



# 自然语言处理与认知科学发展

## An introduction to natural language processing and cognitive sciences

徐杨 PhD

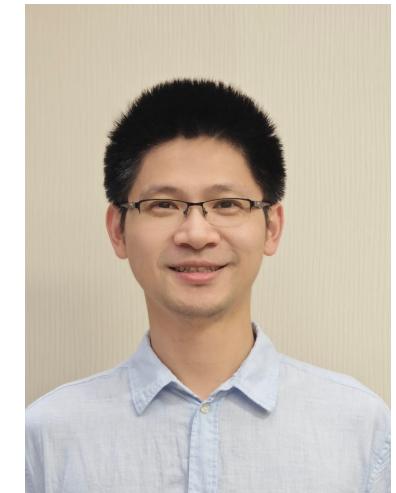
计算机科学与工程系副教授

南方科技大学



# 个人简历

- 徐杨，南方科技大学计算机科学与工程系副教授（研究员）
- 2023年6月加入南科大，正在组建 **计算语言学与意識科学实验室**
- 研究方向：自然语言处理，计算语言学，认知科学，心理语言学
- 2018-2023：任美国圣地亚哥州立大学计算机系助理教授
- 2018：美国宾夕法尼亚州立大学信息科学与技术博士
- 2013：清华大学心理学系硕士
- 2010：清华大学电子工程系学士



# 大纲：自然语言处理 + 认知科学

- 何为NLP?
- NLP技术简史
- 前沿技术 (ChatGPT)
- 未来展望

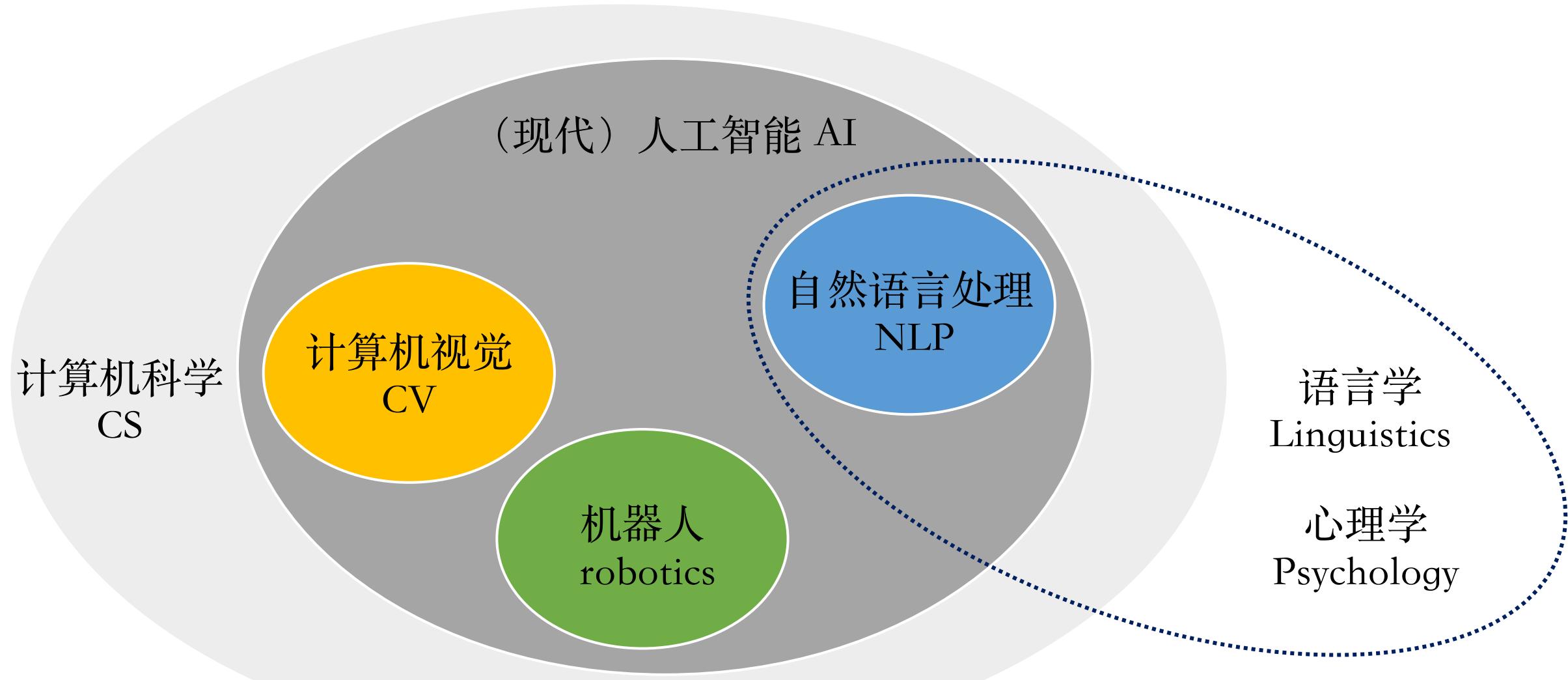


- 认知科学简介
- 与人工智能的联系





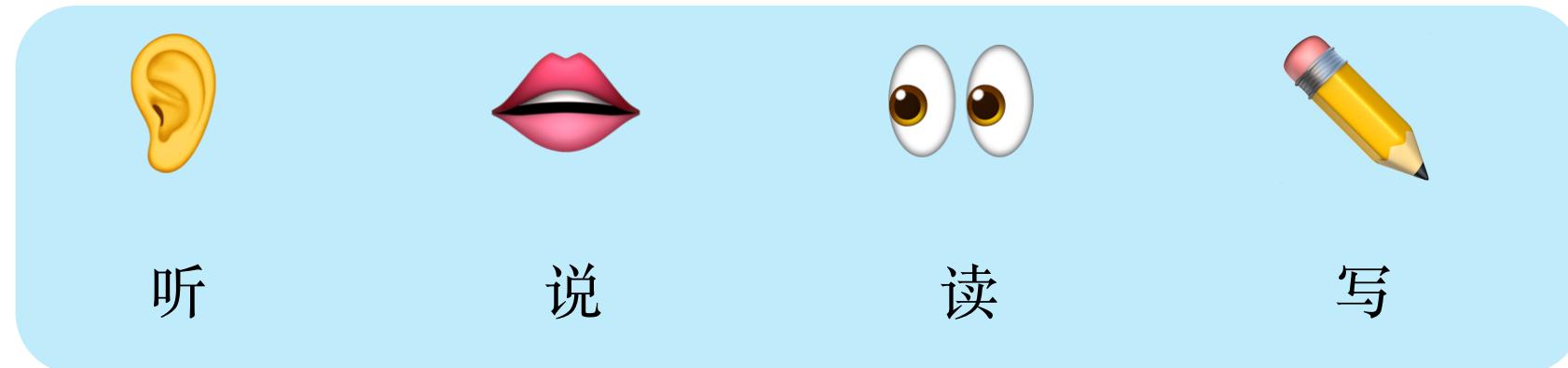
# 何为自然语言处理 Natural Language Processing?





# 何为自然语言处理 Natural Language Processing?

根本问题：让电脑能够像人类般分析，理解和处理自然语言。



“自然” vs. 非“自然”?

# “自然” vs. 非“自然”?

南方科技大学（简称南科大）是深圳在中国高等教育改革发展的宏观背景下，创建的一所高起点、高定位的公办创新型大学。

学而时习之，不亦说乎。有朋自远方来，不亦乐乎。

Southern University of Science and Technology (SUSTech) is a public university, founded in the lush hills of Nanshan District, Shenzhen. It is working towards becoming a world-class university excelling in interdisciplinary research, nurturing innovative talents and delivering new knowledge to the world.

```
#include <stdio.h>
int main() {
    // printf() displays the string inside quotation
    printf("Hello, World!");
    return 0;
}
```

```
org 0x100
mov dx, msg
mov ah, 9
int 0x21
mov ah, 0x4c
int 0x21
msg db 'Hello, World!', 0xd, 0xa, '$'
```

1 吾有一數。曰三。名之曰「甲」。

2 為是「甲」遍。

3 吾有一言。曰「「問天地好在。」」。書之。

4 云云。

Link: <https://ide.wy-lang.org/?file=helloworld%2B>

# “自然” vs. 非“自然”?



NLP的目标：  
近似地“理解”  
和“模拟”

