

2022年全球人工智能产业研究报告

2022 Global Artificial Intelligence Industry Research Report

尚普研究院

SHANGPU Institute



北京：海淀区北四环中路229号海泰大厦1118、20、27、29室 Tel: 15033256189

上海：黄浦区河南南路33号新上海城市广场1809室 Tel: 13611322396

深圳：福田区金田路3038号现代国际大厦1栋11层1101B室 Tel: 18148563314

CAAI 中国人工智能学会会员单位

报告参与机构

尚普研究院



孙硕, FRM, FMVA 研究总监

电话:18611624278 (微信同号)

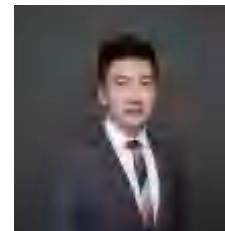
电邮:sunshuo@shangpu-china.com



宿硕 研究经理

电话:13717962127 (微信同号)

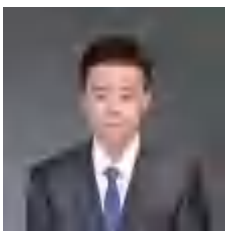
电邮:sushuo@shangpu-china.com



刘宇 高级分析师

电话:17694801358 (微信同号)

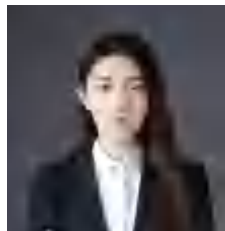
电邮:liuyu@shangpu-china.com



张祎 高级分析师

电话:13701067098 (微信同号)

电邮:zhangyi@shangpu-china.com



周文青 商务经理

电话:15033256189 (微信同号)

电邮:zhouwq@shangpu-china.com

合作机构



中科创星: 创始合伙人/硬科技概念提出者 米磊



北京智谱华章科技有限公司: 研究咨询部负责人 邬克、研究员 张淼



紫荆数字经济研究院: 院长 付喆



中国风险投资有限公司: 投资二部总经理 谷柏



中电海康科创基金: 副总经理 魏强



容亿投资: 创始合伙人 刘宏春、合伙人 崔骅

特别鸣谢

高合咨询: 创始人 高振星、合伙人 邹璐

北京贤林律师事务所: 合伙人 伍慧群

前言：

P R E F A C E

伴随全球数字化进程的加快，人工智能成为引领未来世界发展的关键技术。近年来，各国政府、科研教育机构、科技企业及专家学者纷纷加入到推动人工智能产业发展的过程中，人工智能技术与产业融合程度不断加深。

2022年，尚普研究院从全球视角出发，对于人工智能的发展历程、产业链、核心技术、应用领域及发展趋势等方面进行全面梳理，为政府部门、从业人员、教育工作者、社会公众更好地了解人工智能的过去、现状及未来提供参考。

报告还有许多不足之处，希望读者与尚普研究院分析师团队多沟通交流，以便为今后的报告撰写提供宝贵建议。





尚普咨询集团 尚普研究院
2022年5月

尚普研究院：2022年全球人工智能产业研究报告框架

人工智能产业概况

<h3>基本概念</h3> <p>在各类机器载体上模拟并拥有类似生物/超越生物的智能。</p> 	<h3>发展历程</h3> <p>人工智能历经三次发展浪潮，在技术和应用等方面实现快速发展。</p> 	<h3>技术发展路线</h3> <p>人工智能核心技术历经萌芽期、稳定爬升期和成熟期三个发展阶段。</p> 	<h3>产业发展驱动因素</h3> <p>数据、算法、算力以及政策、顶尖人才/机构共同推动产业发展。</p> <p>数据 政策 算法 人才 算力 机构</p> 	<h3>全球人工智能产业链</h3> <p>全球人工智能产业链由底层基础、核心技术及应用领域三大部分构成。</p> 	<h3>全球人工智能产业图谱</h3> <p>图谱中企业LOGO排列顺序并不代表相关企业综合竞争力强弱，仅供参考。</p> 
---	--	--	---	---	---

人工智能底层基础

<h3>AI芯片</h3> <p>AI芯片类型主要涵盖GPU、FPGA及各类ASIC，主要应用在云端、边缘端和终端等各类场景。评价AI芯片性能需重点关注TOPS/W、时延、功耗等相关指标。</p> 	<h3>开源框架</h3> <p>深度学习开源框架是AI算法开发的必备工具。</p> 	<h3>数据服务</h3> <p>数据库设计、数据采集、数据清洗、数据标注与数据质检。</p> 	<h3>云计算服务</h3> <p>提供智能服务器等算力设施以及人工智能功能模块等产品。</p> 
--	---	---	--

人工智能核心技术

<h3>人工智能核心技术框架</h3>			<h3>AI核心技术总览</h3> <p>AI核心技术主要涉及机器学习、计算机视觉、自然语言处理、语音处理、知识图谱等，机器学习作为人工智能的核心，与计算机视觉、自然语言处理、语音处理和知识图谱密切相关。</p> 	<h3>机器学习总览</h3> <p>机器学习主要包括有监督学习、无监督学习和强化学习三类。机器学习专门研究计算机如何模拟/实现生物体的学习行为，获取新的知识技能，利用经验来改善特定算法的性能。</p> 
<h4>机器学习</h4> <ul style="list-style-type: none"> ✓ 逻辑回归 ✓ 支持向量机 ✓ 决策树 ✓ 随机森林 ✓ 朴素贝叶斯 ✓ 无监督学习 ✓ 强化学习 ✓ 	<h4>神经网络</h4> <ul style="list-style-type: none"> ✓ 神经网络 ✓ 卷积神经网络CNN ✓ 循环神经网络RNN ✓ 图神经网络GNN ✓ 长短期记忆神经网络LSTM ✓ 自编码器AE ✓ 生成对抗网络GAN ✓ 	<h4>计算机视觉</h4> <ul style="list-style-type: none"> ✓ U-Net ✓ R-CNN ✓ Yolo ✓ <h4>自然语言处理</h4> <ul style="list-style-type: none"> ✓ Word2Vec ✓ Seq2Seq ✓ 	<h4>语音处理</h4> <ul style="list-style-type: none"> ✓ 语音识别 ✓ 语音合成 ✓ <h4>知识图谱</h4>	

尚普研究院：2022年全球人工智能产业研究报告框架



报告框架

人工智能核心技术

机器学习典型算法（节选）

支持向量机

以间隔最大化为基准，学习得到尽可能远离数据的决策边界算法，支持向量是确定决策边界的重要数据。

随机森林

利用多棵决策树对样本进行训练并预测的一种分类器。从每个决策树收集输出，通过多数表决得到最终的分类结果。



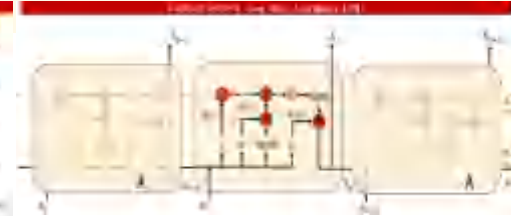
深度学习典型算法（节选）

卷积神经网络

由数千甚至数百万个紧密互连的简单处理节点组成，包括输入层、卷积层、池化层、全连接层和输出层。

长短期记忆神经网络

由状态单元、输入门、遗忘门、输出门组成，以此令长期记忆与短期记忆相结合，达到序列学习的目的。



人工智能应用领域

人工智能下游应用成熟度模型



人工智能趋势展望

前沿技术

- ✓ Transformer模型
- ✓ BERT模型
- ✓ ViT模型
- ✓ 自监督学习
- ✓ 类脑计算
- ✓ AI大模型
- ✓

产业融合

- ✓ 元宇宙
- ✓ 生命科学
- ✓ 新冠疫情
- ✓ 半导体
- ✓ 碳中和
- ✓ 冬奥会
- ✓

热点问题

- ✓ 学术与商业化
- ✓ 伦理与安全
- ✓ 就业
- ✓ 国家间技术限制
- ✓

前沿技术（节选）

Transformer模型

该模型主要由编码器和解码器构成，模型本身并行度较高，在精度和性能上均要优于传统的循环神经网络（RNN）和卷积神经网络（CNN）。



产业融合（节选）

元宇宙（Metaverse）

元宇宙本质上是对现实世界的虚拟化、数字化过程，其主要包括基础设施、人机交互等七层架构。



开篇：全球四次工业革命发展历程

18世纪以来，在全球政治格局变化背景下，工业革命带来的科技革命推动经济发展和社会进步

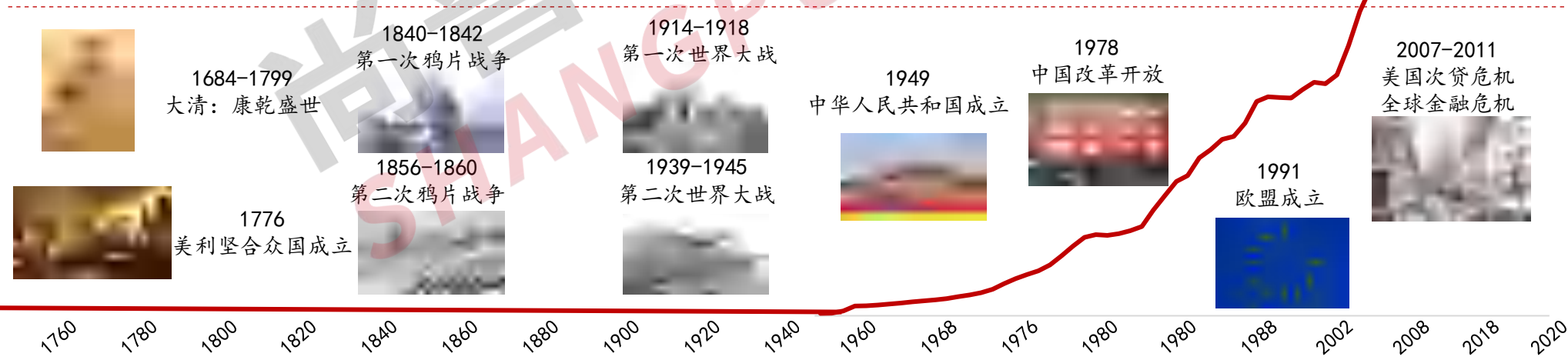
报告开篇

世界GDP走势
单位：亿美元
1,000,000

900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0

第一次工业革命	第二次工业革命	第三次工业革命	第四次工业革命
---------	---------	---------	---------

蒸汽时代 (1760-1840)	电气时代 (1860-1950)	信息时代 (1950-2010S)	智能时代 (2010S至今)
---------------------	---------------------	----------------------	-------------------



技术

政治经济

资料来源：World Bank世界银行，The World Economy: Historical Statistics, OECD, Paris, 北京大学出版社《全球通史—从史前到21世纪》2020年1月第1版，中信出版集团《第四次工业革命》2016年6月第1版，部分图片来自百度百科等，尚普研究院结合公开资料整理绘制

目录：

CONTENTS

1. 人工智能产业概况

2. 人工智能底层基础

3. 人工智能核心技术

4. 人工智能应用领域

5. 人工智能趋势展望

1

人工智能产业概况

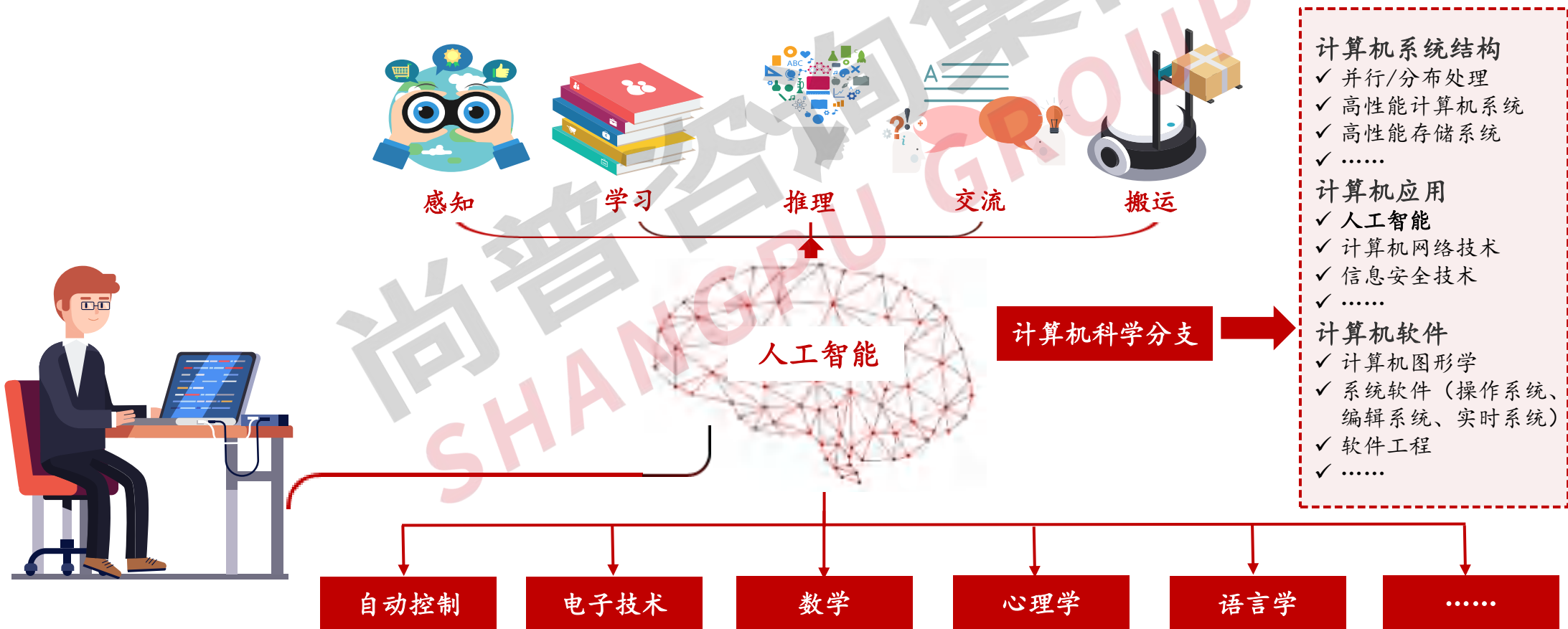
Artificial Intelligence Industry Overview

人工智能概念

在各类机器载体模拟并拥有类似生物或超越生物的智能

人工智能(Artificial Intelligence):主要是指在各类机器载体(手机、电脑、交通工具、机器人、机械设备等)上模拟并拥有类似生物/超越生物的智能(感知、学习、推理、交流等)。人工智能初期属于计算机科学的分支,研究领域涉及计算机视觉、自然语言处理、机器学习、语音处理等,同时又与多个学科紧密相关,包括自动控制、电子技术、数学、心理学、语言学、哲学等。

人工智能概念(Artificial Intelligence)



人工智能三大学派

人工智能三大学派：符号主义、联结主义、行为主义

从学术角度来看，人工智能主要包括符号主义、联结主义和行为主义三大学派。符号主义主张通过人工赋予机器智能，联结主义依托机器自行习得智能，行为主义则在与环境的作用和反馈中获得智能。人工智能三大学派的发展与算法、算力和数据三要素密切相关，未来三大学派将呈现逐步融合的趋势，发挥其各自特点持续推动人工智能产业发展。

学派分类	符号主义 (Symbolicism)	联结主义 (Connectionism)	行为主义 (Actionism)
思想起源	数理逻辑	仿生学	控制论
基本思想	将智能形式化为符号、知识、规则和算法，并用计算机实现符号、知识、规则和算法的表征和计算，从而实现用计算机来模拟人的智能行为。	生物智能是由神经网络产生的，通过人工方式构造神经网络，再训练人工神经网络产生智能。	智能取决于感知和行为，取决于对外界复杂环境的适应，而不是表示和推理，不同的行为表现出不同的功能和不同的控制结构。生物智能是自然进化的产物，生物通过与环境及其他生物之间的相互作用，从而发展出越来越强的智能，人工智能也可以沿这个途径发展。
主要原理	符号操作系统假设、有限合理性原理	神经网络及神经网络间的连接机制与学习算法	控制论及感知-动作型控制系统
代表成果	启发式程序逻辑理论家 (LT)、专家系统、知识工程理论技术等	脑模型 (MP)、感知机、多层网络中的反向传播算法 (BP) 和人工神经网络 (ANN) 等	布鲁克斯六足行走机器人
常用算法	规则和决策树	神经网络	反馈控制、遗传算法、强化学习等
代表人物	纽厄尔 (Newell) 西蒙 (Simon) 尼尔逊 (Nilsson)	霍普菲尔德 (Hopfield) 鲁梅尔哈特 (Rumelhart) 罗森布拉特 (Rosenblatt)	维纳 (Wiener) 麦克洛克 (McCulloch) 布鲁克斯 (Brooks)

