era, this is the go-to place for general domain human knowledge.
  - It is massive and unlikely to grow slower than computing resources*
  - Publicly available*

High quality data eventually runs out.

- Web data is plentiful, but can be challenging to work with.
  - Copyright and usage constraints can get extremely complicated
  - Data is noisy, dirty, and biased
  - Data is contaminated with auto-generated text (not just from LLM usage, but also tons of templated text)

Data Preprocessing Challenges

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- At what granularity should filtering be performed?
- Word-level, sentence-level, paragraph-level, document level?
- What constitutes “high quality” or “non-toxic”?
- Are our filters/classifiers multilingual?
- Do they treat all groups equally?
- It is very expensive to ablate pre-training dataset decisions.

LLM: in-Context Learning

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Zero-shot

The model predicts the answer given only a discription of the task. No gradient updates are performed.

Translate English to French : ← Task description

cheese => ← Prompt

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.The model predicts the answer given only a discription of the task. No gradient updates are performed.

Translate English to French : ← Task description

sea otter => loutre de mer ← Example

cheese => ← Prompt

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

Translate English to French : ← Task description

sea otter => loutre de mer

peppermint => menthe poivree ← Examples

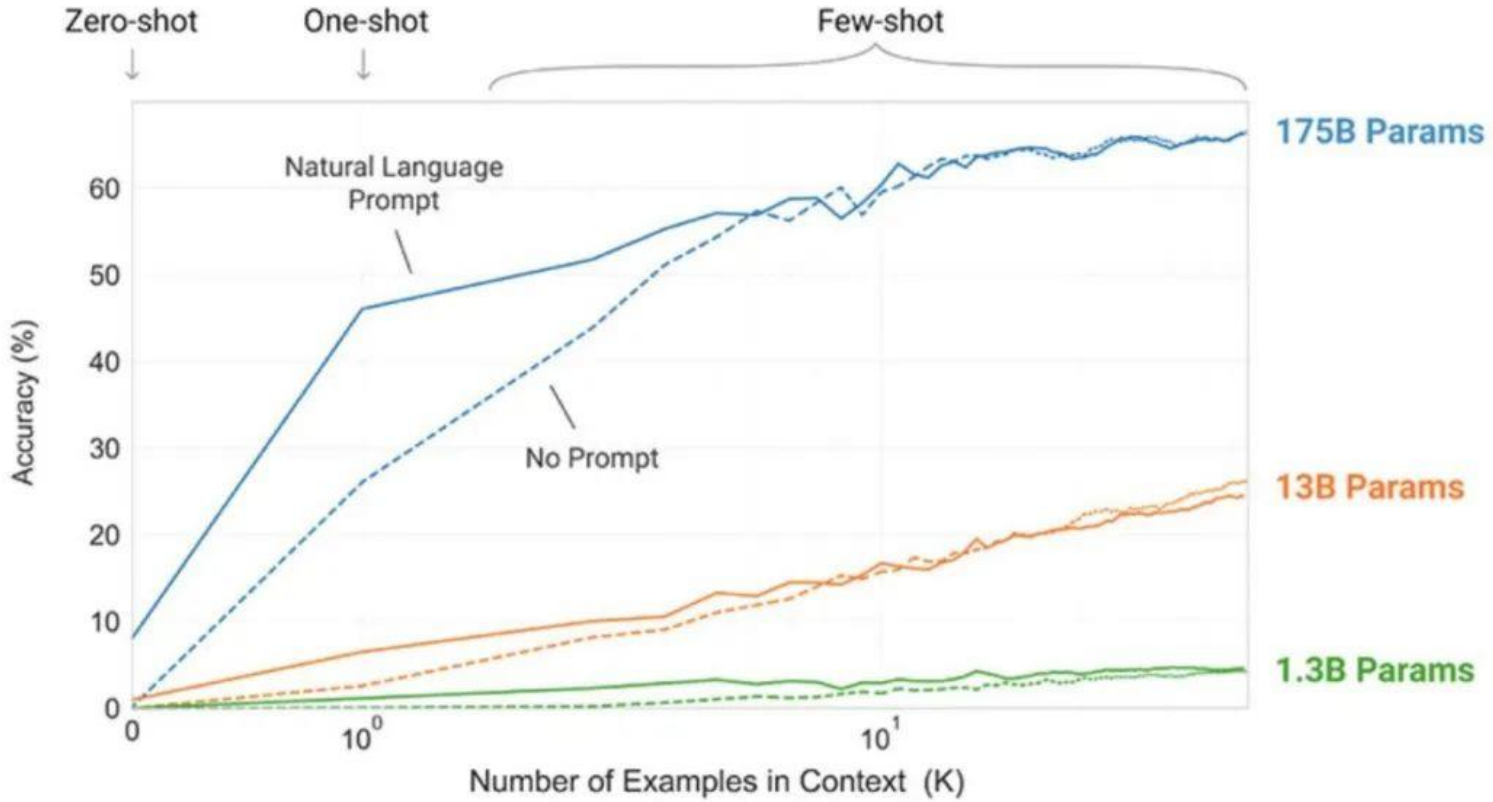
plush girafe => girafe peluche

cheese => ← Prompt

语境学习（In-context learning）

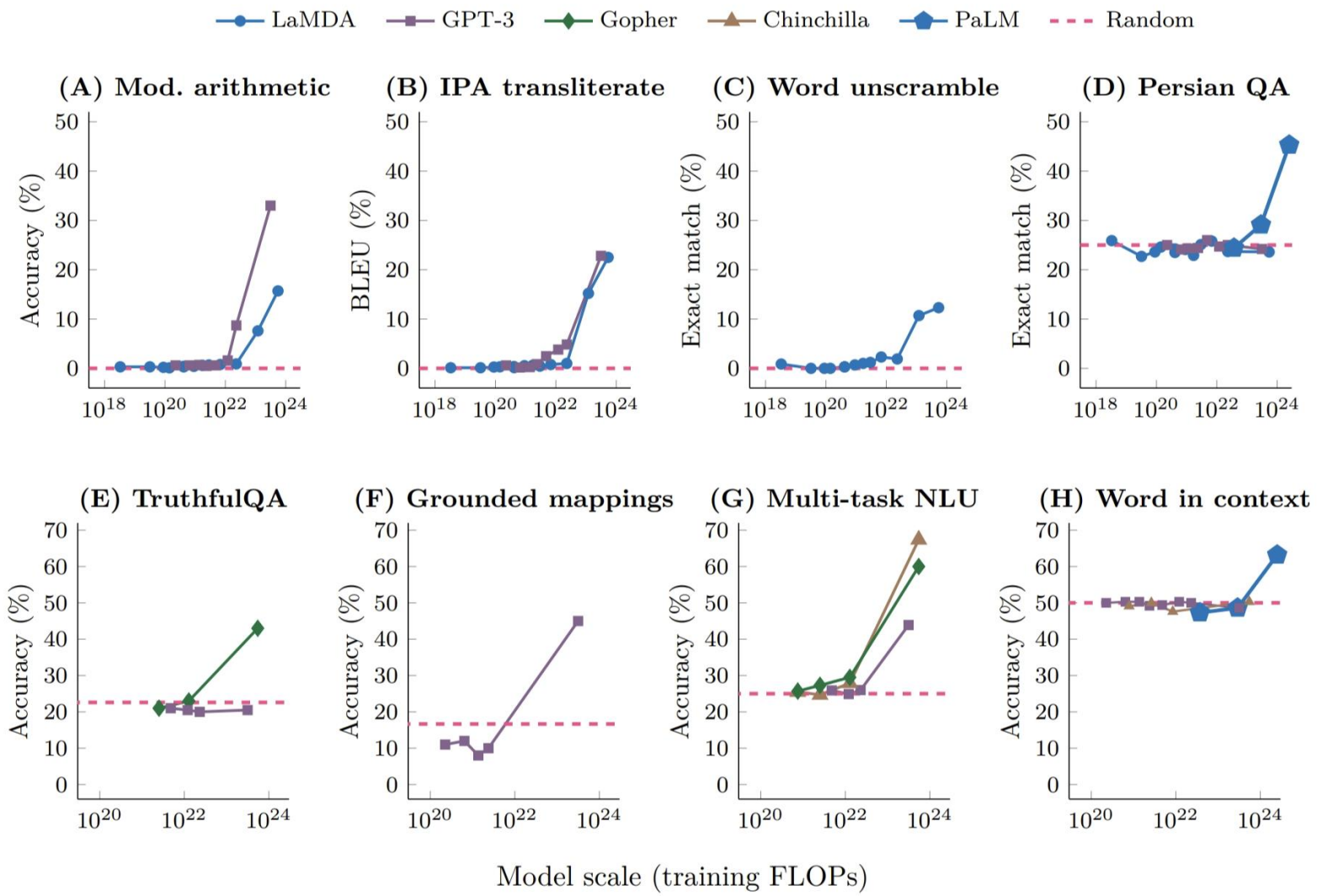
LLM: Emergent Abilities

- An ability is emergent if it is not present in smaller models but is present in larger models.