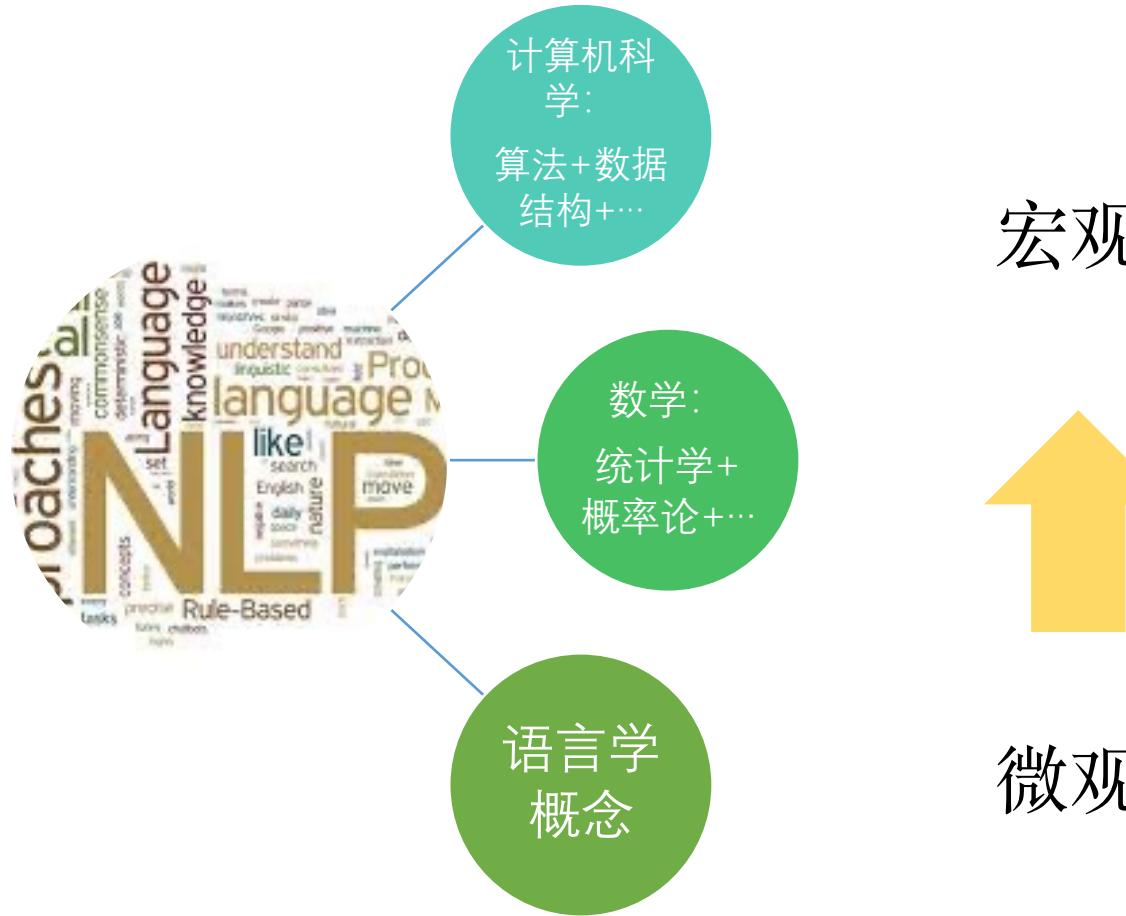


机器语言（非自然）：  
规则明确，有限



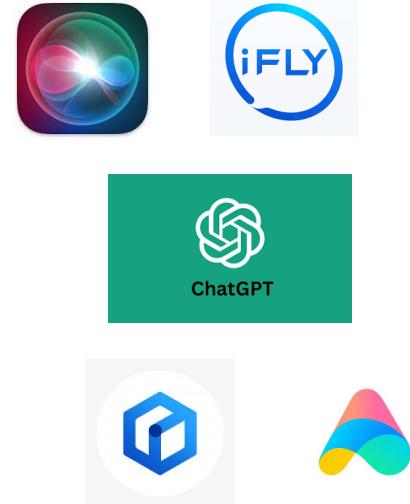
自然语言：  
规则模糊，无限

# 自然语言处理（NLP）技术



## 代表性技术：

- 语音助手
  - 文本信息检索
  - 问答、对话系统
  - 机器翻译……



核心任务：

- 分词、词性
  - 语法、句法
  - 文字蕴含、语意
  - 生成……



# 基础任务-例1：分词

Link: <https://hanlp.hankcs.com/demos/tok.html>

南方科技大学（简称南科大）是深圳在中国高等教育改革发展的宏观背景下，创建的一所高起点、高定位的公办创新型大学。



南方 科技 大学 （ 简称 南科大 ） 是 深圳 在 中国 高等 教育 改革 发展 的 宏观 背景 下 ， 创建 的 一  
所 高 起点 、 高 定位 的 公办 创新型 大学 。



# 基础任务-例2：词性标注

Link: <https://hanlp.hankcs.com/demos/pos.html>

南方科技大学（简称南科大）是深圳在中国高等教育改革发展的宏观背景下，创建的一所高起点、高定位的公办创新型大学。



南方 科技 大学 （ 简称 南科大 ） 是 深圳 在 中国 高等 教育 改革 发展 的 宏观 背景 下 ， 创建 的 一 所  
高 起点 、 高 定位 的 公办 创新型 大学 。

NN → 名词

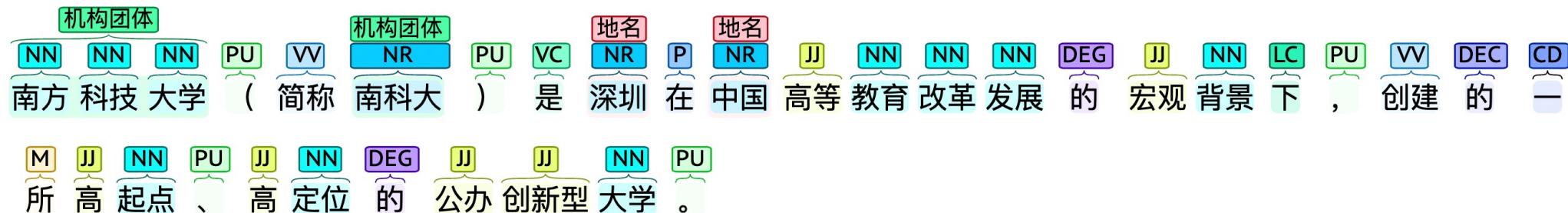
VV, VC → 动词

JJ → 形容词

NR → 专有名词

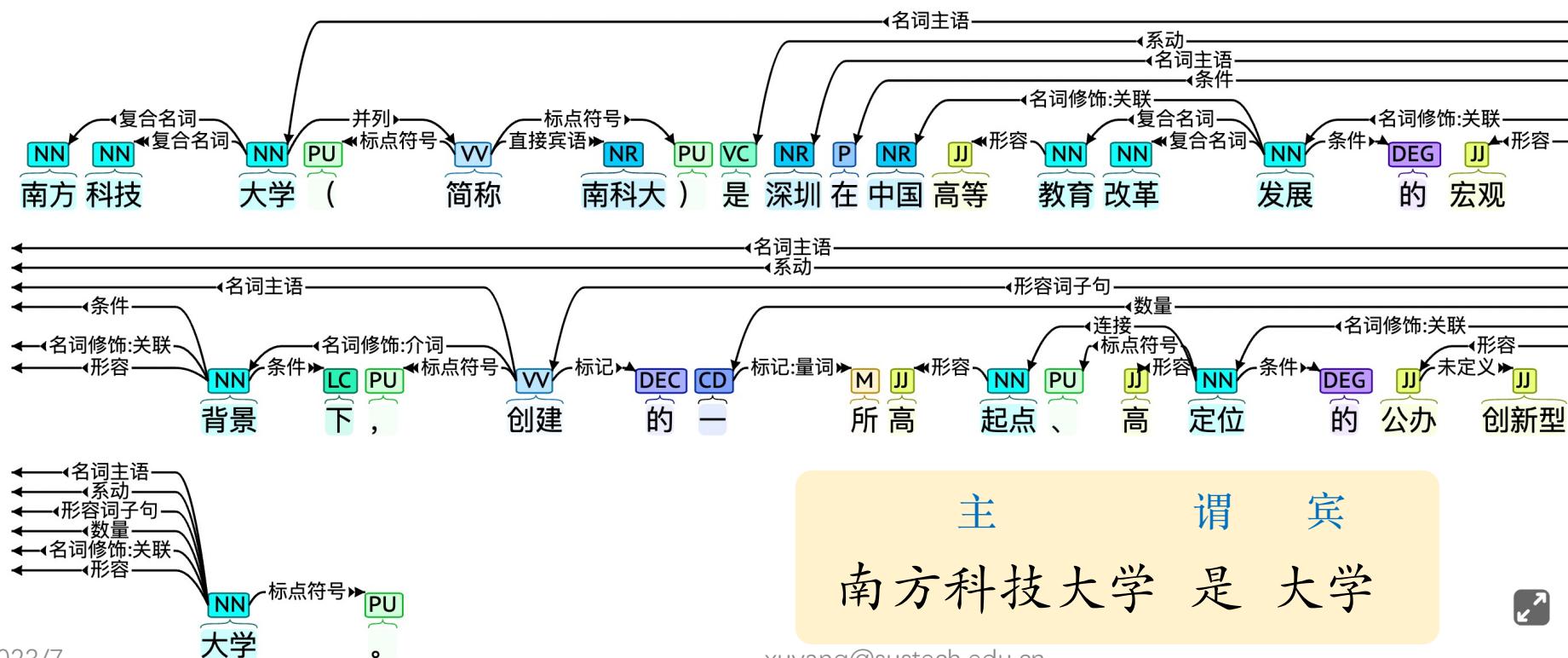
# 基础任务-例3：命名实体识别

南方科技大学（简称南科大）是深圳在中国高等教育改革发展的宏观背景下，创建的一所高起点、高定位的公办创新型大学。



# 基础任务-例4：句法分析

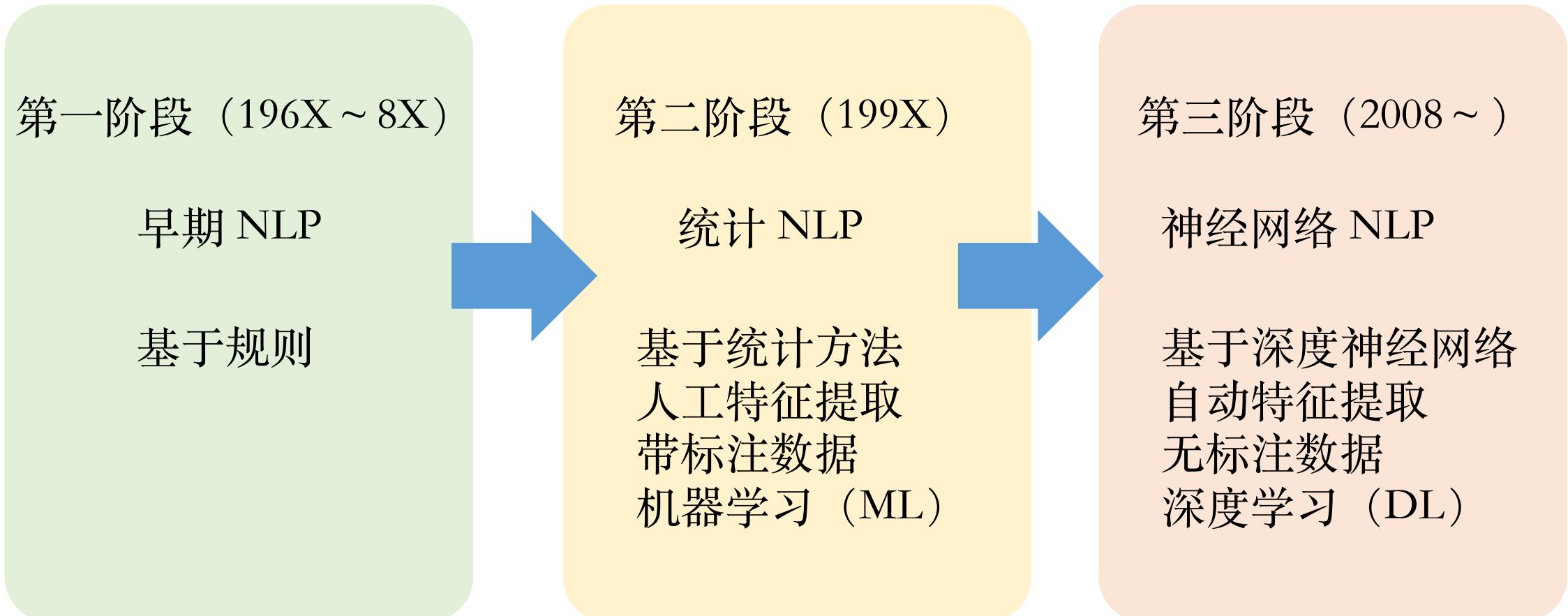
南方科技大学（简称南科大）是深圳在中国高等教育改革发展的宏观背景下，创建的一所高起点、高定位的公办创新型大学。