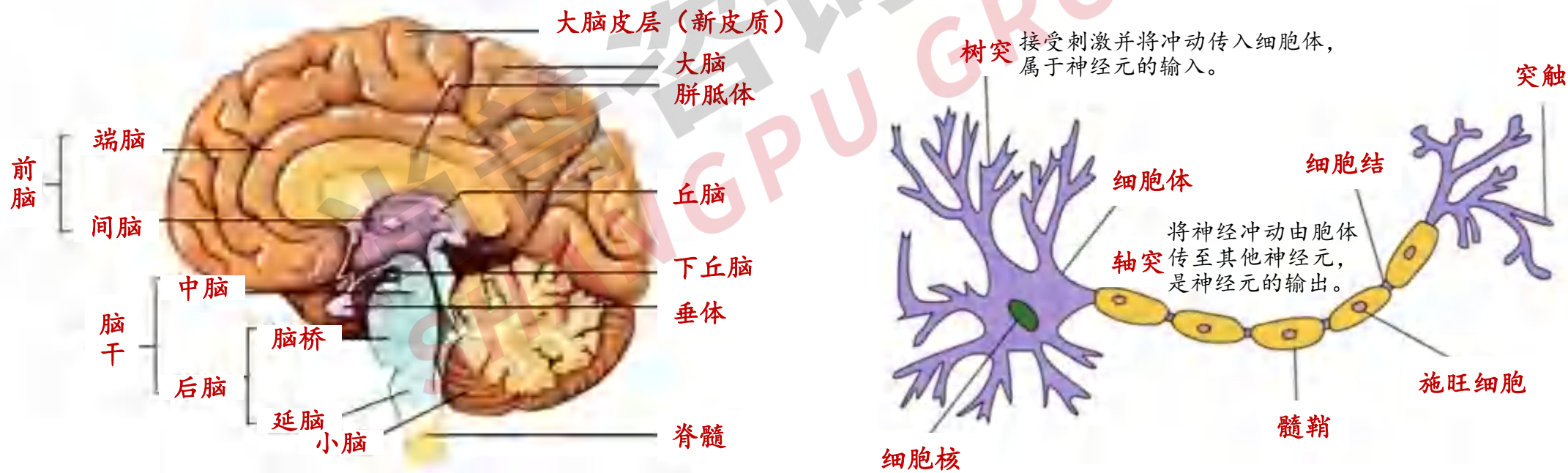
人工智能与大脑的关联性

大脑新皮质负责人类语言、推理等高级功能，赫布理论奠定人工神经网络基础

端脑表面所覆盖的灰质称为大脑皮质，即新皮质（neocortex）。新皮质是进化程度较高级的皮质，是哺乳动物大脑皮质的大部分，在脑半球顶层，其主要分为六层（①分子层；②外颗粒层；③外锥体细胞层；④内颗粒层；⑤内锥体细胞层；⑥多形细胞层），与一些高等功能如知觉、运动指令的产生、空间推理、意识及人类语言有关系。神经元作为新皮质学习的基本单元，学习是由单元之间联系的突触力量所决定的。神经科学为思维的模式识别理论提供依据，加拿大心理学家Donald O. Hebb于1949年首次尝试解释学习的神经原理，即“**Hebbian theory赫布理论**”：突触前神经元向突触后神经元持续重复的刺激可以导致突触传递效能的增加，人工神经网络的建立就是基于赫布神经学习模型。

人工智能产业概况

大脑与神经元结构示意图



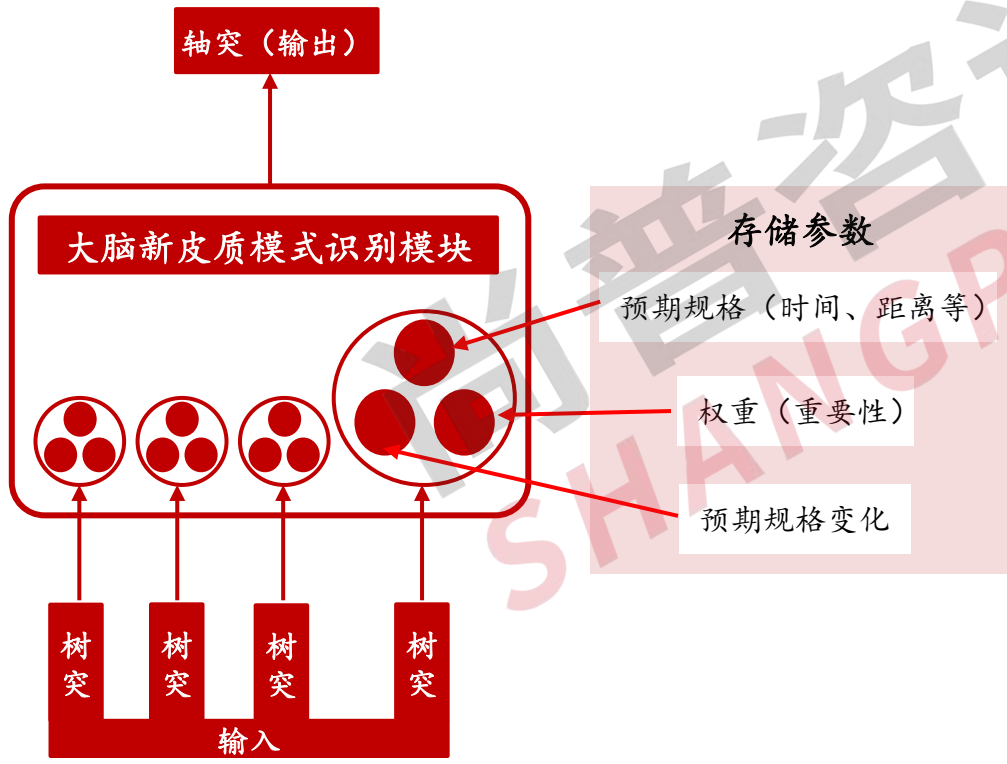
人工智能与大脑的关联性

人工智能相关算法源自大脑新皮质模式识别机制

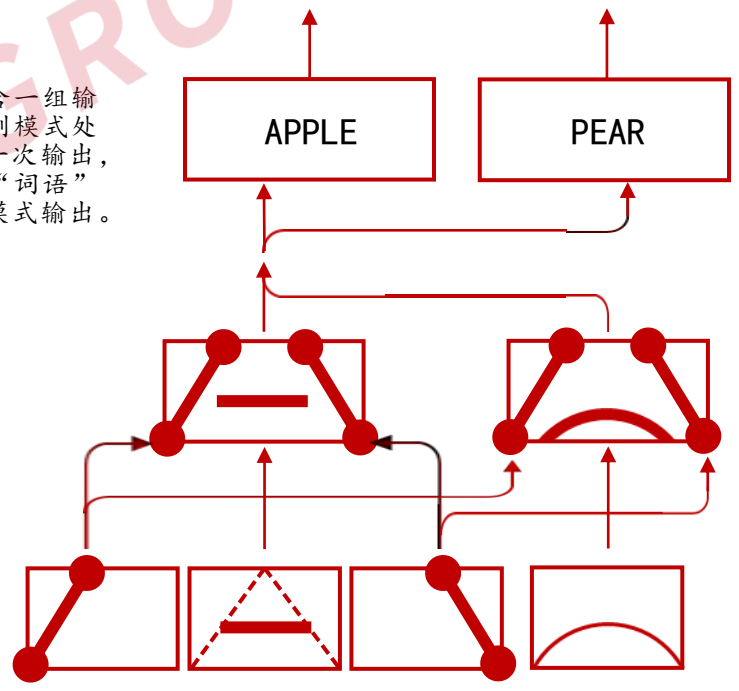
雷·库兹韦尔(Ray Kurzweil)与杰夫·霍金斯(Jeff Hawkins)、迪利普·乔治(Dileep George)共同提出“大脑新皮质模型”。大脑新皮质模式识别主要由三部分组成：①输入(低层级模式)；②模式名称；③较高层级模式集合。每个模式都具有输入信息、识别模式的处理程序以及一次输出过程。在输入环节，模式(图形、字母、词语等)需按照模式识别器的连续顺序出现，才可以被识别；当模式识别器识别到相关模式时，该模式识别器的轴突也会被激活。举例来看，信息沿着概念层级向上流动，从基本的字母特征到字母再到词语，之后识别会继续向上流动到短语，最终形成更为复杂的语言结构。

人工智能产业概况

大脑新皮质模式识别示例



字母模式包含一组输入信息、识别模式处理程序以及一次输出，字母模式向“词语”的更高层级模式输出。



人工智能发展历程

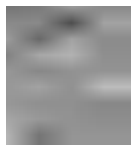
20世纪50年代以来，人工智能历经三次发展浪潮，在技术和应用等方面实现快速发展

1950年，艾伦·图灵 (Alan Turing) 发表了一篇划时代的论文，文中预言了创造出具有真正智能的机器的可能性。1956年夏，约翰·麦卡锡 (John McCarthy)、马文·明斯基 (Marvin Minsky) 等科学家在美国达特茅斯学院开会研讨“如何用机器模拟人的智能”，首次提出“人工智能 (Artificial Intelligence, 简称AI)”这一概念，标志着人工智能学科的诞生。2006年以来，随着大数据、云计算、物联网等信息技术的发展，泛在感知数据和通用图形处理器推动以深度神经网络为代表的人工智能技术飞速发展，大幅跨越了科学与应用之间的“技术鸿沟”，迎来爆发式增长的新高潮。以下是全球人工智能发展历程中的代表性事件：

人工智能第一次浪潮

(1956-1976)

1957: Frank Rosenblatt发明神经网络感知机



1968: 斯坦福研究所发明第一台智能机器人



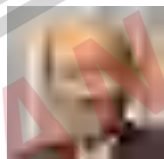
1956: 达特茅斯会议上宣告“人工智能”学科的诞生



1966: Leonard Baum发明隐马尔可夫模型



1979: 福岛邦彦提出卷积神经网络



1951: Marvin Minsky建立第一个神经网络机器SNARC



1966: MIT Joseph Weizenbaum建立第一个自然语言处理程序，聊天机器人ELIZA



1975: 人工智能在化学领域首个科学发现



1950: Alan Turing提出了图灵测试



1965: 专家系统诞生

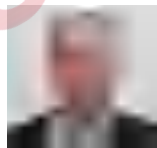
人工智能第二次浪潮

(1976-2006)

1985: Geoffrey Hinton 和 Terry Sejnowski 发明玻尔兹曼机



1982: John Hopfield发明了新型神经网络 (Hopfield 网络)



1989: 卡耐基梅隆大学Alex Waibel开发时延神经网络



1990: Suzanna Becker 提出“自动编码器”概念

1986: 反向传播算法诞生

1987: 人工智能软件在企业战略领域首次商业化应用



1998: MNIST数据库中的图像集建立



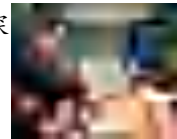
2010: ImageNet 第一届大规模视觉识别挑战赛 (ILSVRC)



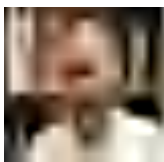
2009: 谷歌推出自动驾驶汽车



1997: IBM深蓝战胜国际象棋冠军卡斯帕罗夫



2006: Geoffery Hinton等人提出深度学习



1996: 康奈尔大学David Field和Bruno Olshausen提出稀疏编码



人工智能第三次浪潮

(2006至今)

2022: AI在北京冬奥会广泛应用



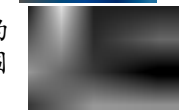
2020: AI助力疫情防控



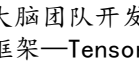
2016: 华盛顿大学Carlos Guestrin等开发LIME算法



2016: Google的Alpha Go战胜围棋冠军李世石



2015: 谷歌大脑团队开发出机器学习开源框架—TensorFlow



2014: 蒙特利尔大学Ian Goodfellow发明“生成对抗性网络”



1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022

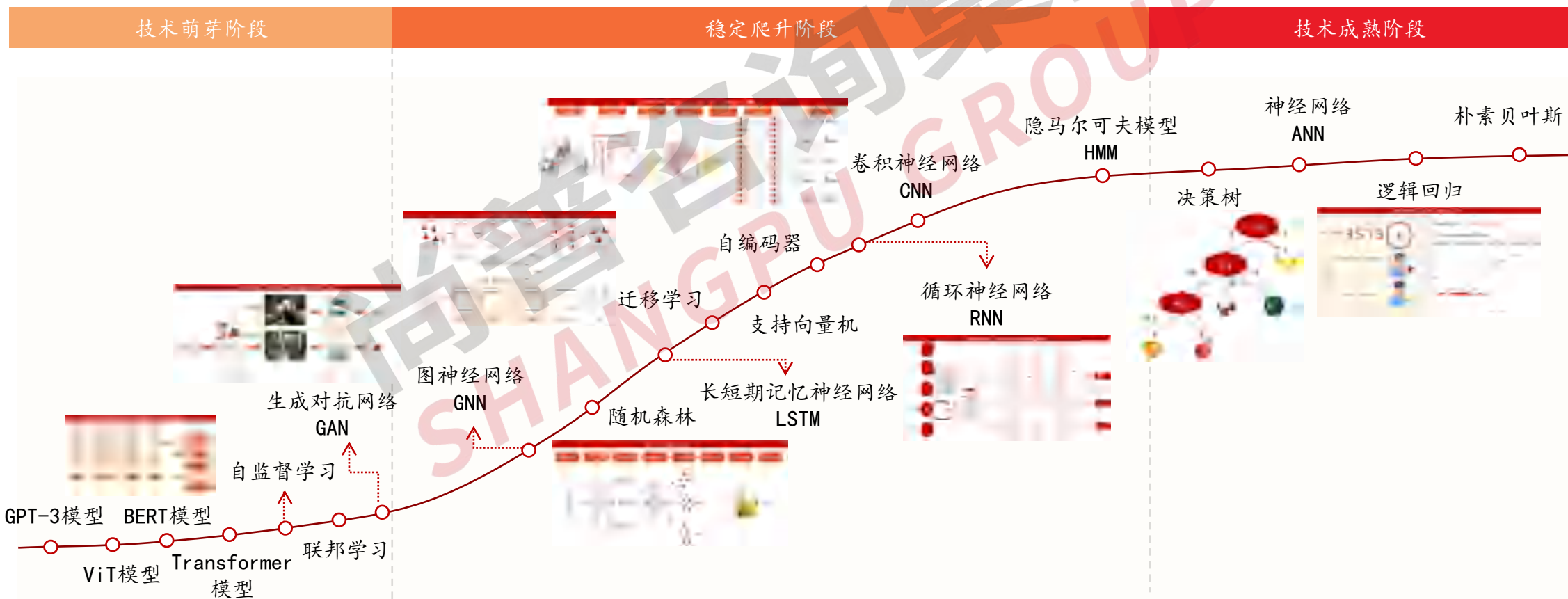
资料来源：人民邮电出版社《人工智能通识课》2020年6月第1版，部分图片来自Google和Wikipedia等，尚普研究院结合公开资料整理绘制 13

人工智能典型算法成熟度路线图

人工智能典型算法历经萌芽期、稳定爬升期和成熟期三个发展阶段

尚普研究院依据人工智能典型算法的发展阶段，将近年来出现的GPT-3模型、ViT模型、BERT模型等划分至技术萌芽阶段；将具有一定技术沉淀但仍需持续迭代的卷积神经网络、长短期记忆神经网络、循环神经网络等划分至技术稳定爬升阶段；将拥有较长发展历史的决策树、逻辑回归、朴素贝叶斯等划分至技术成熟阶段，具体如下图所示。人工智能典型算法成熟度路线图可作为本报告核心技术和趋势展望章节概览，供相关读者参考。伴随人工智能技术的不断发展，典型算法成熟度路线图有待不断优化。

人工智能典型算法成熟度路线图



产业发展驱动因素：数据、算力、算法

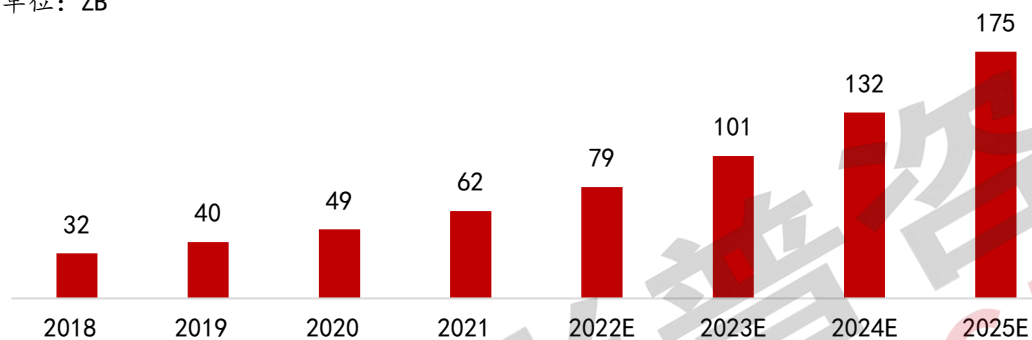
全球数据量及芯片算力持续增长，AI算法不断优化

数据层面：近年来全球数据规模持续增长，IDC预计到2025年全球数据量将达到175ZB；**算力层面：**1971年芯片算力从0.06MIPS快速增长至15.8TOPS以上；**算法层面：**伴随技术的不断发展，机器学习、计算机视觉、自然语言处理、语音处理、知识图谱等AI核心技术相关算法持续迭代优化。在数据、算力和算法的共同推动下，全球人工智能产业和相关技术实现快速发展，下游应用也不断丰富。

数据 (Data)

2018-2025年全球数据量

单位：ZB

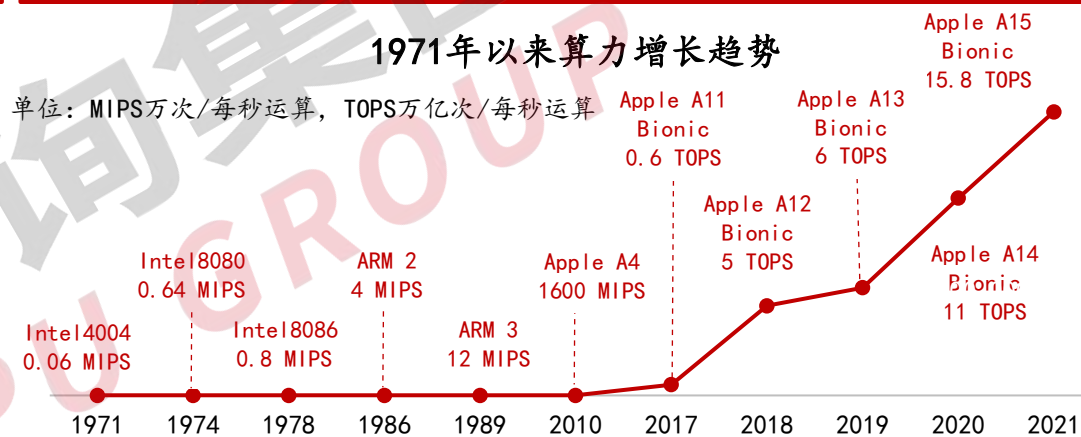


资料来源：IDC, 尚普研究院

算力 (Computing Power)