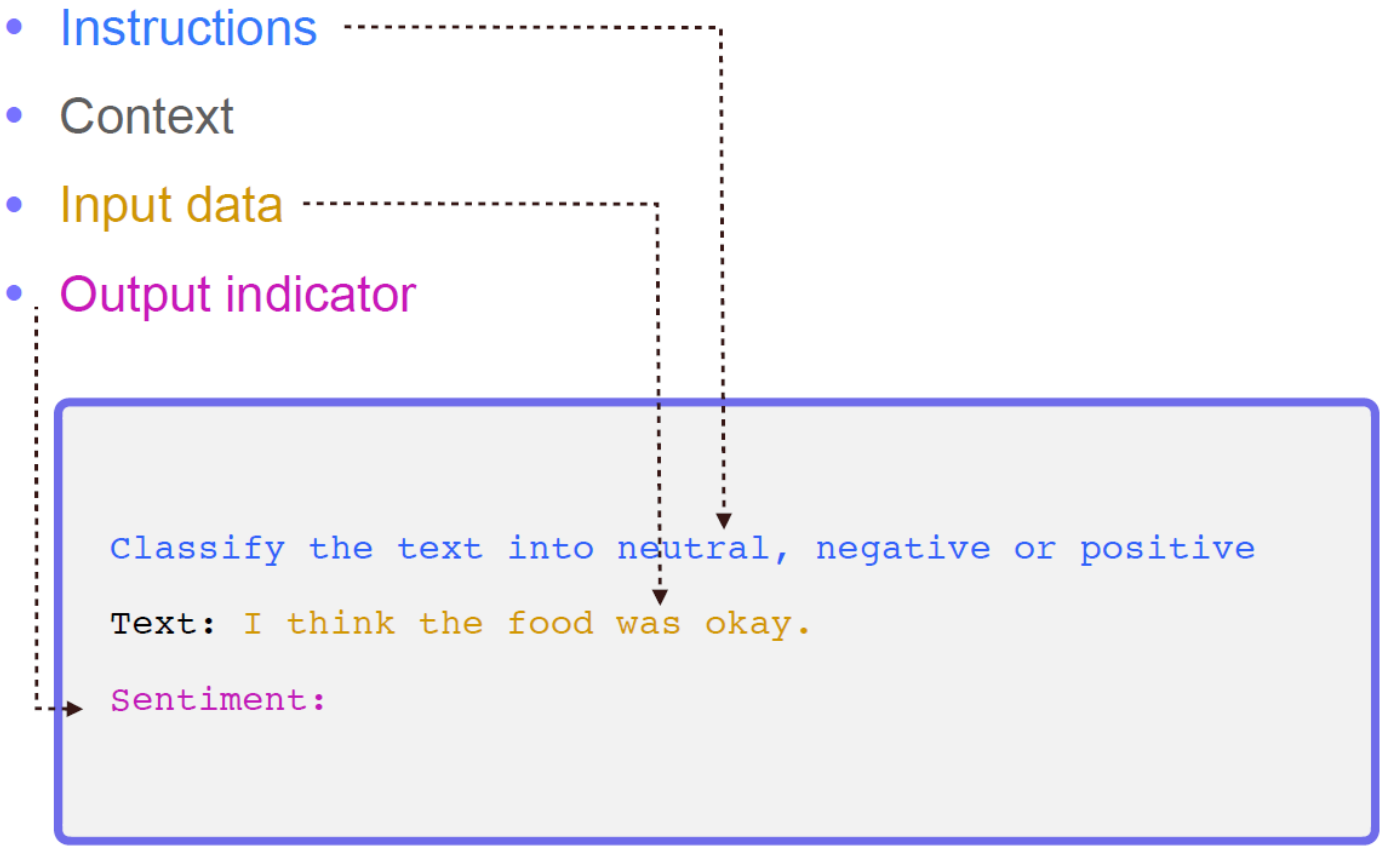
[2206.07682] Emergent Abilities of Large Language Models (arxiv.org)

LLM: Emergent Abilities



Interact with LLM

- Prompt



Chain-of-Thought (CoT)

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- A toy machine learning problem: last-letter-concatenation

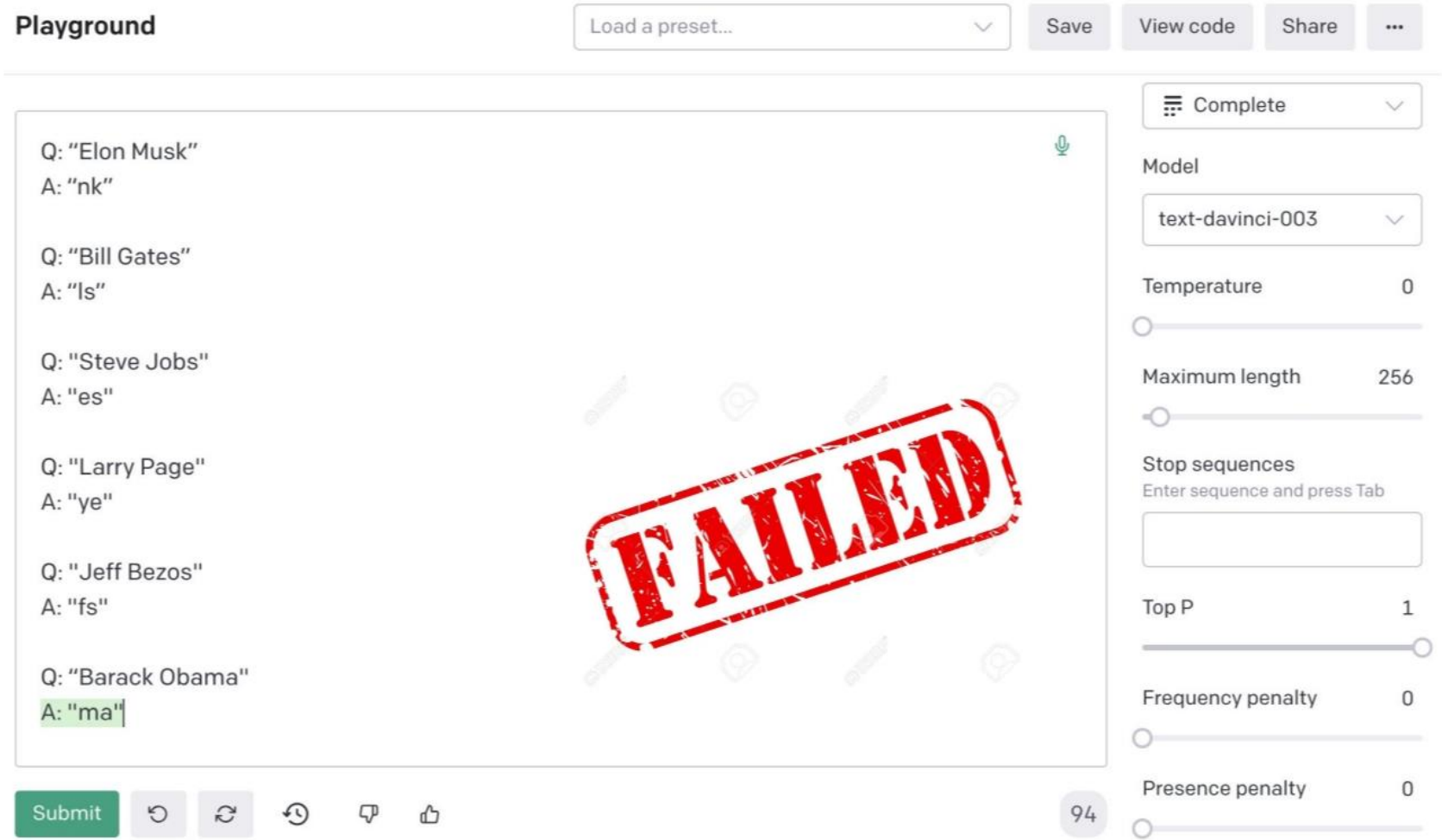
Input	Output
“Elon Musk”	“nk”
“Bill Gates”	“ls”

Teach Language Models to Reason by Denny Zhou

Chain-of-Thought (CoT)

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Simply use in-context learning does not work



Chain-of-Thought (CoT)

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- CoT: Adding “thought” before “answer”

Q: “Elon Musk”
A: the last letter of "Elon" is "n". the last letter of "Musk" is "k". Concatenating "n", "k" leads to "nk". so the output is "nk".

thought

Q: “Bill Gates”
A: the last letter of "Bill" is "l". the last letter of "Gates" is "s". Concatenating "l", "s" leads to "ls". so the output is "ls".

