

知识驱动的数据合成

—— 人工智能时代计算语言学的功能转型

詹卫东（北京大学中文系）

zwd@pku.edu.cn

<https://ccl.pku.edu.cn/doubtfire>

爱尔兰都柏林大学学院
尤利西斯奖章

2013年得主



Noam Chomsky

1928.12.7 —

现代语言科学、认知科学奠基人
生成语言学、普遍语法理论提出者



2024年得主



Geoffrey Hinton

1947.12.6 —

“深度学习”之父
2018年图灵奖得主，2024年诺贝尔物理学奖得主

“

语言学家被一个叫乔姆斯基的人误导了好几代，……
他有一个偏执古怪的理论，即语言不是学会的。他成功地说服很多人相信这一点。但这一看就知道纯粹是胡言乱语。语言显然是学会的。大型神经网络学习语言，学习句法和语义，不需要任何先天结构，只是从随机权重和大量数据开始……”

辛顿（2024年4月）

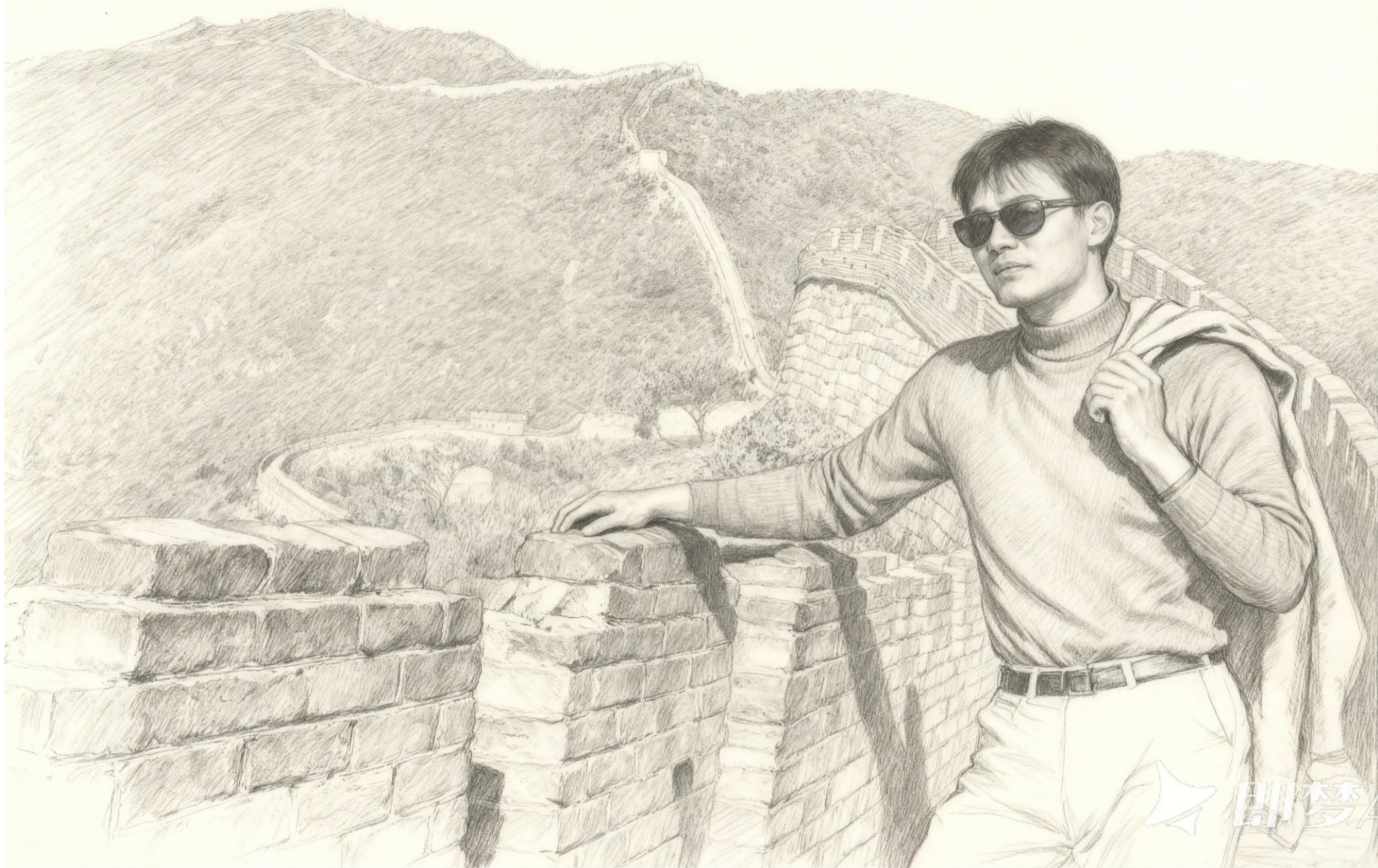


AI生成



即梦

AI生成



即梦



2023/3/9 09:35

Opinion | Noam Chomsky: The False Promise of ChatGPT - The New York Times

The New York Times

<https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

By Noam Chomsky, Ian Roberts and Jeffrey Watumull



语言战略研究

Chinese Journal of Language Policy and Planning

ISSN 2096-1014 CN 10-1361/H

标题 / 作者 / 摘要 / 关键词 / DOI / 全文



高级检索

[首页](#) [期刊介绍](#) [编委会](#) [投稿指南](#) [征订启事](#) [在线期刊](#) [信息动态](#) [联系我们](#) [English](#)

语言战略研究 >> 2024, Vol. 9 >> Issue (6) : 5-17.

语言战略研究

“辛顿·乔姆斯基·语言学发展”多人谈

2024年10月17日 商务印书馆 座谈会

冯志伟¹, 陈国华², 李宇明³, 陈保亚⁴, 袁毓林⁵, 孙茂松⁶, 徐杰⁷, 司富珍⁸, 詹卫东⁹, 完权¹⁰, 王伟¹⁰, 刘畅¹¹

作者信息

Multi-perspective Discussion: Hinton, Chomsky and the Development of Linguistics

Feng Zhiwei, Chen Guohua, Li Yuming, Chen Baoya, Yuan Yulin, Sun Maosong, Xu Jie, Si Fuzhen, Zhan Weidong, Wan Quan, Wang Wei and Liu Chang



2024年10月12日

如果辛顿是对的.....

阅读 7616 赞 91



2024年10月18日

"母鸡语言学"与"鸡蛋语言学"

阅读 1.3万 赞 300



2024年11月21日

辛顿与乔姆斯基大战"中文屋"

阅读 3703 赞 59



2025年2月10日

语言学与人工智能

阅读 7400 赞 172



无题·叹AI迭代太急

原创 詹卫东 语言学光标 2025年11月21日 18:43



惯看新模变旧时，烧卡炼丹鬓生丝。
元月惊呼 R-1出，云端旋见五代旗。
忍看海枯成凡骨，又向双子觅长诗。
测罢低眉无写处，羞提拙笔补天衣。



提 纲

1. 语言学的范式嬗变
2. 人工智能的范式嬗变
3. 语言学与人工智能的融合之路
4. 结语：与AI共舞

- **AI真的学会人类语言了吗？**
- **人到底是如何运用语言的？**
- **AI到底该如何向人学语言？**
- **语言学能为AI干点什么？**

1. 语言学的范式嬗变

1. 语言学的范式嬗变

- 看作**法律**的语言学
(Prescriptive grammar)
- 看作**生物学**的语言学
(Historical/Comparative linguistics)
- 看作**化学**的语言学
(Structural linguistics / Descriptive grammar)
- 看作**数学**的语言学
(Generative linguistics / Formal grammar)

这句话是什么意思？

语言如何演化？

语句的意思从何而来？

语言为何能无限生成？

语言观念的变迁

研究问题的变迁

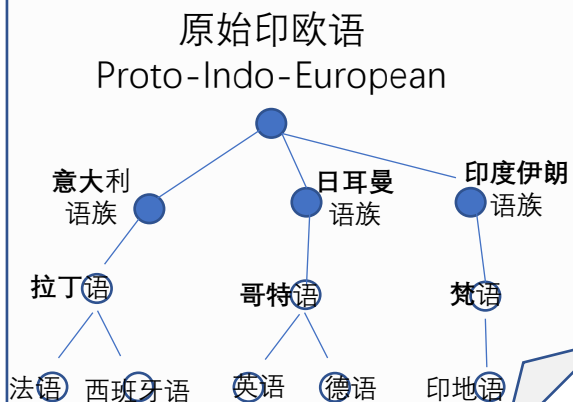


“二月，公**侵**宋。”
(《春秋·庄公十年》)

曷为或言**侵**? 或言**伐**?
恠者曰**侵**，精者曰**伐**。
战不言伐，**围**不言战，
入不言围，**灭**不言入，
书其重者也。
(《春秋公羊传》)

规范语言学范式
看作法律的语言学

拉丁语: **p**ater
印地语: **p**itā
梵语: **p**itar
哥特语: **f**adar
英语: **f**ather



历史比较语言学范式
看作生物学的语言学

阿Q 没 理由 怕 吴妈 。



阿Q 没 理由 **地** 怕 吴妈 。



结构主义语言学范式
看作化学的语言学

The boy is happy.

→ Is the boy happy? ✓

The boy who is tall is happy.

✗ **Is** the boy who tall is happy?

✓ **Is** the boy who is tall happy?

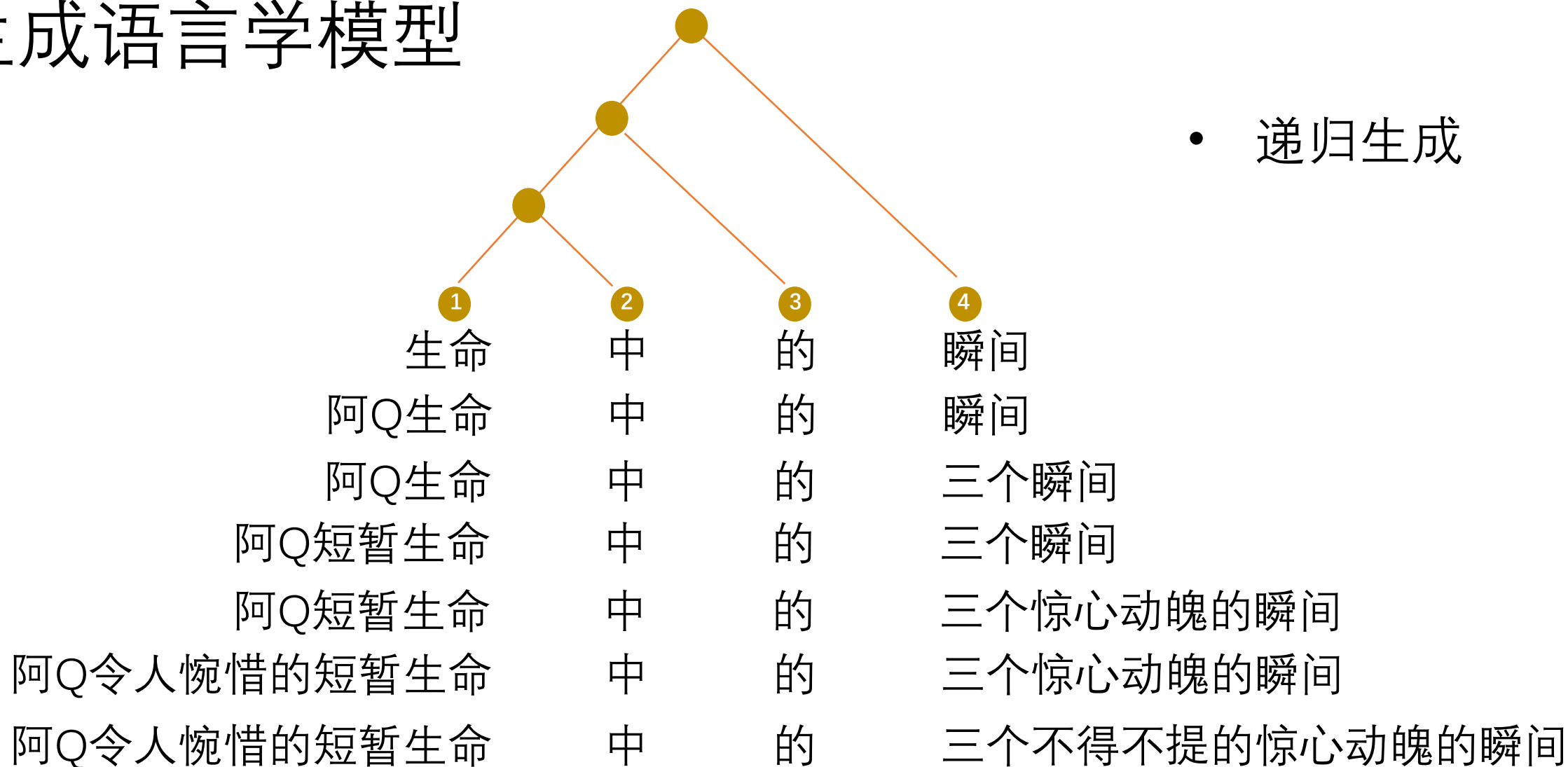
- CP → C S
- S → NP AP
- NP → Det N RC
- AP → adj
- Det → the
- N → boy
- RC → Rel V Adj
- V → be