主              谓              宾  
 南方科技大学 是 大学



# NLP 技术思路简史



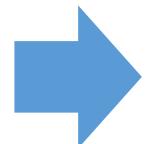
# NLP 技术简史：案例

- A: 你们家千金多大啦?
- B: 千金散尽还复来
- C: 千金难买少年穷

问题：哪两句的“千金”意思更近？

你们 家 千金 多大 啦?  
千金 散尽 还 复 来  
千金 难 买 少年 穷

早期 NLP



统计 NLP



神经网络 NLP

# NLP 技术案例：早期

规则 1：

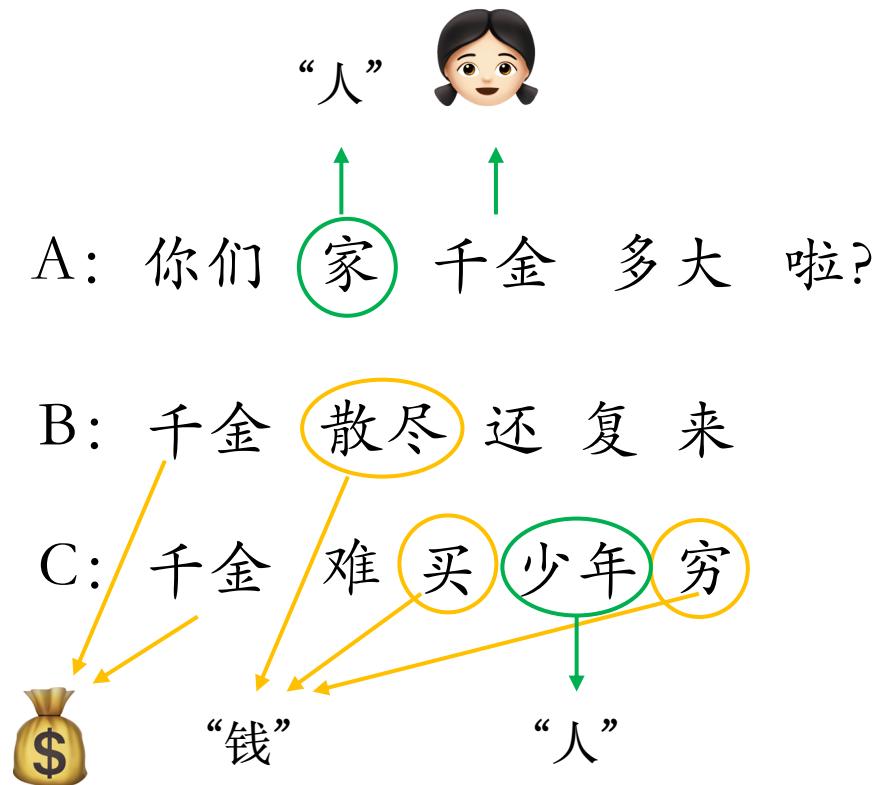
如果 句子中出现和“人”有关的词 则 认为“千金”代表 

规则 2：

如果 句子中出现和“钱”有关的词 则 认为“千金”代表 

早期 NLP

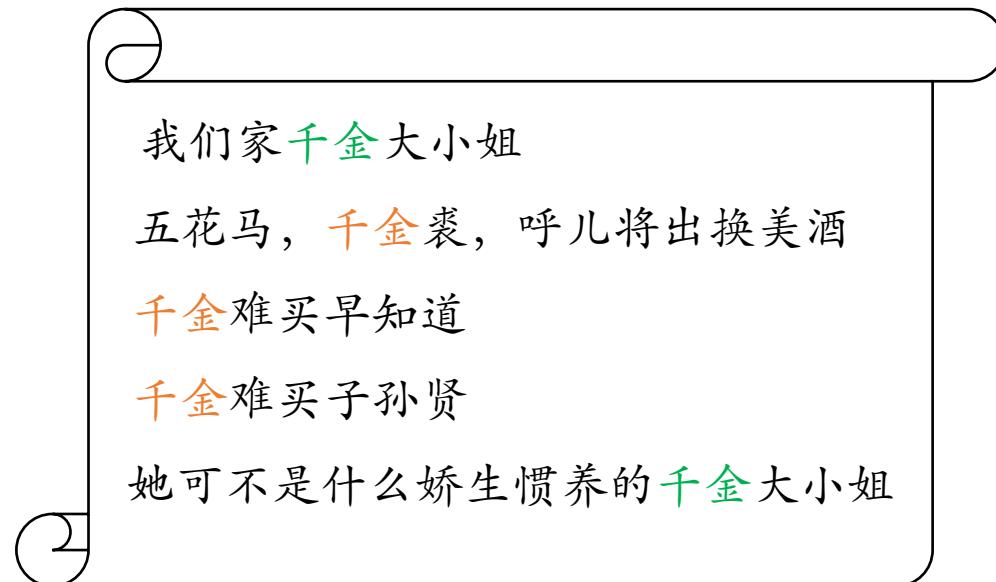
基于规则  
2023/7



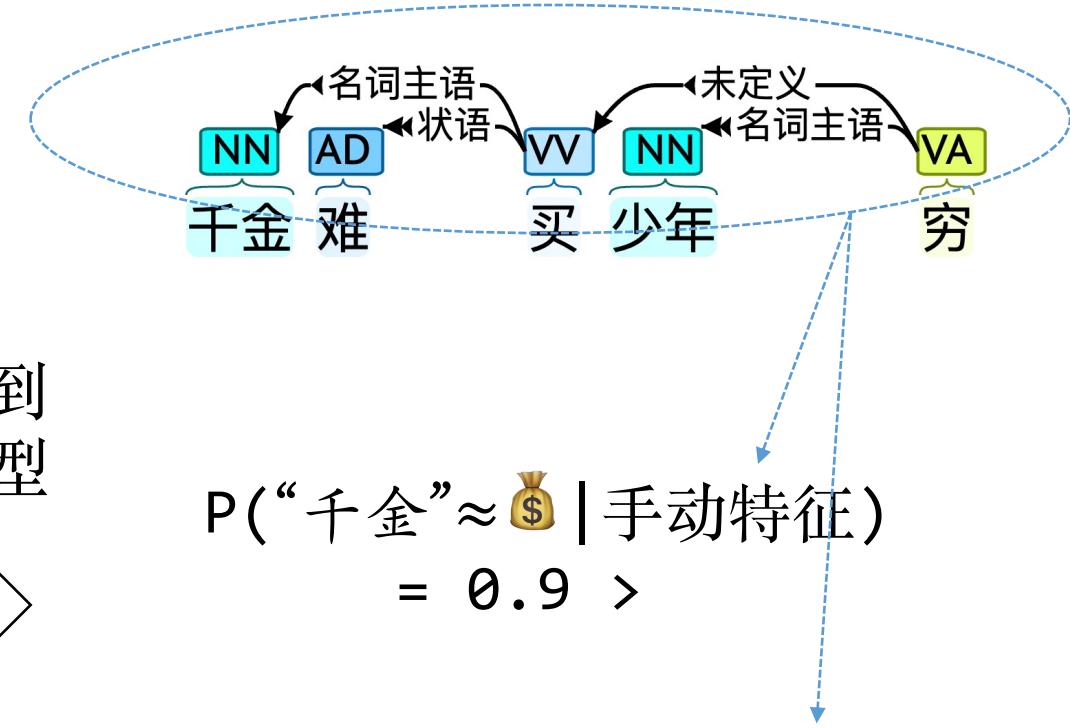
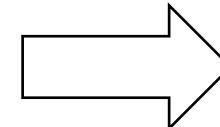
 or  ?