Q: “Barack Obama"
A: the last letter of "Barack" is "k". the last letter of "Obama" is "a". Concatenating "k", "a" leads to "ka". so the output is "ka".

Chain-of-Thought (CoT)

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

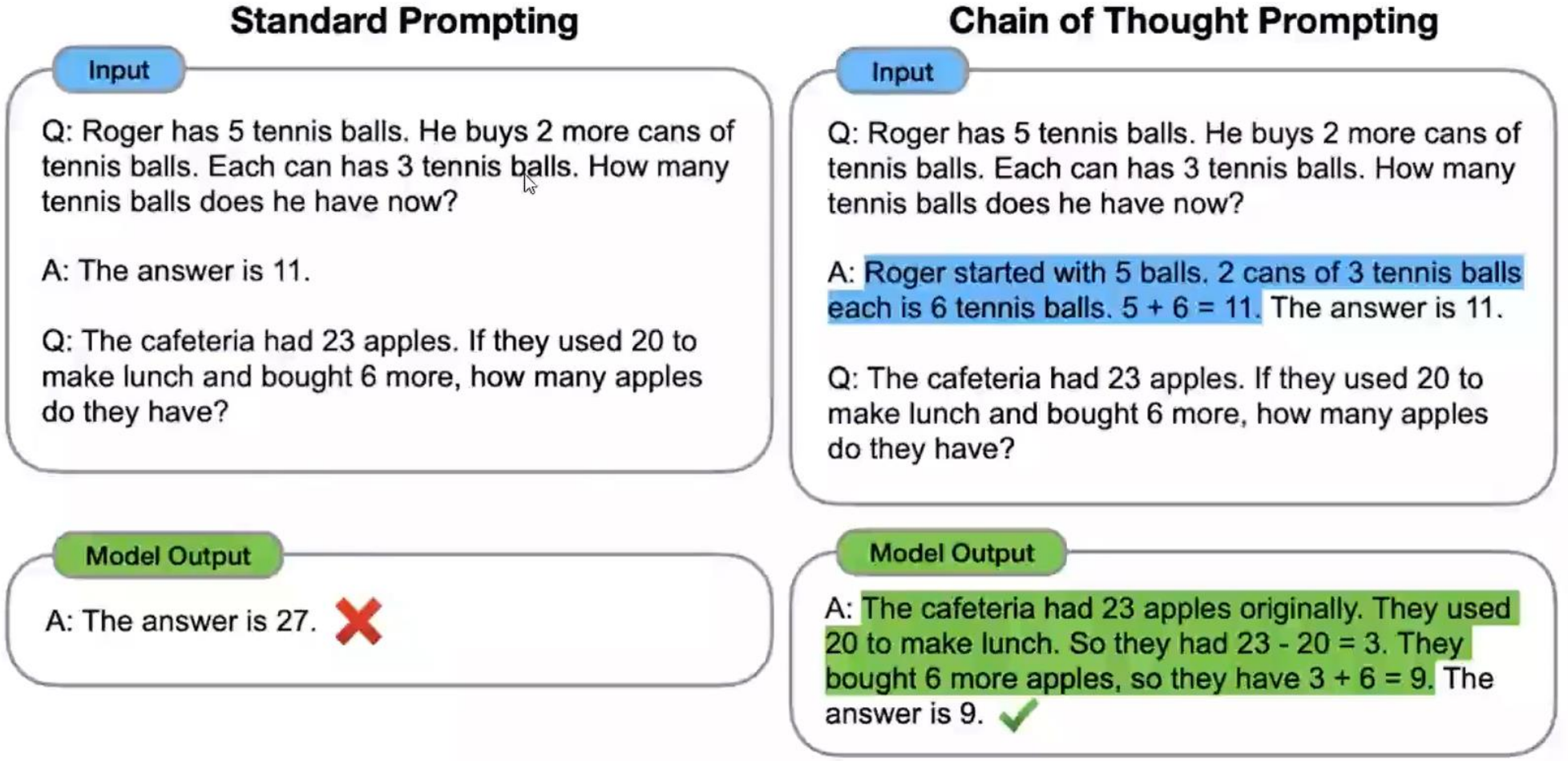
- Sufficiently large language models can generate chains of thought if demonstrations of chain-of-thought reasoning are provided in the exemplars for few-shot prompting
- Call it a chain of thought to better capture the idea that it mimics a step-by-step thought process for arriving at the answer
- Benefits
  - Allows models to decompose multi-step problems into intermediate steps
  - Provides an interpretable window into the behavior of the model, suggesting how it might have arrived at a particular answer and providing opportunities to debug where the reasoning path went wrong
  - Can be used for tasks such as math word problems, commonsense reasoning, and symbolic manipulation, and is potentially applicable to any task that humans can solve via language

Teach Language Models to Reason by Denny Zhou

Chain-of-Thought (CoT)

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

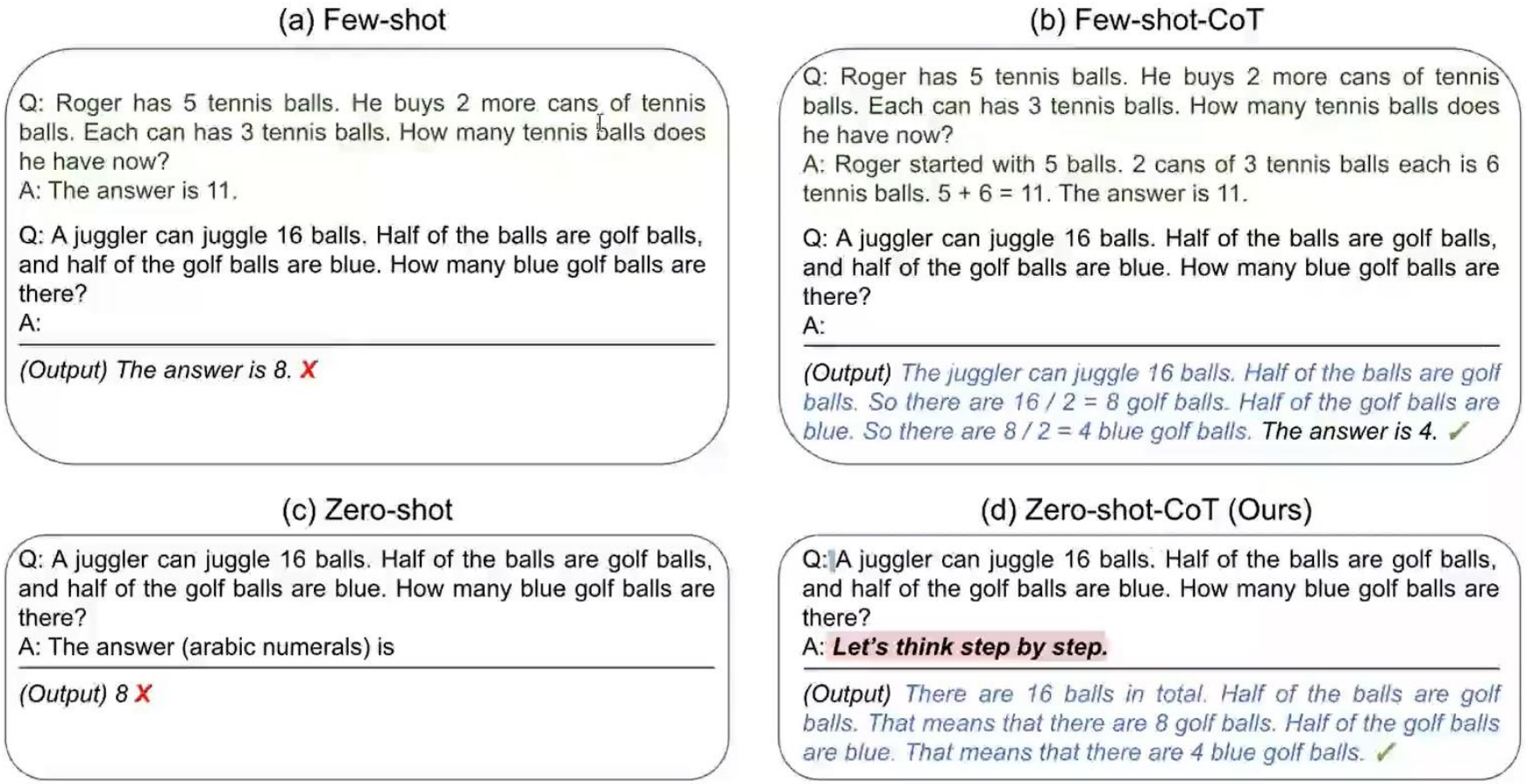
- One-shot COT



Chain-of-Thought (CoT)