Context Free Grammar

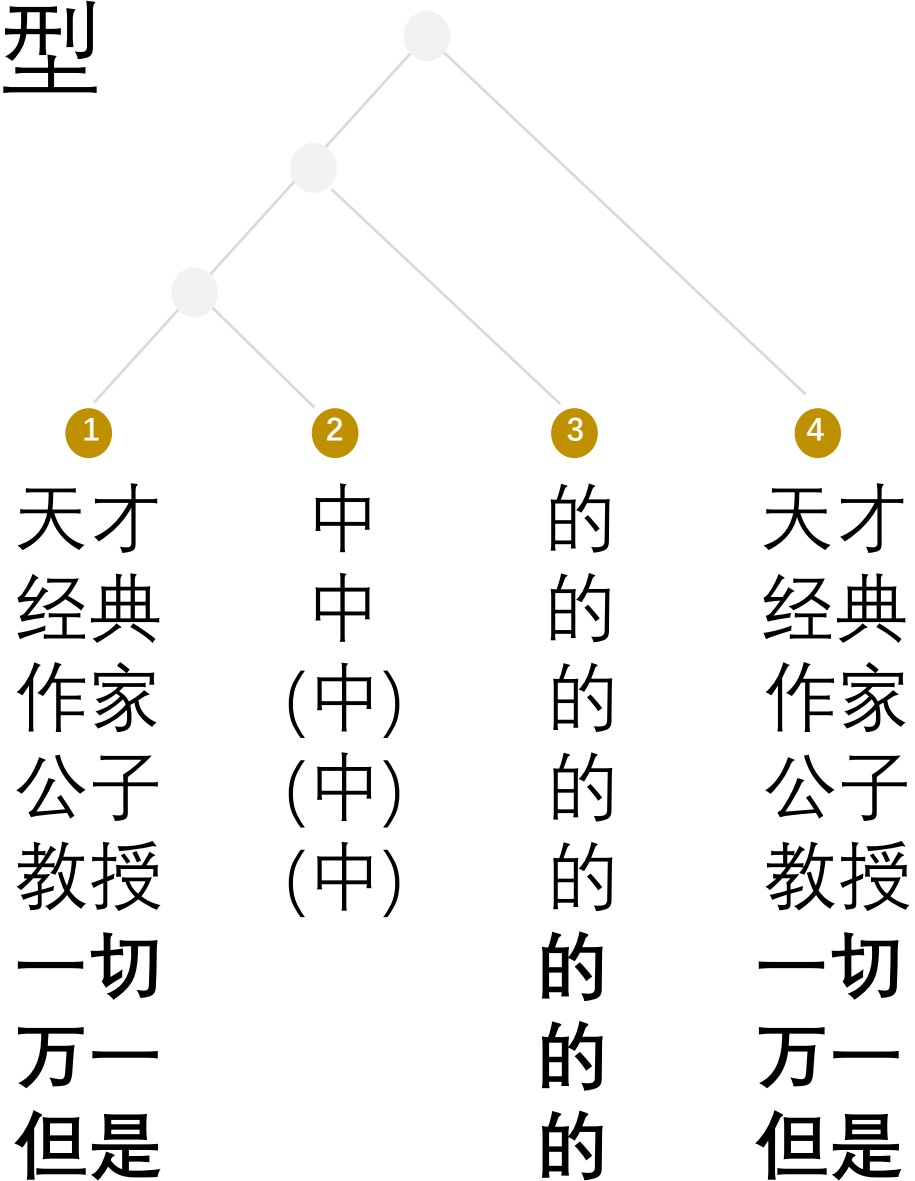
生成语言学范式
看作数学的语言学

生成语言学模型

- 递归生成

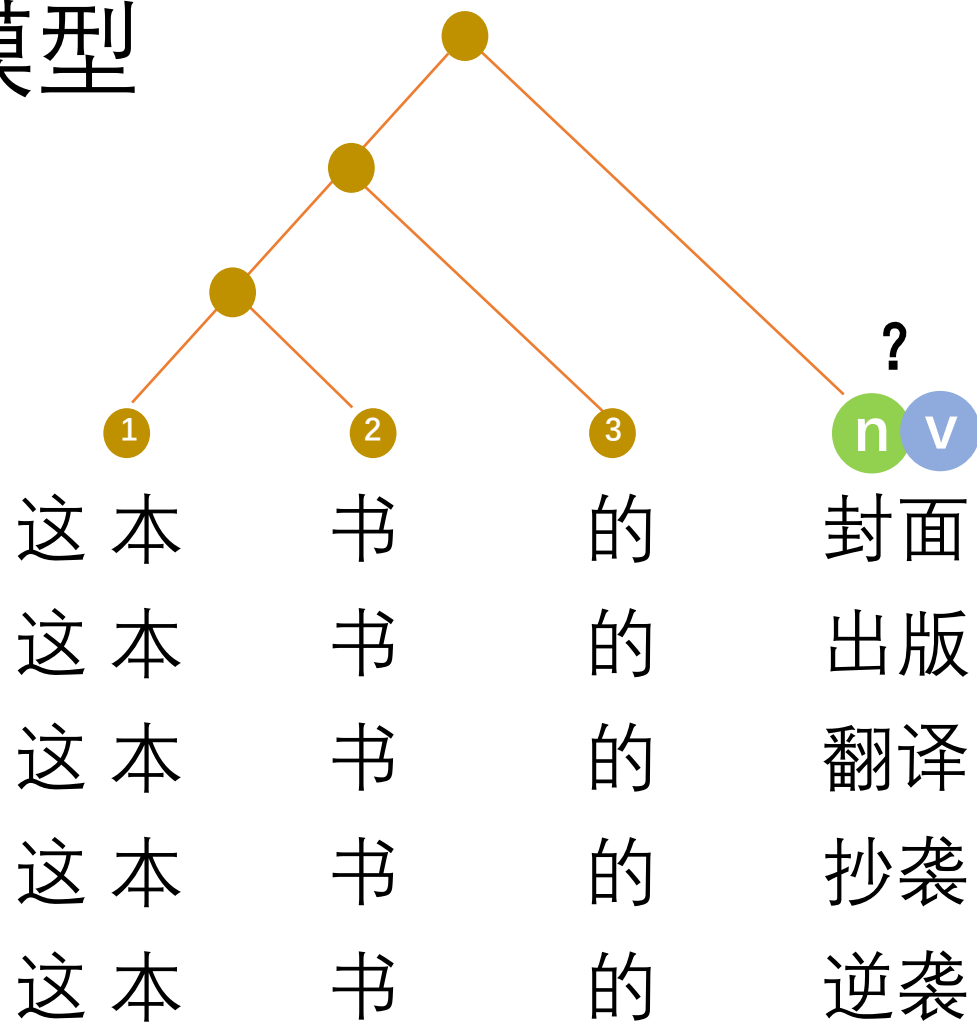


生成语言学模型



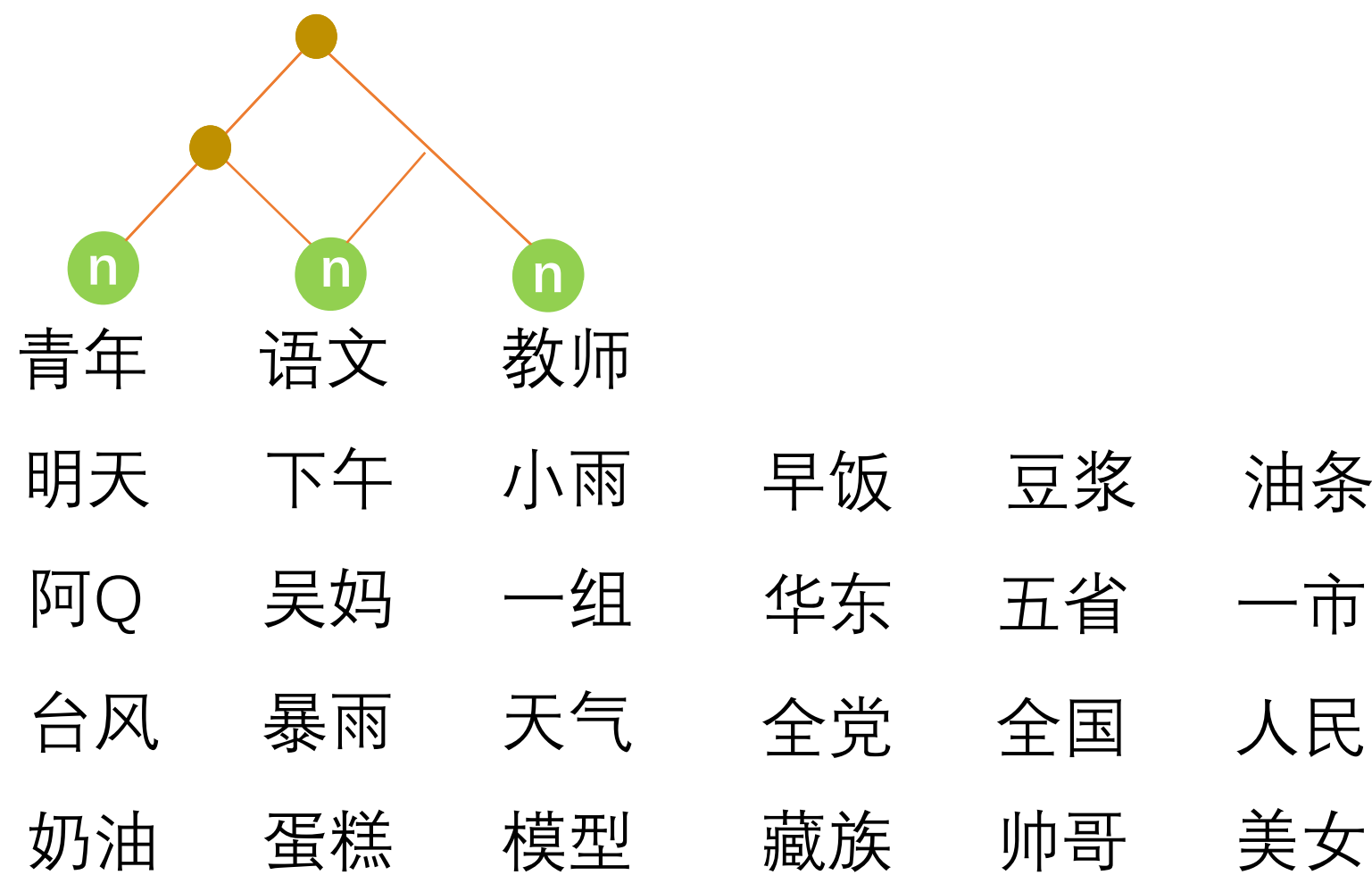
- 非递归生成

生成语言学模型



- 同类/异类生成

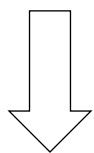
生成语言学模型



生成语言学模型

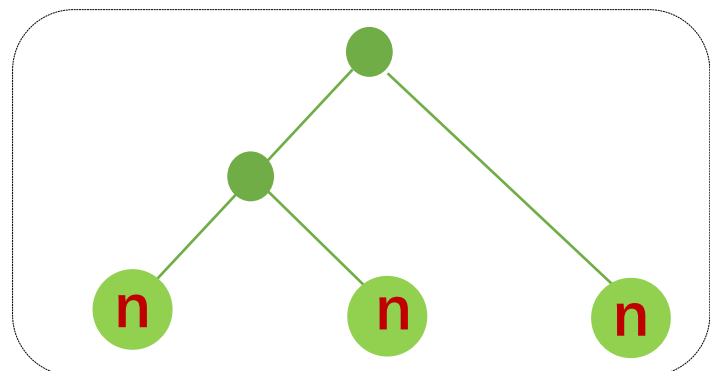
内涵

语言学知识



外延

真实语符序列



羊皮	领子	大衣
明天	下午	小雨
阿Q	吴妈	一组
台风	暴雨	天气
奶油	蛋糕	模型

$$X_{n+1} = \frac{X_n + \frac{\mathbf{a}}{X_n}}{2}$$

压缩

$\sqrt{\mathbf{a}}$

还原

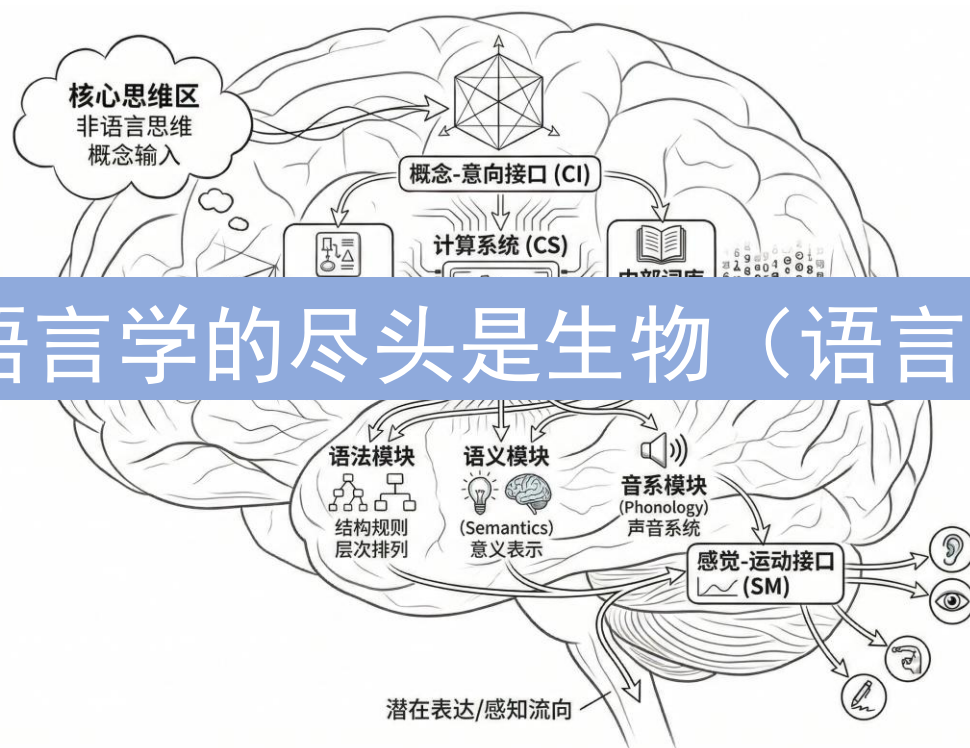
1.4142135623730950...