# NLP 技术案例：统计

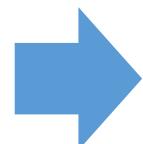
数据集：包含“千金”的 1000 个句子



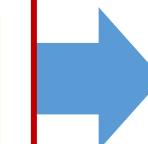
学习得到  
概率模型



早期 NLP

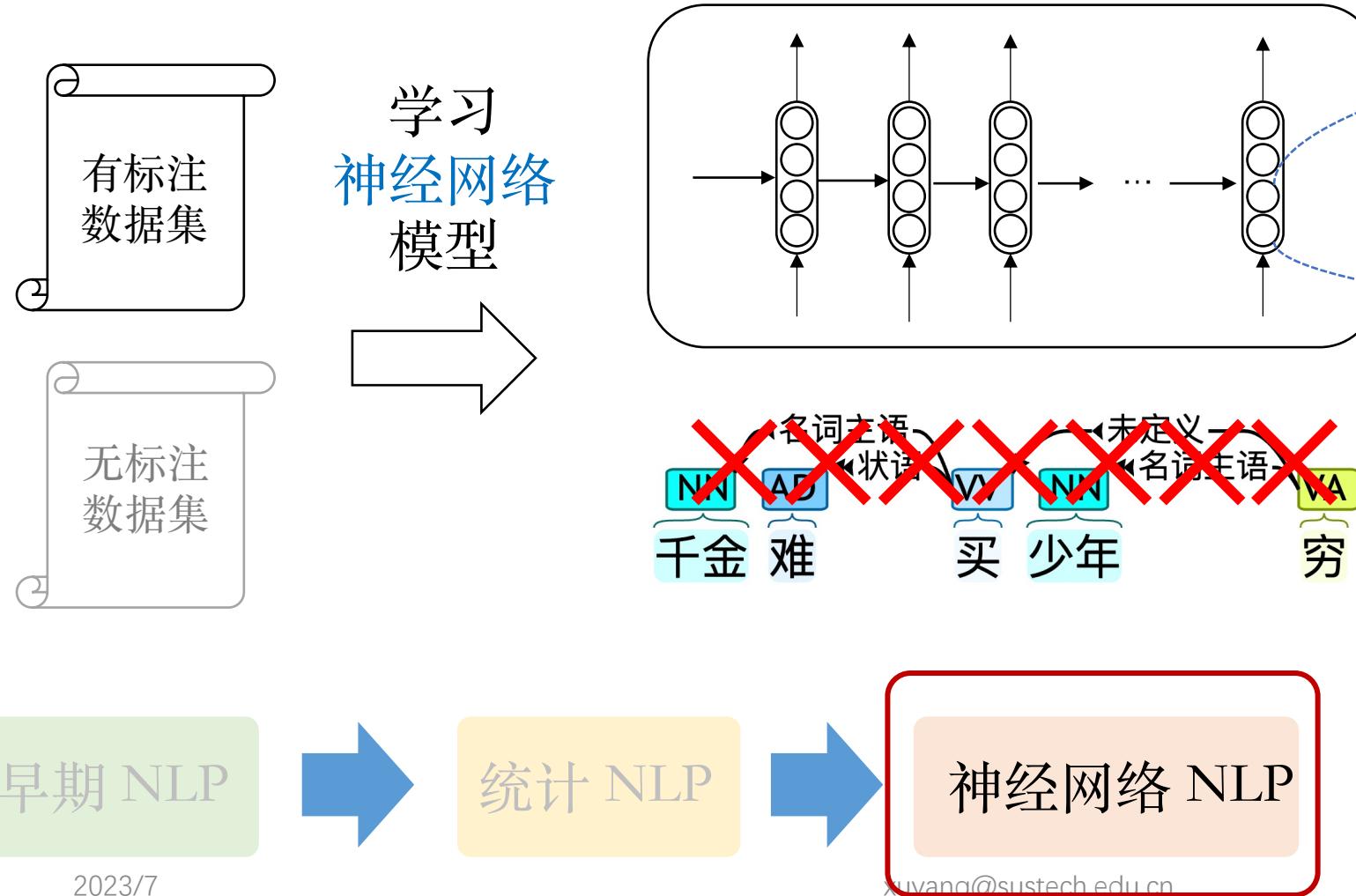


统计 NLP



神经网络 NLP

# NLP 技术案例：神经网络



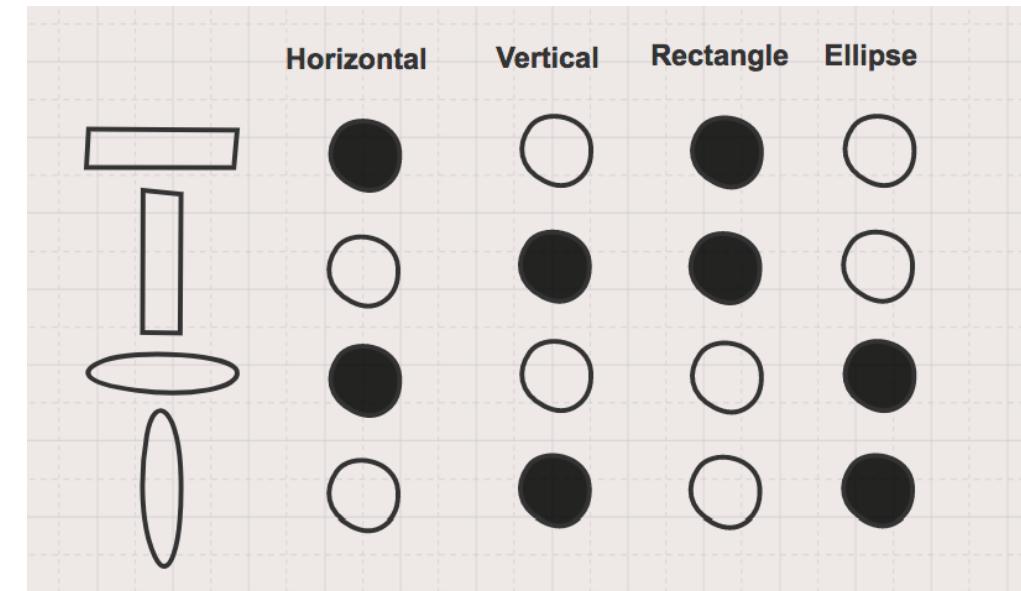
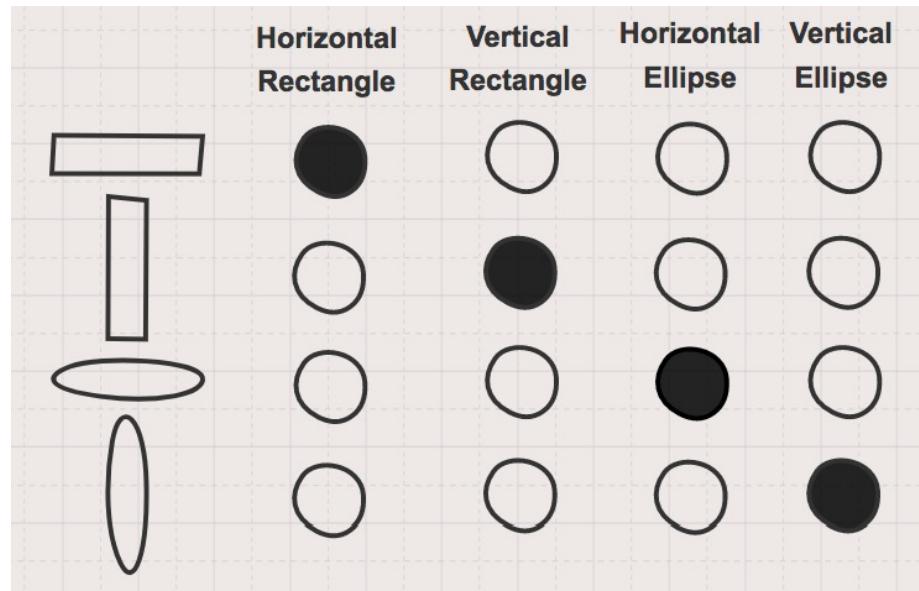
$$P(\text{“千金”} \approx \$ | \text{自动特征}) = 0.95 >$$

$$P(\text{“千金”} \approx \text{小女孩} | \text{自动特征}) = 0.05$$

# 为什么神经网络更有效？

分布式表征：万物皆可向量化！(包括语言)

Distributed representation



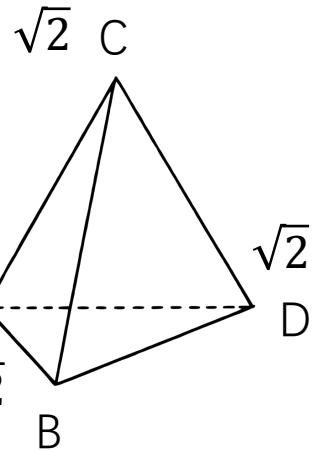
# 非分布式表征

特征1: 横向矩形    特征2: 纵向矩形    特征3: 横向椭圆    特征4: 纵向椭圆

物体A  
物体B  
物体C  
物体D

	Horizontal Rectangle	Vertical Rectangle	Horizontal Ellipse	Vertical Ellipse
物体A				
物体B				
物体C				
物体D				

等距  
不可区分



物体A → (1, 0, 0, 0) ] ] ] ]  
物体B → (0, 1, 0, 0) ] ] ] ]  
物体C → (0, 0, 1, 0) ] ] ] ]  
物体D → (0, 0, 0, 1) ] ] ] ]

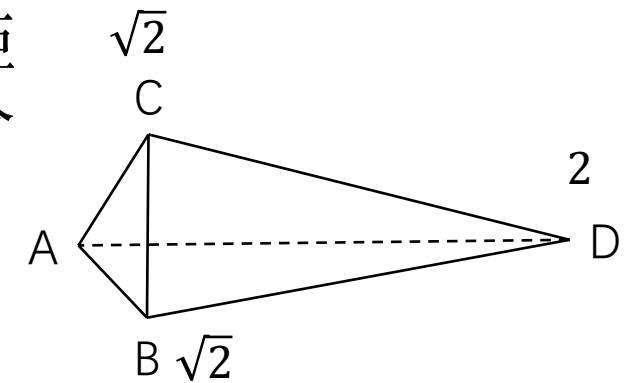
# 分布式表征

特征1: 横      特征2: 纵      特征3: 方      特征4: 圆

物体A  
物体B  
物体C  
物体D

	Horizontal	Vertical	Rectangle	Ellipse
物体A				
物体B				
物体C				
物体D				

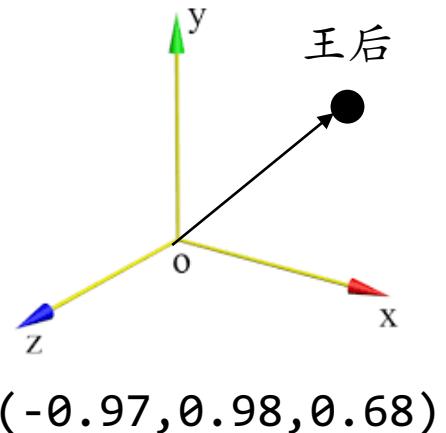
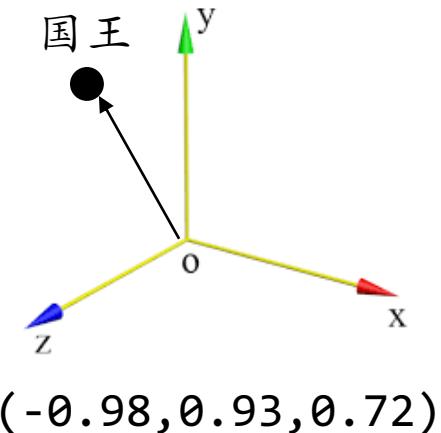
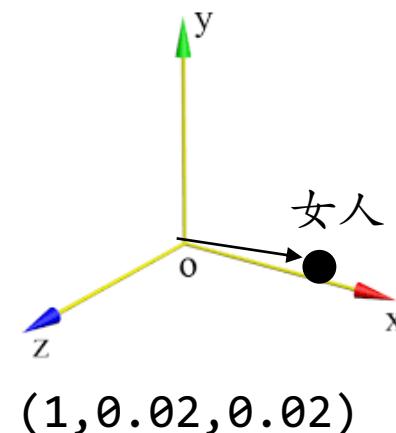
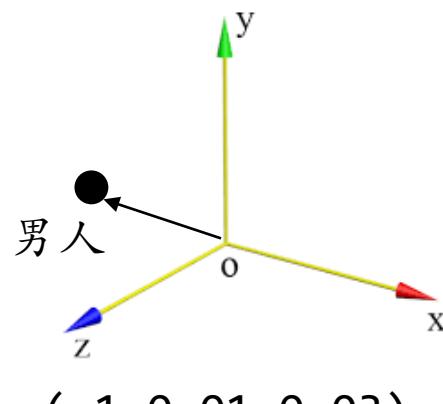
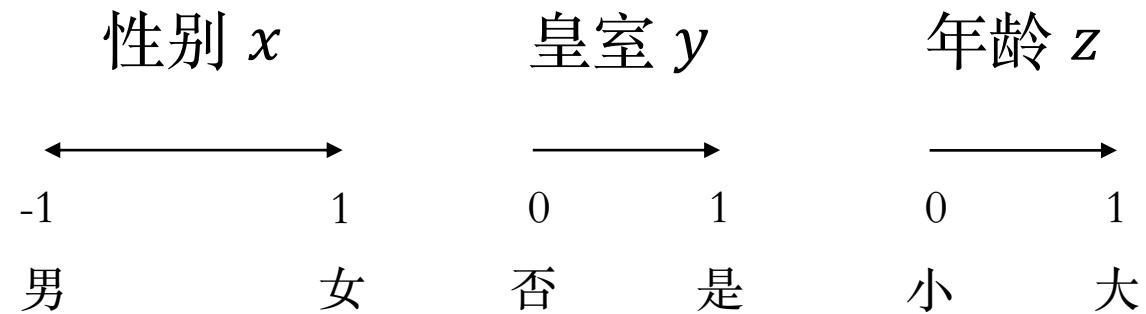
不等距  
可区分



物体A → (1, 0, 1, 0) }  
物体B → (0, 1, 1, 0) }  
物体C → (1, 0, 0, 1) }  
物体D → (0, 1, 0, 1) }

# 词语 $\Rightarrow$ 分布式向量

	性别	皇室	年龄	...
男人	-1	0.01	0.03	...
女人	1	0.02	0.02	...
国王	-0.98	0.93	0.72	...
王后	0.97	0.98	0.68	...