1971年以来算力增长趋势

单位：MIPS万次/每秒运算，TOPS万亿次/每秒运算



资料来源：尚普研究院结合公开资料整理绘制

算法 (Algorithm)

机器学习算法



计算机视觉算法



自然语言处理&语音处理算法



知识图谱算法



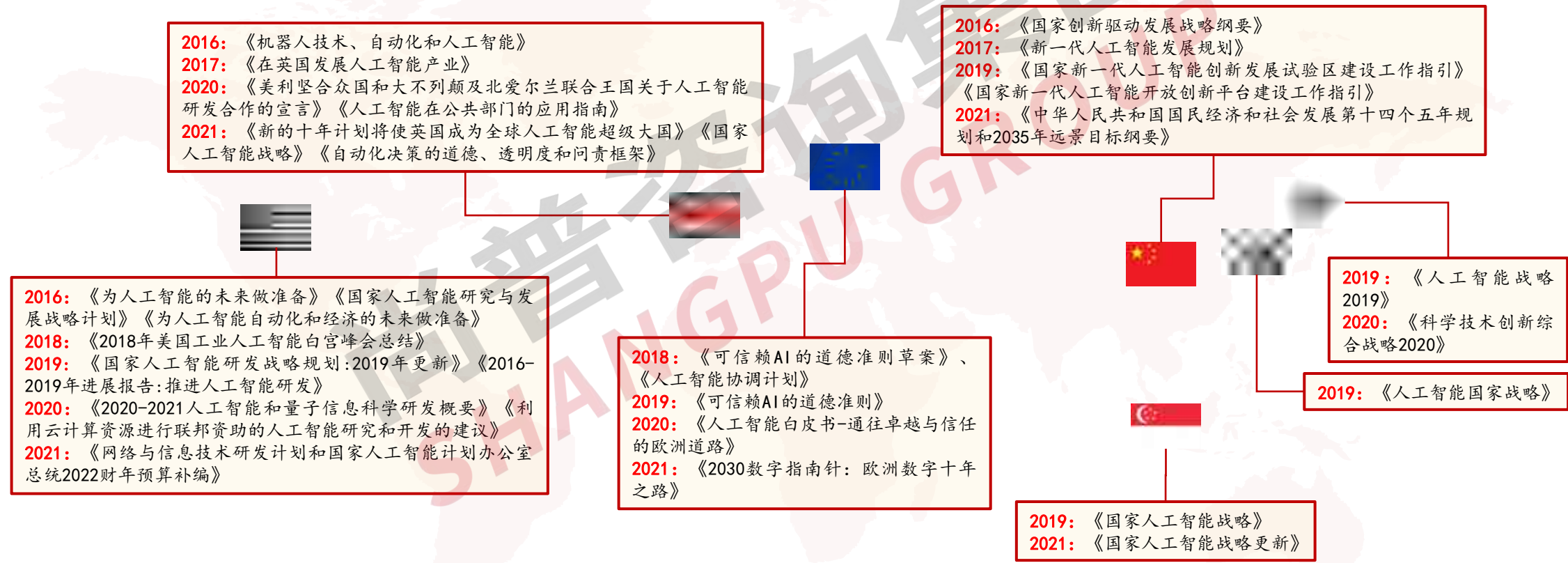
资料来源：尚普研究院结合公开资料整理绘制

产业发展驱动因素：人工智能产业政策

全球主要国家和地区相继出台人工智能政策，推动人工智能产业发展

近年来，主要国家和地区相继出台了人工智能相关战略和规划文件，将政策重点聚焦在加强投资和人才培养、促进合作开放以及完善监管和标准建设上，全球人工智能产业发展进入加速落地阶段。具体而言，美国加大人工智能研发方面的投资；欧盟致力于推进新的人工智能立法提案；英国打造世界人工智能创新中心；日本则强化人工智能应用，加快数字化转型；韩国加强人工智能战略引领，助力本国经济复苏；新加坡为人工智能研究追加投资，推进政府的数字化转型；中国则更加注重推动人工智能与传统产业融合。

人工智能产业概况



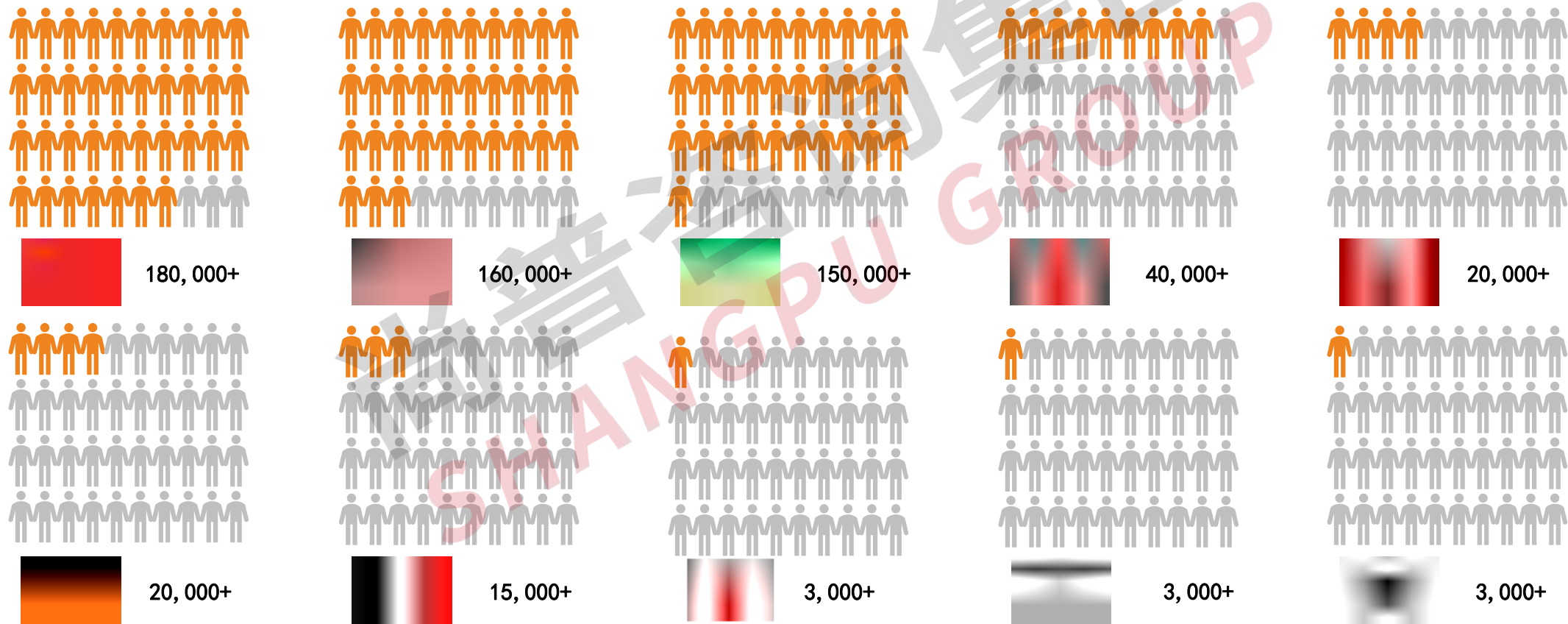
产业发展驱动因素：人工智能人才规模

中、美、印AI人才数量位居世界前三，人才聚集现象凸显

尚普研究院根据LinkedIn和猎聘数据统计，全球当前累计AI人才突破100万人。其中，中国AI人才数量超过18万人位居世界首位，占全球AI人才总数的18%。美国和印度AI人才数量分居全球第二、三位，并且均超过15万。中、美、印三国AI人才合计占比接近全球50%。英国、加拿大、德国、法国等国家AI人才数量也超过万人。AI人才数量排名前10的国家占全球总数超过60%，AI人才聚集现象凸显。

人工智能产业概况

全球主要国家AI人才数量



注：中国AI人才数量来自猎聘，其他国家AI人才数量来自LinkedIn，相关数据统计截至2022年4月

资料来源：尚普研究院结合LinkedIn和猎聘数据整理绘制 17

产业发展驱动因素：人工智能顶尖人才

清华大学AMiner团队推出AI 2000榜单，遴选全球最具影响力的AI顶尖学者

清华大学AMiner团队联合北京智源研究院、清华-中国工程院知识智能联合研究中心，连续3年共同发布《最具影响力学者人工智能全球——AI 2000榜单》，该榜单通过AMiner系统中所收录的学术发表数据及Google Scholar的引用数据用计算机算法自动生成榜单排名，其旨在全球范围内遴选过去十年人工智能学科最有影响力、最具活力的顶级学者。自2020年首届AI 2000榜单发布以来，榜单受到各界的高度认可，已成为AI领域的重要风向标之一。

人工智能产业概况

AI 2000遴选方法概述

- AMiner系统收集过去10年间出现在AI所有子领域顶级期刊和会议的出版论文，对于每位学者，其在一个特定子领域的评分分别计算了AI 2000指数、发表次数、引用次数、引用数权重系数、CSRanking等指标。
- 发表次数：**顶级期刊和会议中发表文章的总数量（单个子领域）
- 引用次数：**顶级期刊和会议中发表所有论文的总引用次数（单个子领域）
- 引用数权重系数：**

$$C_k = \frac{2}{n} \left(1 - \frac{k}{n+1} \right) C$$

- 其中 n 代表一篇文章中的所有作者数量； k 代表论文数量中作者的展示顺序，即文章中第 k 个作者； C 代表引用这篇文章的文献数量。该公式针对某一个领域的一篇文章，针对该学者在该领域所有出版文章的 C_k 的总和即为该学者的引用数权重系数。
- CSRanking：**是一种调整计数方法，表示每一篇文章的功劳平均分配给每一位作者。此处将作者进行排名时，将其所有文章的功劳平均计算后累计叠加得到作者的文章整体贡献度。
- AI 2000指数：**AI 2000指数表示所有作者对该篇文章的贡献度不同，在最新的排名计算中，默认发表的论文中，排名靠前的作者贡献度更高，因此该指数计算的逻辑为，作者的排名顺序与其贡献度成正比。公式定义为：

$$C_k = \frac{1/k}{\sum_m 1/m} C$$

- 其中 n 代表一篇文章中的所有作者数量； k 代表论文数量中作者的展示顺序，即文章中第 k 个作者； C 代表引用这篇文章的文献数量。求和代表所有作者排名次序倒数的求和，从而展示出单个作者排名顺序与贡献度比例关系。该公式针对某一个领域的一篇文章，针对该学者在该领域所有出版文章的 C_k 的总和即为该学者的AI 2000指数。

AI 2000指数实例解析

- 上文提到的AI 2000学者榜单是根据AI 2000指数进行排名所得到的。AI 2000指数：

$$C_k = \frac{1/k}{\sum_m 1/m} C$$

- 假设一篇文章由2名学者共同合作完成，该文章被引用15次，那么第一作者的AI 2000指数即：

$$C_{k_1} = \left(\frac{1}{1 + 1/2} \right) \times 15 = 10$$

- 第二作者的AI 2000指数即：

$$C_{k_2} = \left(\frac{1/2}{1 + 1/2} \right) \times 15 = 5$$

- 由此可见两位作者对文章引用率的贡献度不同，排名次序靠前，贡献度越大。

遴选方法中的其他计算指标也是衡量学者们学术水平的标准，AMiner系统中根据不同的指标对学者进行不同的排名，同样具有重要的参考意义和价值。

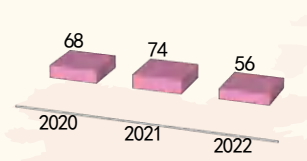
产业发展驱动因素：人工智能顶尖人才

AI 2000上榜学者遍布全球40余个国家，美国AI学者数量位居榜首

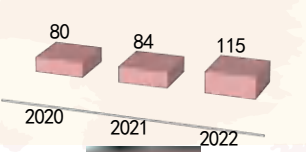
根据AMiner的历年统计数据，美国学者入选AI 2000榜单的数量最多，年均超过1,000人次，远高于其他国家，表明美国在AI人才方面具有全球领军地位。此外，中国人才数量位居第二，但与美国差距仍然较大，英国与中国入选AI 2000榜单的学者数量呈逐年递增趋势。

人工智能产业概况

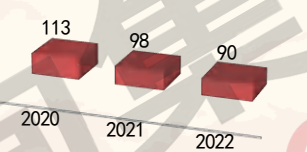
加拿大入选AI 2000学者数量



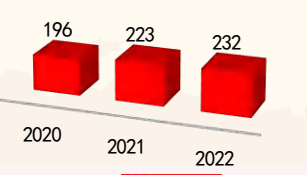
英国入选AI 2000学者数量



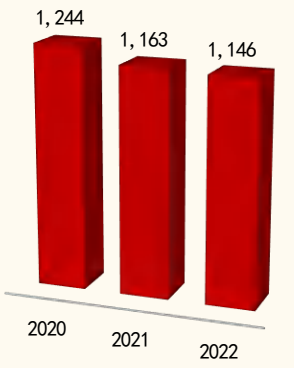
德国入选AI 2000学者数量



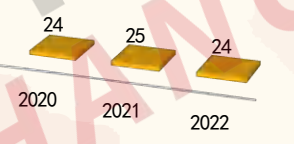
中国入选AI 2000学者数量



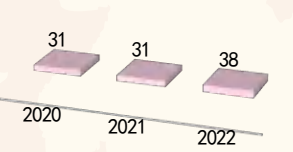
美国入选AI 2000学者数量



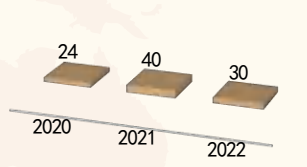
法国入选AI 2000学者数量



瑞士入选AI 2000学者数量



新加坡入选AI 2000学者数量



注：以上图表单位均为人次

资料来源：AMiner团队发布的AI 2000学者榜单，智谱研究，尚普研究院 19

产业发展驱动因素：人工智能顶尖机构

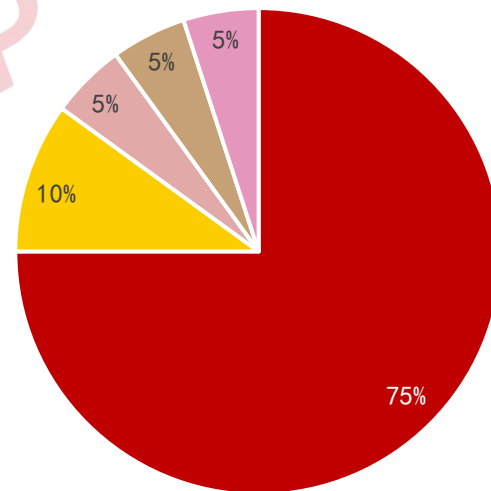
Google成为入选AI 2000学者最多机构，清华大学和阿里巴巴跻身前20强

根据AMiner发布的2022年AI 2000榜单，Google以181人次的入选数量名列全球入选学者最多机构，Meta、Microsoft入选人次分别为87和65位列第二、三位。从全球AI 2000前20强机构来看，美国机构数量达到15家。上榜的两家中国机构为清华大学和阿里巴巴，两家机构入选数量分别为20人次和14人次。

2022年AI 2000全球前20强机构

名次	机构名称	入选人次	名次	机构名称	入选人次
1	谷歌 (美)	181	11	加州大学伯克利分校 (美)	27
2	Meta (美)	87	12	纽约大学 (美)	21
3	微软 (美)	65	13	清华大学 (中)	20
4	麻省理工学院 (美)	47	14	IBM (美)	18
5	卡内基梅隆大学 (美)	44	15	多伦多大学 (加)	17
6	斯坦福大学 (美)	41	16	密歇根大学 (美)	17
7	DeepMind (英)	32	17	加州大学圣迭戈分校 (美)	16
8	苹果 (美)	31	18	新加坡国立大学 (新)	15
9	华盛顿大学 (美)	28	19	佐治亚理工学院 (美)	15
10	亚马逊 (美)	28	20	阿里巴巴 (中)	14

AI 2000机构前20强国家分布



■ 美国 ■ 中国 ■ 英国 ■ 加拿大 ■ 新加坡

在全球AI 2000前20强的机构中，美国机构占15家，中国2家机构上榜，英国、加拿大和新加坡各占1家。由此可见，美国仍具有遥遥领先的AI顶尖人才和AI科研实力。