I-Language
内部语言

理想

现实

E-Language
外部语言



学习能力

语言无限生成能力

局部微观

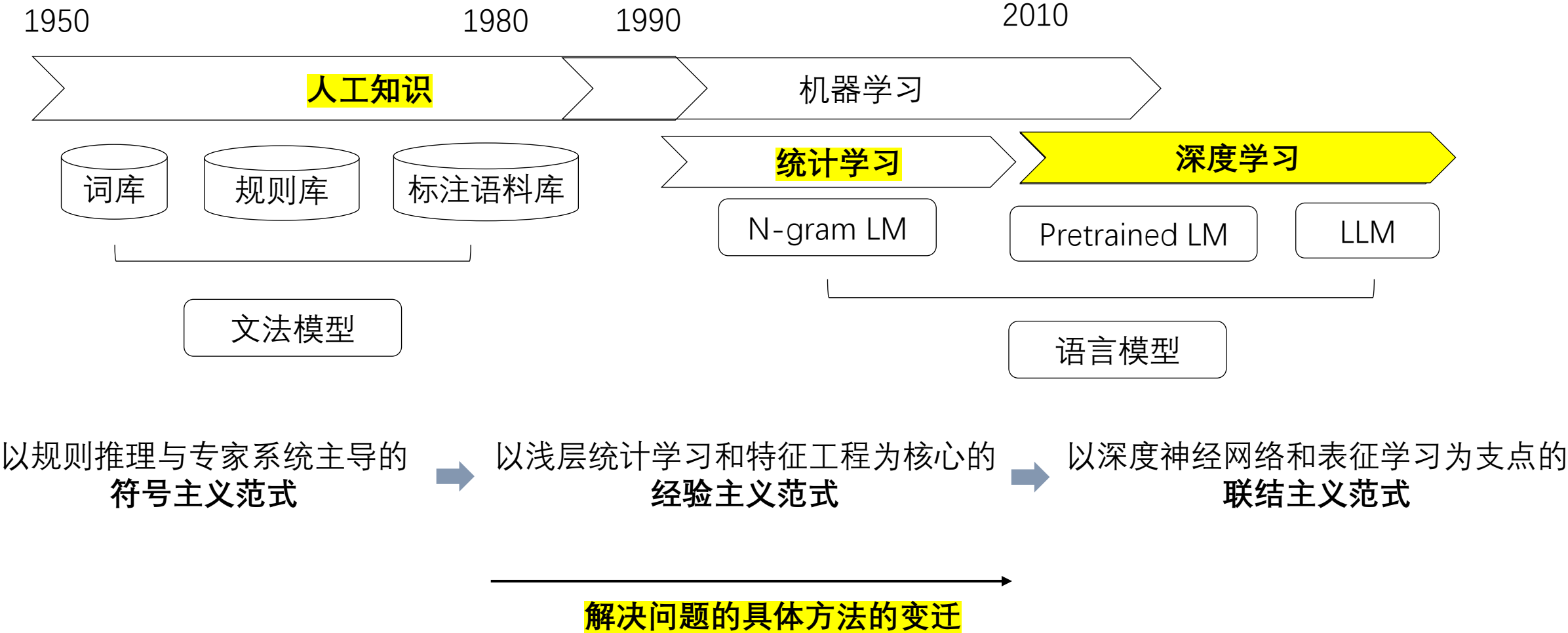
语言现象分析

语言的社会性

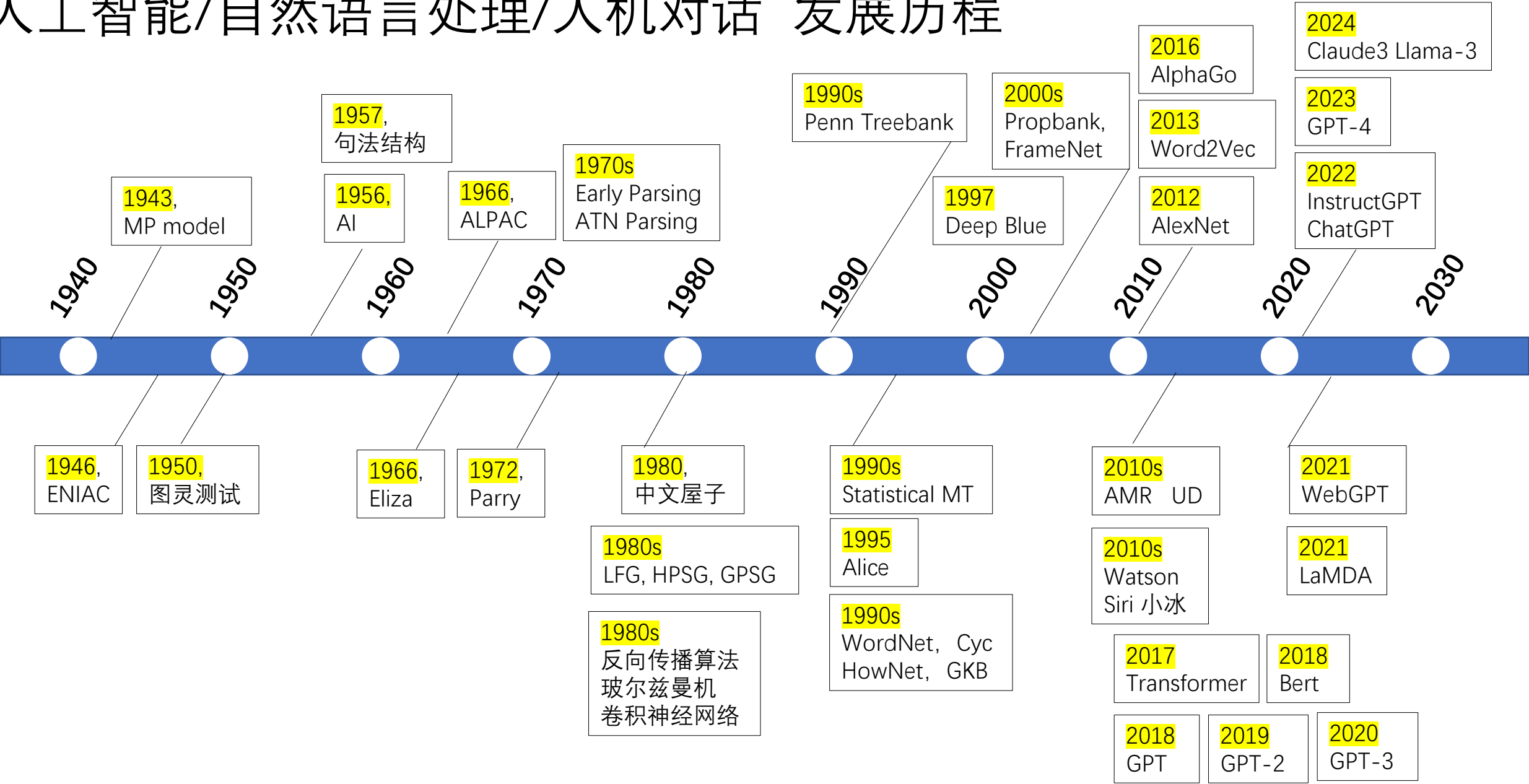
语言知识+世界知识

2. 人工智能的范式嬗变

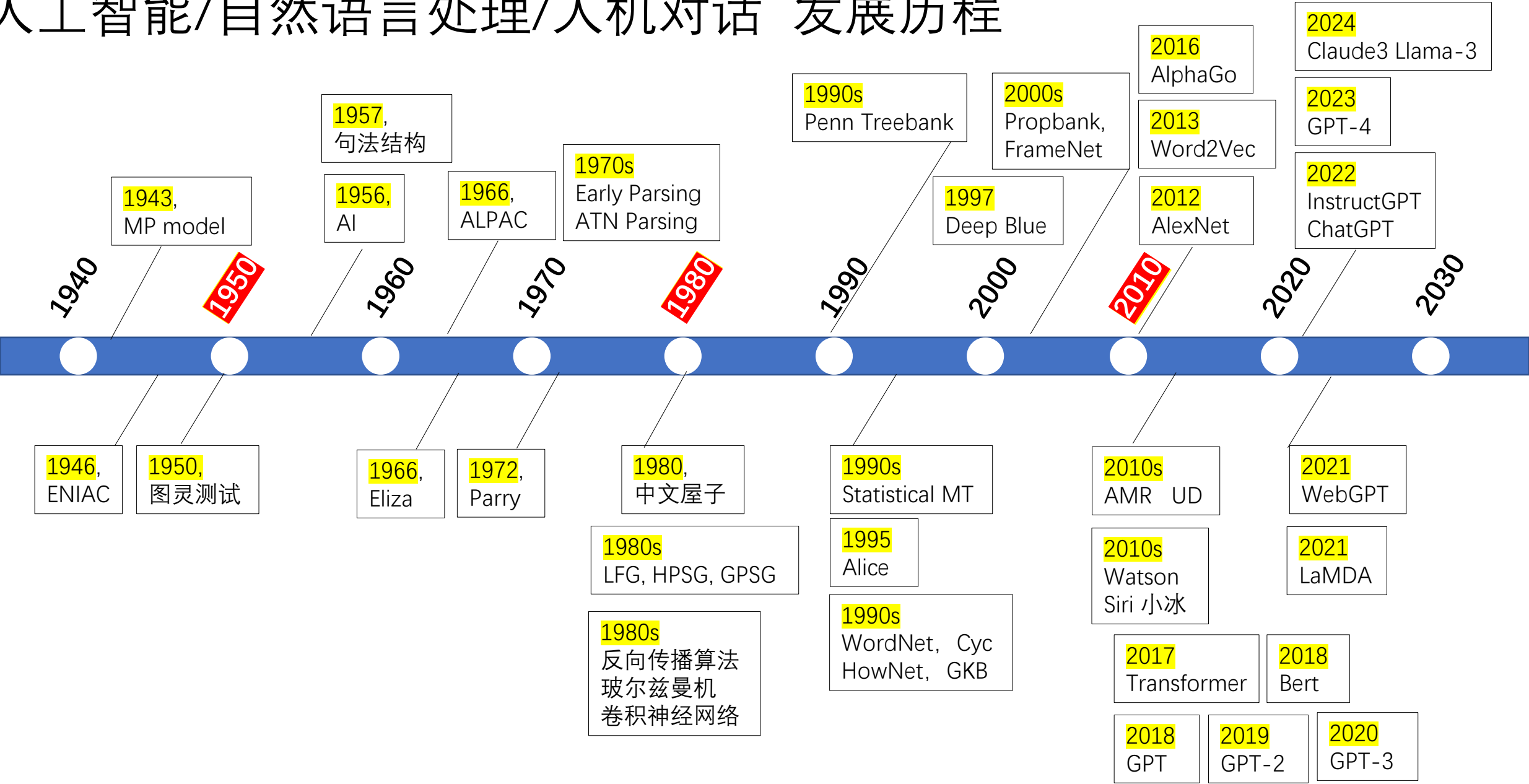
人工智能/自然语言处理/人机对话 发展历程



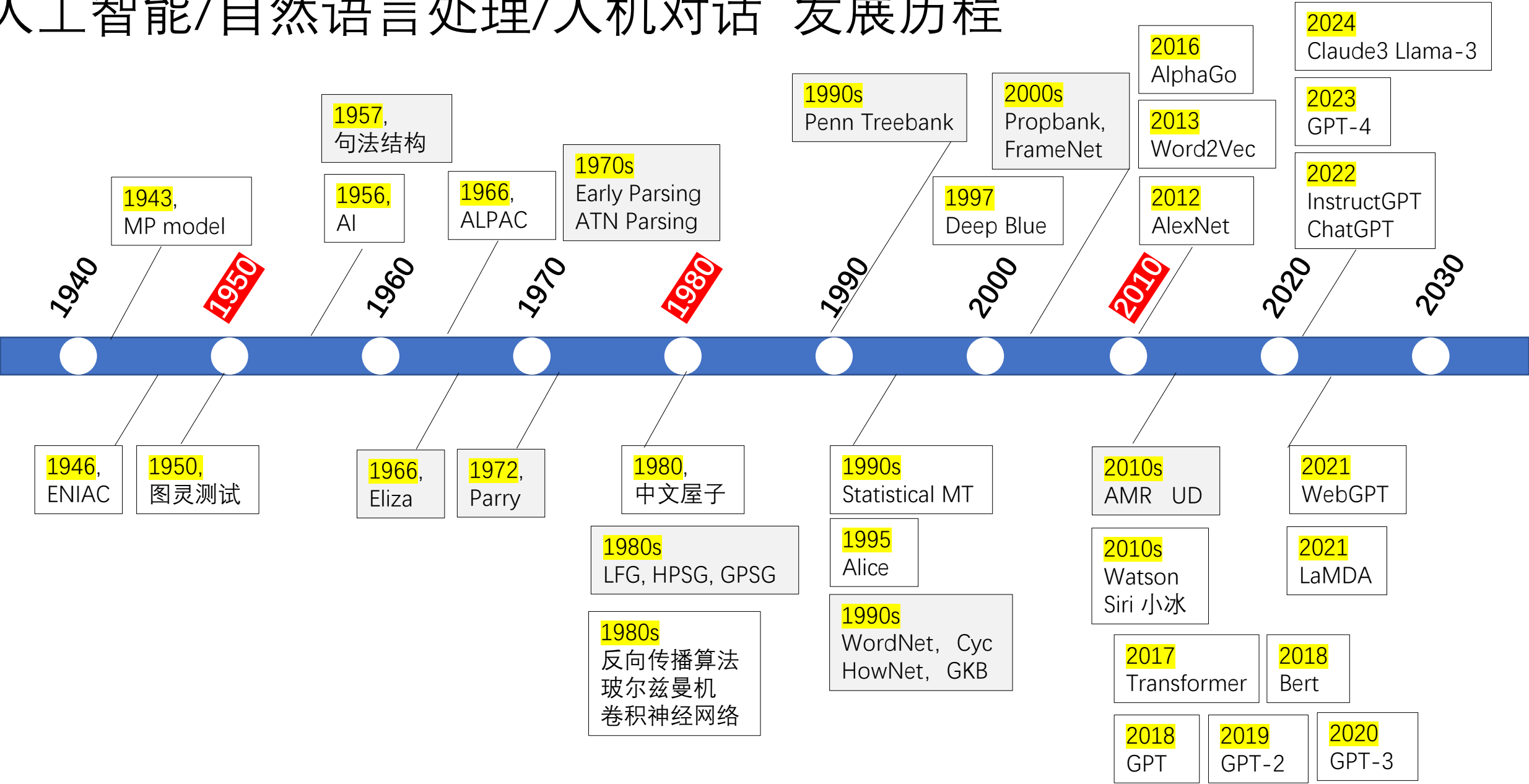
人工智能/自然语言处理/人机对话 发展历程



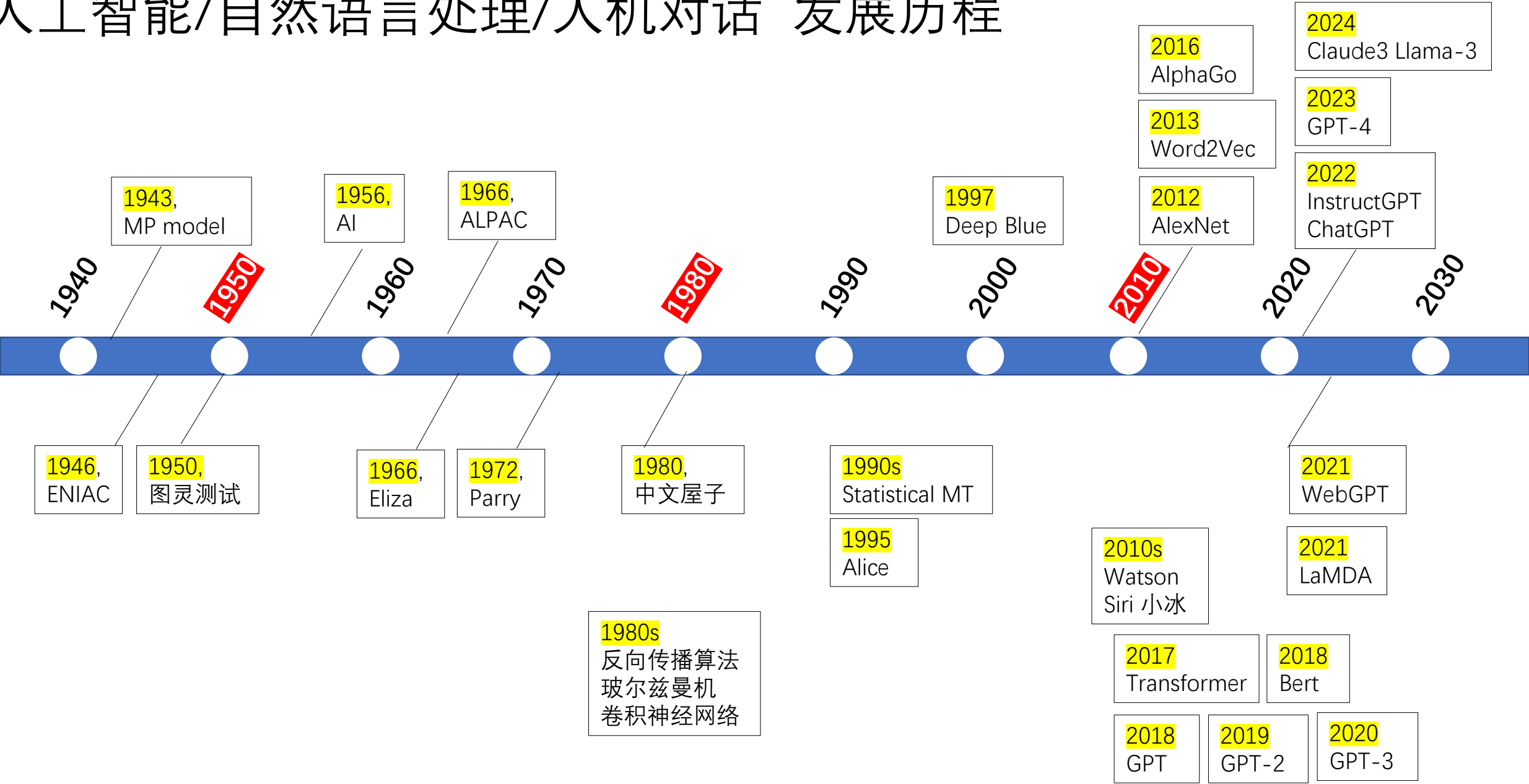
人工智能/自然语言处理/人机对话 发展历程



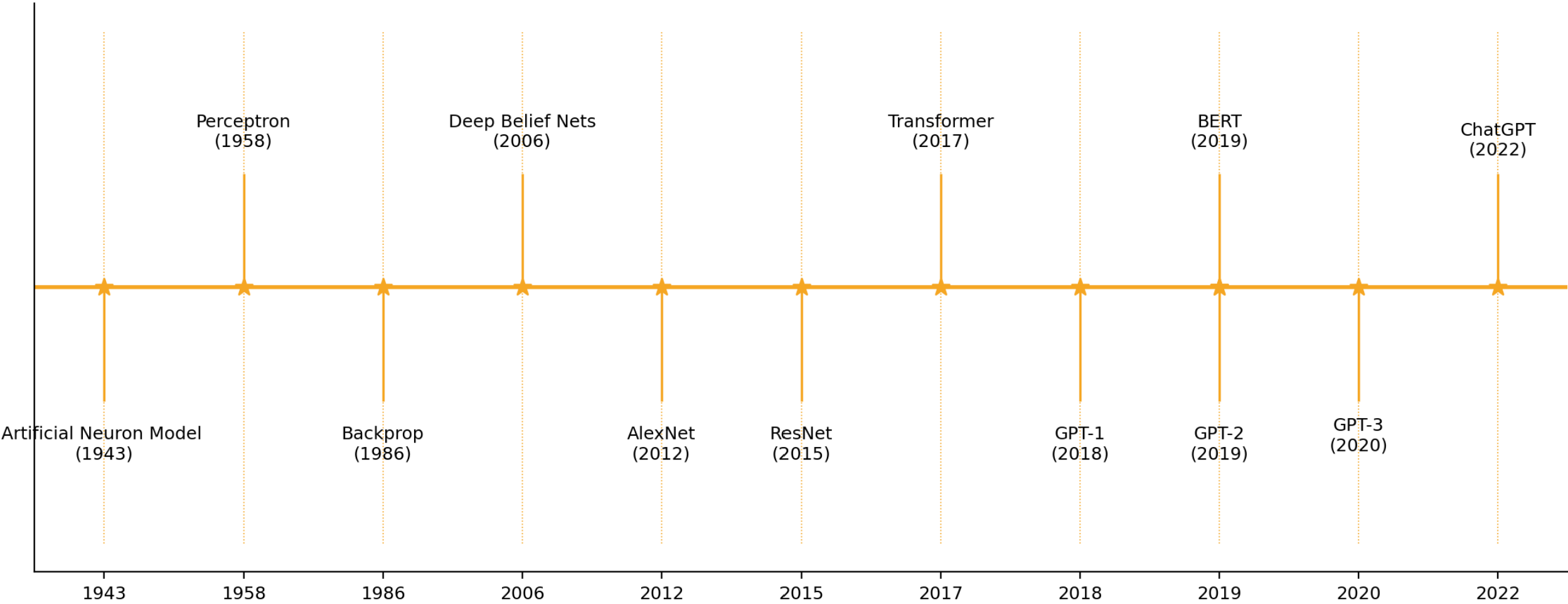
人工智能/自然语言处理/人机对话 发展历程



人工智能/自然语言处理/人机对话 发展历程



深度学习（神经网络）技术路线发展里程碑



AI范式变迁之路

—— 从规则知识到自主进化 ——



专家系统（规则知识）
1950 - 1990

1



深度学习神经网络学习
2010 - 2017



大语言模型
2022 - 2024

5



AI智能体
2026

7



思维链推理大模型
2025



预训练语言模型
2018 - 2022

6



浅层机器学习
1990 - 2010

2

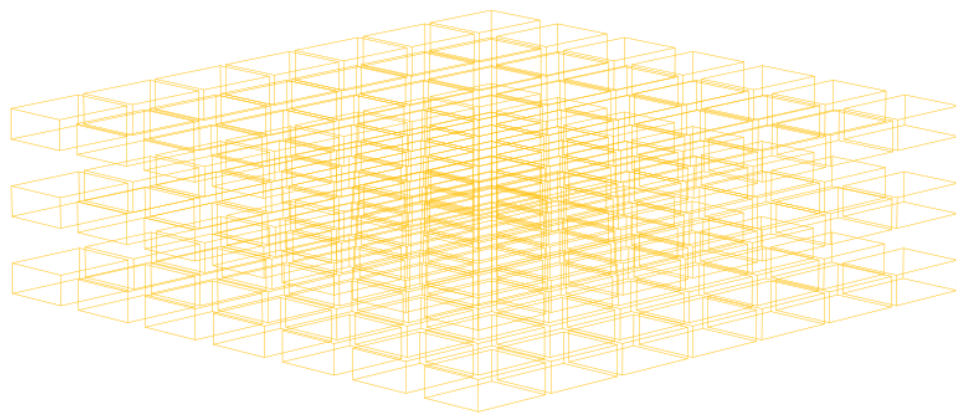
4



自主进化AI
2026 ...

8

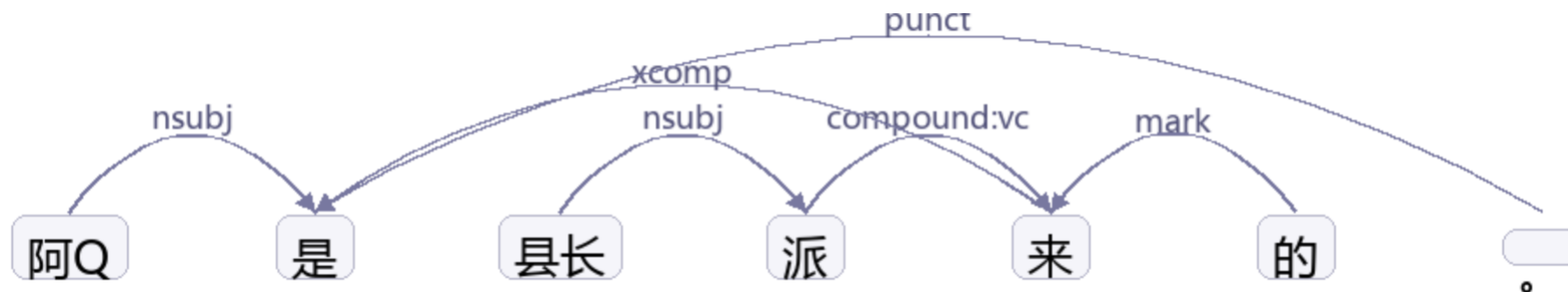
阶段一：语言学微观研究汇聚



机器怎么 理解 / 生成 一句话？

例如：阿Q是县长派来的。

① 单个词的意义 + ② 词之间的关系意义



输入 →

阿Q是县长派来的。



→ 输出

- (a) 阿Q不是县长。
- (b) 阿Q来了。
- (c) 派阿Q来的是县长。
- (d) 阿Q是被县长派来的。

模型1：语言学规则推理

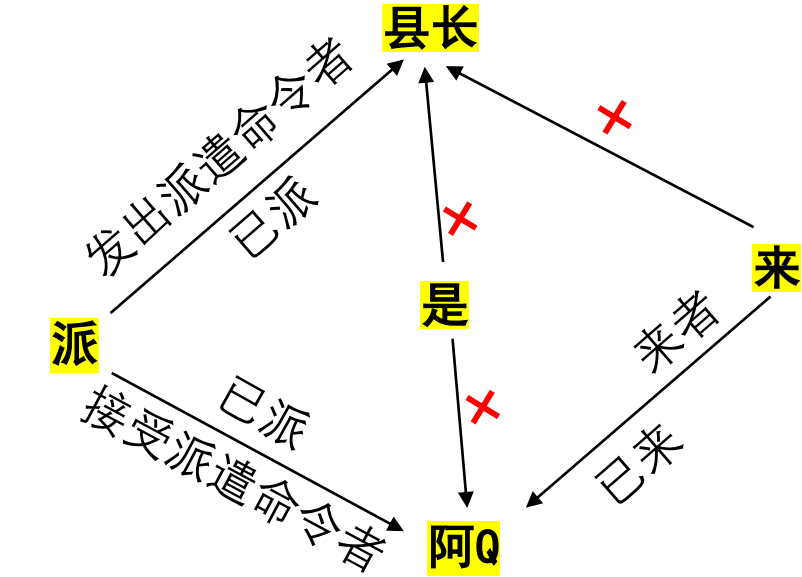
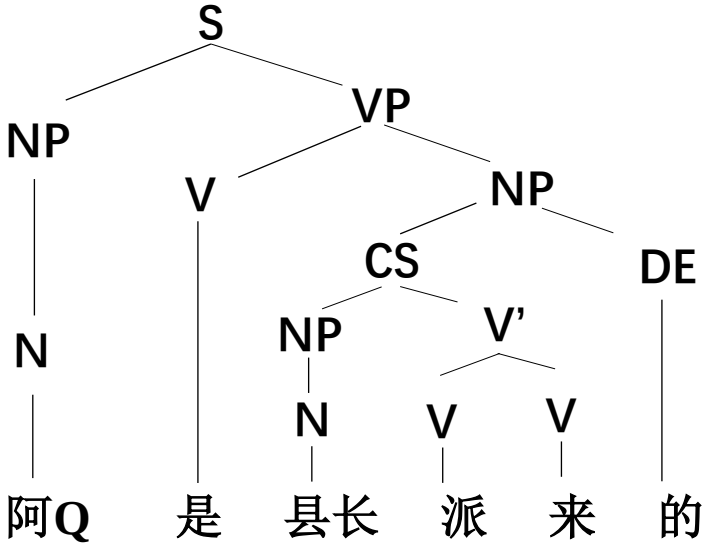
- (1) $S \rightarrow NP \ VP$
- (2) $NP \rightarrow N$
- (3) $NP \rightarrow CS \ 的$
- (4) $CS \rightarrow NP \ V'$
- (5) $VP \rightarrow V \ NP$
- (6) $V' \rightarrow V \ V$

输入

阿Q是县长派来的。

词语	特征1	特征2	特征3	特征4	...
阿Q	人名	男	绍兴人	无业	
是	判断动词	二元	及物		
县长	职务	领导			
派	兼语动词	三元	及物	施事[...]	
来	趋向动词	二元	及物	当事[...]	
的	结构助词	指称化标记			
。	完句标点				

Chomsky 文法模型



输出

- (a) 阿Q不是县长。
- (b) 阿Q来了。
- (c) 派阿Q来的是县长。
- (d) 阿Q是被县长派来的。

模型2：神经网络深度学习

Step 1/40 Loss $\approx$ 2.00

输入序列

阿Q
是
县长
()
来
的

隐藏层 1

隐藏层 2

隐藏层 3

隐藏层 n

输出候选与概率

派

请

安排

弯

胖

汽车

“阿Q是县长派来的。”这句话包含了下面哪些意思？

(a) 阿Q是县长。

(b) 阿Q来了。

(c) 派阿Q来的是县长。

(d) 阿Q是被县长派来的。

.....

小D	是	小张	请	来	的
王胡	是	吴妈	气	走	的
秀才	是	老爷	撞	伤	的
阿Q	是	尼姑	赶	出去	的

N1

V1

N2

V2

V3

DE

县长	是	老爷	花钱	买	的
王胡	是	吴妈	旅游	认识	的
秀才	是	老爷	出资	留学	的
尼姑	是	阿Q	建议	邀请	的

.....