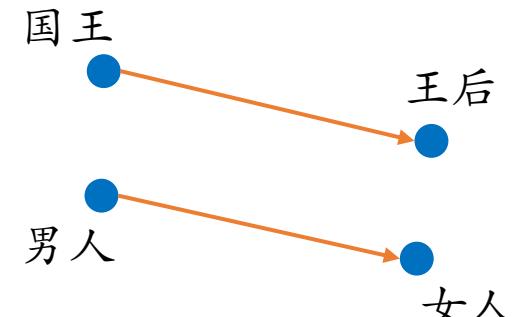
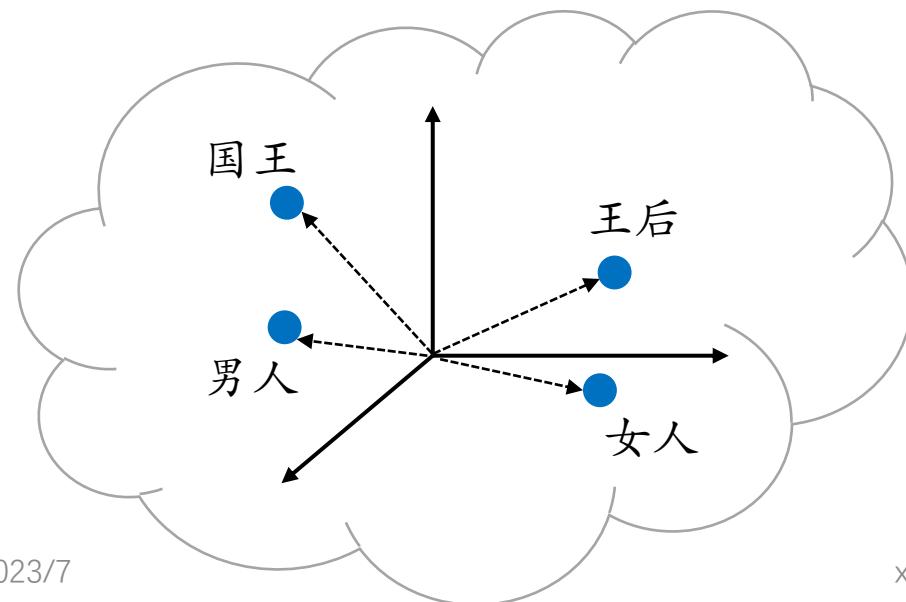
# 分布式向量 $\Rightarrow$ 高维语义空间

	$d_1$	$d_2$	$d_3$	$\cdots$	$d_{100}$	$\cdots$
男人	-1	0.01	0.03	$\cdots$		$\cdots$
女人	1	0.02	0.02	$\cdots$		$\cdots$
国王	-0.98	0.93	0.72	$\cdots$		$\cdots$
王后	0.97	0.98	0.68	$\cdots$		$\cdots$

高维语义空间:  $\{d_1, d_2, \dots, d_N\}$  ( $N > 100$ )

单个维度无明确含义

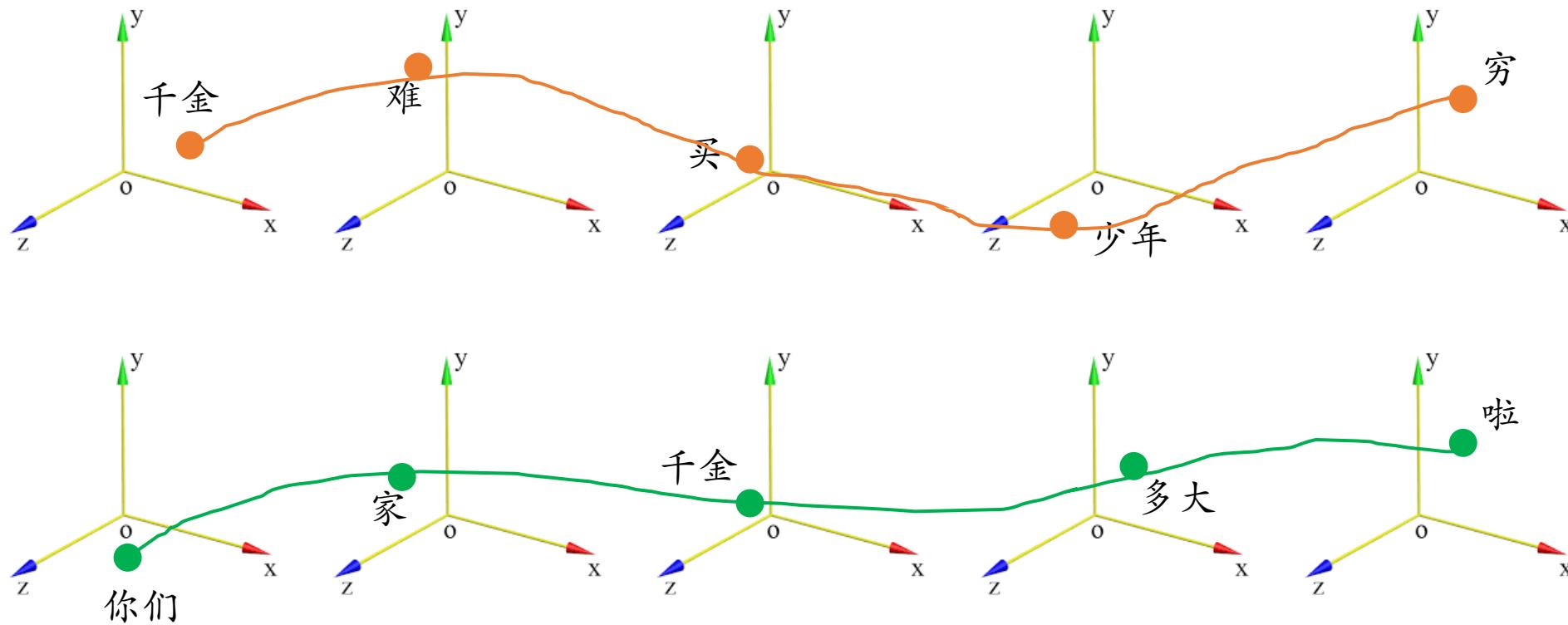
空间中的位置代表了语义



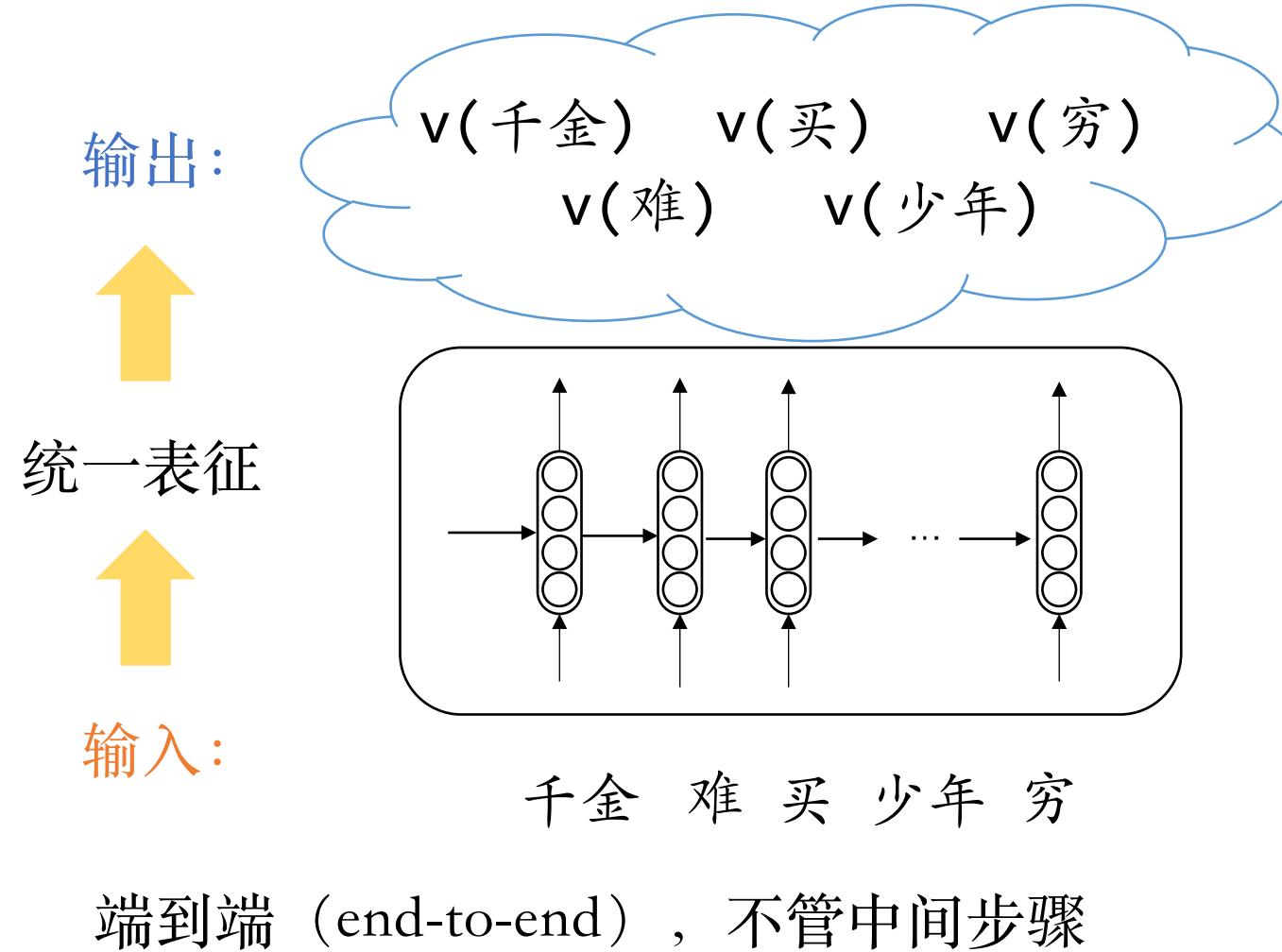
$$\begin{aligned} d &= v(\text{国王}) - v(\text{男人}) \\ v(\text{女人}) + d &\approx v(\text{王后}) \end{aligned}$$