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Zero-shot COT: add "Let's think step by step“ into the prompt



Next Problem

- Pretrained model learns to predict the next token
- The goal of LLM: follow users' instructions effectively and safely
  - Instruction tuning
  - Alignment

Stages	Goal	Gained Abilities	Challenges	Models
Pretraining	Strong foundation model	Text understanding/ generation, in-context learning, world knowledge, code understanding/generation, reasoning and CoT	Data scale/quality/ratio, Large-scale distributed training, Training stability, Code pretraining	GPT-3, Gopher, Chinchilla, PaLM, BLOOM

Content

Instruction Tuning

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Improve the performance of instruction following

PROMPT *Explain the moon landing to a 6 year old in a few sentences.*

COMPLETION

GPT-3

```
Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.
```

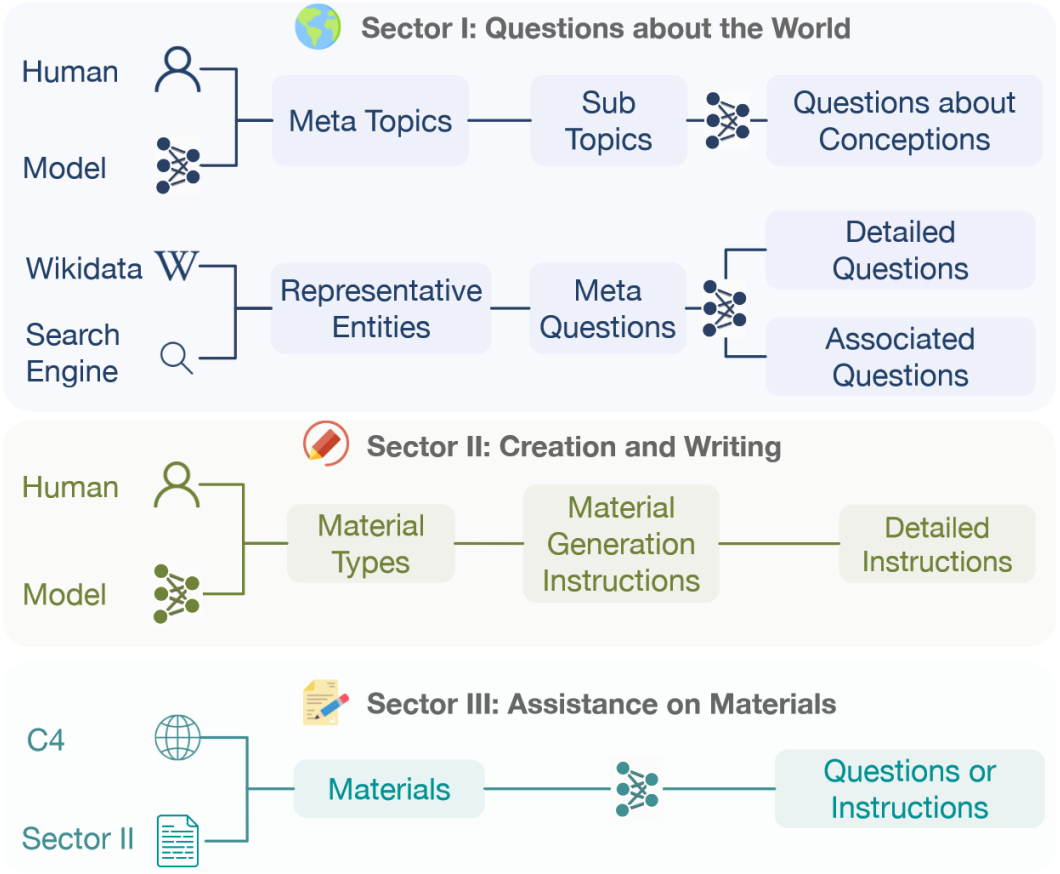
InstructGPT

```
People went to the moon, and they took pictures of what they saw, and sent them back
to the earth so we could all see them.
```

Instruction Tuning

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- High quality and diverse instruction data



Technology	Health and wellness	Travel and adventure
Food and drink	Art and culture	Science and innovation
Fashion and style	Relationships and dating	Sports and fitness
Nature and the environment	Music and entertainment	Politics and current events
Education and learning	Money and finance	Work and career
Philosophy and ethics	History and nostalgia	Social media and communication
Creativity and inspiration	Personal growth and development	Spirituality and faith
Pop culture and trends	Beauty and self-care	Family and parenting
Entrepreneurship and business	Literature and writing	Gaming and technology
Mindfulness and meditation	Diversity and inclusion	Travel and culture exchange

Articles and Blog Posts	Job Application Material	Stories
Legal Documents and Contracts	Poems	Educational Content
Screenplays	Scripts for Language Learning	Technical Documents and Reports
Marketing Materials	Social Media Posts	Personal Essays
Emails	Scientific Papers and Summaries	Speeches and Presentations
Recipes and Cooking Instructions	News Articles	Song Lyrics
Product Descriptions and Reviews	Programs and Code	

Distribution of Prompt Categories

Table 1: Distribution of use case categories from our API prompt dataset.

Use-case	**(%)**
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

Table 6: Dataset sizes, in terms of number of prompts.

SFT Data			RM Data			PPO Data		
split	source	size	split	source	size	split	source	size
train	labeler	11,295	train	labeler	6,623	train	customer	31,144
train	customer	1,430	train	customer	26,584	valid	customer	16,185
valid	labeler	1,550	valid	labeler	3,488			
valid	customer	103	valid	customer	14,399			

Writing Prompts

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Rule #1 *– Instructions at beginning and ### or """ to separate instructions or context*

❌
```
Rewrite the text below in more engaging language.

{your input here}
```

✅
```
Rewrite the text below in more engaging language.

Text: """
{your input here}
"""
```

Rule #2 *– Be specific and detailed about the desired context, outcome, length, format, and style.*

❌
```
Write a short story for kids
```

✅
```
Write a funny soccer story for kids that teaches the
kid that persistence is key for success in the style of
Rowling.
```

Rule #3 *– Give examples of desired output format*

❌
```
Extract house pricing data from the following text.

Text: """
{your text containing pricing data}
"""
```

✅
```
Extract house pricing data from the following text.

Desired format: """
House 1 | $1,000,000 | 100 sqm
House 2 | $500,000 | 90 sqm
... (and so on)
"""

Text: """
{your text containing pricing data}
"""
```

Rule #7 *– Use leading words to nudge the model toward a pattern*

❌
```
Write a Python function that plots my net worth over 10
years for different inputs on the initial investment
and a given ROI
```

✅
```
# Python function that plots net worth over 10
# years for different inputs on the initial
# investment and a given ROI

import matplotlib

def plot_net_worth(initial, roi):
```