(a) ×

(b) ✓

(c) ✓

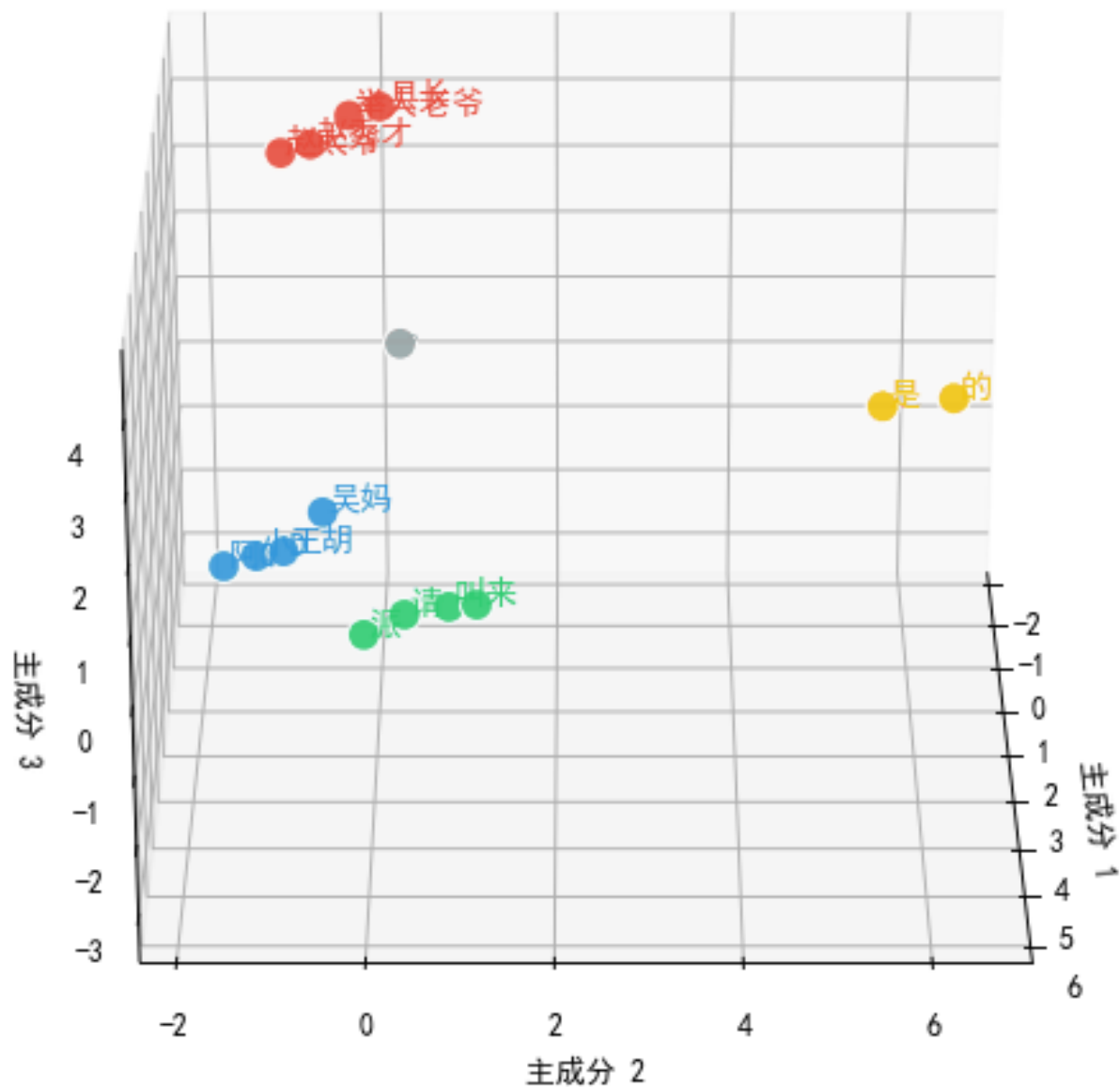
(d) ✓

① “词义”在机器中的存在形式： 向量（Vector）

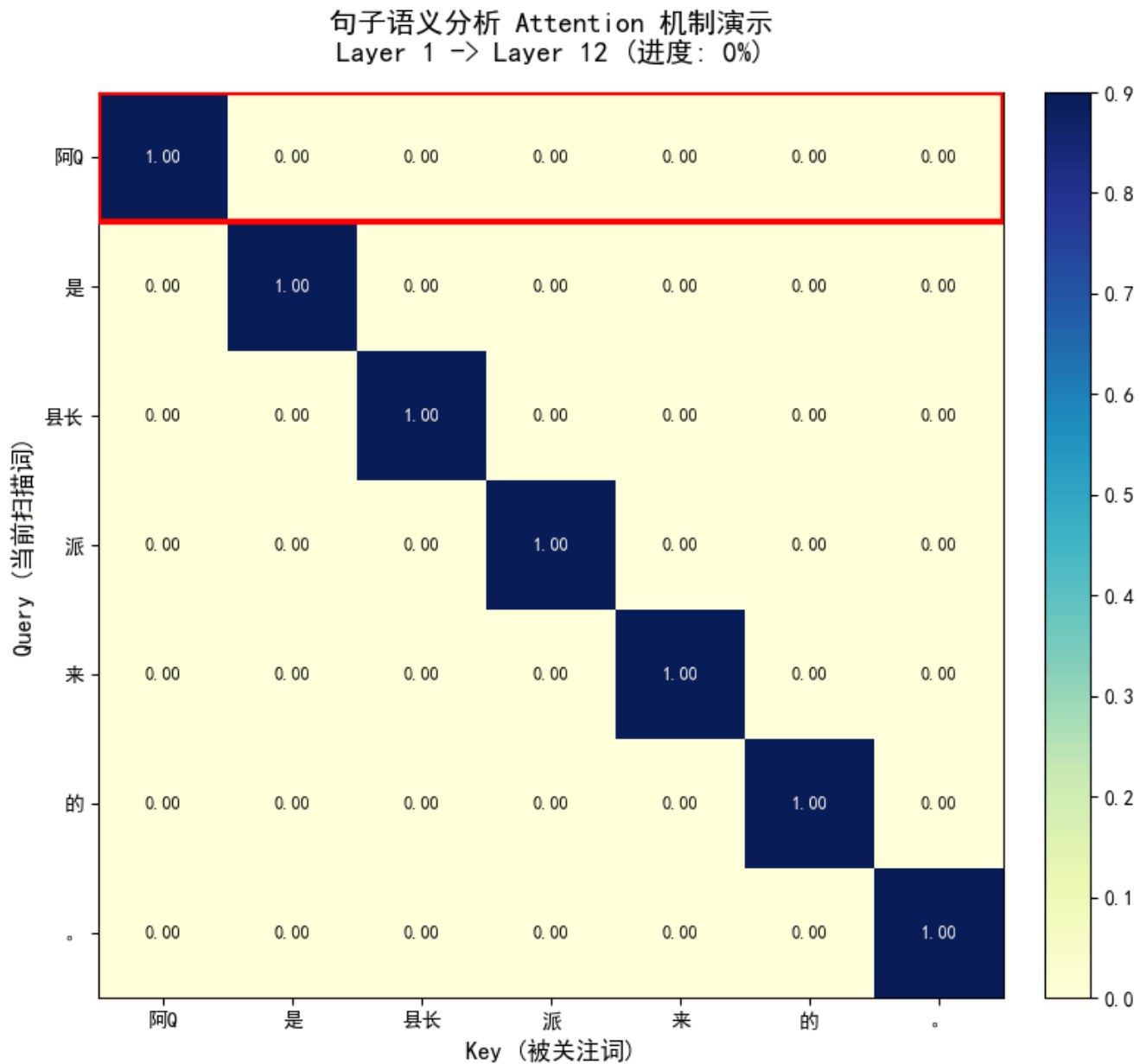
模型2

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
赵太爷	0. 805	0. 794	0. 671	0. 696	0. 826	0. 684	0. 825	0. 642	0. 853	0. 705
县长	0. 847	0. 81	0. 732	0. 681	0. 77	0. 609	0. 866	0. 828	0. 886	0. 808
秀才	0. 731	0. 75	0. 588	0. 804	0. 769	0. 823	0. 801	0. 892	0. 72	0. 804
阿Q	-0. 257	-0. 212	-0. 352	-0. 571	-0. 647	-0. 575	-0. 427	-0. 522	-0. 394	-0. 644
小D	-0. 551	-0. 504	-0. 331	-0. 648	-0. 51	-0. 514	-0. 381	-0. 404	-0. 52	-0. 47
王胡	-0. 536	-0. 489	-0. 75	-0. 391	-0. 47	-0. 506	-0. 577	-0. 47	-0. 553	-0. 474
吴妈	-0. 366	-0. 439	-0. 449	-0. 585	-0. 367	-0. 328	-0. 523	-0. 444	-0. 585	-0. 59
派	0. 082	-0. 181	0. 066	0. 083	-0. 259	0. 054	0. 111	-0. 101	0. 066	0. 026
请	0. 002	-0. 015	-0. 039	0. 168	-0. 14	0. 054	-0. 064	-0. 013	-0. 027	-0. 019
叫	-0. 054	0. 024	0. 032	-0. 079	0. 196	-0. 141	-0. 061	0. 043	0. 02	0. 117
是	0. 032	0. 053	-0. 005	-0. 101	-0. 06	-0. 112	0. 119	-0. 101	0. 049	-0. 031
来	-0. 02	-0. 037	0. 036	0. 12	-0. 182	0. 133	-0. 021	0. 181	-0. 099	-0. 072
的	-0. 05	-0. 156	-0. 065	-0. 005	-0. 161	-0. 047	-0. 086	0. 056	-0. 1	0. 046
。	-0. 066	-0. 052	-0. 148	0. 083	0. 079	0. 109	-0. 084	0. 14	0. 04	-0. 024

① “词义”在机器中的存在形式：向量（Vector）



② 词之间关系的意义：注意力权重（Attention）



县长 派 阿Q 来

命令者：县长

受命者：阿Q

来 者：阿Q

模型2

阿Q是县长请来的。

阿Q是县长？别逗了，他连村长都不是。

阿Q是县长，也是书记。

阿Q是县长，也要依法办事。

阿Q是县长，所以大家都怕他。

阿Q是县长，上点名批评过的。

阿Q是县长，这个决定就算数。

阿Q是县长，可一点官架子都没有。

阿Q是县长，这话是档案里写明白的。

阿Q是县长，因此有决定权。

阿Q是县长拿来当替罪羊的。

有人以为阿Q是县长，其实连正式编制都没有，要是阿Q是县长就好了，

阿Q是县长才怪呢。

阿Q是县长暗中盯上的。

阿Q是县长，而且连任了三届。

阿Q是县长抓走的。

就算阿Q是县长，这也属谣言。

就算阿Q是县长，我也不会给他开后门。

即便阿Q是县长，你也不能阿谀奉承。

难道阿Q是县长？你也信这种鬼话？

阿Q是县长，不是省长。

一旦阿Q是县长，他就有权拍板这件事。

其实他只是县长家的长工。

大家都说阿Q是县长。

阿Q是县长，也不管用。

阿Q是县长，却经常自己下乡调研。

阿Q是县长，这点没人怀疑。

阿Q是县长批评过的。

阿Q是县长最近重点整治的对象。

一旦阿Q是县长，他就有权利批评阿Q。

	基于语言学的符号规则方法	基于神经网络的深度学习方法
优势	精准、透明 Know-Why	鲁棒、泛化 Know-How
劣势	易碎、磨人	玄学、耗能

3. 语言学与人工智能的融合之路

语言学与AI的互动：优势互补

AI

- 依赖**大数据** $\implies$ 由知识驱动生成高质量**小数据**
- 表层数据**盲区** $\implies$ 以**深度认知**弥补表层数据盲区

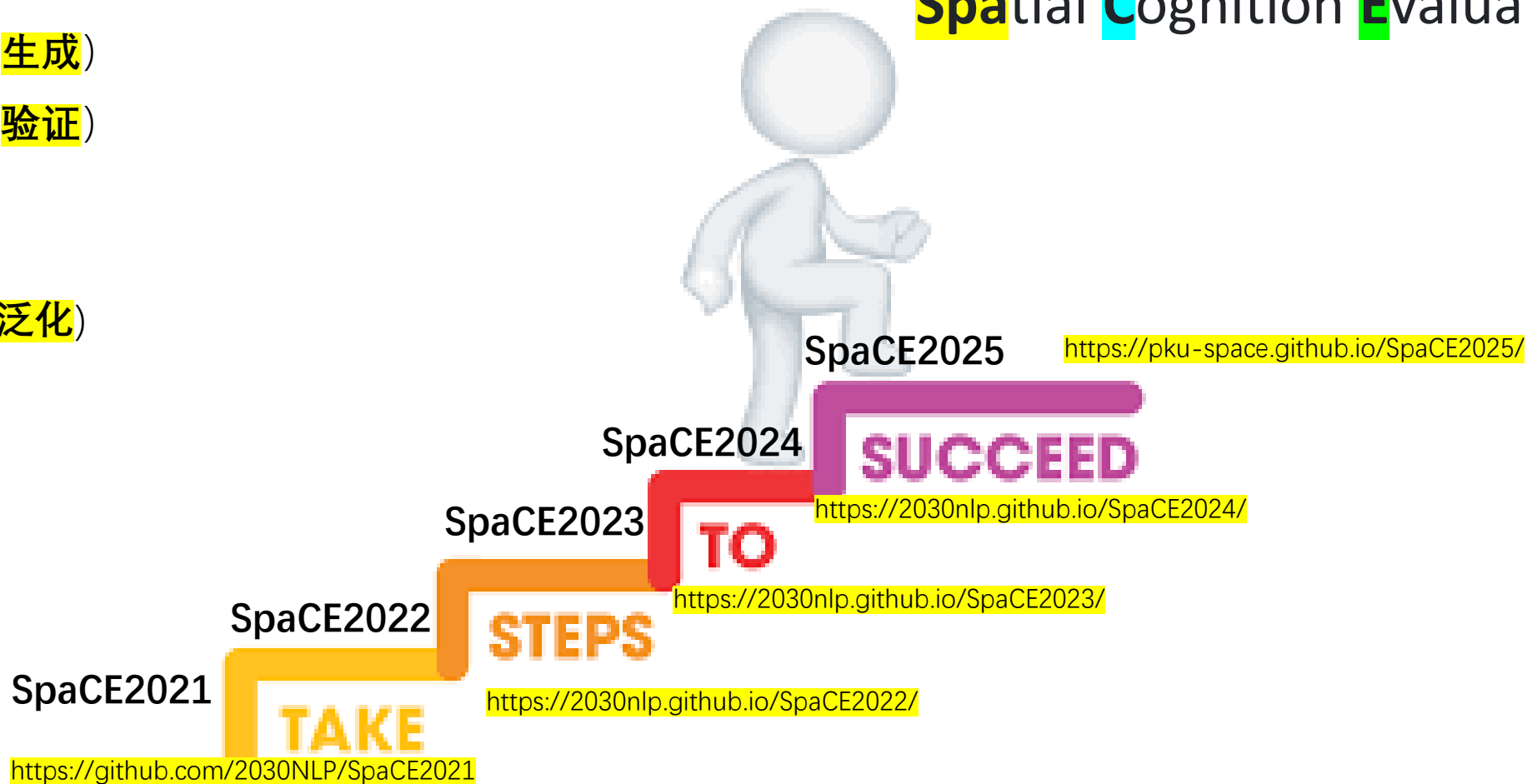
语言学

研究目标

- 可**生成**
- 可**验证**
- 可**泛化**

Spatial Cognition Evaluation

- 知识数据化 (可生成)
 - 数据任务化 (可验证)
 - 规模扩展
 - 类型扩展
- (可泛化)



北京大学中文系（詹卫东课题组）

北京大学计算机学院计算语言所（穗志方课题组）

“科技创新2030-‘新一代人工智能’重大项目”——“以自然语言为核心的语义理解理论、模型与方法”（2020AAA0106701）

空间语义理解能力

- 空间信息正误判别

- 异常空间信息识别

- 空间语义角色标注

- 空间表达异形同义判别

- 缺失参照成分找回

- 空间方位关系推理

大客车沿新源路由北向南行驶至曹安公路路口处遇绿灯，遂右转弯由东向西行驶，适逢被害人李红英骑电动自行车沿新源路西侧非机动车道由北向南行驶，两车相撞。

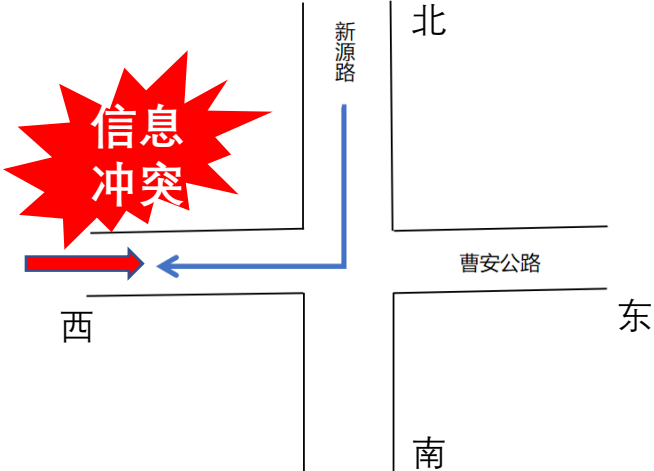


大客车沿新源路由北向南行驶至曹安公路路口处遇绿灯，遂右转弯由西向东行驶，适逢被害人李红英骑电动自行车沿新源路西侧非机动车道由北向南行驶，两车相撞。



空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理



大客车沿新源路由北向南行驶至曹安公路路口处遇绿灯，遂右转弯由西向东行驶，适逢被害人李红英骑电动自行车沿新源路西侧非机动车道由北向南行驶，两车相撞。



信息1	信息2
大客车由北向南行驶	大客车右转弯由西向东行驶
语 义	冲 突

空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

“这就是我从水里捞出来的美国人手上戴着的戒指，里面刻着他的名字。”塞科高高地举起戒指说，“上面的名字是，”他一边看一边念出上面刻着的字：“约翰·罗伯森·邓纳姆。”

美国人	从水里出来	捞
戒指	美国人手上	戴
他的名字	戒指里面	刻
戒指	塞科手上	举起
名字	戒指上面	
字	戒指上面	刻
他	= 塞科	
字	= 名字 = 约翰·罗伯森·邓纳姆	

空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

至今菲律宾的土著居民在见面时，握过手后还要转身向**后**走几步，意思是向对方表明背后没有藏刀，是真诚地迎接对方。

转身向**后**走几步 = 转身向**前**走几步 = 转身 走几步



空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

中间

等大家都坐好，聂赫留朵夫也在他们对面坐下来，臂肘搁在桌上，面前摆着一张纸，他就根据纸上的提纲开始说明他的方案。

异义

铁路要经过很多高山，不得不开凿隧道，其中居庸关和八达岭两条隧道的工程最艰巨。居庸关山势高，岩层厚，詹天佑决定采用从两端同时向中间凿进的办法……把工期缩短了一半。

同义

对面

空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

芳林嫂挑着沉重的担子，颤颤巍巍地走到了大院的门岗前。门岗里的保安见她辛苦，就喊她把担子放下去，这样可以休息一下。

异义

进去

进去

人的咽喉和食管同胃是相通的，喝下去的醋只会与鱼刺接触，醋的脱钙作用无法进行。

同义

空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

在一座小**县城**的一间教室里，工人们正在安装一块电子**白板**。“借助**网课**，我们的学生坐在教室里，就可以跟着**上面**的名师学习，享受优质的教育资源。”校长兴奋地说。

空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

场景：汉钟离、铁拐李、姜子牙、韩湘子四人来到火锅店吃火锅，选了四人卡座坐下。卡座分列一张长方形桌子长边两侧，每排卡座上坐两人。**面对面**而坐。已知：

汉钟离在铁拐李的**右手边**

姜子牙在汉钟离的**正对面**

韩湘子在姜子牙的**右手边**

问题：_____在铁拐李的**斜对面**？

选项：A. 汉钟离 B. 韩湘子
C. 姜子牙 D. 以上选项都不是

答案：C. 姜子牙



空间语义理解能力

- 空间信息正误判别
- 异常空间信息识别
- 空间语义角色标注
- 空间表达异形同义判别
- 缺失参照成分找回
- 空间方位关系推理

SpaCE2021

SpaCE2022

SpaCE2023

SpaCE2024

SpaCE2025

√

√

—

—

√

√

√

√

√

—

—

√

√

√

—

—

—

√

√

√

—

—

—

√

√

—

—

—

√

√

中文/英文

<https://pku-space.github.io/SpaCE2025/>

<https://2030nlp.github.io/SpaCE2024/>

<https://2030nlp.github.io/SpaCE2023/>

<https://2030nlp.github.io/SpaCE2022/>

<https://github.com/2030NLP/SpaCE2021>

SpaCE2021、2022、2023、2024、2025

从原始语料到数据集

	SpaCE2021	SpaCE2022	SpaCE2023	SpaCE2024	SpaCE2025
初始语料池字数	42,000,000	89,740,377	-	-	-
标注原句句数	463	6,643	-	-	-
标注原句字数	45,574	698,476	-	-	-
数据集文段数	402	5,095	-	6,868	6,423
数据集文段字数	40,091	543,316	-	1,009,658	840,539
标注语料不重复句数	16,692	47,920	-	-	-
标注语料不重复句字数	2,088,037	5,557,501	-	-	-
数据集不重复文段数	7,738	18,907	-	7,654	5,130
数据集不重复文段字数	863,669	2,158,803	-	1,115,583	664,436
数据集题数	18,236	24,947	9,565	10,373	11,540
数据集总字数	2,034,101	2,849,795	1,079,826	1,859,261	2,256,441

基于知识库的空间方位关系推理数据集制作



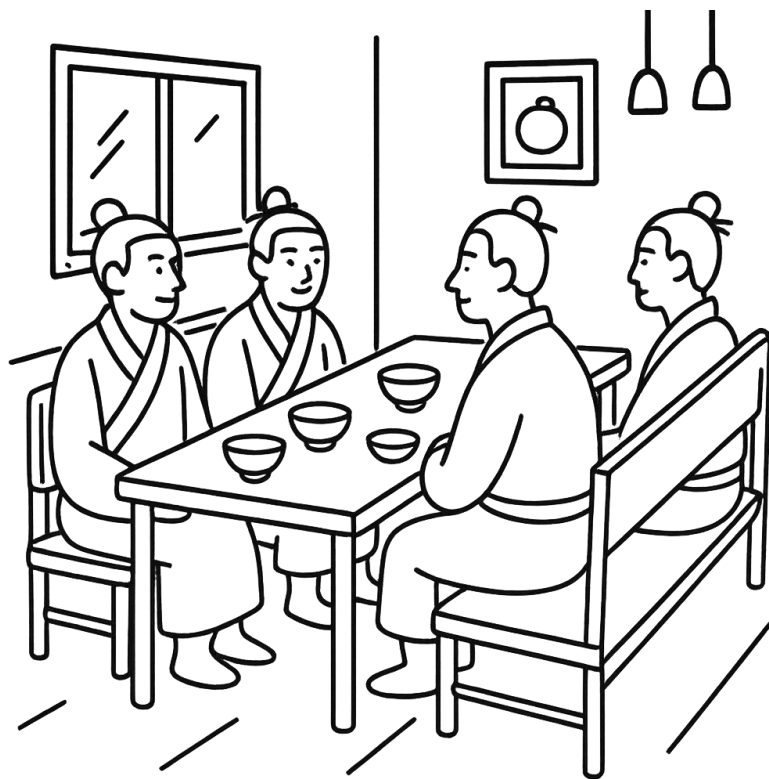
空间布局图式

空间关系知识库

空间推理题题库

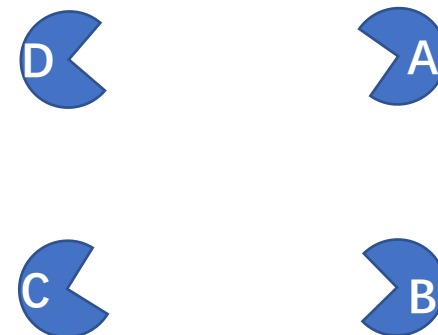
空间布局（场景）

四人卡座示意图

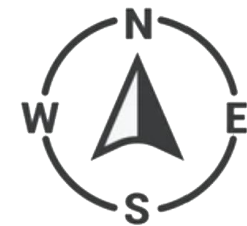


图源：GPT4o

东西向对坐



空间布局（场景）



四人卡座示意图



南北向对坐



图源：GPT4o

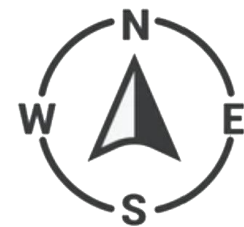
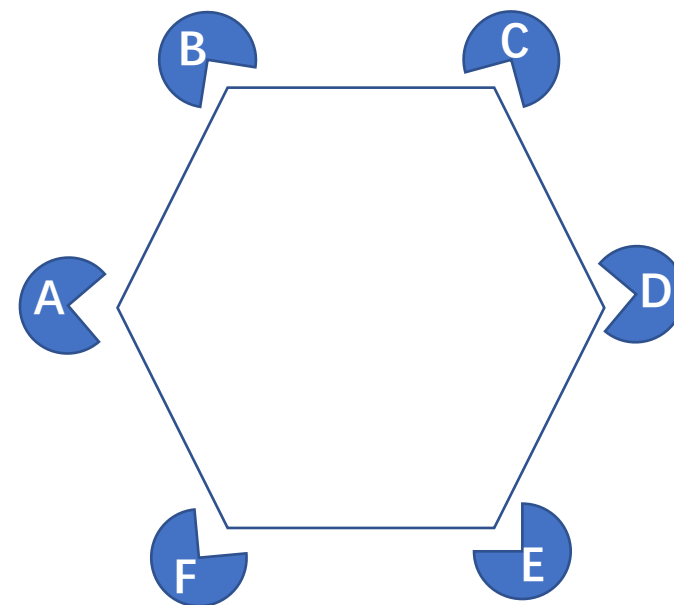
空间布局（场景）