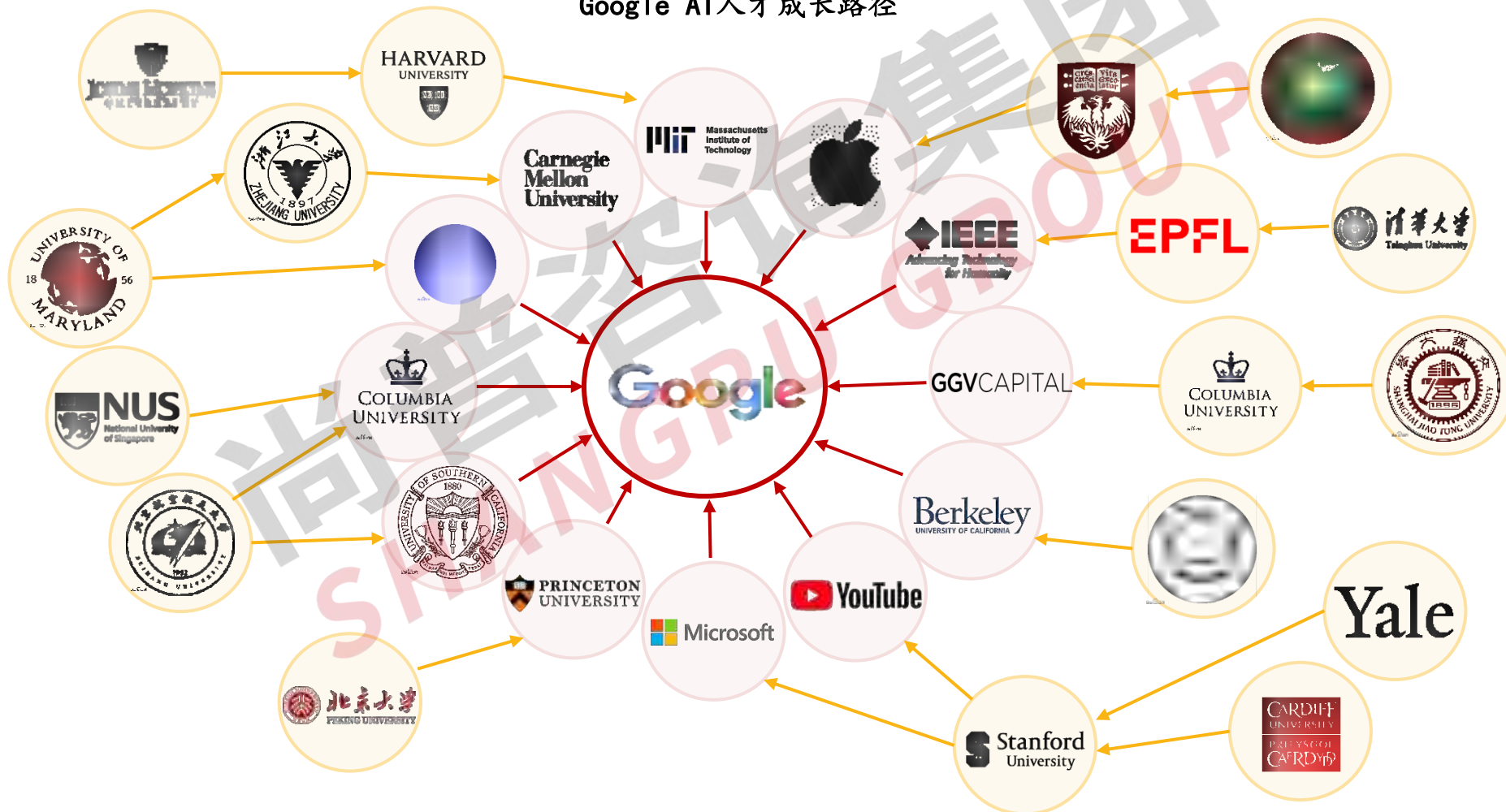
产业发展驱动因素：人工智能顶尖机构

Google AI人才普遍拥有世界顶尖高校教育背景，具有多年科技巨头从业经历

尚普研究院结合LinkedIn相关人才数据，对于Google AI人才的教育及从业背景进行梳理。Google AI人才大多拥有Stanford、MIT、CMU、UC Berkeley等世界知名高校教育背景，同时也曾在Apple、Microsoft、Facebook等头部科技巨头从业多年，AI人才优势显著。

人工智能产业概况

Google AI人才成长路径



资料来源：尚普研究院结合LinkedIn相关人才数据整理绘制 21

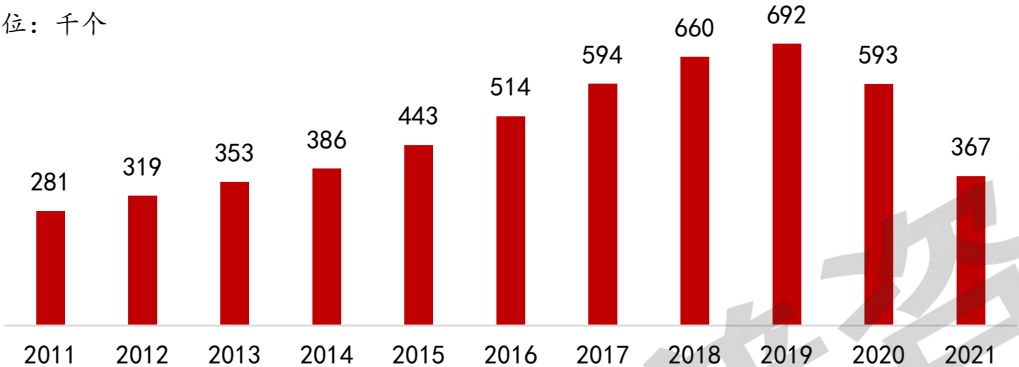
全球人工智能专利情况

中国、美国、日本在人工智能专利申请数量方面整体居前

尚普研究院结合智谱研究数据，2011-2021年全球人工智能专利申请数量呈现先升后降的趋势。按地区分布看，2011-2021年中国、美国和日本累计申请人工智能专利数量最多，中国位居世界首位，累计持有专利数量191万个；按申请机构来看，人工智能专利申请数量前十大机构集中在韩国、日本、美国和中国。其中，三星电子株式会社持有的人工智能专利数量超过10万个。

2011-2021年全球人工智能专利申请数量变化

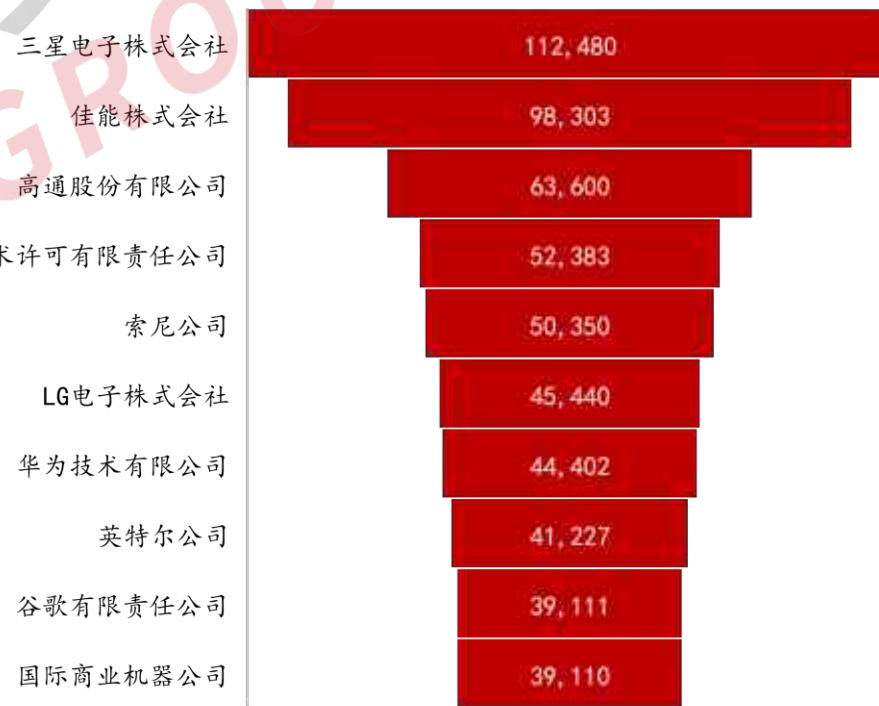
单位：千个



资料来源：智谱研究，尚普研究院

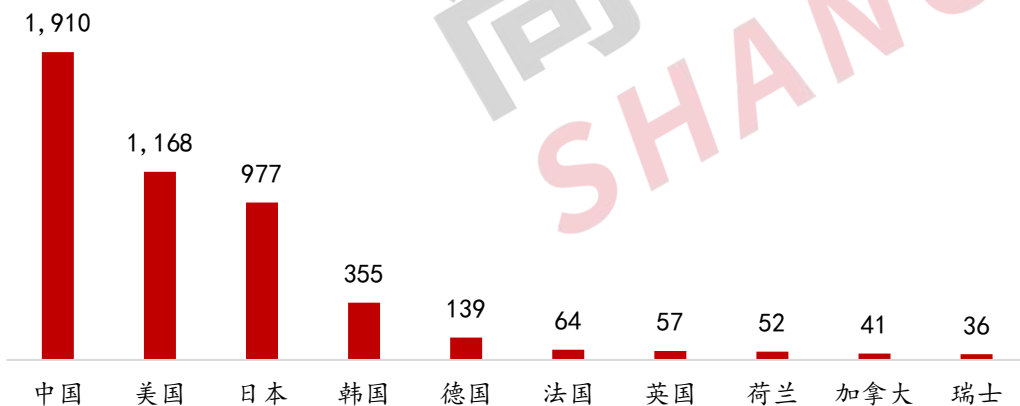
2011-2021年全球人工智能累计专利申请机构TOP10

单位：个



2011-2021年全球人工智能累计专利申请国家TOP10

单位：千个



资料来源：智谱研究，尚普研究院

资料来源：智谱研究，尚普研究院

中美人工智能产业发展状况对比

美国在企业数量、融资规模、算力方面领先，中国则在数据量方面具有明显优势

结合CB Insights, IDC、浪潮信息与清华全球产业院联合发布《2021-2022全球算力指数评估报告》等相关资料，尚普研究院对中美两国在人工智能产业发展和基础方面进行相关对比。当前美国在人工智能企业数量、融资规模、算力方面均占据显著优势。其中，美国AI企业数量超4,100个，累计融资规模超1,600亿，算力指数77；中国则在数据量方面具有优势，IDC预计到2025年中国数据量规模达到48.6ZB。此外，在开源框架方面，中美两国也拥有多个开源框架，服务于全球广大AI开发者。

中美人工智能产业发展状况对比

衡量指标	美国人工智能产业发展状况	中国人工智能产业发展状况
累计企业数量	4,100+个	1,200+个
累计融资规模	1,600+亿	470+亿
算力指数	77	70
2025年数据量	30.6ZB	48.6ZB
企业所在地	加州旧金山、加州山景城、加州圣何塞、纽约等地 	北京、上海、广州、杭州、深圳等地 
开源框架		

注：累计企业数量、累计融资规模截至2022年4月

资料来源：CB Insights, IDC、浪潮信息、清华全球产业院《2021-2022全球算力指数评估报告》，尚普研究院结合公开资料整理绘制

中美人工智能人才培养情况

中美两国高校通过建立人工智能学院或依托计算机科学系等方式，加强人工智能人才培养

尚普研究院结合2022年U. S. News发布的Best Artificial Intelligence Programs以及软科（ShanghaiRanking）高校排名榜单，对于美国和中国知名高校在人工智能学院以及人工智能学科专业建设情况进行梳理。在人工智能人才培养方面，美国高校主要依托于早期建立的计算机科学学院/电子工程与计算机科学系设立人工智能研究方向或相关专业，如卡内基梅隆大学、麻省理工学院、斯坦福大学等；中国高校则在近些年建立专门的人工智能学院，加强对人工智能高层次人才培养，如南京大学、西安交通大学等。

中美主要高校人工智能教学研究机构设立情况

序号	美国高校名称	所属学院/院系	序号	中国高校名称	所属学院/院系
1	卡内基梅隆大学 	计算机科学学院	1	清华大学 	计算机科学与技术系 交叉信息研究院 自动化系
2	麻省理工学院 	电气工程与计算机科学系	2	北京大学 	计算机学院 智能学院
3	斯坦福大学 	计算机科学系	3	浙江大学 	计算机科学与技术学院
4	加利福尼亚大学-伯克利分校 	电气工程与计算机科学系	4	上海交通大学 	电子信息与电气工程学院
5	康奈尔大学 	计算机科学系	5	复旦大学 	计算机科学技术学院
6	佐治亚理工学院 	计算机科学与工程学院 交互计算学院	6	南京大学 	人工智能学院
7	华盛顿大学 	计算机科学与工程学院	7	中国科学技术大学 	计算机科学与技术学院
8	伊利诺伊大学厄巴纳-香槟分校 	计算机科学与工程学院	8	华中科技大学 	人工智能与自动化学院
9	密歇根大学-安娜堡分校 	电气工程与计算机科学系	9	武汉大学 	计算机学院
10	得克萨斯大学-奥斯汀分校 	计算机科学系	10	西安交通大学 	人工智能学院 未来技术学院

资料来源：尚普研究院结合U. S. News: 2022 Best Artificial Intelligence Programs, 软科（ShanghaiRanking）：2022中国大学排名, 各高校官网整理绘制 24

美国人工智能顶尖学府

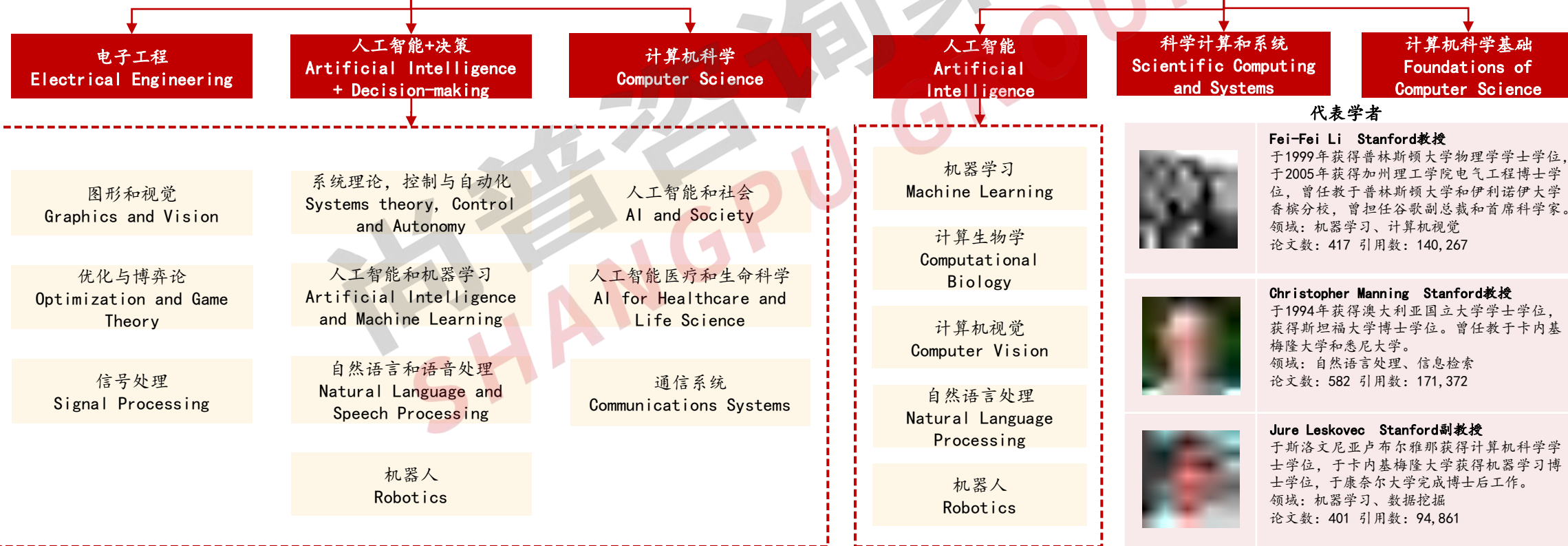
MIT与Stanford作为全球AI顶尖学府，拥有完善且系统的AI研究组织架构

MIT与Stanford作为全球知名的人工智能研究机构，其在学科设置、学者数量等方面拥有领先优势。MIT拥有完善的人工智能研究机构，麻省理工学院电子工程与计算机科学系（MIT EECS）包括图形和视觉、系统理论及控制与自动化、优化与博弈论、人工智能和机器学习、人工智能医疗和生命科学、信号处理、自然语言和语音处理、通信系统和机器人等十大研究领域。斯坦福大学计算机科学系（Stanford DCS）主要划分为机器学习、计算生物学、计算机视觉、自然语言处理、机器人等五大研究领域，其中以Fei-Fei Li、Christopher Manning、Jure Leskovec为代表的顶级学者在机器学习、计算机视觉、自然语言处理等领域分别有所造诣。

MITEECS 电气工程与计算机科学系

Stanford ENGINEERING Computer Science

计算机科学系



资料来源：MIT EECS官网，Stanford DCS官网，尚普研究院结合公开资料整理绘制 25

中美科技巨头人工智能布局情况

中美科技巨头在AI芯片、技术平台等领域广泛布局

尚普研究院对于中美科技巨头在人工智能主要产品及解决方案布局进行梳理。从美国来看，Google、AWS、Microsoft、IBM、Meta等科技巨头在人工智能领域布局较早，并且拥有成熟的AI芯片产品、开源框架及技术平台；从中国来看，以百度、阿里巴巴、腾讯、华为为代表的科技领军企业也在相关领域有所布局，依托子公司、研发平台等渠道，持续开发AI核心技术等底层基础设施，并在多个场景实现商业化落地应用。