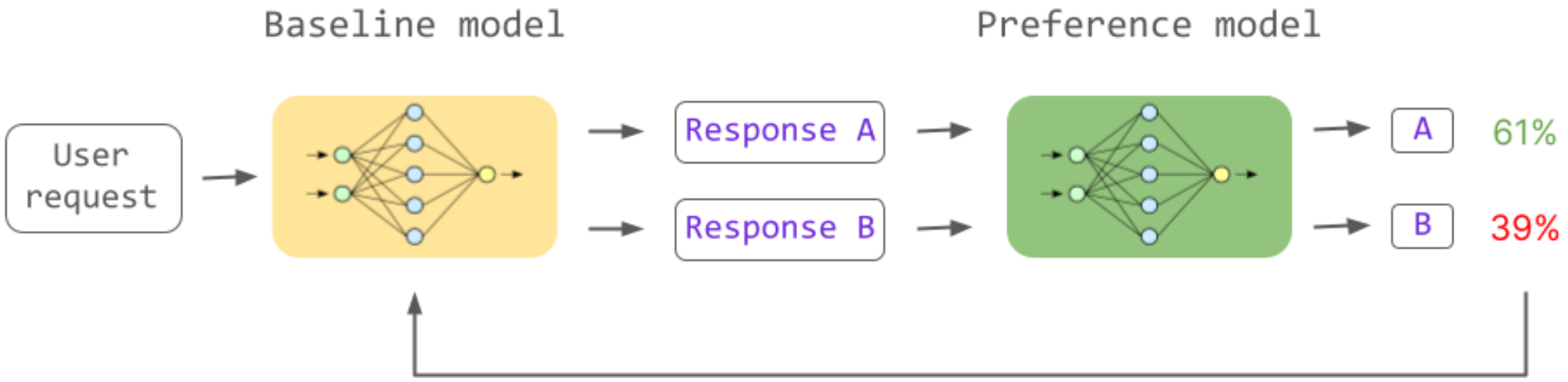
RLHF

Reinforcement Learning from Human Feedback

- Baseline model
  - Unaligned
- Reward model
  - Determine which action a human would prefer within a given list of possibilities
  - Assign a numerical score to each action, effectively ranking them according to human preferences
- Be refined iteratively, altering its internal text distribution to prioritize sequences favored by humans

RLHF

- Human preferences dataset
  - Used to learn the reward function that represents the desired outcome for a particular task
  - Preference orderings, demonstrations, corrections, natural language input



RLHF

Step 1

Collect demonstration data and train a supervised policy.

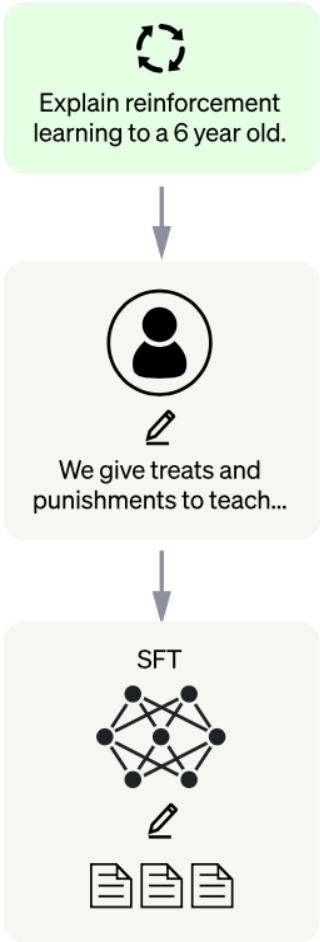
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: """ {summary} """ This is the outline of the commercial for that play: """

A relatively small dataset

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



Prompt Creation

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- **Plain:** We simply ask the labelers to come up with an arbitrary task, while ensuring the tasks had sufficient diversity.
- **Few-shot:** We ask the labelers to come up with an instruction, and multiple query/response pairs for that instruction.
- **User-based:** We had a number of use-cases stated in waitlist applications to the OpenAI API. We asked labelers to come up with prompts corresponding to these use cases.

RLHF

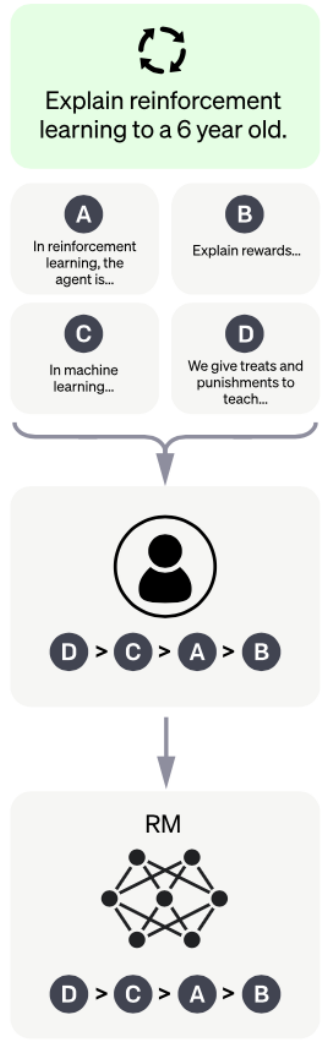
Step 2

Collect comparison data and train a reward model.

$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} E_{(x,y_w,y_l)\sim D}\left[\log\left(\sigma\left(r_\theta(x,y_w) - r_\theta(x,y_l)\right)\right)\right]$$

- A list of prompts is chosen
- SFT model generates multiple outputs (between 4 and 9) for each prompt.
- Annotators rank these outputs from best to worst, forming a new labeled dataset with rankings serving as labels.

A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

RLHF

Step 3

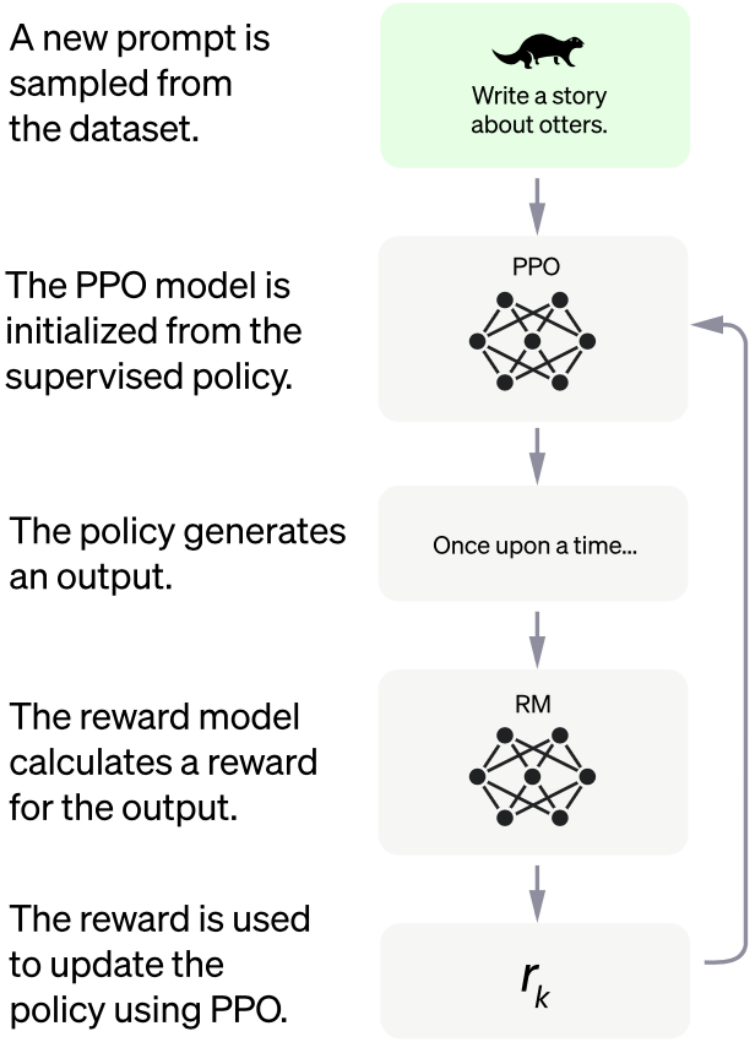
Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

$$\text{objective}\,(\phi) = E_{(x,y)\sim D_{\pi_\phi^{\text{RL}}}}\left[r_\theta(x,y) - \beta \log\left(\pi_\phi^{\text{RL}}(y \mid x)/\pi^{\text{SFT}}(y \mid x)\right)\right] + \gamma E_{x\sim D_{\text{pretrain}}}\left[\log(\pi_\phi^{\text{RL}}(x))\right]$$

The SFT model is fine-tuned via the reward model.
The outcome is the so-called policy model.

Proximal Policy Optimization (PPO)

lvwerra/trl: Train transformer language models with reinforcement learning.