多人围坐示意图



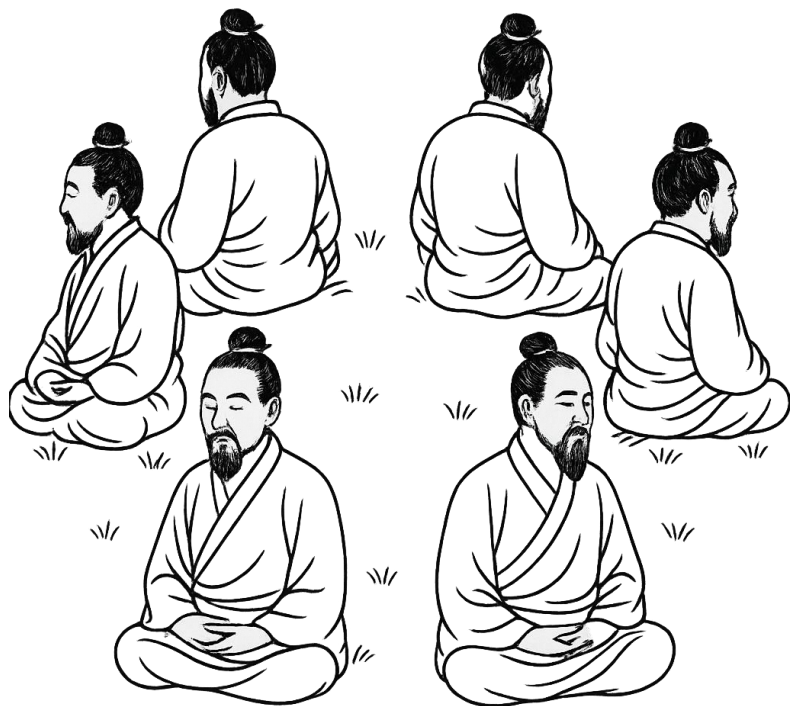
图源：GPT4o

向心六角布局



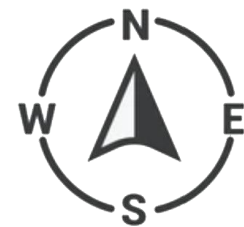
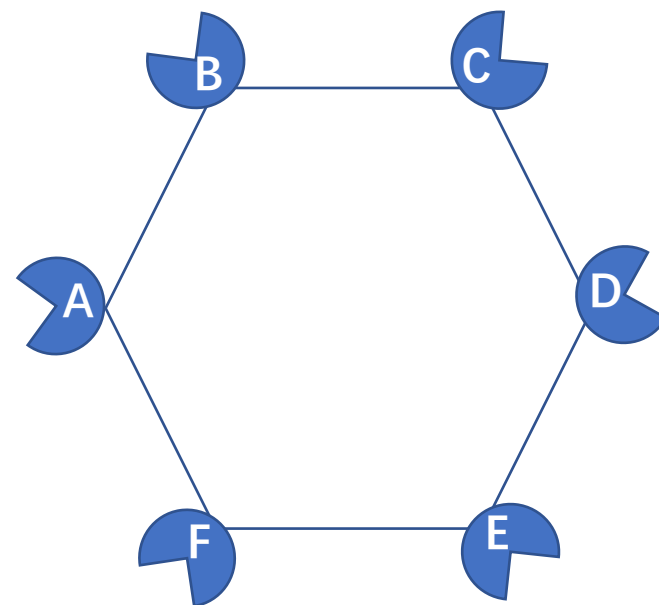
空间布局（场景）

多人围坐示意图

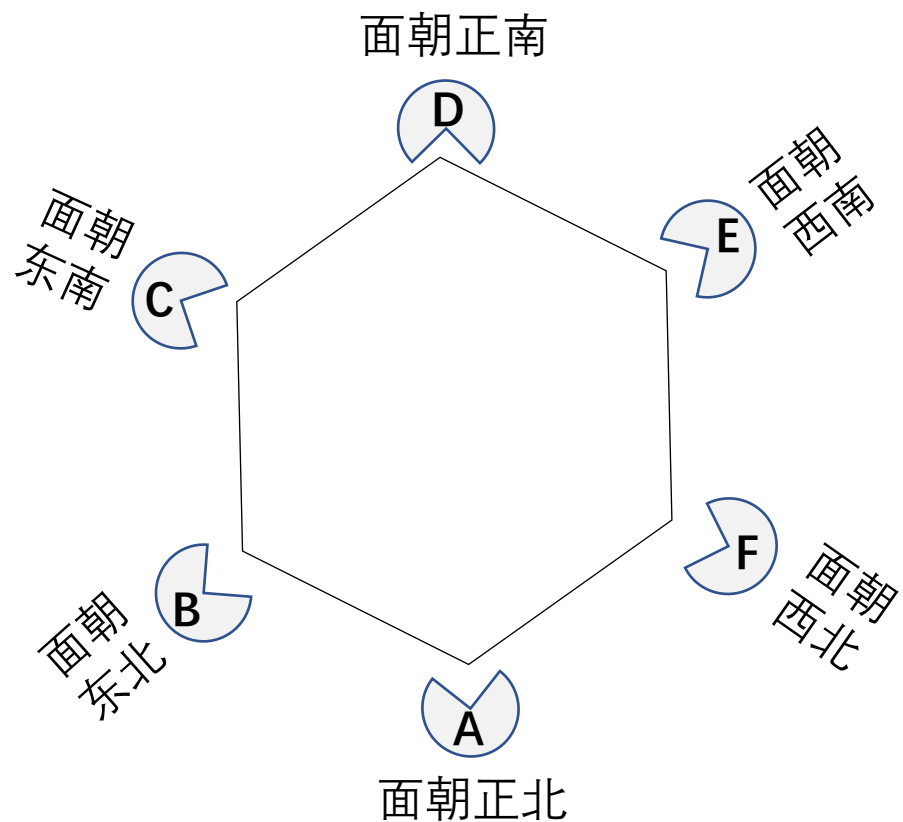


图源：GPT4o

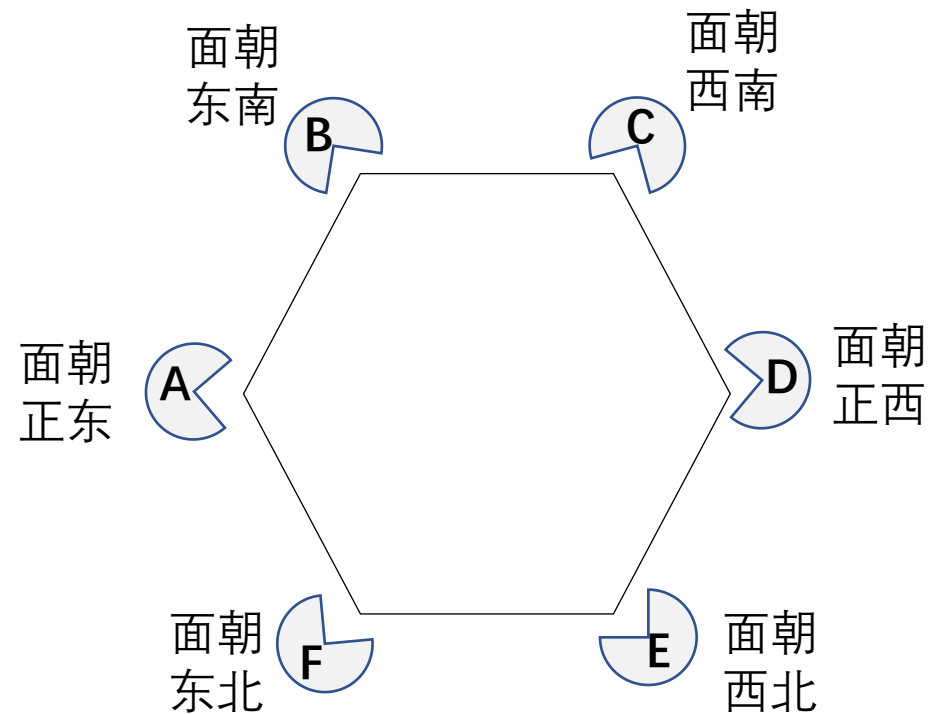
离心六角布局



向心六角布局（增加朝向）



(1) 正南正北



(2) 正东正西

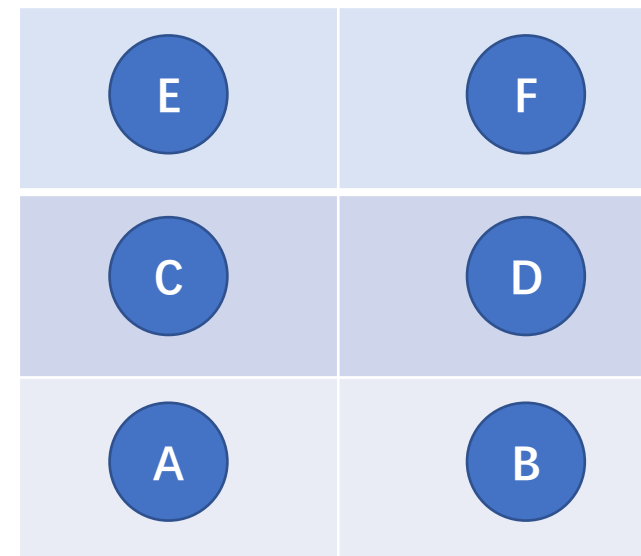
空间布局（场景）

三层置物架示意图

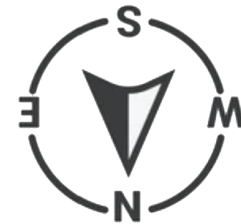


图源：GPT4o

三层两列布局



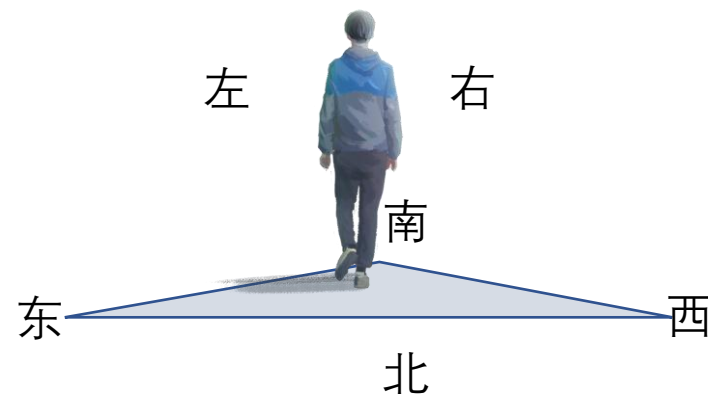
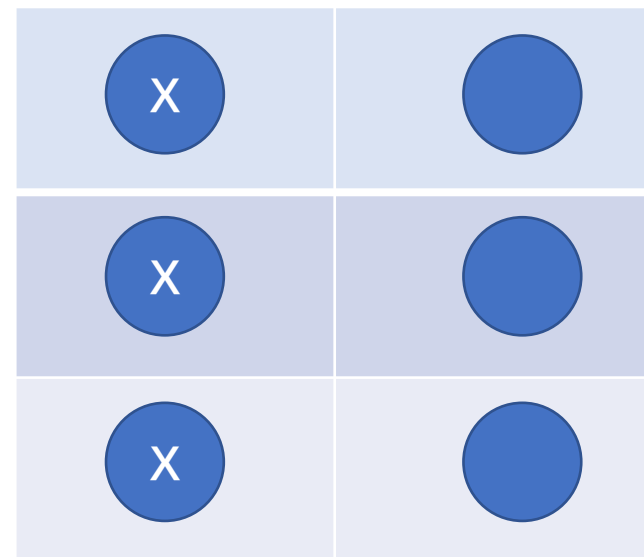
从空间布局到空间关系知识库



1

templates

```
template_X在东侧: {  
  text: "(X)在东侧",  
  precise: null,  
  symmetric: null,  
  question: true,  
},
```



从空间布局到空间关系知识库

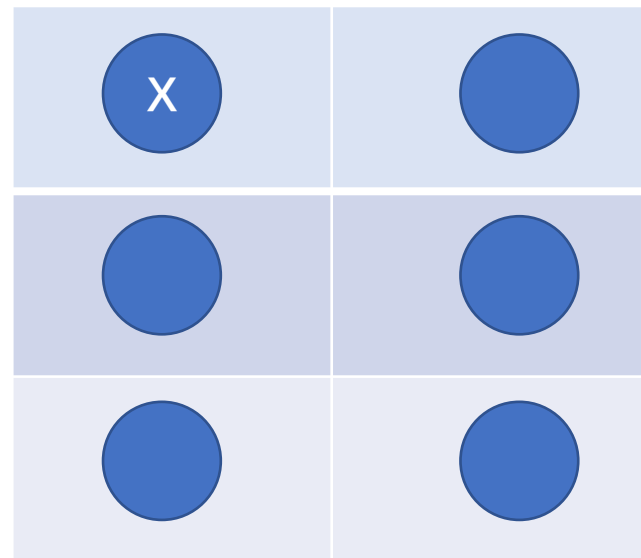


1

templates

```
template_X在东侧: {  
  text: "(X)在东侧",  
  precise: null,  
  symmetric: null,  
  question: true,  
},
```

```
template_X在三层东侧: {  
  text: "(X)在三层东侧",  
  precise: true,  
  symmetric: null,  
  question: true,  
},
```



从空间布局到空间关系知识库



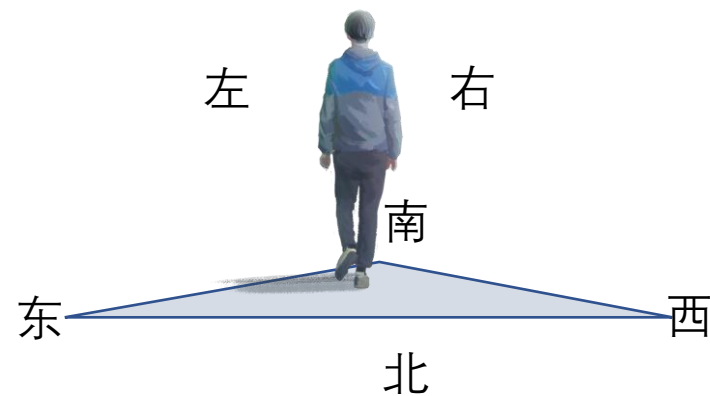
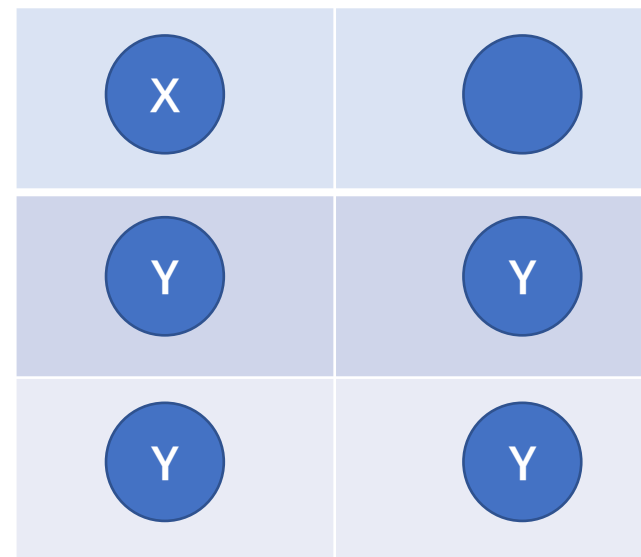
1

templates

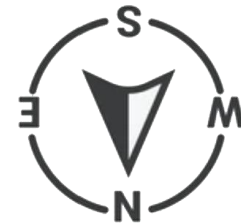
```
template_X在东侧: {  
  text: "(X)在东侧",  
  precise: null,  
  symmetric: null,  
  question: true,  
},
```

```
template_X和Y不在同一层: {  
  text: "(X)和(Y)不在同一层",  
  precise: null,  
  symmetric: true,  
  question: true,  
},
```

```
template_X在三层东侧: {  
  text: "(X)在三层东侧",  
  precise: true,  
  symmetric: null,  
  question: true,  
},
```



从空间布局到空间关系知识库



1

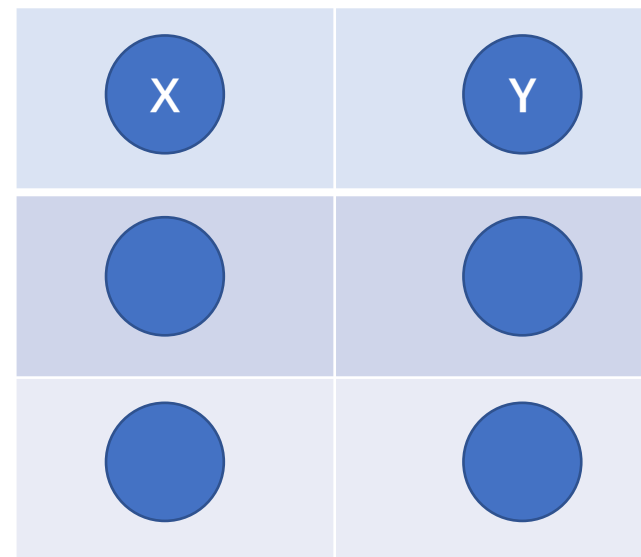
templates

```
template_X在东侧: {  
  text: "(X)在东侧",  
  precise: null,  
  symmetric: null,  
  question: true,  
},
```

```
template_X和Y不在同一层: {  
  text: "(X)和(Y)不在同一层",  
  precise: null,  
  symmetric: true,  
  question: true,  
},
```

```
template_X在三层东侧: {  
  text: "(X)在三层东侧",  
  precise: true,  
  symmetric: null,  
  question: true,  
},
```

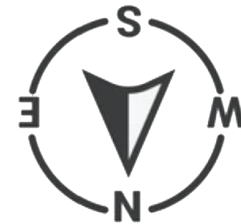
```
template_X和Y在同一层: {  
  text: "(X)和(Y)在同一层",  
  precise: true,  
  symmetric: true,  
  question: true,  
},
```



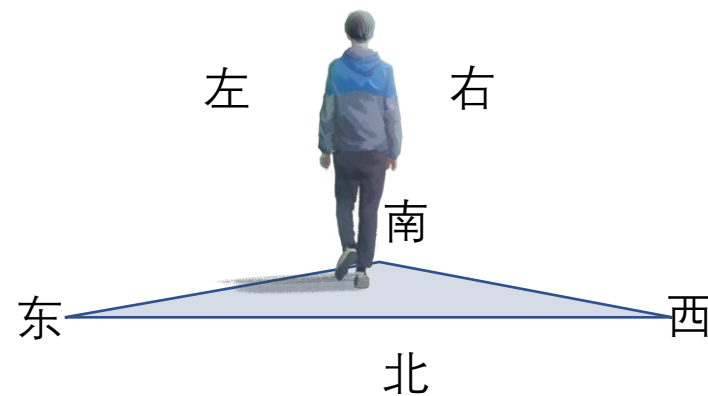
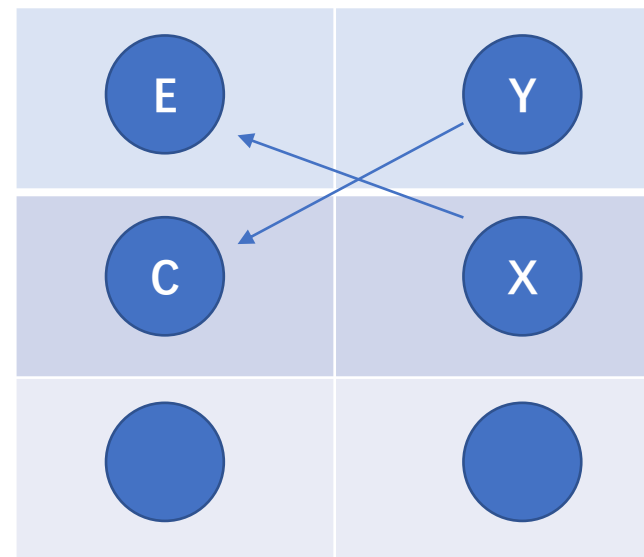
从空间布局到空间关系知识库