人工智能产业概况

中美科技巨头人工智能布局情况

AI芯片	开源框架	技术平台	预训练模型	AI芯片	开源框架	技术平台	预训练模型
<ul style="list-style-type: none"> ✓ Edge TPU 	<ul style="list-style-type: none"> ✓ TensorFlow 	<ul style="list-style-type: none"> ✓ Vertex AI 	<ul style="list-style-type: none"> ✓ Switch Transformer ✓ BERT ✓ T5 	<ul style="list-style-type: none"> ✓ 昆仑芯2代 	<ul style="list-style-type: none"> ✓ PaddlePaddle 	<ul style="list-style-type: none"> ✓ EasyDL ✓ BML ✓ UNIT ✓ 	<ul style="list-style-type: none"> ✓ 文心大模型
<ul style="list-style-type: none"> ✓ AWS Trainium ✓ AWS Inferentia ✓ AWS Graviton3 	<ul style="list-style-type: none"> ✓ MXNet 	<ul style="list-style-type: none"> ✓ Amazon ML ✓ Amazon Lex ✓ Amazon Rekognition ✓ 	<ul style="list-style-type: none"> ✓ Turing-NLG ✓ MT-NLG 	<ul style="list-style-type: none"> ✓ 含光800 		<ul style="list-style-type: none"> ✓ 机器学习平台PAI ✓ 视觉智能开放平台 ✓ NLP自学习平台 ✓ 	<ul style="list-style-type: none"> ✓ M6中文预训练模型
	<ul style="list-style-type: none"> ✓ CNTK 	<ul style="list-style-type: none"> ✓ Azure AI 	<ul style="list-style-type: none"> Microsoft 	<ul style="list-style-type: none"> ✓ 紫霄 		<ul style="list-style-type: none"> ✓ 机器学习平台Angel 3.0 ✓ 腾讯云TI平台 ✓ 腾讯智能对话平台 ✓ 	
<ul style="list-style-type: none"> ✓ Power9 processor chip ✓ IBM Telum 	<ul style="list-style-type: none"> ✓ SystemML 	<ul style="list-style-type: none"> ✓ IBM z16 ✓ IBM Cloud Pak for Data 				<ul style="list-style-type: none"> ✓ NeuFoundry ✓ QuickAI ✓ 制造云平台 ✓ 	
	<ul style="list-style-type: none"> ✓ PyTorch 	<ul style="list-style-type: none"> ✓ ClassyVision ✓ Fairseq 	<ul style="list-style-type: none"> ✓ RoBERTa 				
<ul style="list-style-type: none"> ✓ H100 ✓ Xavier ✓ Orin ✓ 	<ul style="list-style-type: none"> ✓ MONAI 		<ul style="list-style-type: none"> Meta 	<ul style="list-style-type: none"> ✓ 昇腾710 ✓ 昇腾910 ✓ 麒麟980 ✓ 	<ul style="list-style-type: none"> ✓ MindSpore 	<ul style="list-style-type: none"> ✓ 分布式训练平台 ✓ 智能边缘平台 ✓ ModelArts ✓ 	<ul style="list-style-type: none"> ✓ Pangu NLP
			<ul style="list-style-type: none"> NVIDIA 	<ul style="list-style-type: none"> HUAWEI 			

全球人工智能专业组织

IEEE、ACM、AAAI、CAAI、CCF为全球及中国重要的人工智能专业组织



IEEE电气与电子工程师协会

电气与电子工程师学会（IEEE）成立于1963年，IEEE总部位于美国纽约州纽约市。鼓励技术创新和卓越，以造福人类。IEEE定位在“科学和教育，并直接面向电子电气工程、通讯、计算机工程、计算机科学理论和原理研究的组织，以及相关工程分支的艺术和科学”。



ACM国际计算机学会

国际计算机学会（ACM）成立于1947年，总部位于美国纽约州纽约市，是一个全球性的科学和教育组织，致力于促进艺术、科学、工程和计算应用，通过促进信息的公开交流和促进最高的专业和道德标准，为专业和公共利益服务。ACM设有8个主要奖项，来表彰计算机领域的技术和专业成就。最高奖项为图灵奖Turing Award，常被形容为计算机领域的诺贝尔奖。



AAAI国际先进人工智能协会

国际先进人工智能协会（AAAI）成立于1979年，总部位于美国加州，致力于促进对思想和智能行为背后的机制及其在机器中的体现。AAAI旨在促进人工智能的研究和负责任的使用，增加公众对人工智能的理解，改进人工智能从业者的教学和培训，并为研究规划者和资助者提供有关当前人工智能发展的重要性及潜力以及未来方向的指导。



CAAI中国人工智能学会

中国人工智能学会（CAAI）成立于1981年，是经国家民政部正式注册的我国智能科学技术领域唯一的国家级学会，拥有51个分支机构，覆盖了智能科学与技术领域。基本任务是团结全国智能科学技术工作者和积极分子，通过学术研究、国内外学术交流、科学普及、学术教育、科技会展、学术出版、人才推荐、学术评价、学术咨询、技术评审与奖励等活动，促进我国智能科学技术的发展。



CCF中国计算机学会

中国计算机学会（CCF）成立于1962年，全国一级学会，独立社团法人，中国科学技术协会成员。CCF是中国计算机及相关领域的学术团体，宗旨是为本领域专业人士的学术和职业发展提供服务；推动学术进步和技术成果的应用；进行学术评价，引领学术方向；促进技术和产业应用一线的交流 and 互动；对在学术和技术方面有突出成就的个人、企业和单位给予认可和表彰。

全球人工智能产业规模

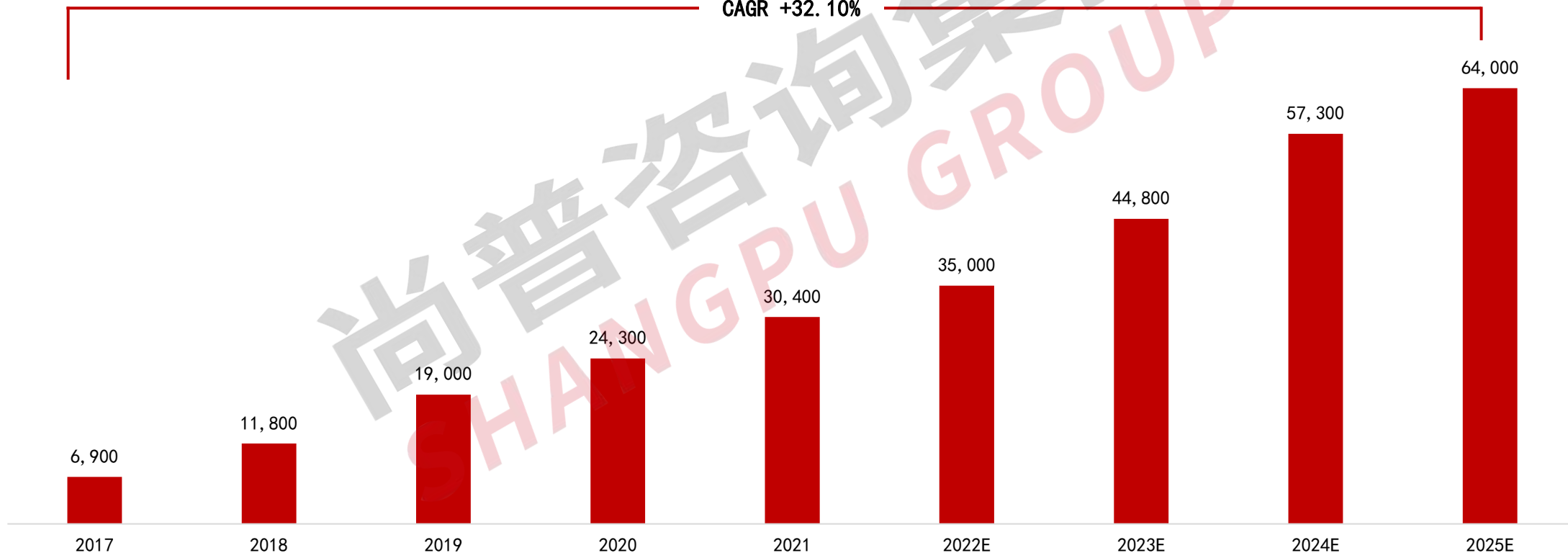
预计到2025年，全球人工智能产业规模将达到6.4万亿美元

近年来，人工智能技术得到快速发展，其对经济社会发展以及生产生活方式变革将产生重大影响。全球范围内美国、欧盟、英国、日本、中国等国家和地区均大力支持人工智能产业发展，相关新兴应用不断落地。根据Deloitte报告预测，全球人工智能产业规模从2017年的6,900亿美元增长至2025年的64,000亿美元，2017-2025年复合增长率32.10%，呈现较快增长走势。

2017-2025年全球人工智能产业规模

单位：亿美元

CAGR +32.10%



全球人工智能产业链&产业图谱

关于全球人工智能产业链及产业图谱的相关说明

1、全球人工智能产业链

- 本报告中的全球人工智能产业链主要分为人工智能底层基础；人工智能核心技术；人工智能应用领域。
- 人工智能底层基础主要包括AI芯片、AI开源框架、传感器、云计算、存储、通信设备、数据服务。
- 人工智能核心技术为产业链的核心环节，人工智能核心技术通常可分为机器学习、计算机视觉、自然语言处理、语音处理、知识图谱。
- 人工智能应用领域主要涉及安防、金融、医疗、教育、零售、工业、农业等场景。

2、全球人工智能产业图谱

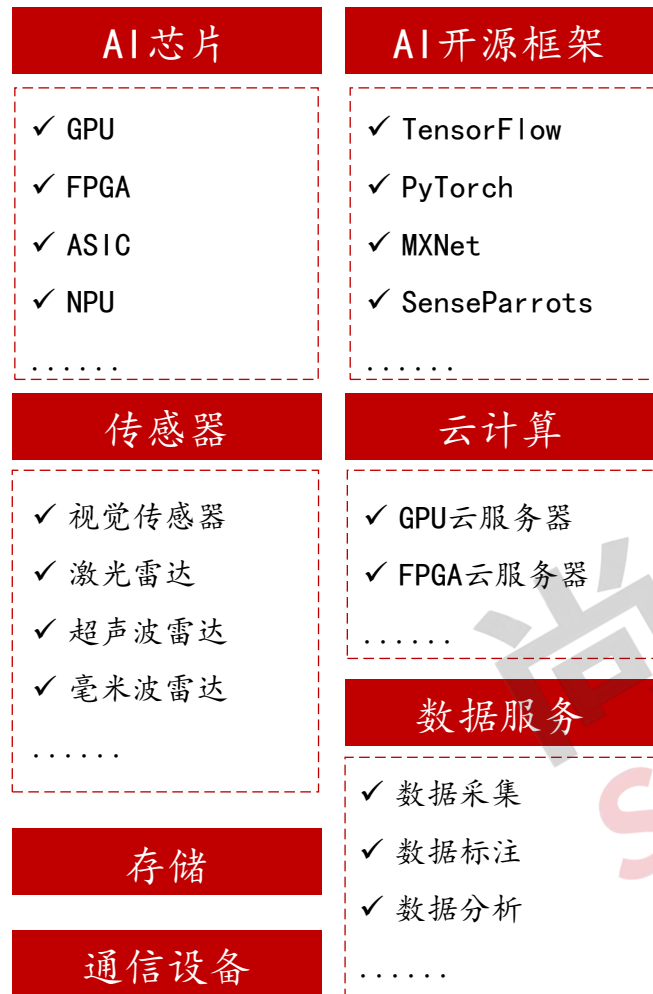
- 本报告中的全球人工智能产业图谱主要基于全球人工智能产业链结构，并按照细分领域体现全球及中国主要人工智能企业LOGO。
- 各细分领域的企业分布主要按照全球跨国企业、中国企业的顺序排列，同时也会综合考虑企业在细分领域的市占率、品牌知名度等因素，企业LOGO的排列顺序并不代表相关企业综合竞争力的强弱，仅供参考。

全球人工智能产业链



人工智能产业概况

人工智能底层基础



人工智能核心技术



人工智能应用领域



全球人工智能产业图谱—人工智能底层基础设施提供商



人工智能产业概况

AI芯片



AI开源框架



云计算



传感器



数据服务



存储



通信设备



注：企业LOGO排列顺序并不代表相关企业综合竞争力的强弱，仅供参考

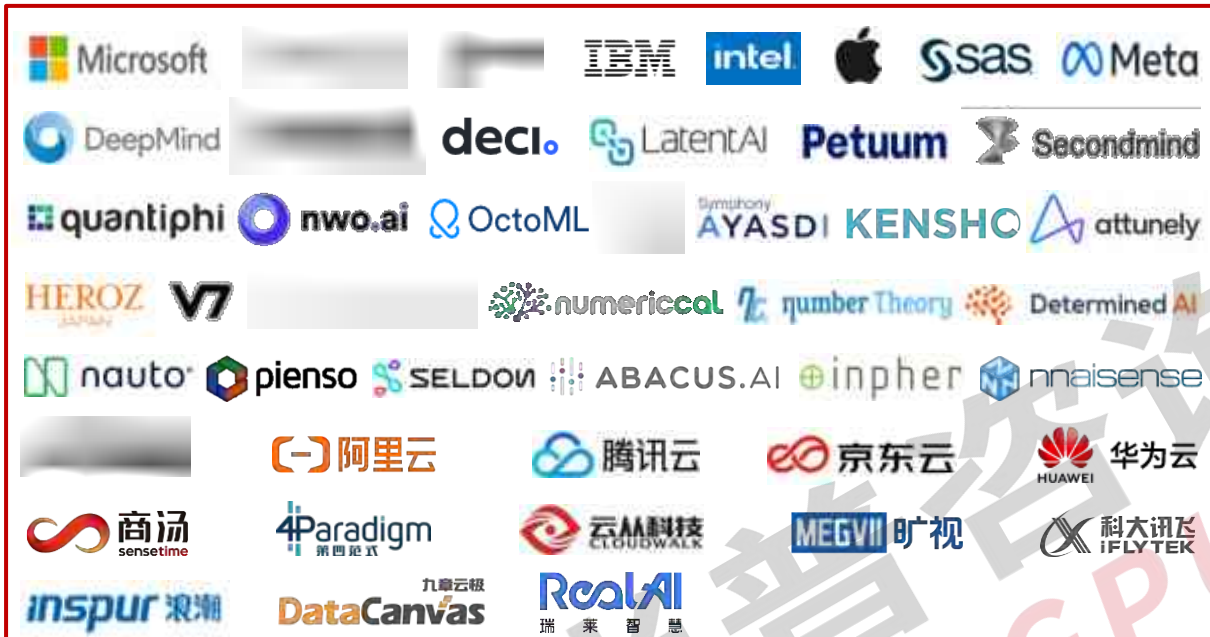
资料来源：尚普研究院结合公开资料整理绘制 31

全球人工智能产业图谱—人工智能核心技术提供商



人工智能产业概况

机器学习



计算机视觉



自然语言处理



知识图谱



语音处理



注：企业LOGO排列顺序并不代表相关企业综合竞争力的强弱，仅供参考

资料来源：尚普研究院结合公开资料整理绘制 32

全球人工智能产业图谱—人工智能应用服务商



人工智能产业概况

智能驾驶

机器人

算法	高精地图	智能座舱	智能驾驶方案	新势力车企

工业机器人	服务机器人	特种机器人

安防

金融

医疗

制造

教育

零售

客服

农业

注：企业LOGO排列顺序并不代表相关企业综合竞争力的强弱，仅供参考

资料来源：尚普研究院结合公开资料整理绘制 33

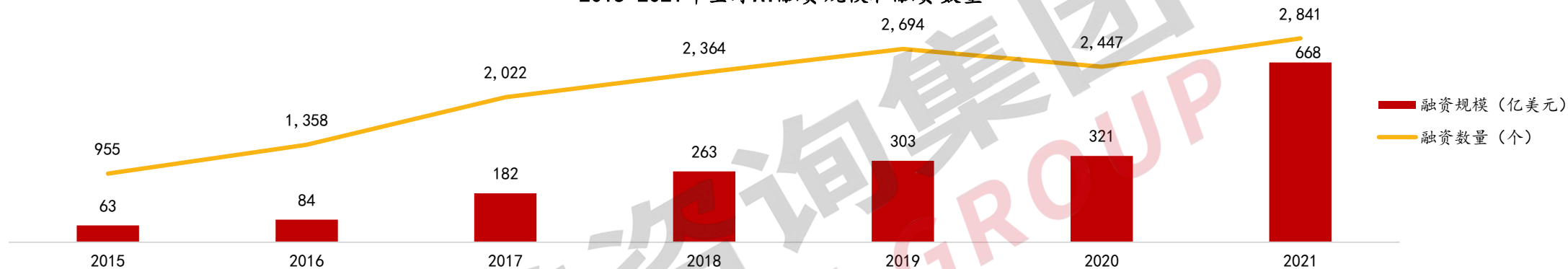
全球人工智能市场融资情况

全球人工智能融资规模和融资数量均呈现增长走势，医疗AI领域增长明显

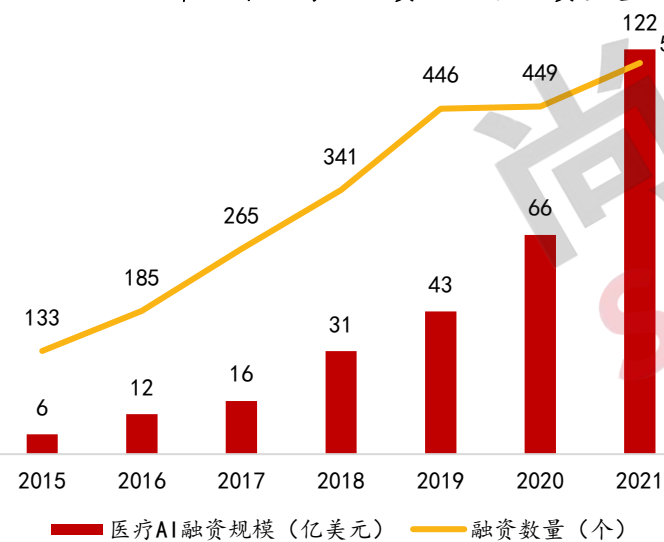
全球人工智能投资市场近年来快速发展，整体融资规模从2015年的63亿美元增长至2021年的668亿美元。2021年全球医疗AI融资规模较高达到122亿美元，金融科技AI融资规模为39亿美元，零售AI融资规模为37亿美元。