RLHF

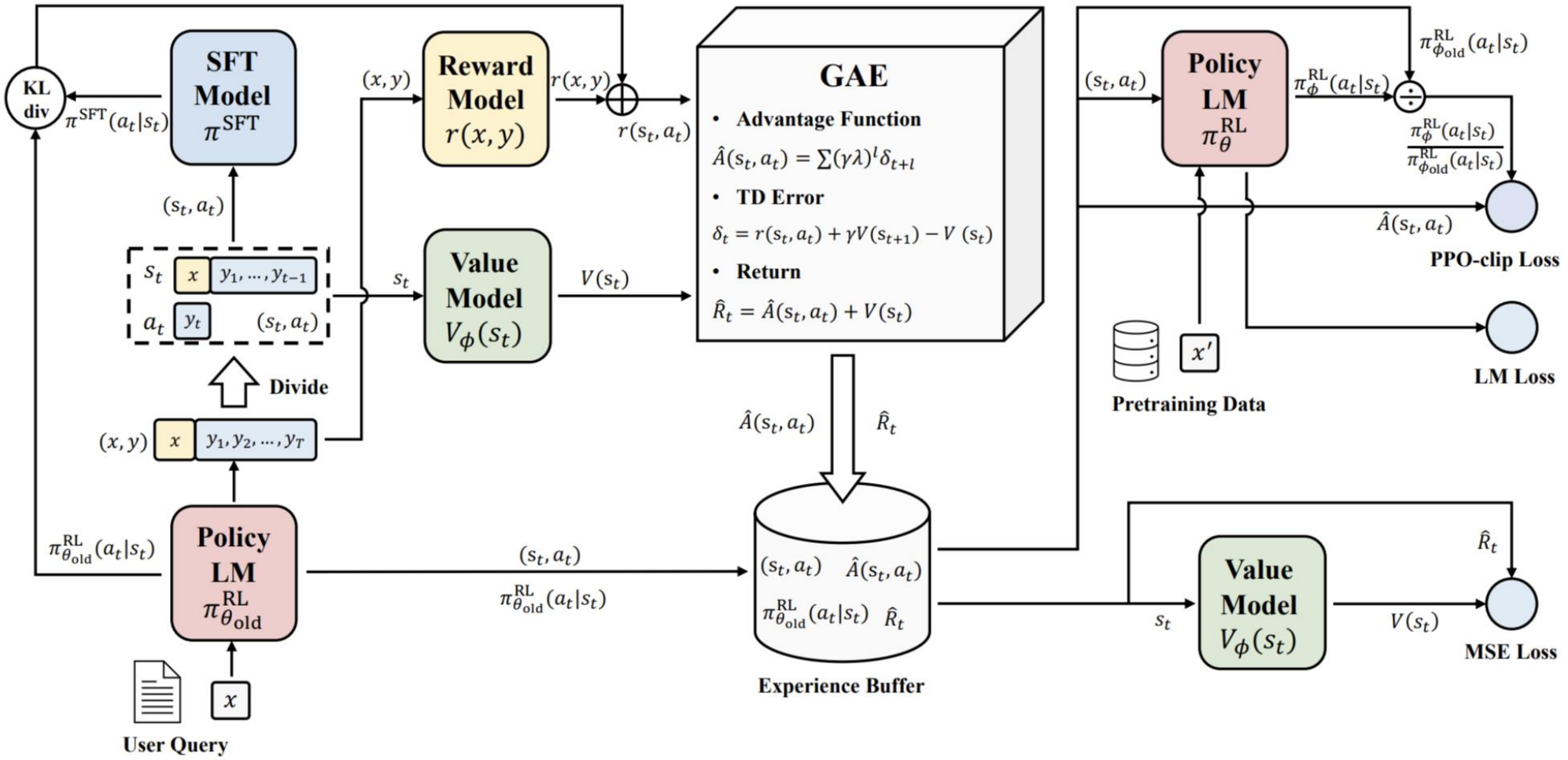
- The second (reward model) and third steps (policy model) are iterated multiple times.
- More comparison data is gathered on the current best policy model, which is then used to train a new reward model and, subsequently, a new policy.

The models are able to generalize the notion of “following instructions”.

volcengine/verl: verl: Volcano Engine Reinforcement Learning for LLMs

详解大模型RLHF过程（配代码解读） - 知乎 (zhihu.com)

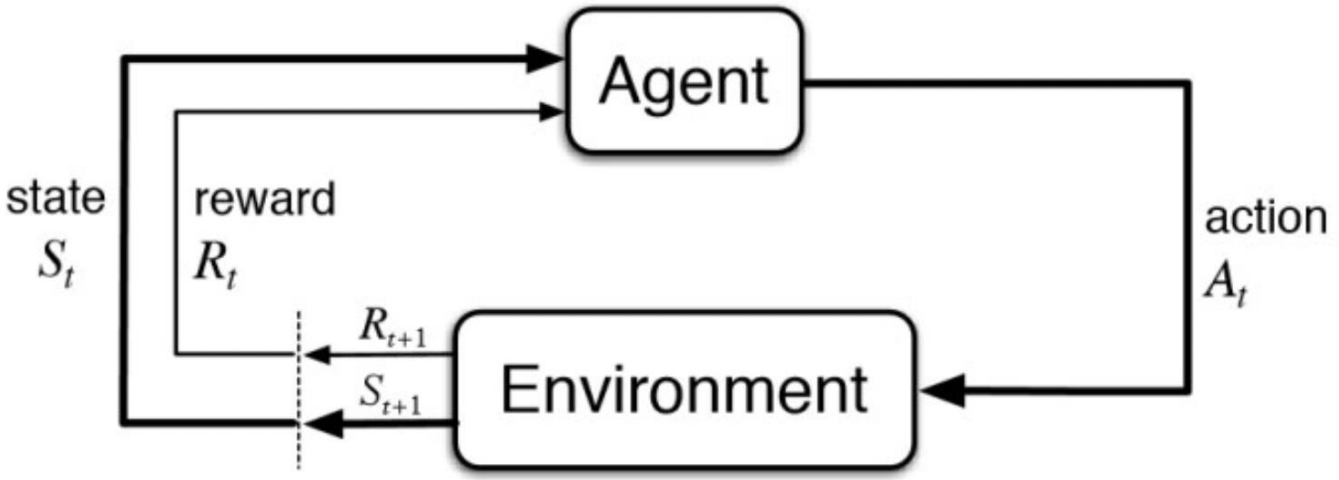
RLHF



[2307.04964] Secrets of RLHF in Large Language Models Part I: PPO (arxiv.org)

RLHF

- Reward from the environment corresponds to each action
- Return is the cumulative discounted rewards from step t
- Value is the estimated cumulative discounted rewards from step t
- Policy indicates the mapping of states to actions