2

templates_logic



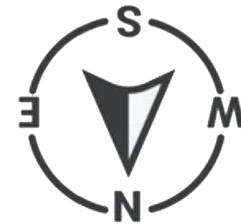
$p \leftrightarrow q$ ["等价", "template_X的左邻在Y的左下方", "template_Y的左邻在X的左上方"]



从空间布局到空间关系知识库

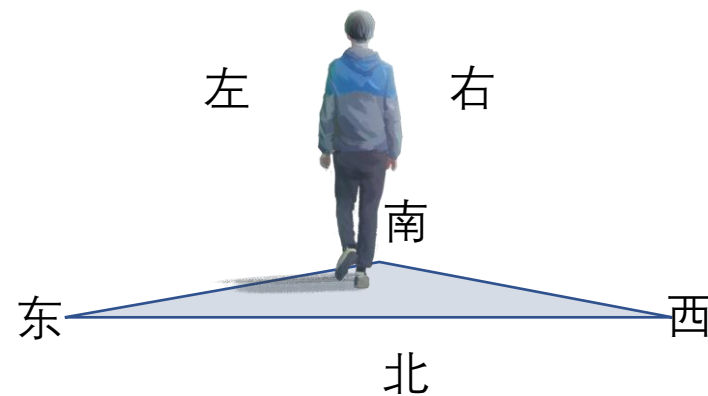
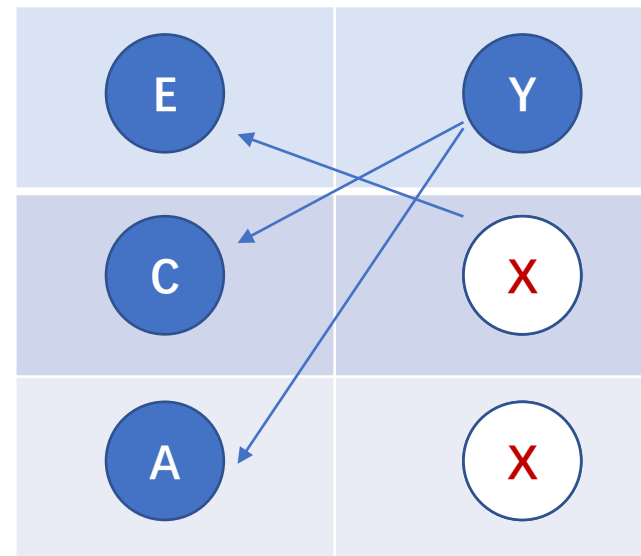
2

templates_logic



$p \leftrightarrow q$ ["等价", "template_X的左邻在Y的左下方", "template_Y的左邻在X的左上方"]

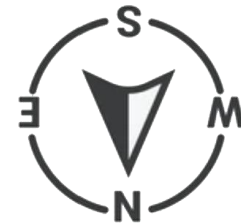
$p \rightarrow q$ ["蕴含", "template_X的左邻在Y的左下方", "template_X在Y正下方"]



从空间布局到空间关系知识库

2

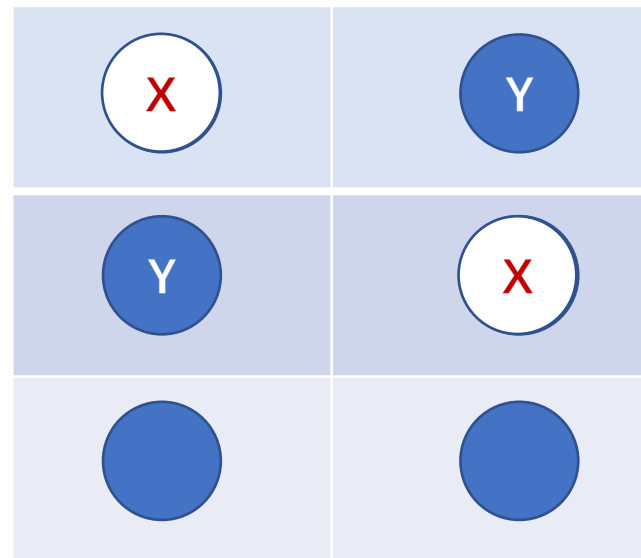
templates_logic



$p \leftrightarrow q$ ["等价", "template_X的左邻在Y的左下方", "template_Y的左邻在X的左上方"]

$p \rightarrow q$ ["蕴含", "template_X的左邻在Y的左下方", "template_X在Y正下方"]

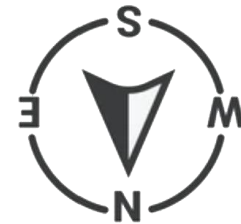
$q \rightarrow p$ ["包括", "template_X与Y同层", "template_X在Y左边"]



从空间布局到空间关系知识库

2

templates_logic

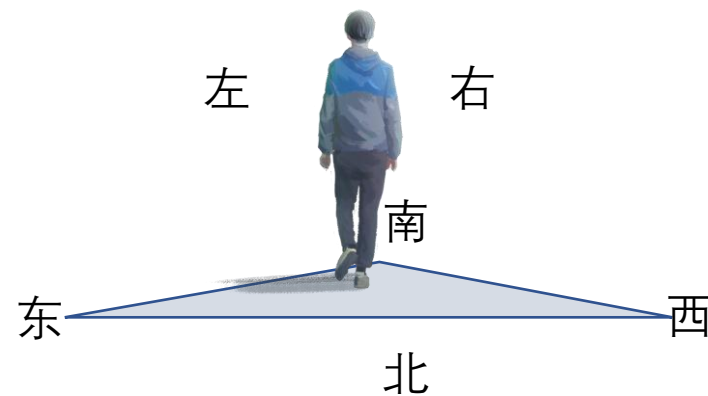
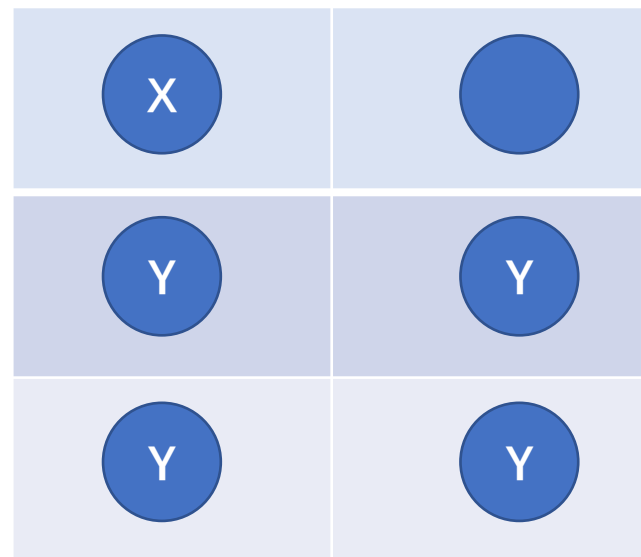


$p \leftrightarrow q$ ["等价", "template_X的左邻在Y的左下方", "template_Y的左邻在X的左上方"]

$p \rightarrow q$ ["蕴含", "template_X的左邻在Y的左下方", "template_X在Y正下方"]

$q \rightarrow p$ ["包括", "template_X与Y同层", "template_X在Y左边"]

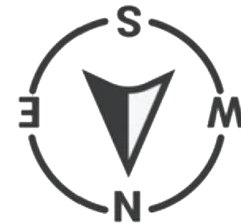
$\neg(p \wedge q)$ ["冲突", "template_X和Y隔了一层", "template_Y所在层和X所在层相邻"]



从空间布局到空间关系知识库

2

templates_logic



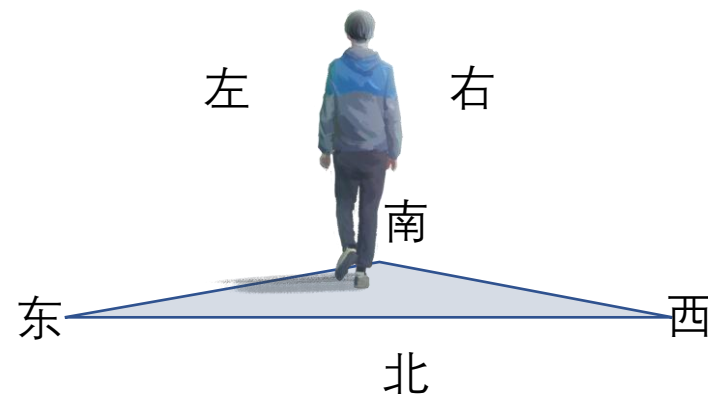
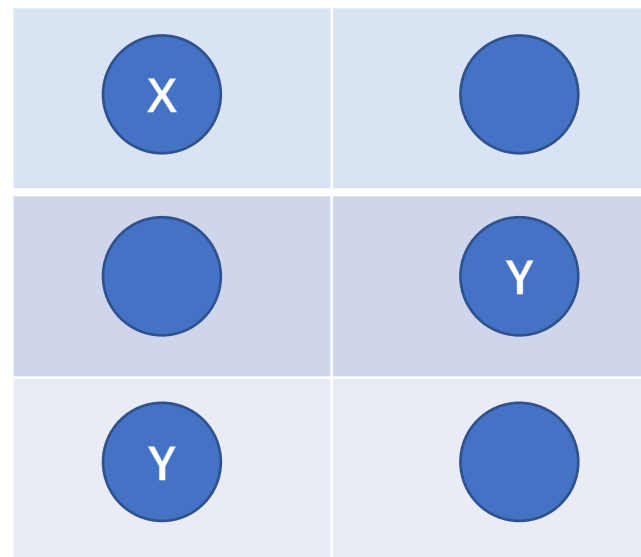
$p \leftrightarrow q$ ["等价", "template_X的左邻在Y的左下方", "template_Y的左邻在X的左上方"]

$p \rightarrow q$ ["蕴含", "template_X的左邻在Y的左下方", "template_X在Y正下方"]

$q \rightarrow p$ ["包括", "template_X与Y同层", "template_X在Y左边"]

$\neg(p \wedge q)$ ["冲突", "template_X和Y隔了一层", "template_Y所在层和X所在层相邻"]

$\neg p \leftrightarrow q$ 或者 $p \leftrightarrow \neg q$ ["互反", "template_X与Y同侧", "template_X与Y不同侧"]



从空间布局到空间关系知识库

3

rules

rule_隔层方位加东西推上方: {

labels: ["J","K"],

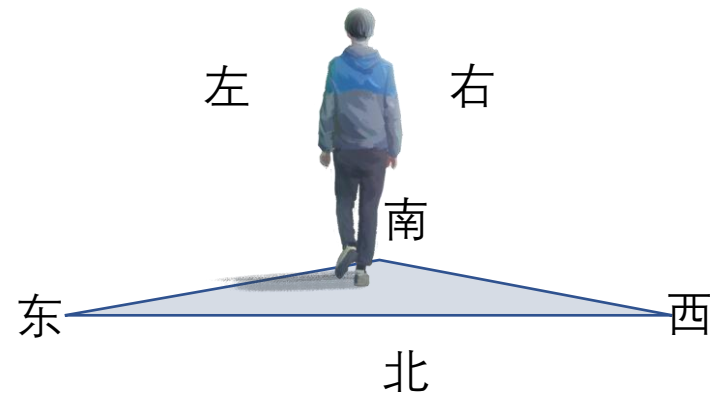
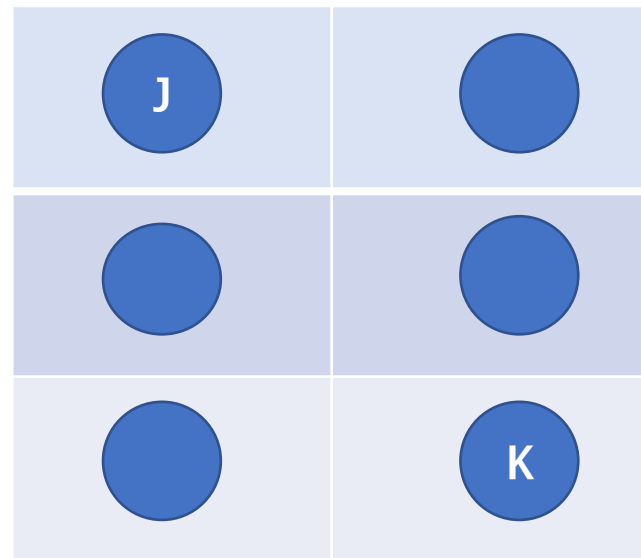
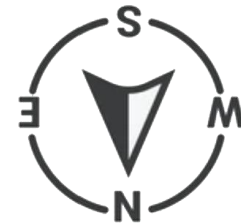
conditions: [

 ["J","empty","template_X在三层东侧"],

 ["K","empty","template_X在一层西侧"],

],

conclusions: ["J","K","template_X在Y左上方且二者隔了一层"]

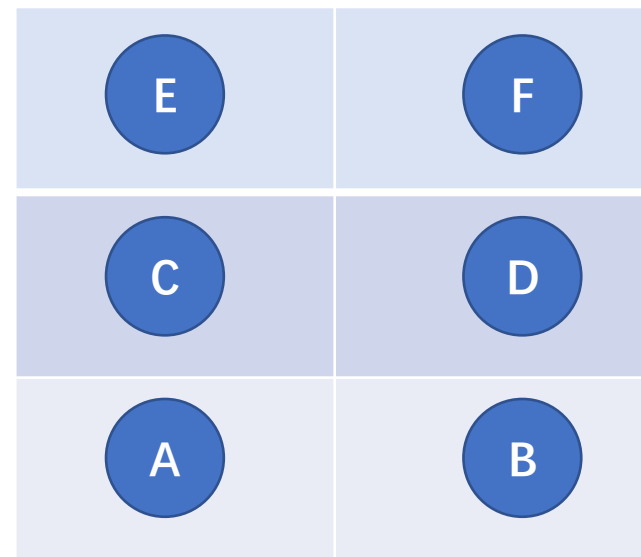
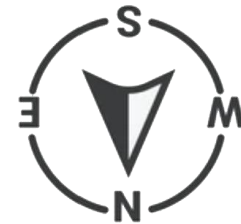


从空间布局到空间关系知识库

4

replacement_entities

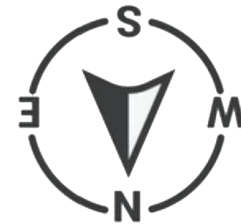
```
target_slots: [  
    "A",  
    "B",  
    "C",  
    "D",  
    "E",  
    "F",  
],
```



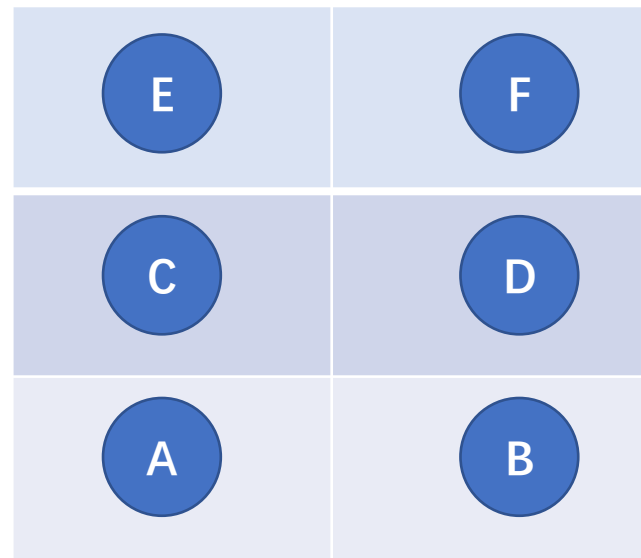
5

initial_props

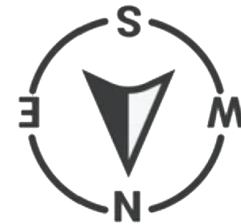
从空间布局到空间关系知识库



```
initial_props: [  
  /* 当前使用的初始命题 */  
  ['A', 'template_X在一层东侧'],  
  ['B', 'template_X在一层西侧'],  
  ['C', 'template_X在二层东侧'],  
  ['D', 'template_X在二层西侧'],  
  ['E', 'template_X在三层东侧'],  
  ['F', 'template_X在三层西侧'],  
],
```



从空间布局到空间关系知识库



6

script

```
script: [  
  {
```

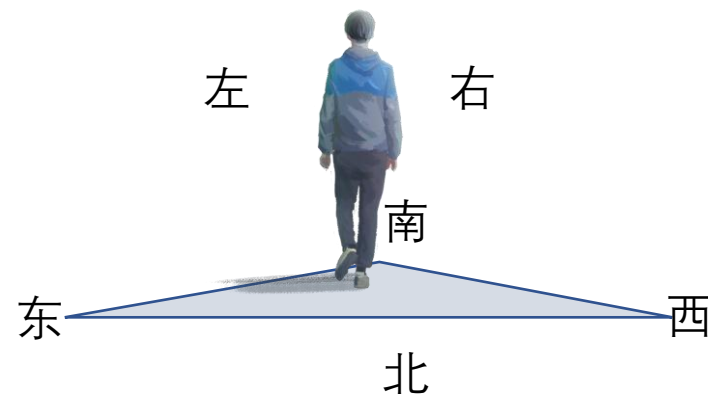
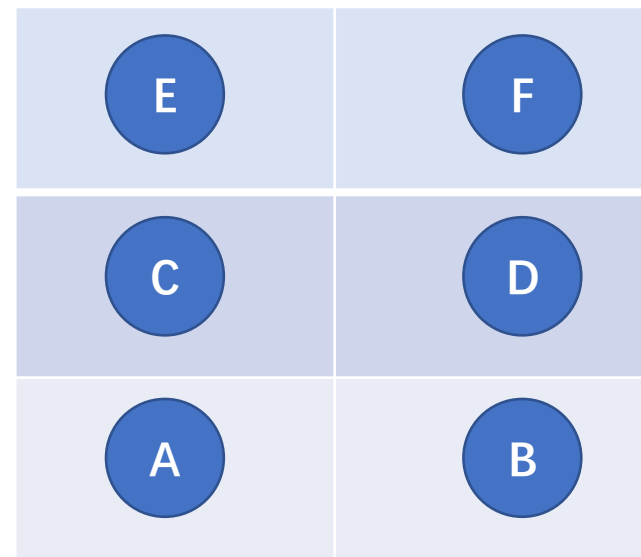
```
    context: "六盆花放置在三层花架上呈列，花架紧靠大厅  
              南墙放置，每层两格，各放一盆花，一在东，一  
              在西。画师站在花架前，面对花架支起画架，为  
              花架中六盆花画素描。在描述各花的方位关系时，  
              约定以画师自身左右方位为参照，即东侧花盆为  
              左，西侧花盆为右。东侧花盆在西侧花盆左边，  
              西侧花盆在东侧花盆右边。已知：",
```

```
    entities: [  
      "牡丹", "月季", "茶花", "茉莉", "百合", "郁金香", "  
      水仙", "君子兰", "天竺葵", "波斯菊",
```

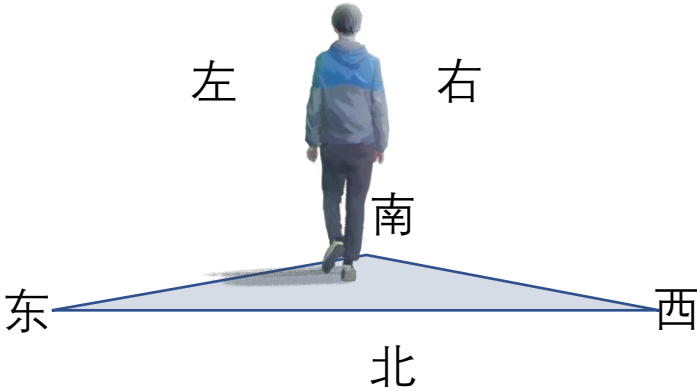
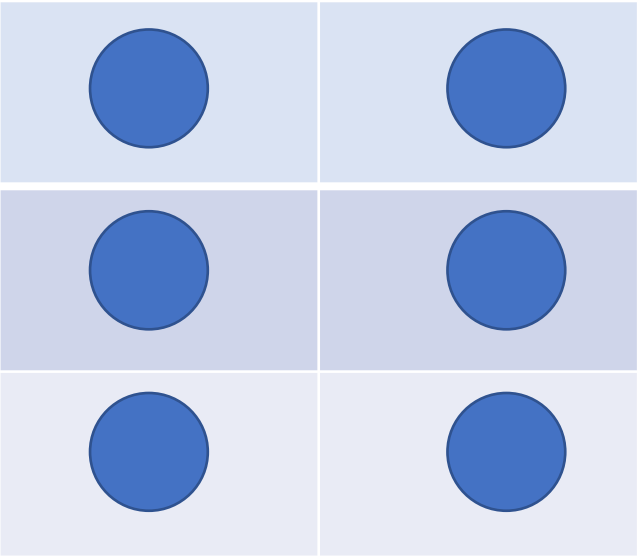
```
    ],
```

```
  },
```

```
]
```