人工智能产业概况

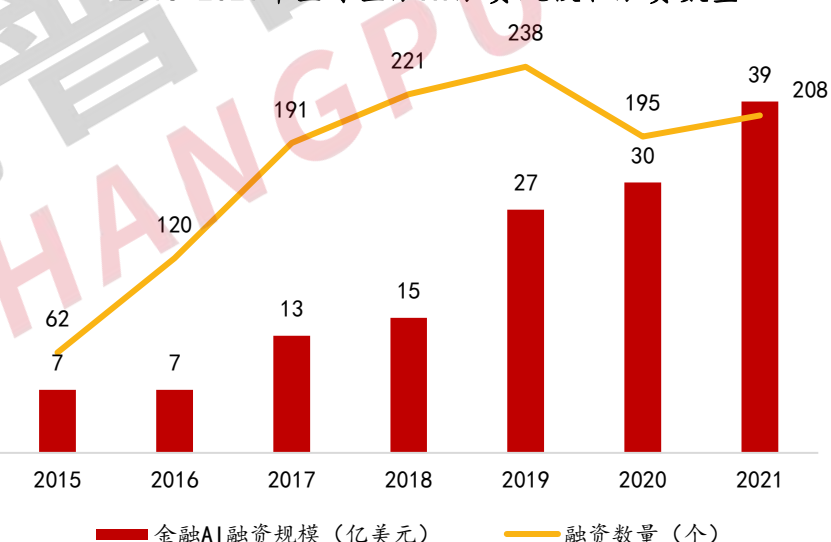
2015-2021年全球AI融资规模和融资数量



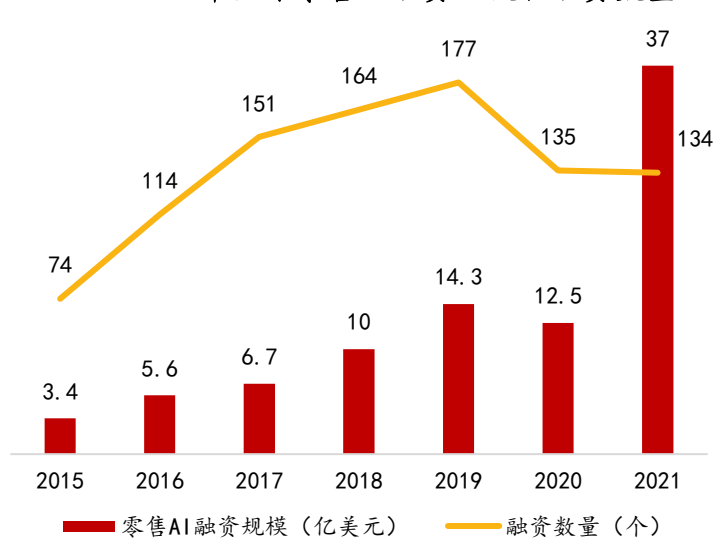
2015-2021年全球医疗AI融资规模和融资数量



2015-2021年全球金融AI融资规模和融资数量



2015-2021年全球零售AI融资规模和融资数量

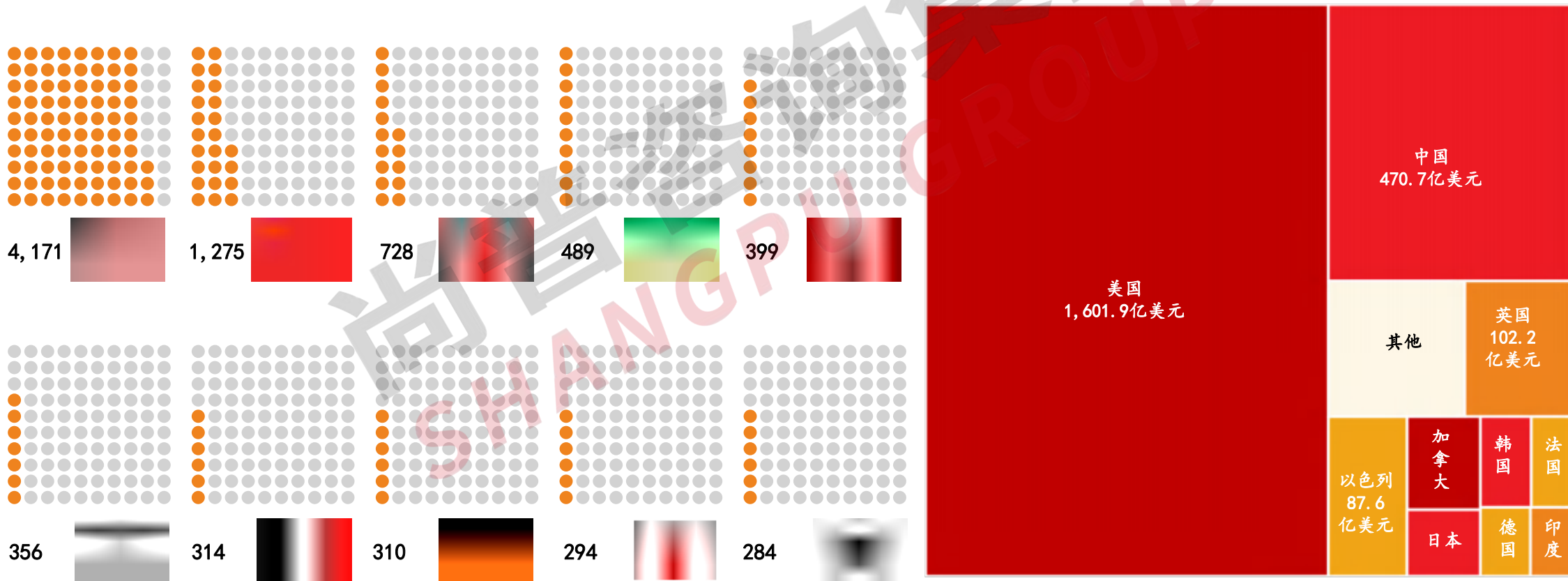


全球主要国家人工智能企业数量与累计融资规模

全球人工智能产业集中度高，美、中两国在AI企业数量与累计融资规模方面处于领先地位

尚普研究院根据CB Insights数据统计，全球现有人工智能企业超过11,000家，累计融资总额超过2,500亿美元。其中，美国拥有AI相关企业达到4,171家，累计融资金额达到1,601.9亿美元，在公司数量和融资规模上均位居世界首位；中国拥有1,275家AI公司，融资总金额为470.7亿美元，位居世界第二位。英国、印度、加拿大等国家也各自拥有百余家AI公司，AI企业数量排名前10位的国家占全球总数的78.3%，累计融资金额则达到全球总额的95%，AI企业和融资活动集中在美、中、英等国家。

全球主要国家AI企业数量与累计融资规模



注：相关数据截至2022年4月

资料来源：CB Insights, 尚普研究院 35

全球人工智能产业链主要上市企业情况

全球AI上市企业以科技巨头和细分领域技术龙头为主

尚普研究院将处于全球人工智能产业链中的典型上市公司进行整理，名单中以具备综合实力的科技巨头为主，国外如Google、Amazon、Microsoft等，国内如百度、阿里巴巴、腾讯等。此外，国内的商汤、科大讯飞等具有强技术属性的AI公司受关注度较高。

全球人工智能产业链主要上市企业名单

单位：亿美元

股票代码	企业名称	总部所在国家	市值	2021年营业收入	2021年净利润	产业链主要细分领域						
						AI芯片	云计算	AI技术	AI解决方案			
									智能驾驶	医疗	金融	制造
全球跨国企业												
GOOG.O	谷歌 (ALPHABET)	美国	15,081.93	2,576.37	760.33	✓	✓	✓	✓	✓	✓	✓
AMZN.O	亚马逊 (AMAZON)	美国	12,644.91	4,698.22	333.64	✓	✓	✓	✓	✓	✓	✓
MSFT.O	微软 (MICROSOFT)	美国	20,755.81	1,680.88	612.71		✓	✓	✓	✓	✓	✓
AAPL.O	苹果 (APPLE)	美国	25,515.94	3,658.17	946.8	✓		✓				
NVDA.O	英伟达 (NVIDIA)	美国	4,655.30	269.14	97.52	✓		✓	✓	✓	✓	✓
INTC.O	英特尔 (INTEL)	美国	1,782.40	790.24	198.68	✓		✓	✓	✓	✓	✓
AMD.O	超威半导体 (AMD)	美国	1,385.86	164.34	31.62	✓		✓	✓	✓		✓
QCOM.O	高通 (QUALCOMM)	美国	1,564.53	335.66	90.43	✓		✓	✓	✓		✓
005930.KS	三星 (SAMSUNG)	韩国	3,166.52	—	—	✓		✓				
IBM.N	国际商业机器公司 (IBM)	美国	1,188.98	573.50	57.43	✓	✓	✓		✓	✓	✓
TSLA.O	特斯拉 (TESLA)	美国	9,021.16	538.23	55.19	✓		✓	✓			
中国大陆企业												
9888.HK	百度集团	中国	460.10	195.26	16.04	✓	✓	✓	✓	✓	✓	✓
9988.HK	阿里巴巴	中国	2,821.94	—	—	✓	✓	✓	✓	✓	✓	✓
0700.HK	腾讯控股	中国	4,623.08	880.45	352.62	✓	✓	✓	✓	✓	✓	✓
9618.HK	京东集团	中国	1,057.27	1,492.53	-5.58		✓	✓	✓		✓	✓
0020.HK	商汤集团	中国	221.20	7.37	-26.88			✓	✓	✓	✓	✓
002230.SZ	科大讯飞	中国	129.95	28.53	2.44			✓		✓	✓	✓
688256.SH	寒武纪	中国	31.41	1.13	-1.29	✓		✓				

注：市值数据截至2022.04.29 企业排列顺序并不代表相关企业综合竞争力的强弱，仅供参考

资料来源：Wind万得资讯，尚普研究院结合公开资料整理绘制 36



2

人工智能底层基础

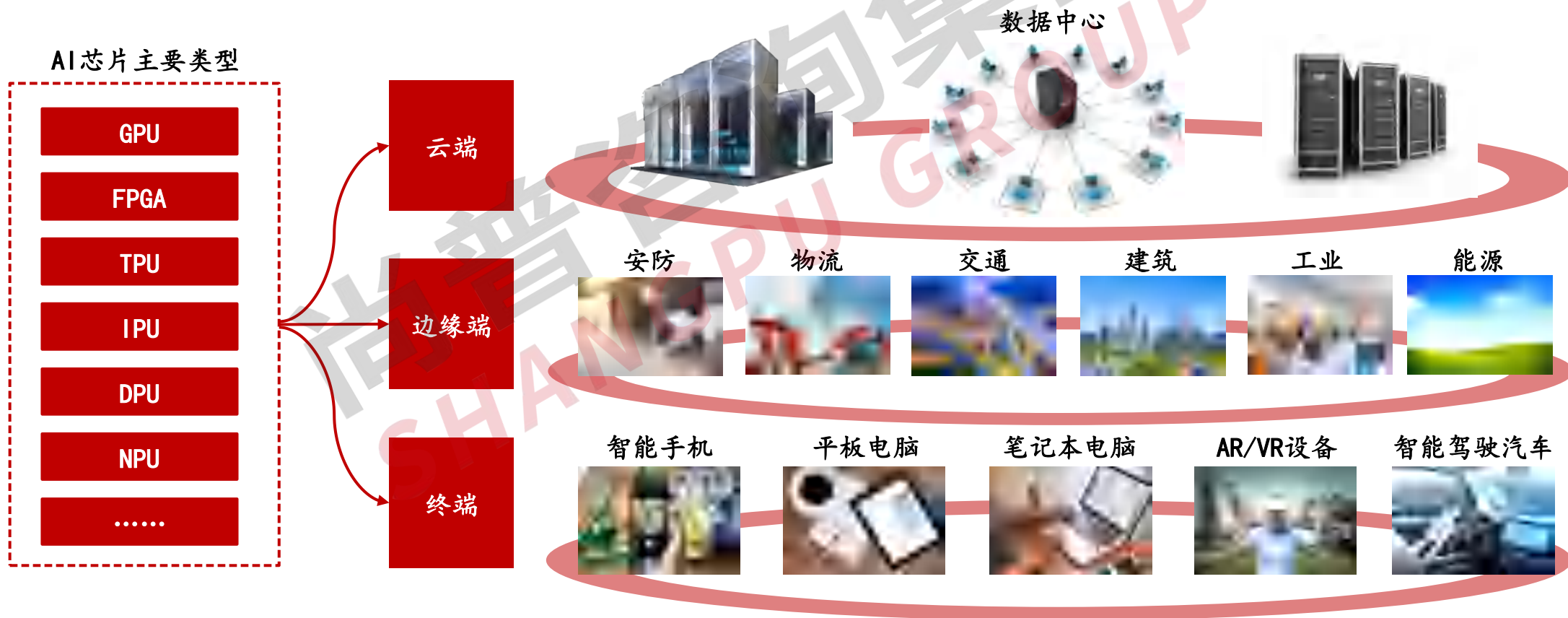
Foundation of Artificial Intelligence

AI芯片总览

AI芯片针对机器学习算法设计开发，广泛应用于云、边、端各类场景

AI芯片 (AI Chip): 专门用于处理人工智能相关的计算任务，其架构针对人工智能算法和应用进行专门优化，具有高效处理大量结构化和非结构化数据的特征，可高效支持视觉、语音、自然语言处理等智能处理任务。当前AI芯片类型主要涉及GPU、FPGA、TPU、IPU、DPU、NPU等类别。从应用场景来看，AI芯片广泛应用于云端、边缘端、终端等各类场景，其中云端AI芯片具备高性能特征，终端AI芯片具有低功耗和高能效特性，而边缘端AI芯片性能介于云端和终端之间。

人工智能底层基础

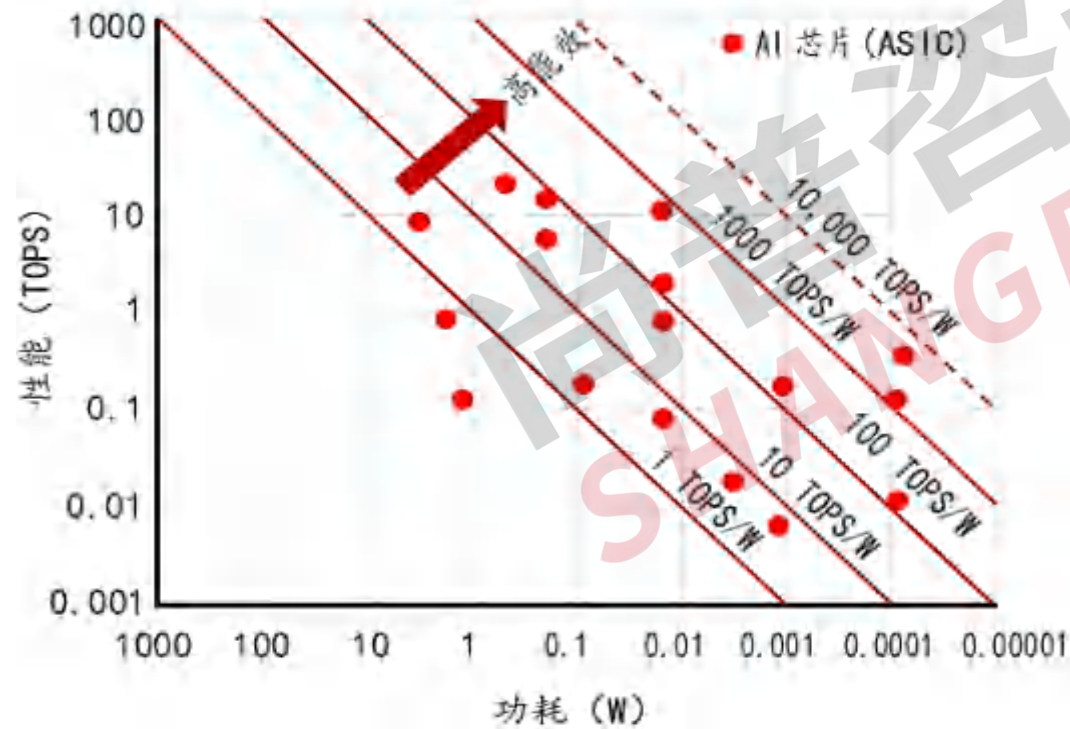


AI芯片性能指标评价

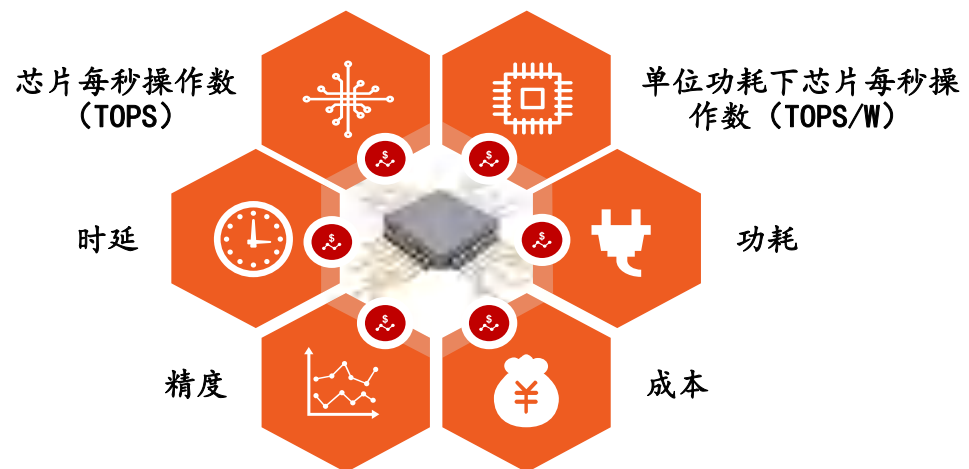
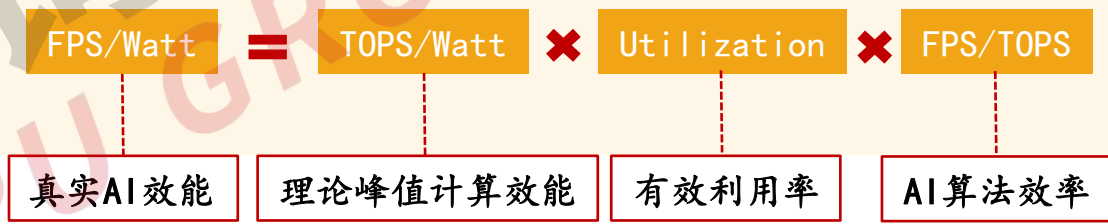
评价AI芯片性能需重点关注TOPS/W、时延、功耗等相关指标