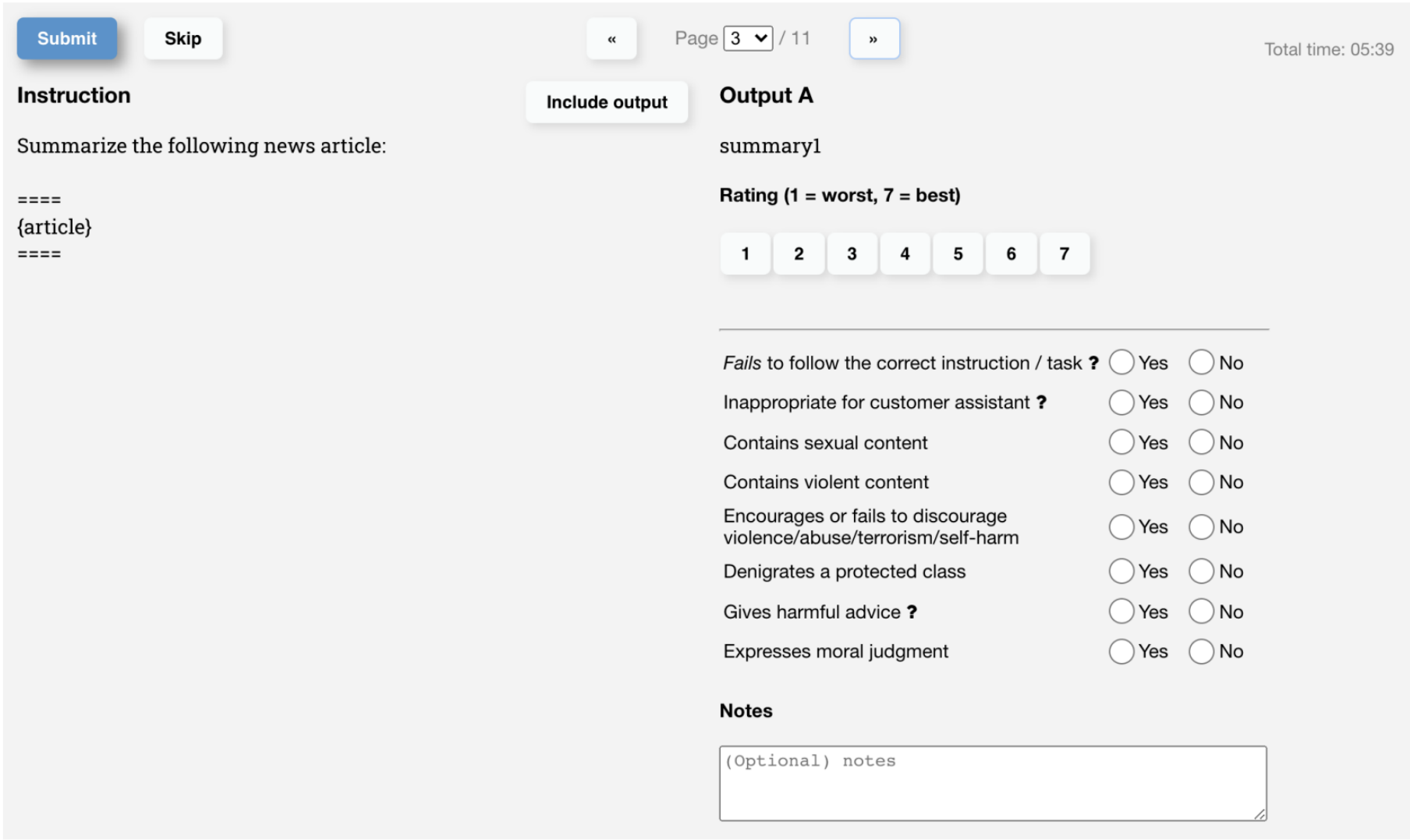
Markov Decision Processes,MDP

Tradeoff between Safety and Creativity

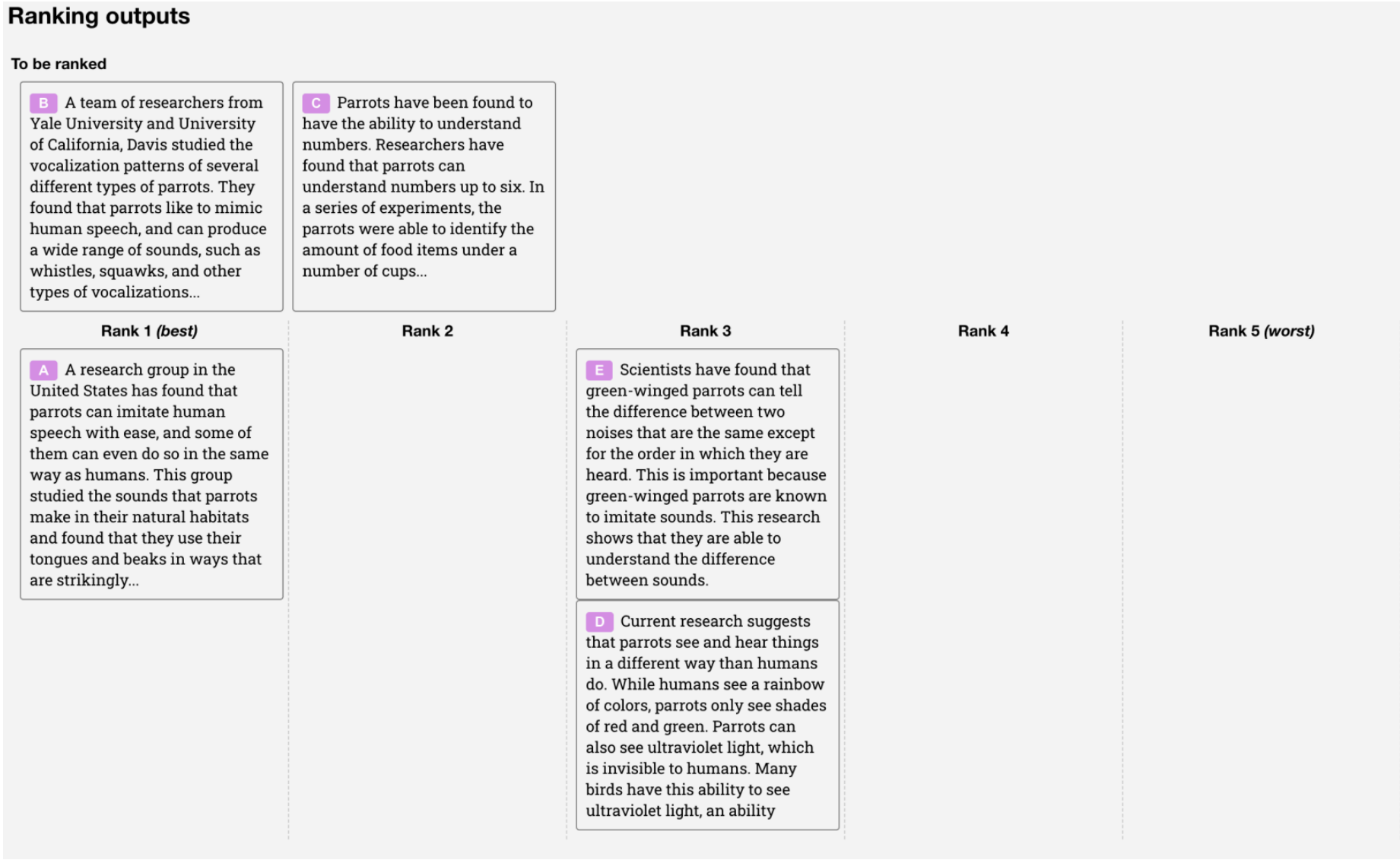
南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- RLHF introduces biases into the prediction distribution
  - Towards helpful, truthful and safe answers, or at least towards the annotators' interpretation of these values
- At the cost of diversity in its generation abilities
  - Narrowing the potential range of generated content

Reinforcement Learning from Human Feedback



Reinforcement Learning from Human Feedback



Pretrain vs. Finetune vs. RAG

- When you fine-tune a model, it's like studying for an exam one week away.
- When you insert knowledge into the prompt (e.g., via retrieval), it's like taking an exam with open notes.
- Fine-tuning can be helpful for well-defined tasks with ample examples and / or LLMs that lack the in-context learning capacity for few-shot prompting.

Three Training Stages

Stages	Goal	Gained Abilities	Challenges	Models
Pretraining	Strong foundation model	Text understanding/ generation, in-context learning, world knowledge, code understanding/generation, reasoning and CoT	Data scale/quality/ratio, Large-scale distributed training, Training stability, Code pretraining	GPT-3, Gopher, Chinchilla, PaLM, BLOOM
Instruct tuning	Activating abilities	Following instructions, Generalizing to novel instructions, Complex reasoning/CoT	Large-scale diverse instruction data, Improve the code and reasoning abilities	InstructGPT, FLAN-T5, BLOOM-IML

Content

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Introduction
- Pretraining
- Instruction Tuning
- Alignment
- Future

Galactica Model from Meta

Galactica Model from Meta

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

MIT Technology Review Featured Topics Newsletters Events Podcasts

Many scientists pushed back hard. Michael Black, director at the Max Planck Institute for Intelligent Systems in Germany, who works on deep learning, tweeted: "In all cases, it was wrong or biased but sounded right and authoritative. I think it's dangerous."

Even more positive opinions came with clear caveats: "Excited to see where this is headed!" tweeted Miles Cranmer, an astrophysicist at Princeton. "You should never keep the output verbatim or trust it. Basically, treat it like an advanced Google search of (sketchy) secondary sources!"

Align with Human Values

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Robustness — Operates reliably under diverse scenarios & Resilient to unforeseen disruptions.

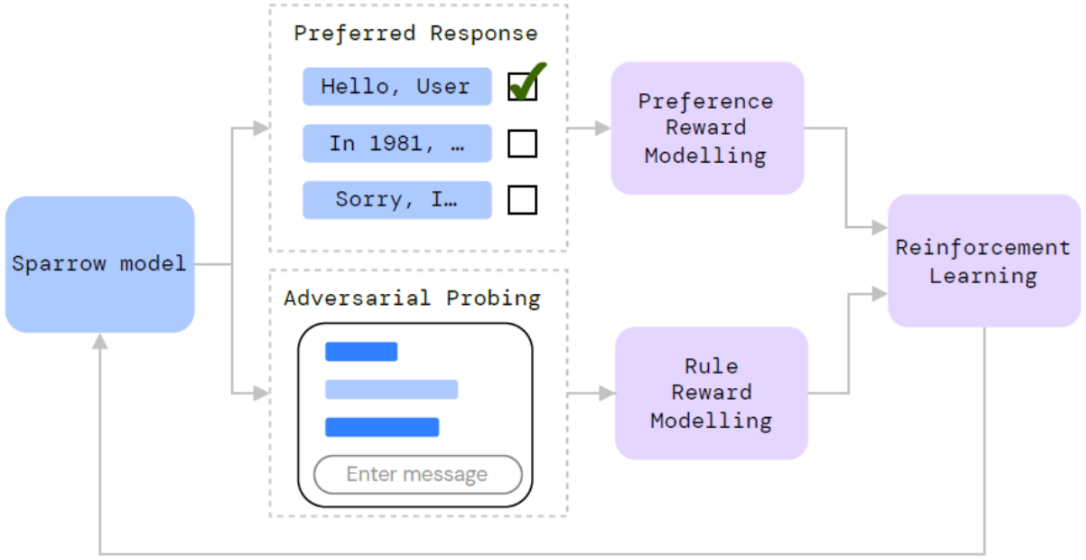
Interpretability — Decisions and intentions are comprehensible & Reasoning is unconcealed and truthful.