基于空间关系知识库自动出题



	text
	question
	option
	answer

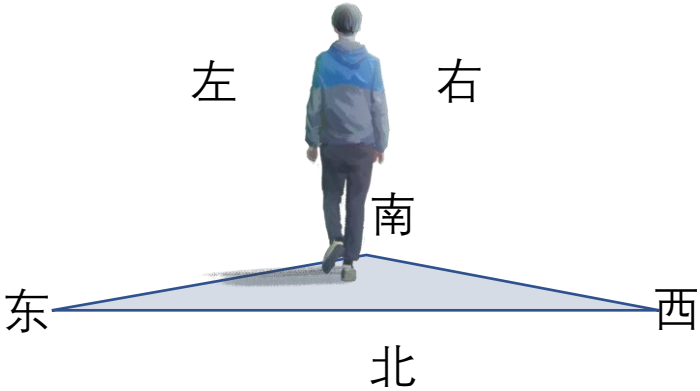
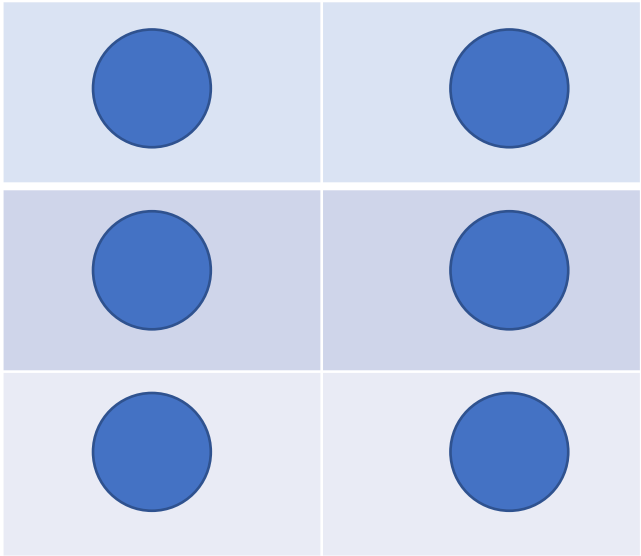
基于空间关系知识库自动出题



茉莉、君子兰、天竺葵、月季、郁金香、水仙六盆花放置在三层花架上呈列，花架紧靠大厅南墙放置，每层两格，各放一盆花，一在东，一在西。画师站在花架前，面对花架支起画架，为花架中六盆花画素描。在描述各花的方位关系时，约定以画师自身左右方位为参照，即东侧花盆为左，西侧花盆为右。东侧花盆在西侧花盆左边，西侧花盆在东侧花盆右边。已知：

text

- (1) 郁金香在顶层，
- (2) 郁金香在茉莉正上方，
- (3) 水仙在月季左下方且二者不隔层，
- (4) 茉莉和郁金香隔着一层，
- (5) 月季在郁金香右边且二者同层。



question

option

answer

基于空间关系知识库自动出题



茉莉、君子兰、天竺葵、月季、郁金香、水仙六盆花放置在三层花架上呈列，花架紧靠大厅南墙放置，每层两格，各放一盆花，一在东，一在西。画师站在花架前，面对花架支起画架，为花架中六盆花画素描。在描述各花的方位关系时，约定以画师自身左右方位为参照，即东侧花盆为左，西侧花盆为右。东侧花盆在西侧花盆左边，西侧花盆在东侧花盆右边。已知：

(1) 郁金香在顶层，

(2) 郁金香在茉莉正上方，

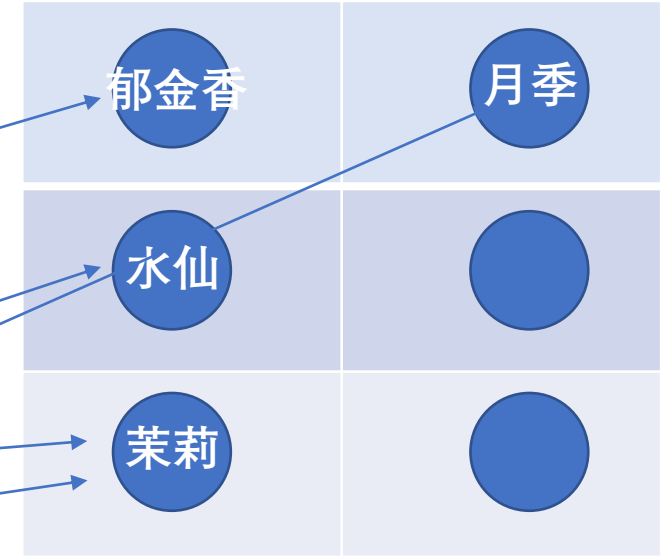
(3) 水仙在月季左下方且二者不隔层，

(4) 茉莉和郁金香隔着一层，

(5) 月季在郁金香右边且二者同层。

请问:二层东侧是__。

A.茉莉 B.郁金香 C.水仙 D.月季



基于空间关系知识库自动出题



茉莉、君子兰、天竺葵、月季、郁金香、水仙六盆花放置在三层花架上呈列，花架紧靠大厅南墙放置，每层两格，各放一盆花，一在东，一在西。画师站在花架前，面对花架支起画架，为花架中六盆花画素描。在描述各花的方位关系时，约定以画师自身左右方位为参照，即东侧花盆为左，西侧花盆为右。东侧花盆在西侧花盆左边，西侧花盆在东侧花盆右边。已知：

text

- (1) 郁金香在顶层，
- (2) 郁金香在茉莉正上方，
- (3) 水仙在月季左下方且二者不隔层，
- (4) 茉莉和郁金香隔着一层，
- (5) 月季在郁金香右边且二者同层。

请问:二层东侧是__。

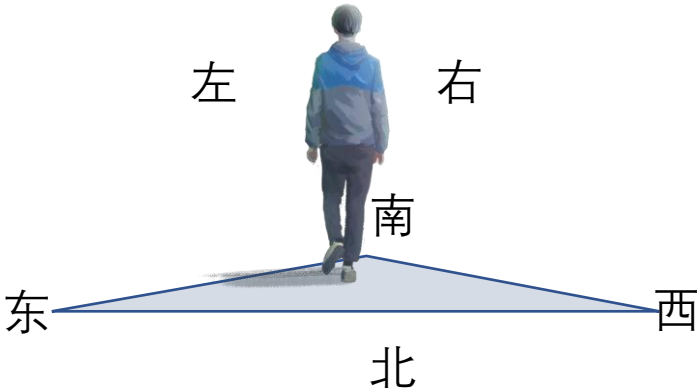
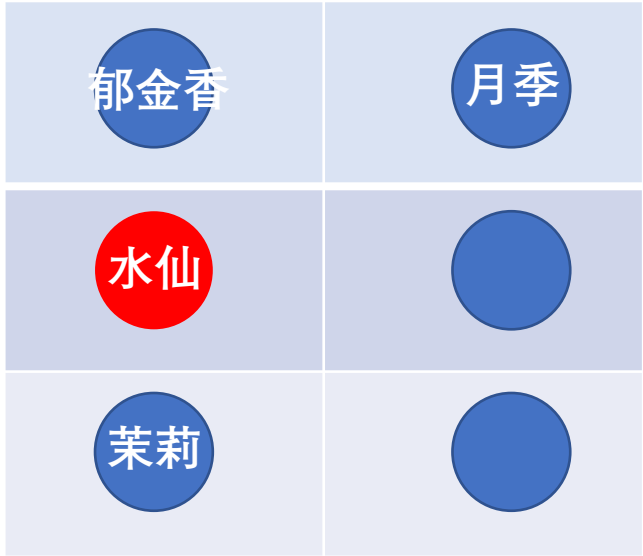
question

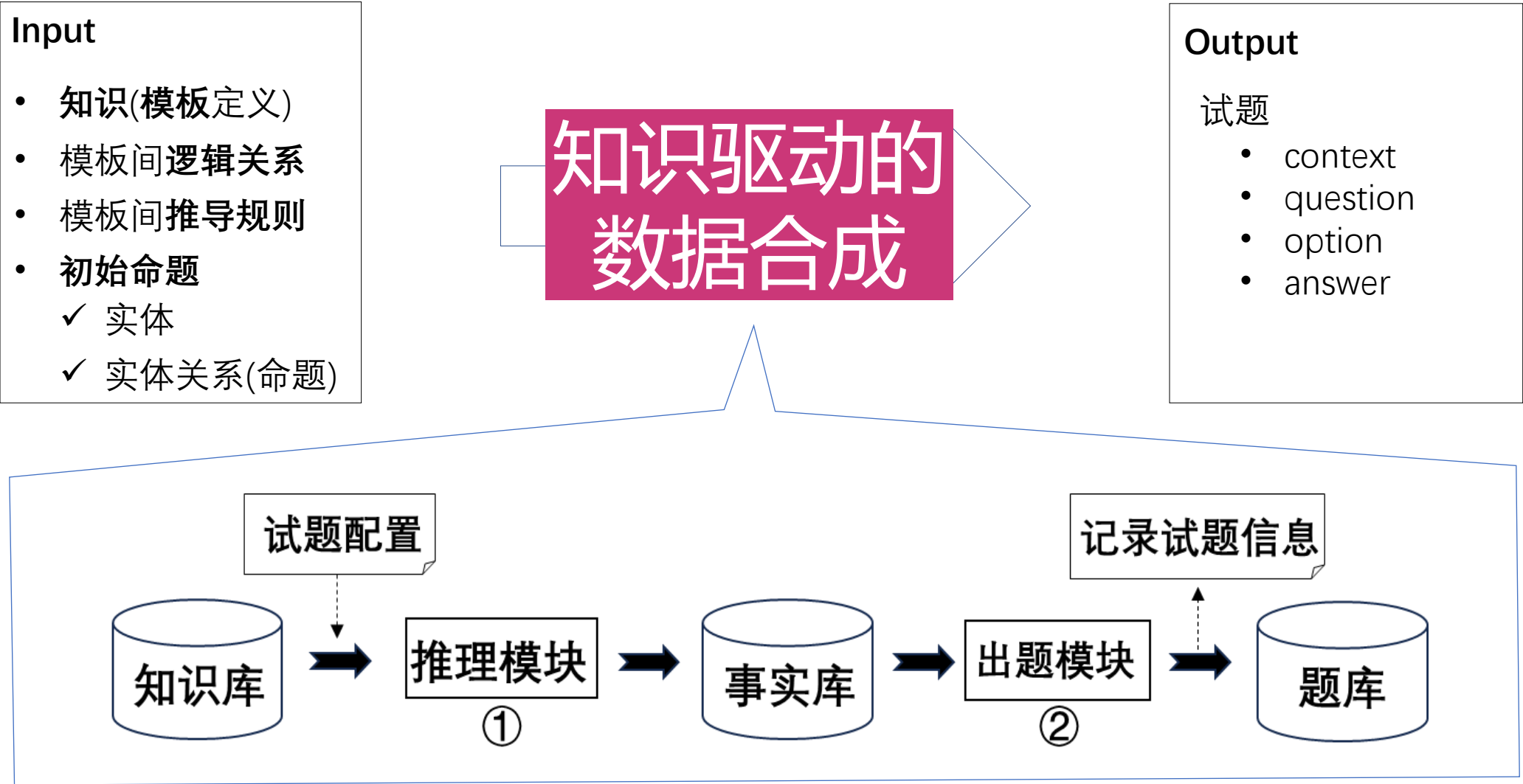
A.茉莉 B.郁金香 C.水仙 D.月季

option

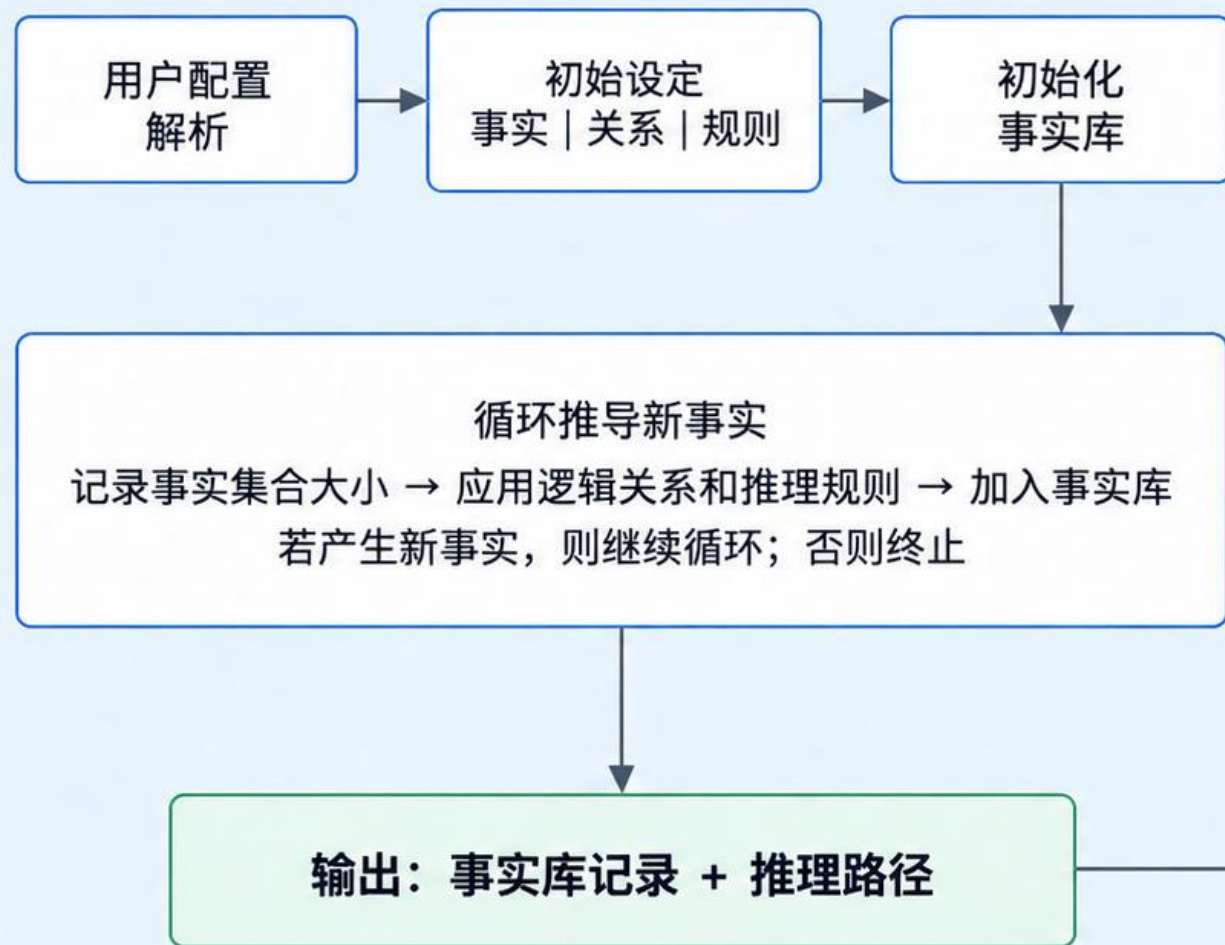
答案：C

answer





① 推理器：构建事实库并记录推理路径



② 出题器：基于事实库生成题目



SpaCE2024获奖队伍的成绩（准确率）

<https://2030nlp.github.io/SpaCE2024/leaderboard.html>

获奖名次	总分	单选	多选	角色识别	参照补回	异常识别	同义判别	方位推理
第1名	60.24	64.90	37.45	93.64	89.47	84.80	56.31	34.71
第2名	59.69	64.34	36.93	91.43	84.91	81.00	54.31	37.16
第3名	59.49	63.55	39.66	94.29	77.19	78.00	58.77	37.11
第4名	56.60	61.71	31.60	91.43	85.26	79.80	49.23	32.11
第5名	56.47	61.26	33.03	92.86	73.68	76.20	58.62	32.40
第6名	56.20	61.79	28.87	94.29	79.65	74.20	56.92	30.64
均值	58.12	62.93	34.59	92.99	81.69	79.00	55.69	34.02
机器基线	47.92	54.37	16.38	88.18	75.09	68.60	42.00	21.96
人类基线	—	—	—	96.00	92.33	91.50	86.33	—

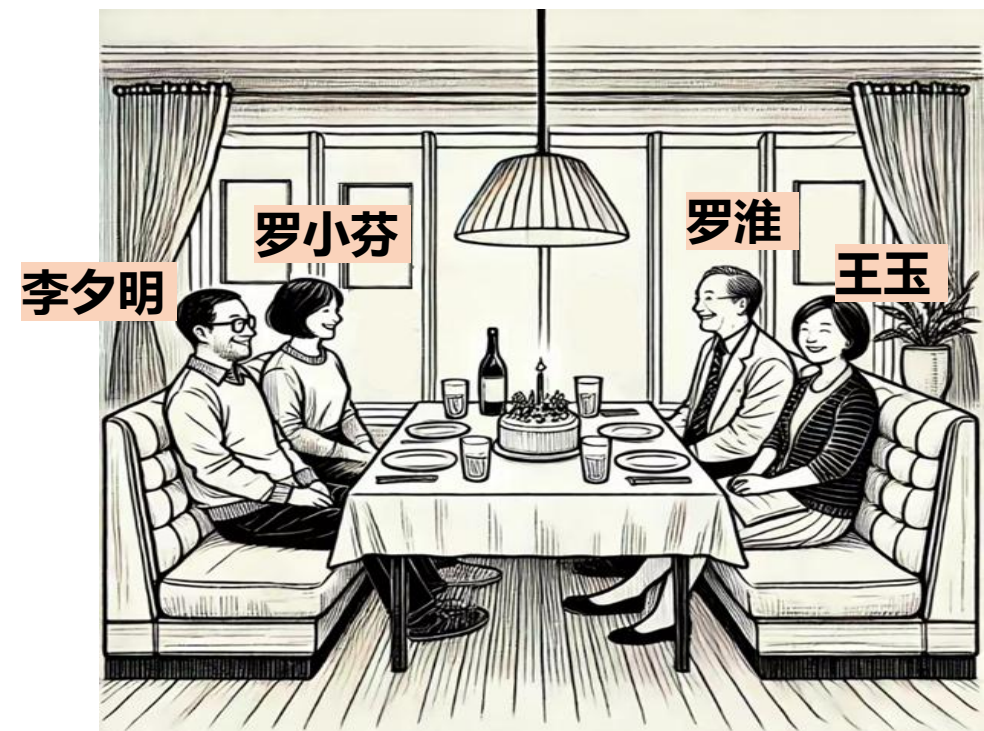
机器基线：训练集微调Qwen1.5-7B-Chat。zero-shot测试。

人类基线：6名人类被试的得分均值。每个任务独立进行100道测试集题目的人类测试。

进一步拓展：空间+社会关系常识推理

今天是罗小芬的生日，她的丈夫和父母特意设宴为她庆祝。四个人进了好客酒家，选了一个靠窗临街的四人卡座坐下。卡座分列一张长方形桌子长边两侧，每排卡座上坐两人，面对面而坐。罗小芬坐在靠窗的位置，脸上洋溢着幸福的笑容。她的父亲罗淮坐在她正对面。坐在罗淮斜对面的是他的女婿李夕明。罗小芬的母亲王玉坐在自己丈夫罗淮的左手边。