# 句子 $\Rightarrow$ 向量化

把词语的向量串起来，句子变成了语义空间中的轨迹



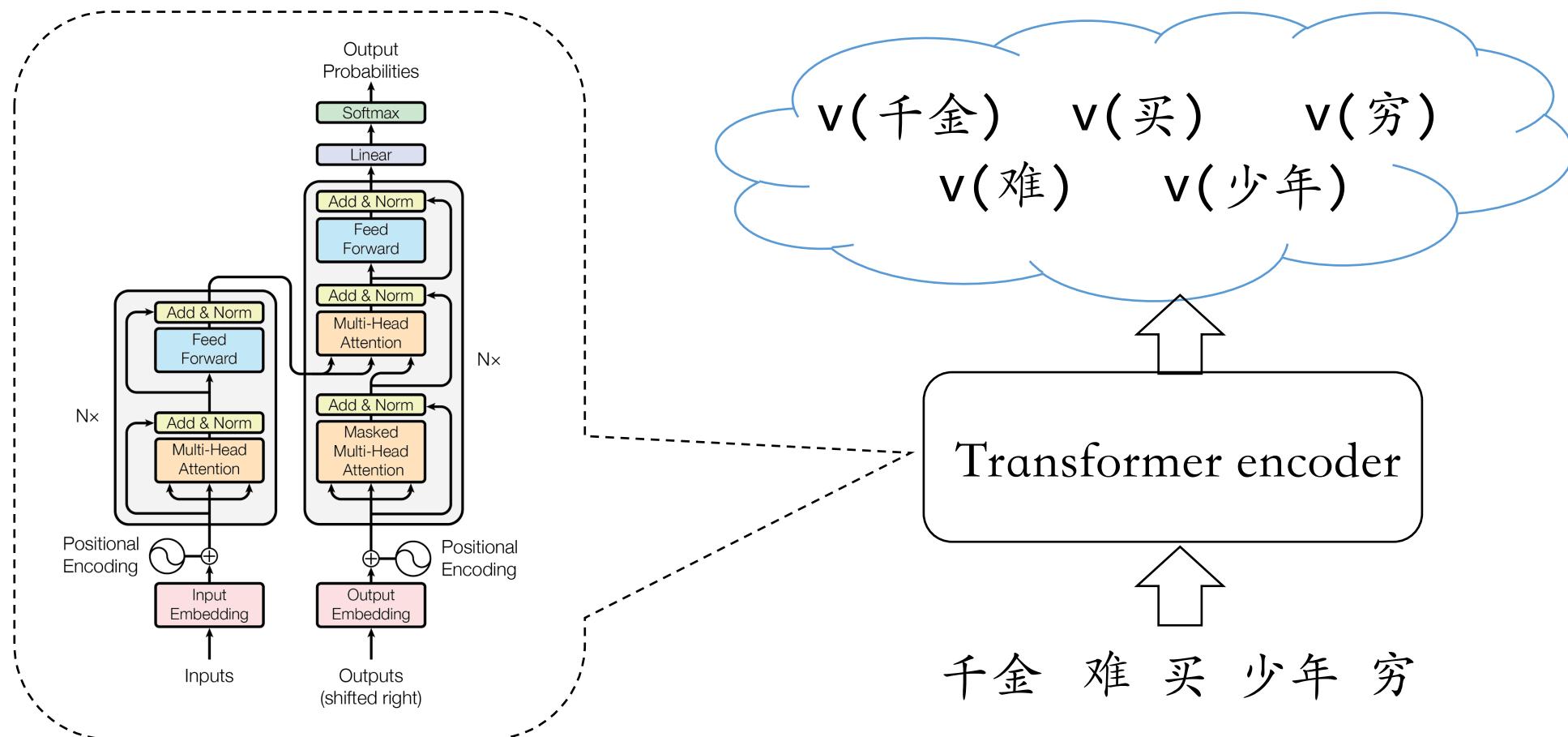
# 输入向量 $\Rightarrow$ 神经网络 $\Rightarrow$ 输出向量



包含了有关“千金”等词语的全部语义信息

问题：用什么构架的神经网络？

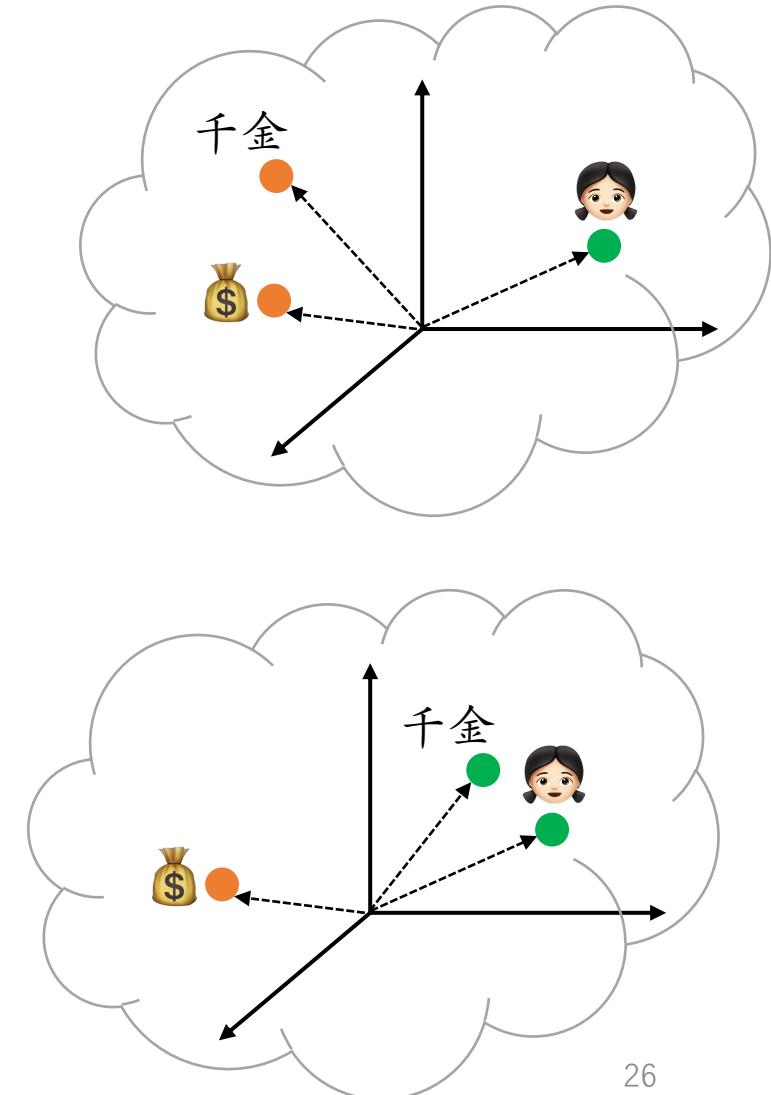
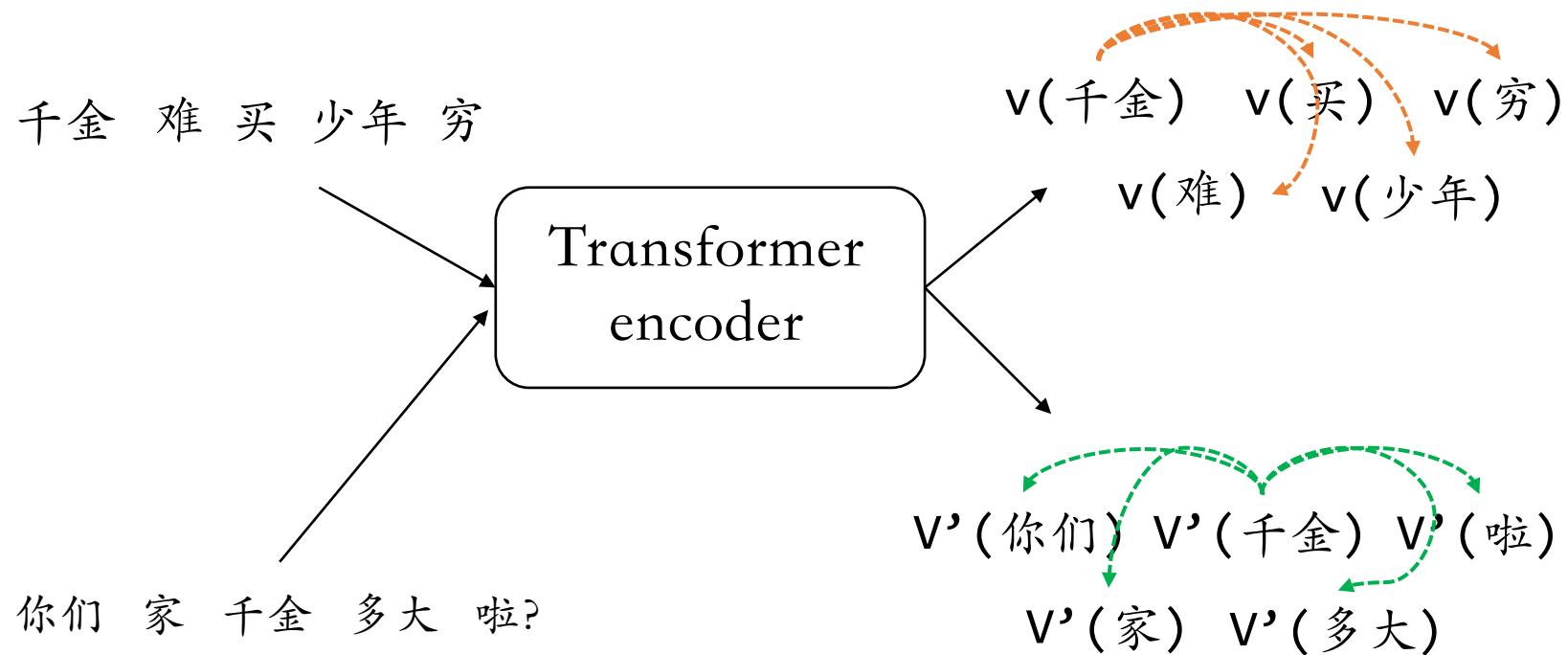
# 神经网络构架的突破：Transformer



Ref: Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

# Transformer 的核心：注意力机制

- “注意力”：一种更好的“向量化”算法





# 前沿NLP技术：ChatGPT/GPT-4

- 技术：Transformer构架 + 大量数据训练 + 人工反馈（强化学习）
- 优势：具有优良的
  - ✓ 交互能力：与物、与人
  - ✓ 视、听觉能力
  - ✓ 数学推理能力
  - ✓ 编程能力
- GPT-4 是 ChatGPT 的进化版本

# GPT-4 案例：多模态交互能力（视觉 + 语言）

## GPT-4 visual input example, Extreme Ironing:

User      What is unusual about this image?



Source: <https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg>

GPT-4      The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

用户提问：这幅图有什么特别之处？

GPT-4回答：这张照片的不同寻常之处在于，一名男子正在一辆移动出租车车顶上的熨衣板上熨烫衣服。

Link: <https://arxiv.org/abs/2303.08774>



# GPT-4 案例：数学推理

**Prompt:** Within the duration of one year, a rabbit population first multiplies itself by a factor  $a$  and on the last day of the year  $b$  rabbits are taken by humans for adoption. Supposing that on first day of the first year there are  $x$  rabbits, we know that exactly 3 years afterwards there will be a population of  $27x - 26$  rabbits. What are the values of  $a$  and  $b$ ?