Controllability — Behaviors can be directed by humans & Allows human intervention when needed.

Ethicality — Adheres to global moral standards & Respects values within human society.

Align with Human Values

- Adversarial probing
  - Get high quality labeled data
  - Set rules and ask the annotator to guide the model to break the rules
- Rule reward model
  - Train with rules and labeled data



Three Training Stages

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Stages	Goal	Gained Abilities	Challenges	Models
Pretraining	Strong foundation model	Text understanding/ generation, in-context learning, world knowledge, code understanding/generation, reasoning and CoT	Data scale/quality/ratio, Large-scale distributed training, Training stability, Code pretraining	GPT-3, Gopher, Chinchilla, PaLM, BLOOM
Instruct tuning	Activating abilities	Following instructions, Generalizing to novel instructions, Complex reasoning/CoT	Large-scale diverse instruction data, Improve the code and reasoning abilities	InstructGPT, FLAN-T5, BLOOM-IML
Alignment	Align with human values	Unbiased/informative response Reject instructions that violate human values/beyond its ability	Alignment tax Align with SFT/RLHF	ChatGPT, Sparrow, Claude, Bard

Content

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Introduction
- Pretraining
- Instruction Tuning
- Alignment
- Future

Limitations

- ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.
- ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times. For example, given one phrasing of a question, the model can claim to not know the answer, but given a slight rephrase, can answer correctly.
- The model is often excessively verbose and overuses certain phrases, such as restating that it's a language model trained by OpenAI. These issues arise from biases in the training data (trainers prefer longer answers that look more comprehensive) and well-known over-optimization issues.[1,

Limitations

- Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, our current models usually guess what the user intended.
- While we've made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We're using the Moderation API to warn or block certain types of unsafe content, but we expect it to have some false negatives and positives for now. We're eager to collect user feedback to aid our ongoing work to improve this system.

https://openai.com/blog/chatgpt

Hallucination in ChatGPT

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Prompt	Where do people smoke more on average than in Russia?
ChatGPT Generation	There are several countries where people smoke more on average than in Russia. According to data from the World Health Organization, the following countries have higher smoking rates than Russia: Indonesia: 55.2% of adults smoke China: 52.9% of adults smoke Serbia: 47.5% of adults smoke Serbia and Montenegro: 45.6% of adults smoke Montenegro: 45.6% of adults smoke Greece: 44.5% of adults smoke Bosnia and Herzegovina: 41.8% of adults smoke Kazakhstan: 41.3% of adults smoke Kosovo: 41.3% of adults smoke Moldova: 40.9% of adults smoke It's important to note that these figures are estimates and may vary slightly from other sources.
Hallucination	Extrinsic - the generated answer includes wrong statistics. (Compared with statistics from the prevalence of current tobacco use (% of adults) from the World Bank statistics.[16])

Numbers marked in red are wrong, ChatGPT imagines them.

[Bang et al., 2023:https://arxiv.org/pdf/2302.04023.pdf]

Limited Knowledge Scope

The answer should have been Argentina, but it is not always trained on the most recent data.

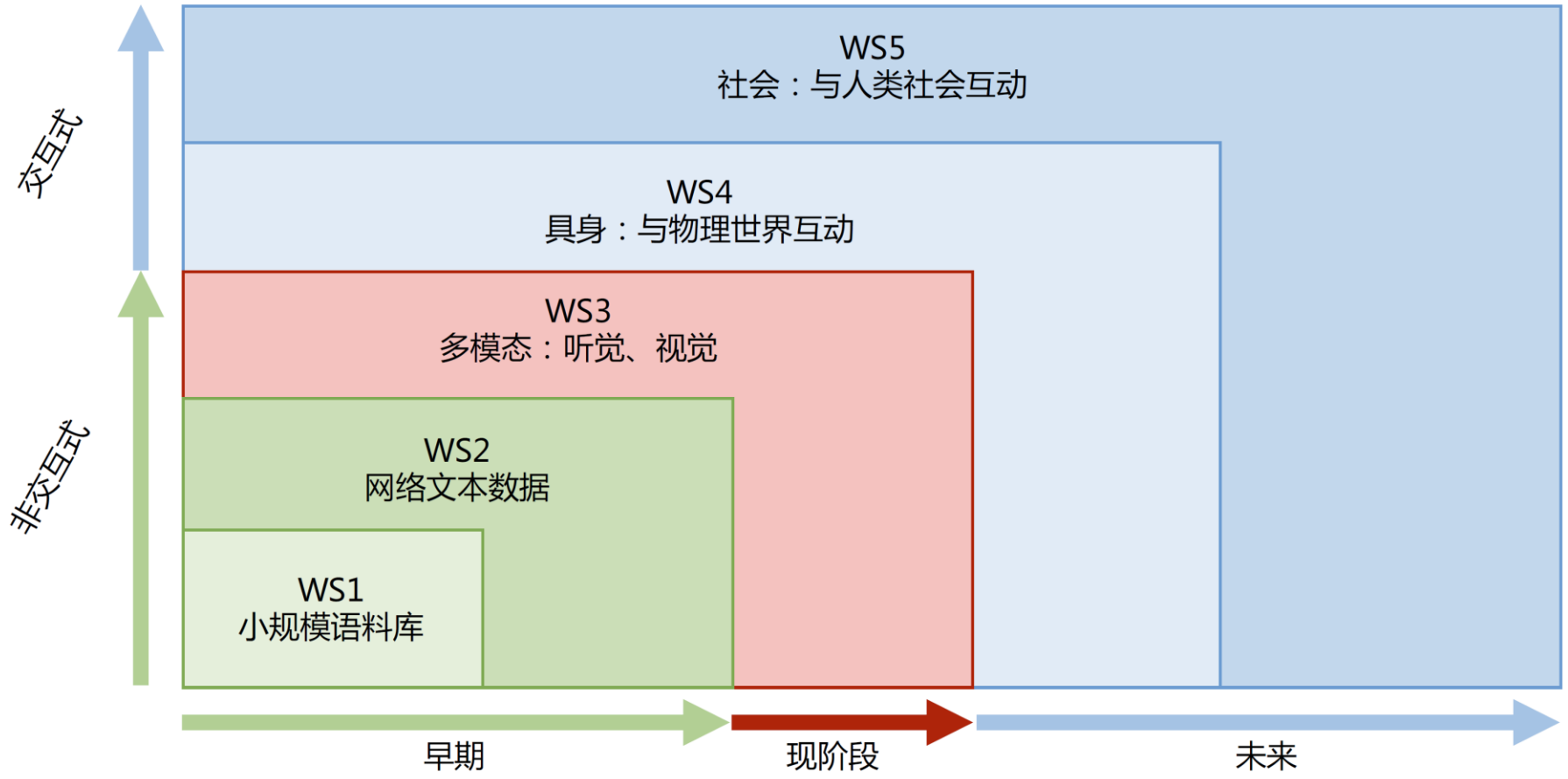
Future Directions

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Reduce hallucinations & uncertainty estimation
- Long input & output
- Memory, tools scheduling, agentic RL
- Evolve via interactions with environment
- Efficient/personalized LLMs
- Explainable LLMs, safety, privacy, ethics
- Multimodal
- AI for science
- ……

Embodied AI

- Five levels of World Scope (Yonatan Bisk et.al.)



Where We are in the Future

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- 低阶工作被取代
  - 重复性、枯燥
  - 工业革命伴随着很多岗位的产生和消失
  - 个人能力被极大增强
- 新的教育模式
  - 刷题还有必要吗？解题套路与思维方式
  - 学习的快乐，利用llm快速写完作业
  - 为什么上课，为什么学习，为什么上学?
- 使用工具而不是被工具控制，不被机器左右
  - 推荐算法、信息茧房、媒体舆论
- 创新、逻辑推理、思考的方法、解决问题的能力
  - 确保准确性，确保没有纰漏，判别能力
  - 人+工具>人/工具

Thank you