目前，评价AI芯片性能的指标主要包括：TOPS、TOPS/W、时延、成本、功耗、可扩展性、精度、适用性、吞吐量、热管理等。其中，TOPS/W是用于度量在1W功耗的情况下，芯片能进行多少万亿次操作的关键指标。近年来，MIT、Nvidia等研发人员开发专门的芯片软件评价工具对于AI芯片的性能做出了系统、全面评价，如Accelergy（评估芯片架构级能耗）、Timeloop（评估芯片运算执行情况）等。MLPerf是由来自学术界、研究实验室和相关行业的AI领导者组成的联盟，旨在“构建公平和有用的基准测试”，可用于衡量深度学习软件框架、AI芯片和云平台性能。

AI芯片的性能、功耗和能效关系



影响AI芯片效能主要因素



资料来源：人民邮电出版社《AI芯片—前沿技术与创新未来》2021年4月第1版，地平线会议，尚普研究院结合公开资料整理绘制 39

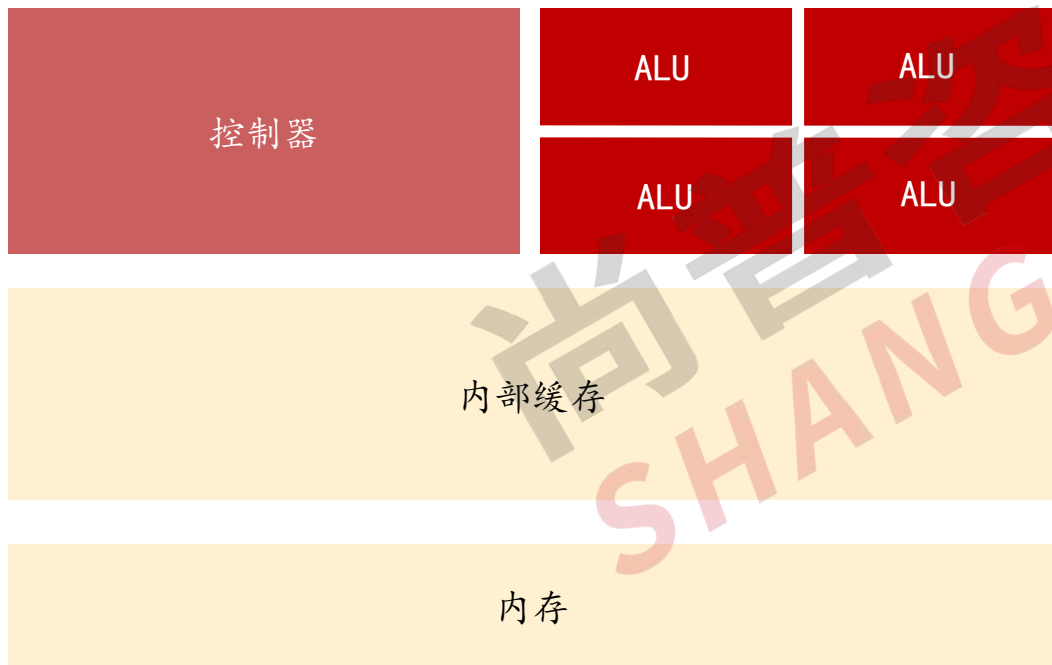
AI芯片：云端场景—数据中心

GPU具备矩阵和大规模并行计算优势，适合数据中心场景

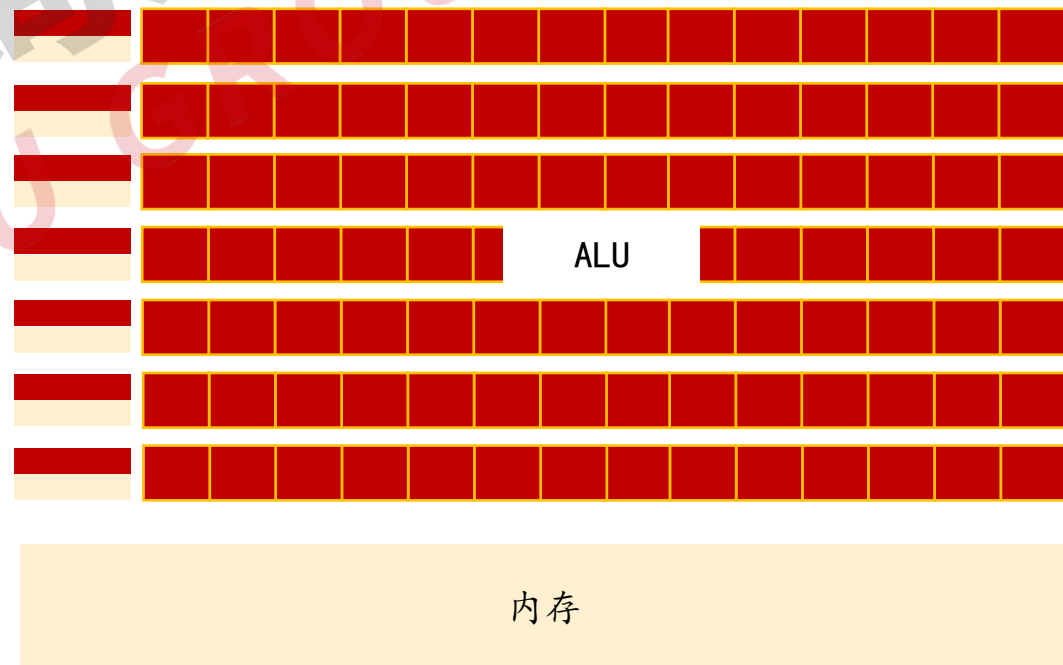
CPU作为通用处理器，包含控制单元（指令读取及指令译码）、存储单元（CPU片内缓存和寄存器）及运算单元（ALU约占20%CPU空间）三个主要模块，但受制于成本/功耗/技术难度/算力瓶颈等问题，目前仍未出现适配AI高算力要求的主流CPU产品。GPU是一种由大量核心组成的大规模并行计算架构，拥有较多运算单元（ALU）和较少缓存（cache），是一款专为同时处理多重任务而设计的芯片，具备良好的矩阵计算能力和并行计算优势，能满足深度学习等AI算法的处理需求，成为主流云端AI芯片。

人工智能底层基础

CPU内部架构



GPU内部架构



AI芯片：云端场景—数据中心

张量计算单元是GPU进行深度学习运算的核心组成部分

英伟达A100芯片为多个SM单元(Streaming Multiprocessors, 流式多处理器)构成的并发多核处理器, 不同SM单元共享L2 Cache存储资源进行数据访存, A100的安培架构中有128个SM核, SM结构是芯片架构升级的核心。Tensor Core是英伟达GPU架构中专为深度学习矩阵运算设置的张量计算单元, 是英伟达GPU系列深度学习运算加速的核心。Tensor Core处理的是大型矩阵运算, 其执行一种专门的矩阵数学运算, 适用于深度学习和某些类型的高性能计算。Tensor Core功能是执行融合乘法和加法的运算, 其中两个4*4 FP16矩阵相乘, 然后将结果添加到4*4 FP32或FP64矩阵中, 最终输出新的4*4 FP32或FP64矩阵。

人工智能底层基础

NVIDIA A100 Tensor Core GPU安培架构



NVIDIA A100 Tensor Core GPU-SM单元



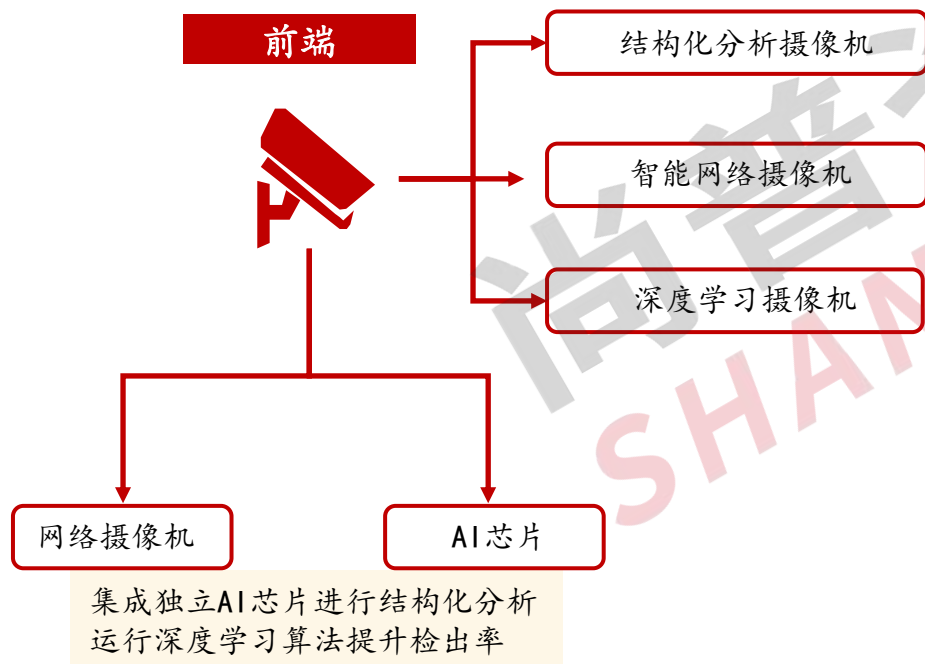
AI芯片：边缘端场景—AIoT

边缘端集成AI芯片可以实现本地化数据的实时处理

AIoT通过融合AI和IoT技术实现万物智联，其主要涉及到安防、移动互联网等多样化场景。在智慧安防方面，由于终端摄像头每天产生大量的视频数据，若全部回传到云数据中心将会对网络带宽和数据中心资源造成极大占用。通过在终端加装AI芯片，可实现数据本地化实时处理，即仅将经过结构化处理、关键信息提取后带有关键信息的数据回传云端，大大降低网络传输带宽压力。当前主流解决方案为前端摄像头设备内集成AI芯片，在边缘端采用智能服务器级产品，后端在边缘服务器中集成智能推理芯片。目前国内外企业正在加大对边缘端AI视觉处理芯片的研发和投入，相关芯片产品如英伟达Jetson AGX Xavier、地平线旭日3、华为海思Hi3559A V100等。

人工智能底层基础

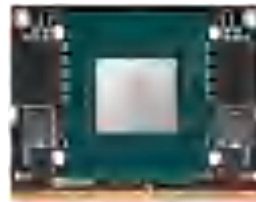
AI芯片在智能安防摄像头应用



典型安防AI芯片



NVIDIA Jetson AGX Xavier



NVIDIA Jetson Xavier NX

典型安防AI芯片指标

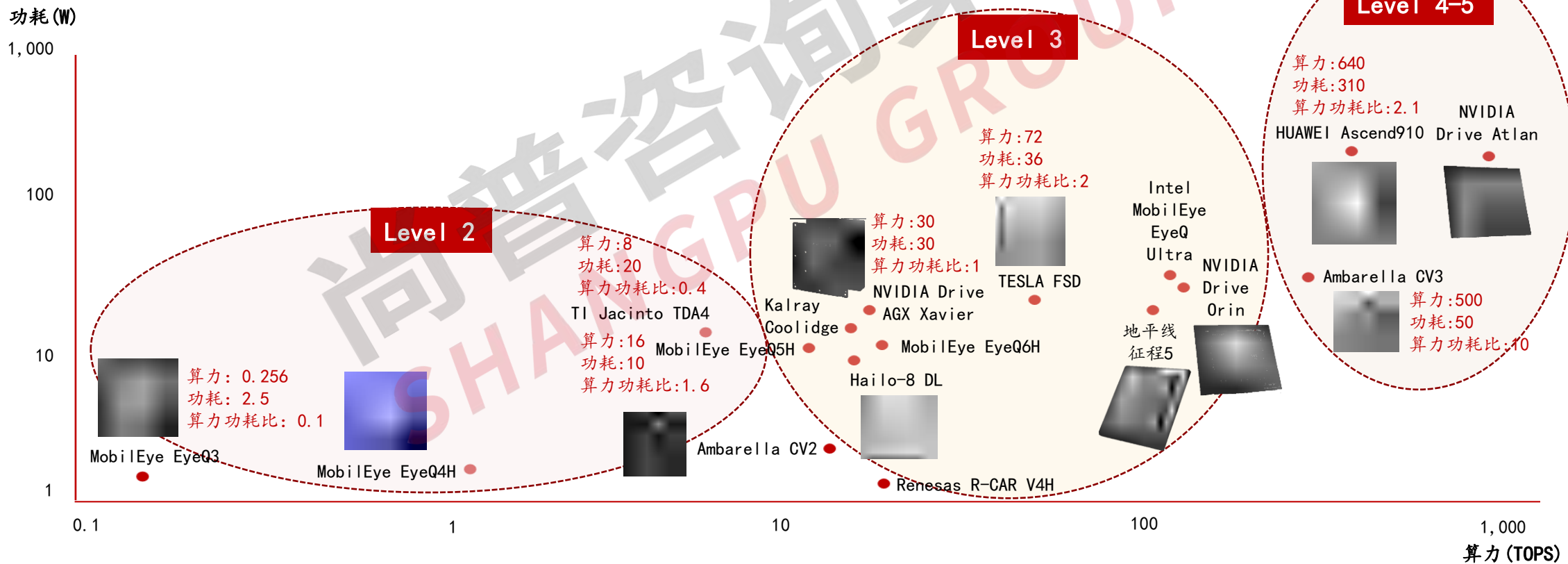
厂商	产品	算力	功耗
英伟达	TESLA P4	22TOPS	50W/75W
英伟达	JETSON XAVIER NX	21TOPS	15W/20W
地平线	旭日3	5TOPS	2.5W
华为海思	Hi3559A V100	4TOPS	3W
芯微电子	AR9341	4TOPS	2W

AI芯片：终端场景—智能驾驶

随着智能驾驶等级的提高，技术不断迭代促使车用AI芯片性能逐步增强

SAE（国际自动机工程师学会）将自动驾驶划分为六个级别：L0（非自动化）、L1（驾驶员辅助）、L2（部分自动化，如交通拥堵辅助和高级紧急制动+转向）、L3（有条件自动化，如高速公路自动驾驶）、L4（高度自动化，如城市自动驾驶）和L5（完全自动化，如全场景自动驾驶）。从L3开始每一级别均需要强大的算力进行实时分析，处理大量数据，执行复杂的逻辑运算，对计算能力有着极高要求。每一级别自动驾驶汽车所需的芯片计算能力一般认为是：L3约为250TOPS，L4超过500TOPS，L5超过1,000TOPS。随着芯片设计和制造工艺的提高，车用AI芯片正朝着高算力、低功耗的方向发展。

全球主要车用AI芯片概览



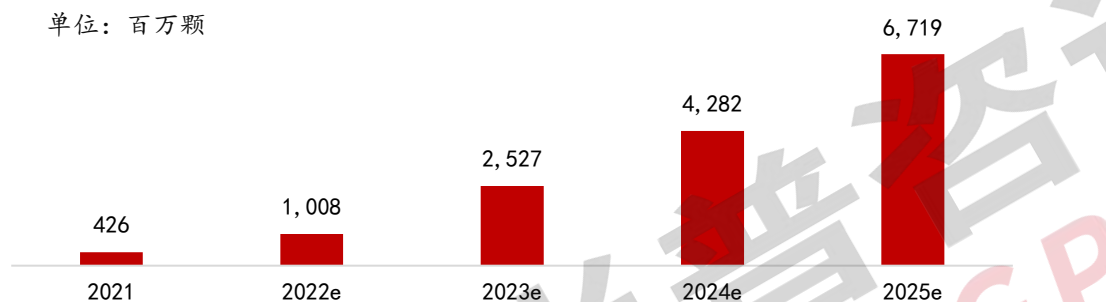
资料来源：各公司官网，尚普研究院结合公开资料整理绘制

AI芯片：终端场景—智能驾驶

预计到2025年，全球车用AI芯片市场规模将突破17亿美元

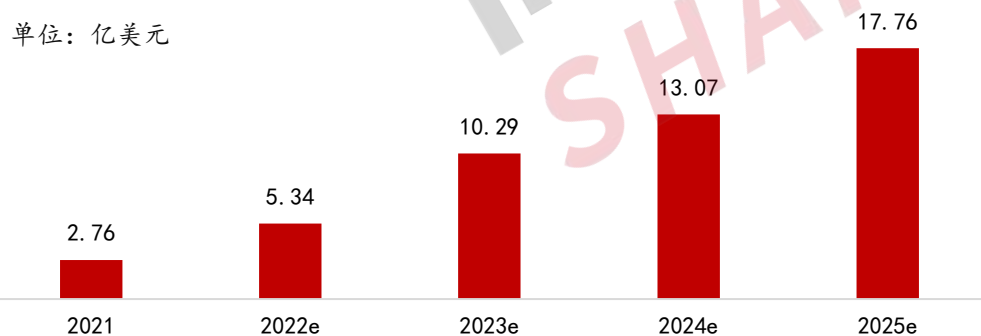
随着汽车控制方式逐渐由机械式转向电子式，每辆汽车对车用AI芯片需求提升，带动车用AI芯片长期发展。根据Yole预测，2025年全球车用AI芯片产量将达到67.19亿颗，市场规模将达到17.76亿美元，年复合增速分别达到99.28%和59.27%。此外，车用AI芯片逐渐往高性能方向发展，如英特尔2022年推出的EyeQ Ultra自动驾驶汽车芯片，基于经过验证的Mobileye EyeQ架构而打造，其含有8个PMA、16个VMP、24个MPC、2个CNN Accelerator视觉处理单元（VPU），通过优化算力和效能以达到176TOPS，可满足L4自动驾驶场景，该产品将于2023年底供货，预计在2025年全面实现车规级量产。