TEXT



_____ 没挨着李夕明的**岳父**坐？

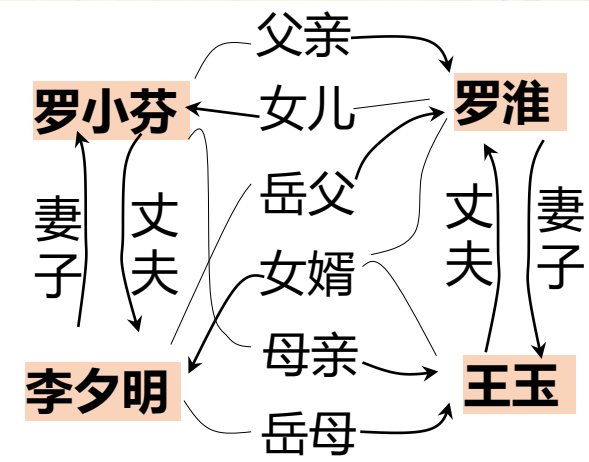
QUESTION

A.李夕明 B.王玉 **C.罗小芬** D.以上选项都不是

OPTION

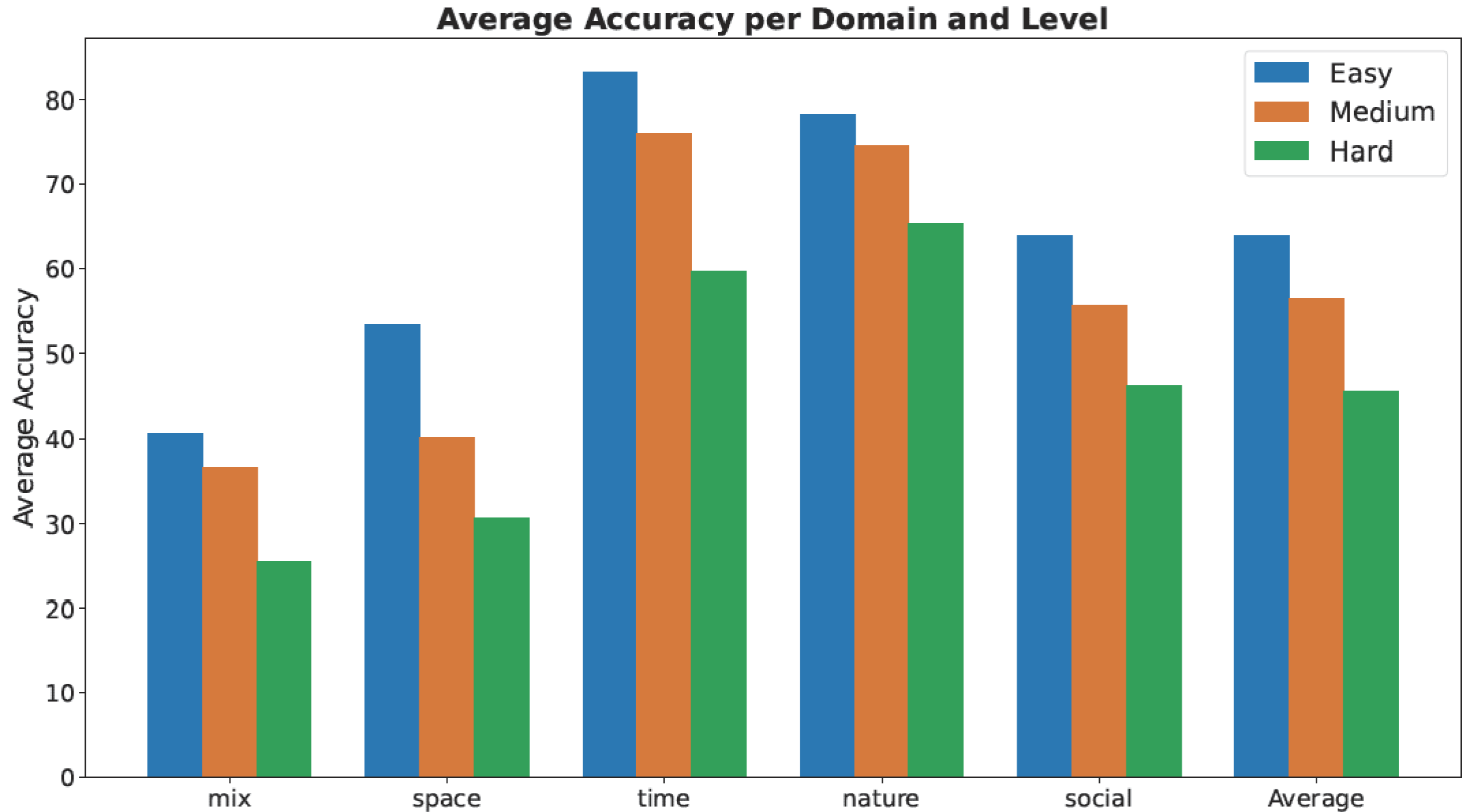
A、 C

ANSWER



Model	Space		Nature		Time		Social		Mix		Avg
	CN	EN	CN	EN	CN	EN	CN	EN	CN	EN	
Closed-Source Models											
o1-preview	67.17	58.83	89.83	84.33	79.67	80.67	49.67	71.67	61.17	54.83	69.78
o1-mini	62.00	56.50	82.00	75.33	88.00	85.67	33.83	56.67	48.00	45.33	63.33
claude-3.5-sonnet	45.33	43.17	76.00	70.83	60.50	71.67	56.83	68.67	36.33	36.17	56.55
glm-zero-preview	39.50	34.50	73.17	71.83	66.83	79.17	38.67	78.33	22.83	27.33	53.22
glm-4-plus	30.33	30.17	74.17	64.50	71.67	70.67	38.67	58.17	25.50	26.00	48.98
gpt-4o	29.50	29.00	68.17	65.67	65.50	69.83	25.17	45.00	23.67	26.83	44.83
qwen-max	15.83	13.17	65.83	64.00	54.17	71.67	47.67	56.33	24.83	26.00	44.50
o3-mini	19.67	22.00	57.33	57.83	58.50	62.33	24.67	70.00	26.67	25.17	42.41
Open-Source Models											
deepseek-r1	54.67	52.17	81.33	75.33	57.67	71.83	80.83	84.00	48.67	39.17	64.65
qwq-32B-preview	40.50	43.50	76.33	75.00	59.67	76.83	55.67	77.50	27.50	32.67	56.52
deepseek-v3	36.83	35.67	67.83	64.00	60.83	65.00	45.00	82.17	24.83	33.67	50.56
deepseek-r1-distill-qwen-32b	45.17	37.50	81.83	60.17	66.00	65.33	54.33	42.83	26.50	23.00	50.27
qwen-2.5-72B	30.67	26.00	68.83	60.83	66.00	77.00	29.33	51.17	22.50	23.17	45.55
Human											
Best	96.67	90.00	100.00	96.67	93.33	93.33	80.00	100.00	96.67	86.67	93.33
Mean	85.71	85.24	95.19	94.07	80.42	80.83	70.83	96.11	80.48	80.00	82.36

Table 2: Models Performance on Different Domains (Accuracy %). Best performers in each domain are **bolded**.



SCoRE2026

第一届基于情景的常识推理评测

首页

最新消息

排行榜

第一届基于情景的常识推理评测任务 (SCoRE2026)

👉 点我立即报名 👉

- 组织者/主席：詹卫东、穗志方
- 工作人员：胡楠，肖力铭，秦宇航，邓思锐，崔香，蔡奇栩，姜秀旻，丁锦坤，张子涵，王希豪，孙春晖，邢丹，杨奕欣，马敬原，李政，王悦
- 主办单位：北京大学
- 联系方式：hunan@stu.pku.edu.cn

• [任务内容](#)

• [评测数据](#)

- [数据字段](#)
- [数据样例](#)
- [数据概况](#)



<https://pku-space.github.io/SCoRE2026/>

4. 结语：与AI共舞

- AI真的学会人类语言了吗？

会了，又没全会

- 人到底是如何运用语言的？

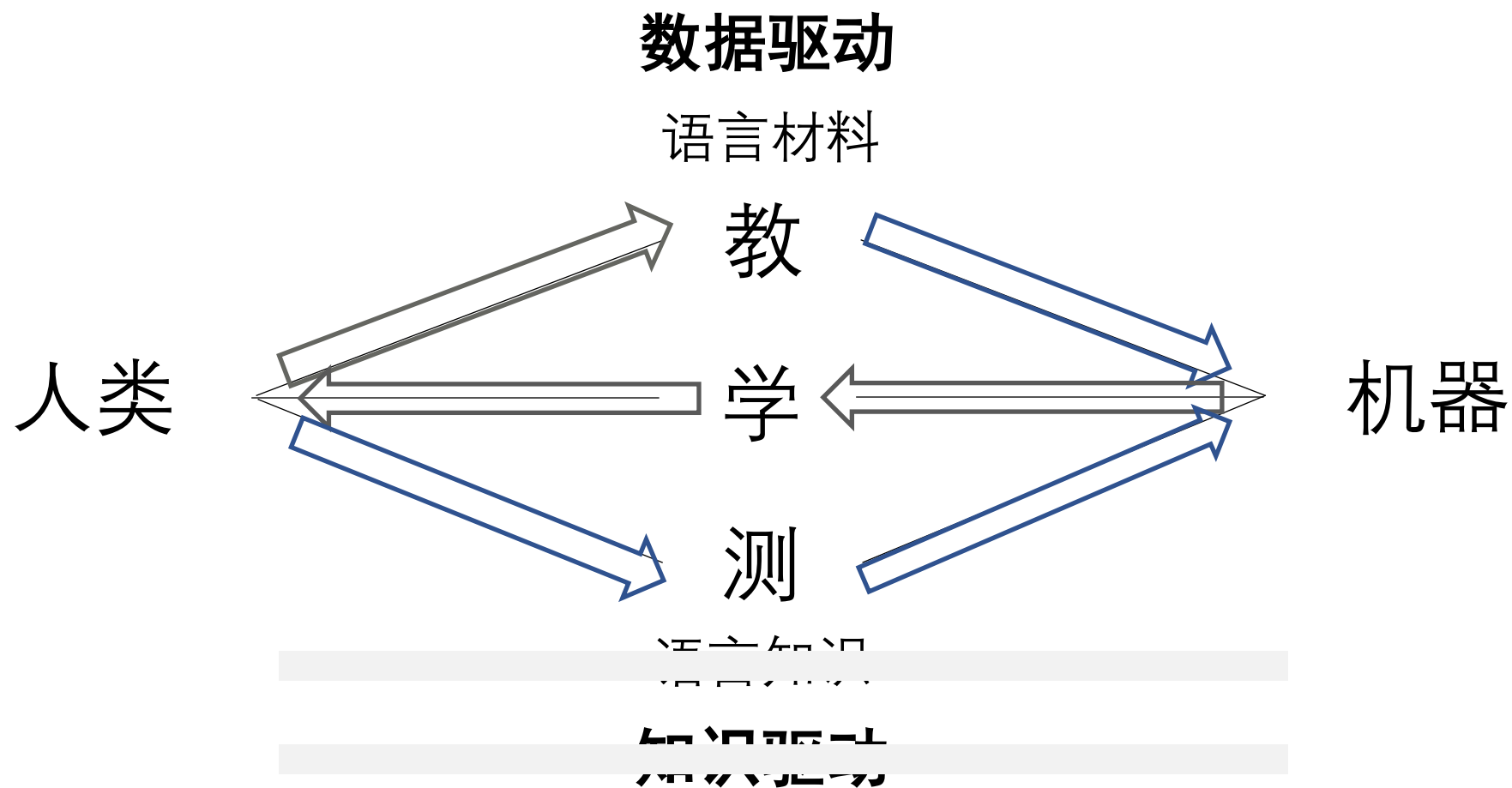
说不清楚，还得摸索

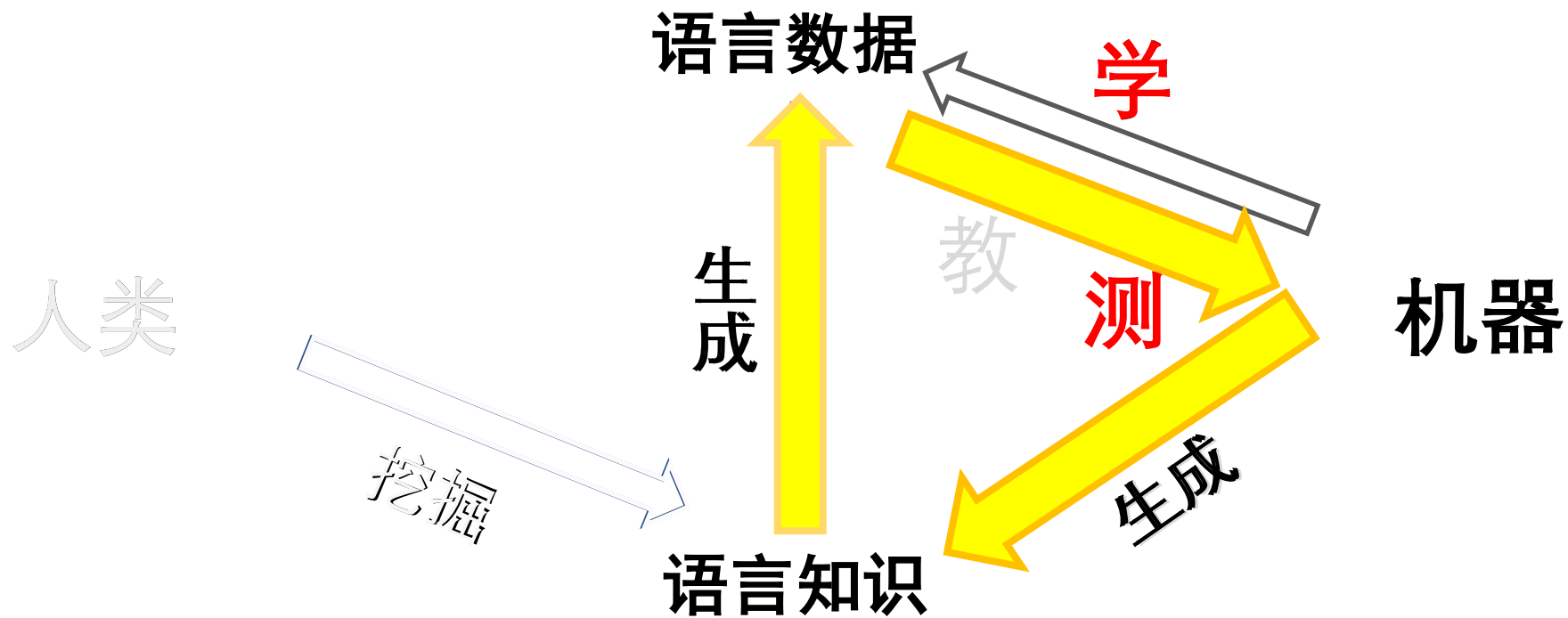
- AI到底该如何向人学语言？

学外部语言 vs. 学内部语言

- 语言学能为AI干点什么？

把可靠知识转化为高质量数据

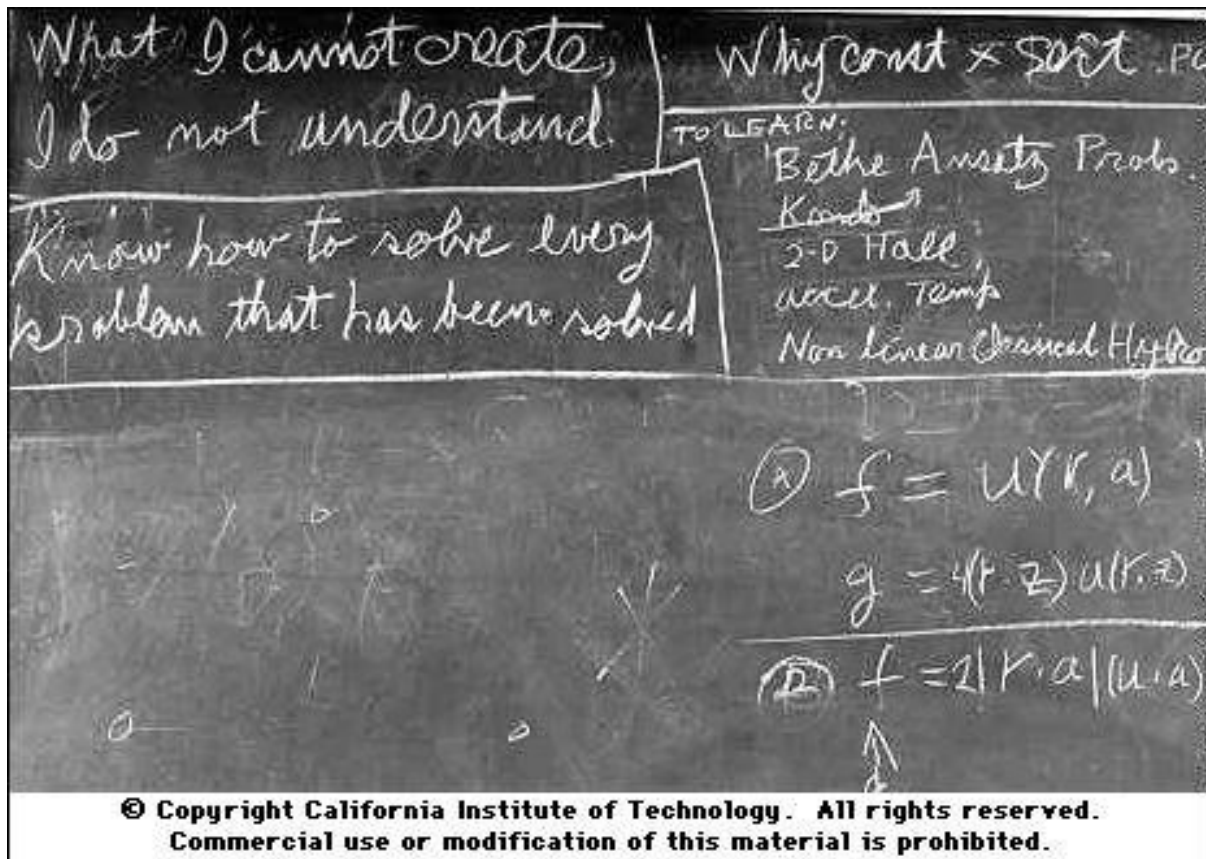




观察充分 → 描写充分 → 解释充分 → 生成充分

observational adequacy descriptive adequacy explanatory adequacy generative adequacy

语言学的目标 Chomsky 1965



费曼最后的板书

Generative models are one of the most promising approaches towards this goal. To train a generative model we first collect a large amount of data in some domain (e.g., think millions of images, sentences, or sounds, etc.) and then train a model to generate data like it. The intuition behind this approach follows a famous quote from Richard Feynman:

非经亲手造，何敢言真知

—— 理查德·费曼

1. 詹卫东, 2026, 从范式嬗变看语言学与人工智能的融合路径, 《语言战略研究》2026年第3期, pp.41-52。
2. Weidong Zhan, Yue Wang, Nan Hu, Liming Xiao, Jingyuan Ma, Yuhang Qin, Zheng Li, Yixin Yang, Sirui Deng, Jinkun Ding, Wenhan Ma, Rui Li, Weilin Luo, Qun Liu, Zhifang Sui, 2026, **SCoRE: Benchmarking Long-Chain Reasoning in Commonsense Scenarios**, Conference paper, AAAI2026 Bridge: Logical and Symbolic Reasoning in Language, 2026, Singapore.
3. Yuhang Qin, Liming Xiao, Nan Hu, Sirui Deng, Jingyuan Ma, Hyang Cui, Zihan Zhang, Chi Hsu Tsai, Jinkun Ding, Sumin Kang, Zhifang Sui, and Weidong Zhan. 2025. **Overview of CCL25-Eval Task 1: The Fifth Spatial Cognition Evaluation (SpaCE2025)**. In Proceedings of the 24th China National Conference on Computational Linguistics (CCL 2025), pages 33–46, Jinan, China. Chinese Information Processing Society of China.
4. 詹卫东、胡楠、肖力铭、孙春晖, 2025, 基于知识库的空间关系推理数据集自动生成研究, 《语言文字应用》2025年第4期, 16-26页。
5. 詹卫东、秦宇航、肖力铭, 2025, 空间图式与方位词异形同义, 《语言研究与应用》2025年第1辑, 27-44页。
6. 詹卫东、孙春晖、肖力铭, 2024, 语言学知识驱动的空间语义理解能力评测数据集研究, 《语言战略研究》, 2024年第5期, Vol.9, No.5, pp.7-21。
7. 詹卫东、秦宇航、肖力铭, 2024, 基于异形同义现象的机器空间语义理解能力评测研究, 《辞书研究》, 2024年第5期, pp.1-19。
8. 詹卫东、孙春晖、岳朋雪、唐乾桐、秦梓巍, 2022, 空间语义理解能力评测任务设计的新思路——SpaCE2021数据集的研制, 《语言文字应用》2022年第2期 (总第122期), pp.99-110。

谢谢!

zwd@pku.edu.cn