求解兔子繁殖问题

## GPT-4:

Let's call the rabbit population after one year  $y$ . We know that  $y = ax - b$ . After two years, the population will be  $ay - b = a(ax - b) - b = a^2x - ab - b$ . After three years, the population will be  $a(a^2x - ab - b) - b = a^3x - a^2b - ab - b$ . We are given that this is equal to  $27x - 26$ , so we have the equation:

$$a^3x - a^2b - ab - b = 27x - 26$$

We can rewrite this equation as:

$$a^3x - 27x = a^2b + ab + b - 26$$

Now we can factor out an  $x$  on the left side and a  $b$  on the right side:

$$x(a^3 - 27) = b(a^2 + a + 1 - 26/b)$$

We can see that  $a = 3$  is a solution for the left side of the equation, so we can substitute that in:

$$x(0) = b(9 + 3 + 1 - 26/b)$$

This simplifies to:

$$0 = 13b - 26$$

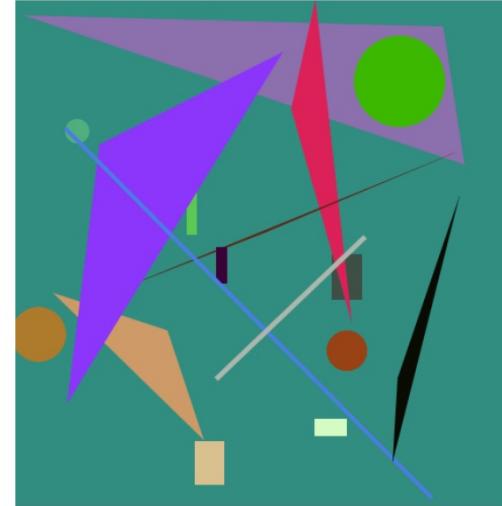
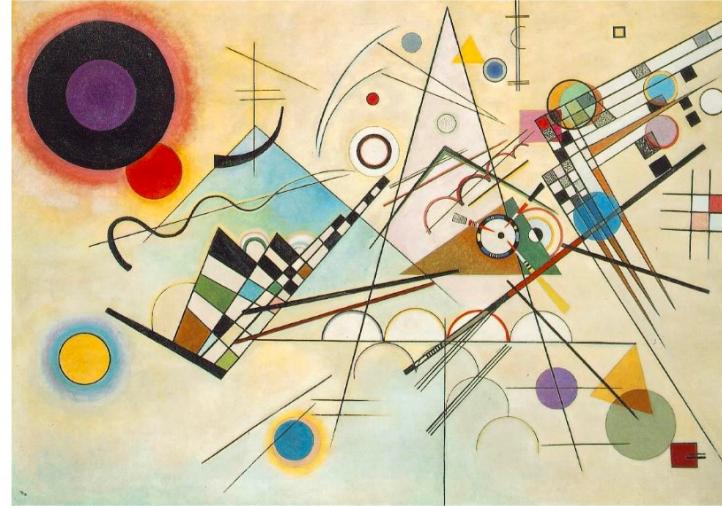
Solving for  $b$ , we find that  $b = 2$ . So  $a = 3$  and  $b = 2$  are the values we are looking for.

列出方程，逐步求解  
(二元三次方程特殊解)

Link: <https://arxiv.org/abs/2303.12712>

# GPT-4 案例：代码生成

Link: <https://arxiv.org/abs/2303.12712>



左：画家 Wassily Kandinsky 的作品

中：GPT-4 生成的绘图代码生成的作品

右：ChatGPT 生成的绘图代码生成的作品

提示词：Produce JavaScript code that creates a random graphical image that looks like a painting of Kandinsky

(写一段JavaScript 代码，用来画一张看起来像Kandinsky的作品的随机图形图像)

# GPT-4 案例：推测意图

Link: <https://arxiv.org/abs/2303.12712>

## GPT-4 and ChatGPT: Realistic Scenario 1

Conversation:

Mark: I don't like how you treated Jack last night

Judy: Did you see what he was doing? He hit his brother on the head!

Mark: That doesn't justify you yelling at him

Judy: What do you want me to do? Let him beat his brother up and say nothing?

Mark: No, I don't

Judy: Then why are you defending him?

--

Q: What does Mark's intent seem to be?

## GPT-4:

Mark's intent seems to be to express his displeasure with Judy's way of handling the situation with Jack, and to suggest that she could have been more calm and respectful.

Mark 的意图似乎是表达他对 Judy 处理 Jack 问题的方式的不满，并建议她可以更冷静和尊重。

M: 我不喜欢你昨晚对待Jack的方式

J: 你看到他做了什么吗？他打他弟弟的脑袋！

M: 你不能因为这样就吼他呀

J: 那你想我怎么做？让他继续打，也不管吗？

M: 不，我不是这意思

J: 那你怎么还替他说话？

提问：M的意图是什么？



# GPT-4 案例：推测意图（续）

Link: <https://arxiv.org/abs/2303.12712>

提问：M的意图是什么？

**ChatGPT:**

Based on the conversation, Mark's intent seems to be to defend Jack's behavior and to express his disapproval of Judy's treatment of Jack.

根据谈话内容，Mark 的意图似乎是为 Jack 的行为辩护并表达他的不赞成 Judy 对待 Jack 的方式。

提问：J 对 M 的回答是什么反应？

**GPT-4:**

Judy received his comment as an attack on her parenting skills and a sign of sympathy for Jack's misbehavior. She reacted defensively and angrily, and questioned Mark's motives.

Judy 认为 Mark 的评论是对她育儿技巧的攻击，也是对 Jack 不当行为的同情。她的反应是防御性和愤怒的，并质疑马克的动机。

**ChatGPT:**

It is not possible for me to accurately determine how Judy received Mark's comment without more context. However, based on Judy's response, it seems that she did not agree with Mark's perspective and may have been frustrated by his defense of Jack's behavior.

如果没有更多的背景信息，我不可能准确地确定 Judy 是如何对 Mark 的评论做出反应的。

# GPT-4 案例：音乐生成

Link: <https://arxiv.org/abs/2303.12712>

GPT-4

Can you compose a short tune (say four to eight bars) using ABC notation?

There are many possible ways to compose a short tune in abc notation, but here is one example:

X:1

T:Example Tune

M:4/4

L:1/8

K:C

| :G2AB c2BA | G2AB c2c2 | edcB c2BA | G4-G2z2 :|  
| :c2de f2ed | c2de f2f2 | g2fe d2cB | c4-c2z2 :|

提示词：你可以用ABC音符（哆来咪）写一段曲子吗？



A musical score for two staves. The top staff is in treble clef and 4/4 time, featuring a melody of eighth and sixteenth notes. The bottom staff is in bass clef and 4/4 time, providing harmonic support with sustained notes and eighth-note chords.



A continuation of the musical score from the previous staff, showing the progression of the melody and harmonic structure over the next set of measures.

# GPT-4/ChatGPT 的局限

- 生成的过程缺少计划 (planning)
- 推理缺少中间步骤
- 有时会“一本正经地胡说八道” (ChatGPT)

参考文献：Sparks of Artificial General Intelligence: Early experiments with GPT-4 (<https://arxiv.org/abs/2303.12712>)



# GPT-4 的局限：案例

Link: <https://arxiv.org/abs/2303.12712>

- GPT-4

$$2 * 8 + 7 * 6 = 58$$

$$7 * 4 + 8 * 8 = 88$$

用 100 个随机样本测试模型，四个数字在 0 到 9 之间均匀生成，结果只有 58% 的准确率。

What is the value of the following expression?  $116 * 114 + 178 * 157 = ?$

- Let's think step by step to solve the expression, write down all the intermediate steps, and only then produce the final solution.

换一种方式提问，准确率上升！



# 未来 NLP 挑战

- 基于 Transformer 的模型本身的局限
- 中文数据集建设：
- <https://github.com/esbatmop/MNBVC>  
(目前总数据量 5039GB，目标是达到 chatGPT 的 40T 数据，目前进度 12.6%)



# 计算机教育的挑战：Code is cheap?

- 过去：*talk* is cheap, show me the *code*
- 现在：*code* is cheap?
- 代码（code）的一大特点——复用（重复使用）
- 比自然语言更容易预测、生成
- 需不需要培养学生的代码能力？
- 设计 design => 实现 implementation

# Part 2: 认知科学 (cognitive sciences)

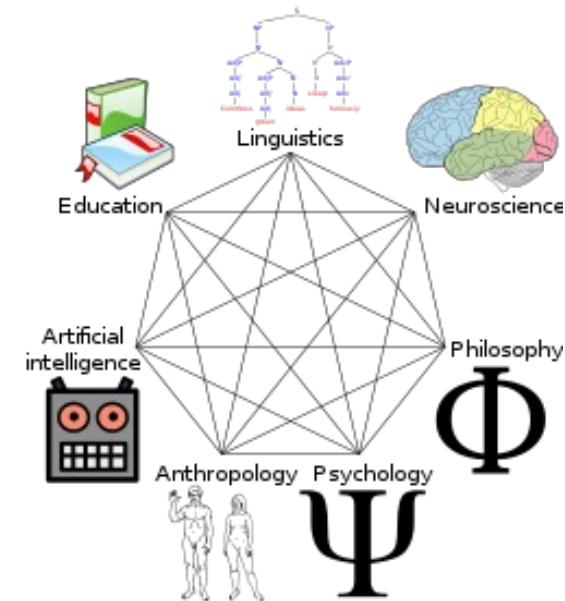
- 认知科学简介
- 与人工智能的联系



研究问题:

- 注意, 语言, 记忆
- 知觉, 行为
- 心智, 意识, .....

跨学科:

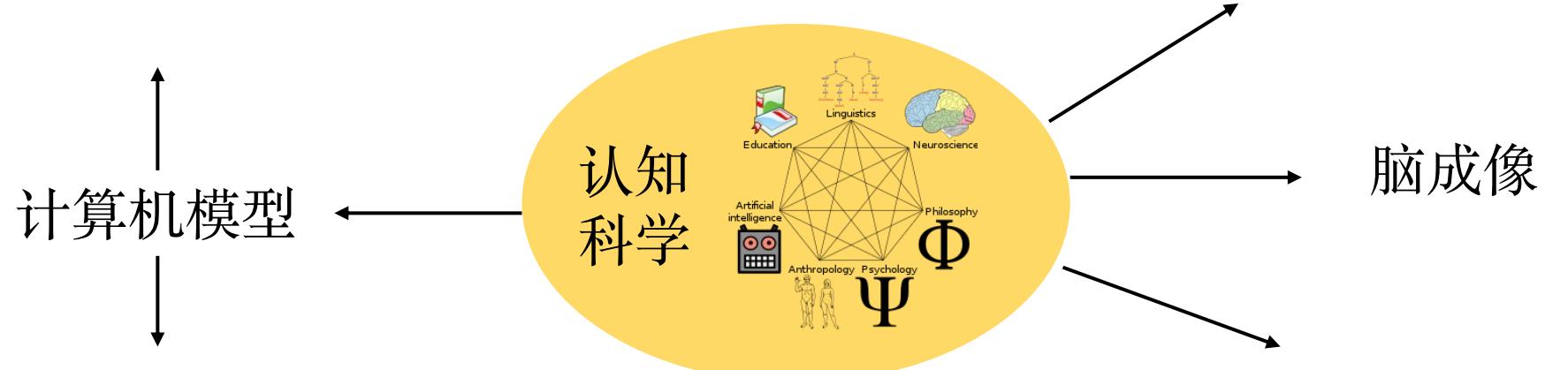


# 认知科学的研究方法

- 百花齐放，莫衷一是，有迹可循

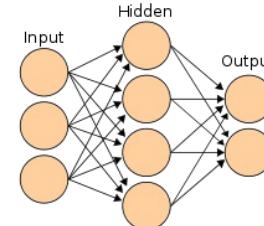
符号主义：

认知模型、逻辑主义、专家系统

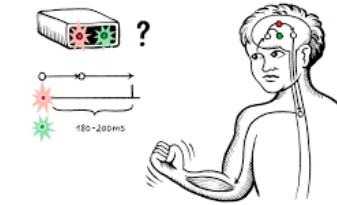


联结主义：

分布式表征、神经网络（现代人工智能的基石）



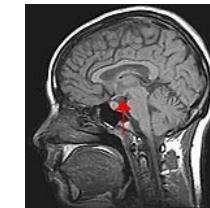
反应时



眼动



行为实验

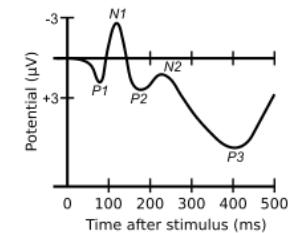
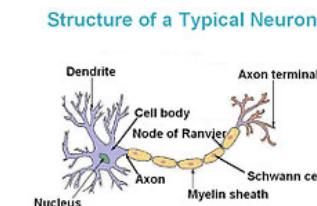


fMRI, PET, EEG

脑成像

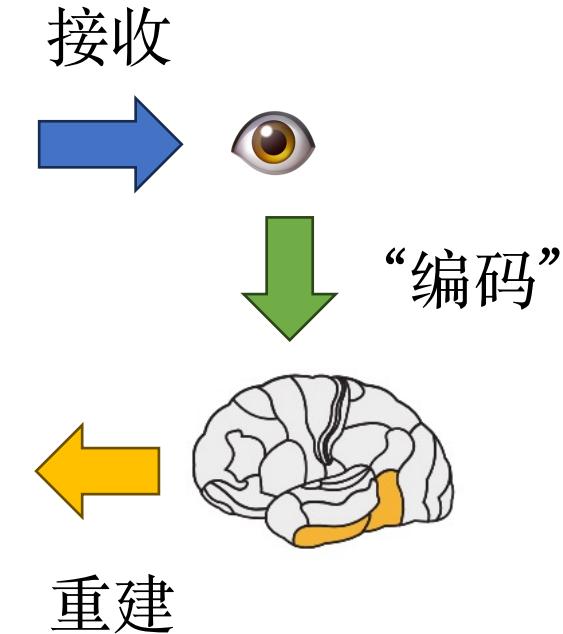
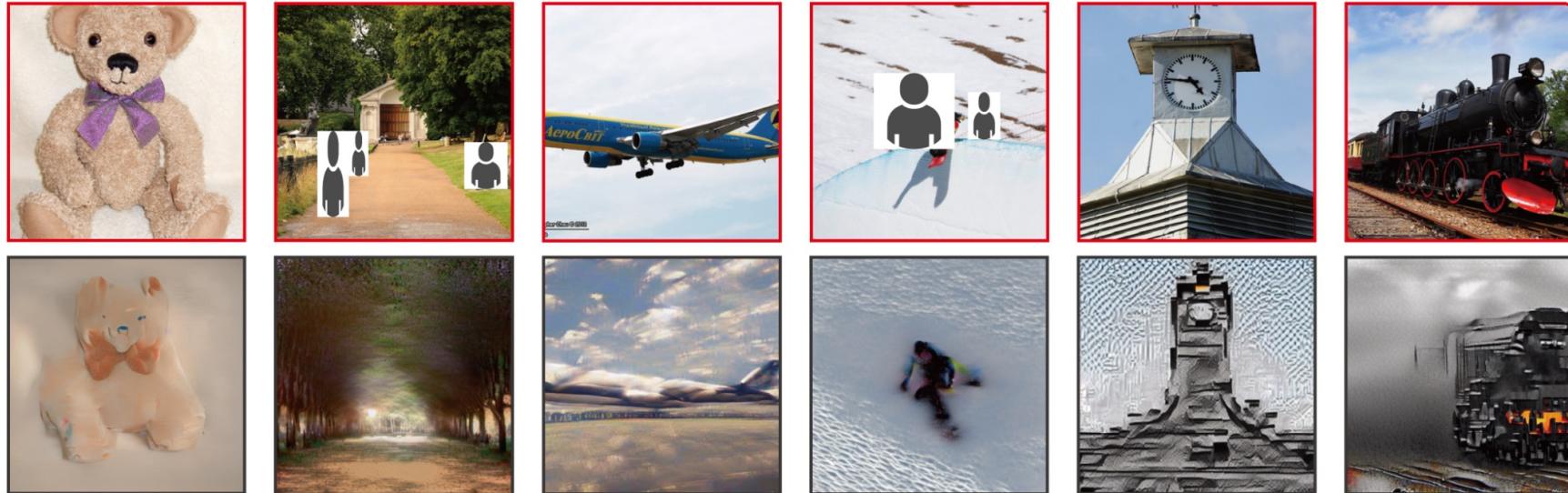


神经生物学



# 现代人工智能中的认知科学

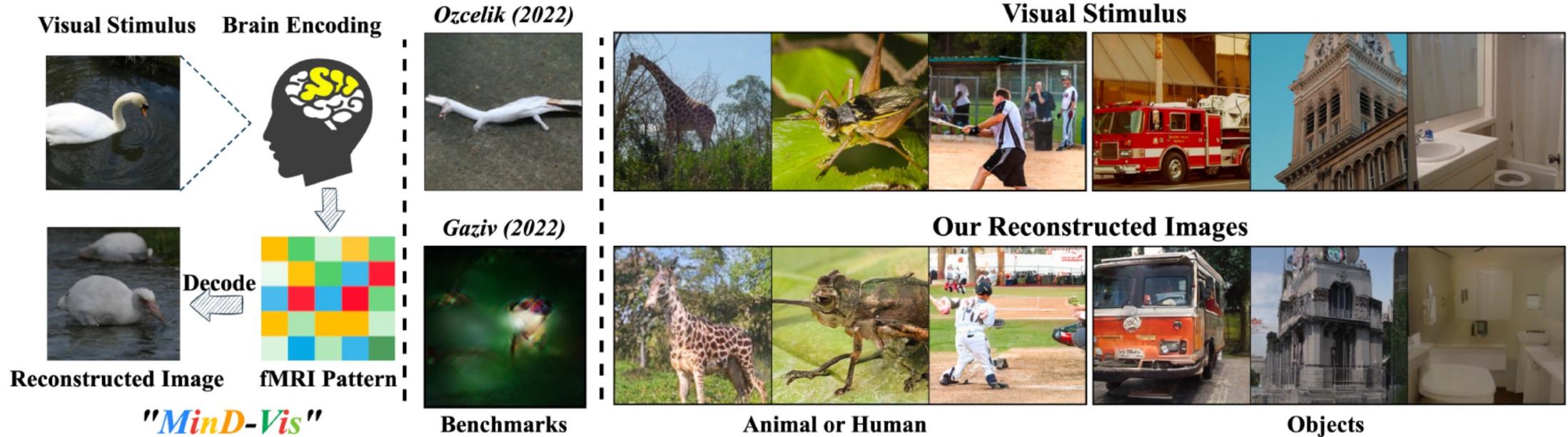
- AI “读脑术”: 从脑信号重建看到的画面



Link: High-resolution image reconstruction with latent diffusion models from human brain activity  
(<https://www.biorxiv.org/content/10.1101/2022.11.18.517004v3.full.pdf> )

# 现代人工智能中的认知科学

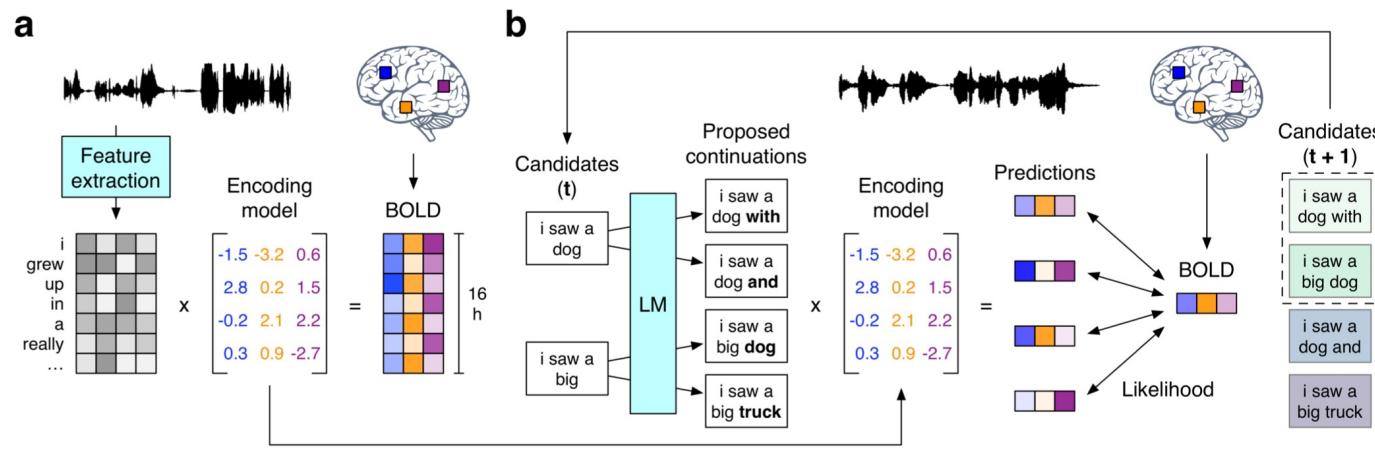
- AI “读脑术”：从脑信号重建看到的画面



Link: High-resolution image reconstruction with latent diffusion models from human brain activity  
<https://www.biorxiv.org/content/10.1101/2022.11.18.517004v3.full.pdf>

# 现代人工智能中的认知科学

- AI “读脑术”：从脑信号重建听到的语句



Link: Semantic reconstruction of continuous language from non-invasive brain recordings  
(<https://www.nature.com/articles/s41593-023-01304-9>)

Exact

Gist

Error

# 未来教育展望

- 大脑  $\Leftrightarrow$  神经网络模型，读书  $\Leftrightarrow$  训练模型：
  - 向语言模型学习  $\Rightarrow$  从原始数据中学习，不加评断  
(朱子读书法：随本文正义，虚心以求其义，不执己见)
  - 学习过程的负反馈（不适感） $\Rightarrow$  更新模型参数  
(有疑则不容不立论)
  - 人类熟悉的思维  $\Rightarrow$  跳出某个局部最优解  
(宁拙毋巧，宁下毋高)
- 不要低估学生获取信息的能力：翻开过的书  $\Leftrightarrow$  没翻开过的书
- 人类要不要训练自己底层的基础能力？到什么程度？