2021-2025年全球车用AI芯片产量



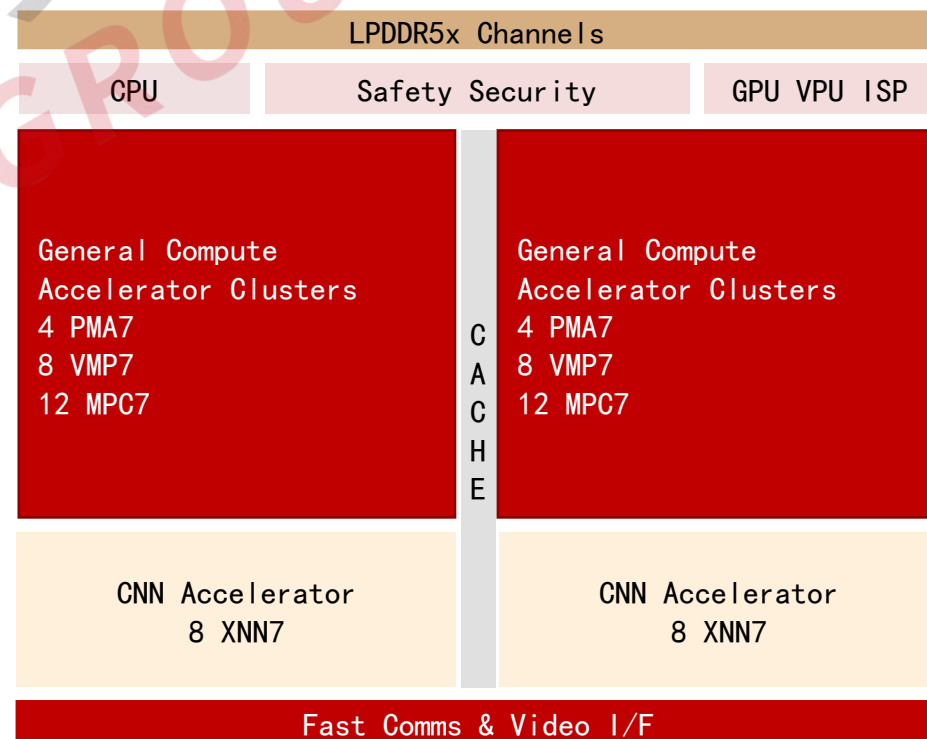
资料来源：Yole, 尚普研究院

2021-2025年全球车用AI芯片市场规模



资料来源：Yole, 尚普研究院

Mobileye EyeQ Ultra芯片结构



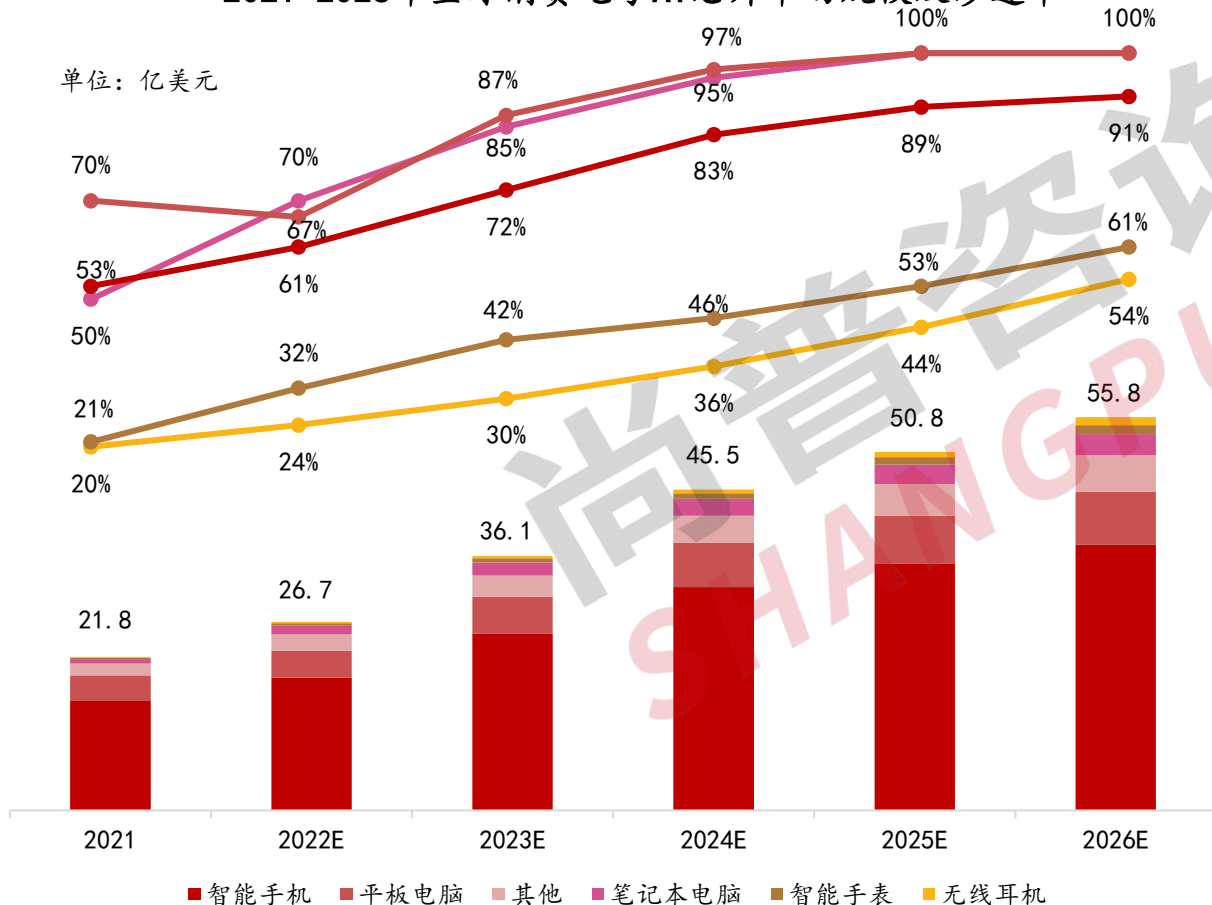
资料来源：Mobileye官网, 尚普研究院

AI芯片：终端场景—消费电子

2026年全球消费电子AI芯片市场规模将突破55亿美元

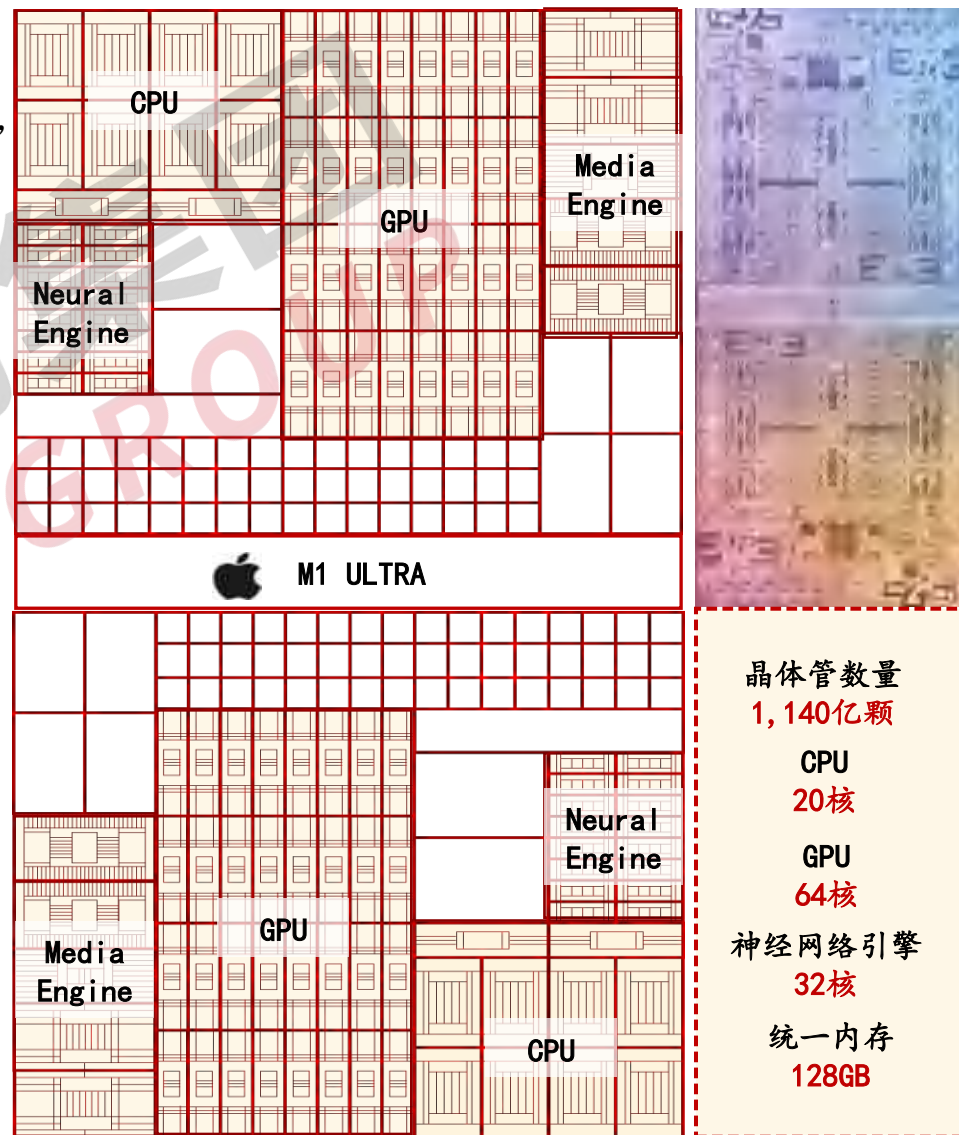
AI芯片在图像识别、语音识别和快速建立用户画像等方面具有重要作用。根据Yole预测，2026年全球消费电子AI芯片市场规模将达到55.8亿美元，其中笔记本电脑、平板电脑和智能手机AI芯片渗透率将分别达到100%、100%和91%，未来全球消费电子AI芯片市场规模和渗透率呈现逐步增长态势。

2021-2026年全球消费电子AI芯片市场规模及渗透率



资料来源：Yole, 尚普研究院

Apple M1 Ultra内部结构图



资料来源：Apple官网, 尚普研究院结合公开资料整理绘制

深度学习开源框架

深度学习开源框架是AI算法开发的基础工具

深度学习开源框架是AI算法模型设计、训练和验证的标准接口、特性库和工具包，集成数据调用、算法封装及计算资源的使用。国际上广泛使用的深度学习开源框架包括Google TensorFlow、Facebook PyTorch、Amazon MXNet和微软CNTK等。目前中国的深度学习开源框架主要包括百度PaddlePaddle、华为MindSpore等，已初步应用于工业、服务业等场景，服务200余万开发者。

典型深度学习开源框架



TensorFlow是谷歌开发的用于深度学习或人工神经网络的开源软件库，支持语言有Python、C++、Java等。



PyTorch由Facebook人工智能研究团队开发，它支持CUDA技术，可以在训练、分析和验证神经网络模型中充分利用专用GPU的功能。支持语言有Python、C++等。



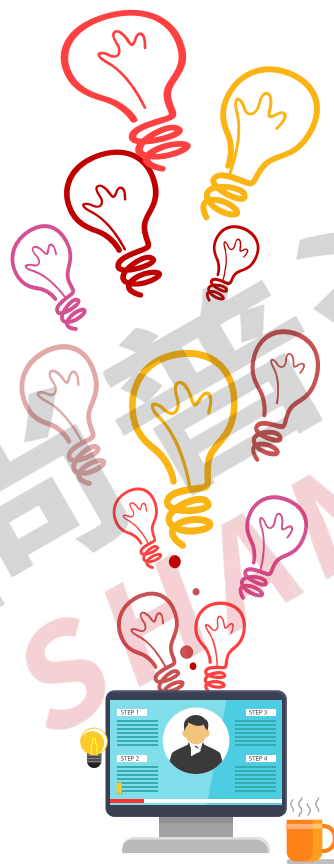
Theano是专为深度学习而设计的Python库，由蒙特利尔学习算法研究所（MILA）开发，支持语言有Python等。



CNTK是由微软研究院开发的深度学习框架，旨在使用神经网络来处理非结构化数据的大型数据集，支持语言涉及C++等。



MXNet由卡内基梅隆大学发起，后捐赠给Apache基金会，现成为Amazon AWS最主要的AI框架，支持语言有Python、C++、Java等。



华为推出昇思MindSpore，在全场景协同、可信赖方面有一定的突破；支持动态图和静态图的转换技术，可实现计算效率和灵活性的平衡。



百度开发的PaddlePaddle飞桨深度学习框架，面向深度学习设计，支持语言有Python和C++等。



旷视推出天元MegEngine，基于计算图的深度神经网络学习框架，在训练推理一体化方面深度布局，支持语言有C++和Python等。



OneFlow是由北京一流科技开发的一款深度学习框架，是专门针对深度学习打造的异构分布式流式系统。



Jittor是由清华大学团队研发的完全基于动态编译的深度学习框架，其内部使用创新的元算子和统一计算图的深度学习框架。



Github社区主流深度学习框架

序号	框架名称	活跃度	关注度	贡献人数
1	TensorFlow	124,494	163,000	3,056
2	PyTorch	43,390	53,700	2,137
3	Theano	28,127	9,500	352
4	CNTK	16,116	17,100	201
5	MXNet	11,776	19,800	868
6	MindSpore	37,308	2,700	267
7	PaddlePaddle	33,753	17,500	524
8	MegEngine	2,282	4,100	32
9	OneFlow	7,621	3,000	99
10	Jittor	1,266	2,300	31

注：相关数据截至2022年1月

资料来源：各开源框架官网，中国信通院《AI框架发展白皮书2022》，尚普研究院结合公开资料整理绘制

人工智能数据服务

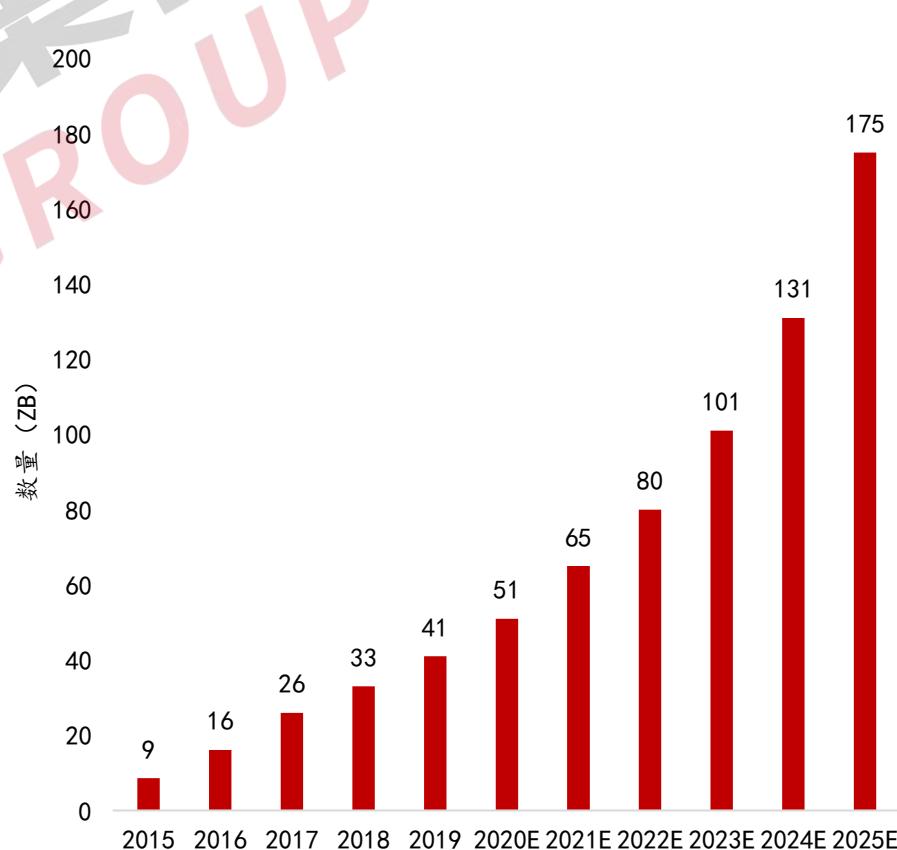
数据服务以AI训练与调优为目的，涉及数据采集、标注与质检等环节

人工智能数据服务包括数据库设计、数据采集、数据清洗、数据标注与数据质检服务。整个数据服务流程围绕客户需求展开，最终产出产品以数据集和数据资源定制服务为主，为AI模型训练提供可靠、可用的数据。随着短视频、直播、社交电商等应用快速兴起，全球数据量快速增长。根据IDC预测，全球数据量将从2015年的9ZB增加到2025年的175ZB，为人工智能技术持续迭代提供重要底层基础。

人工智能数据服务流程与主要产品



全球数据量



资料来源：尚普研究院结合公开资料整理绘制

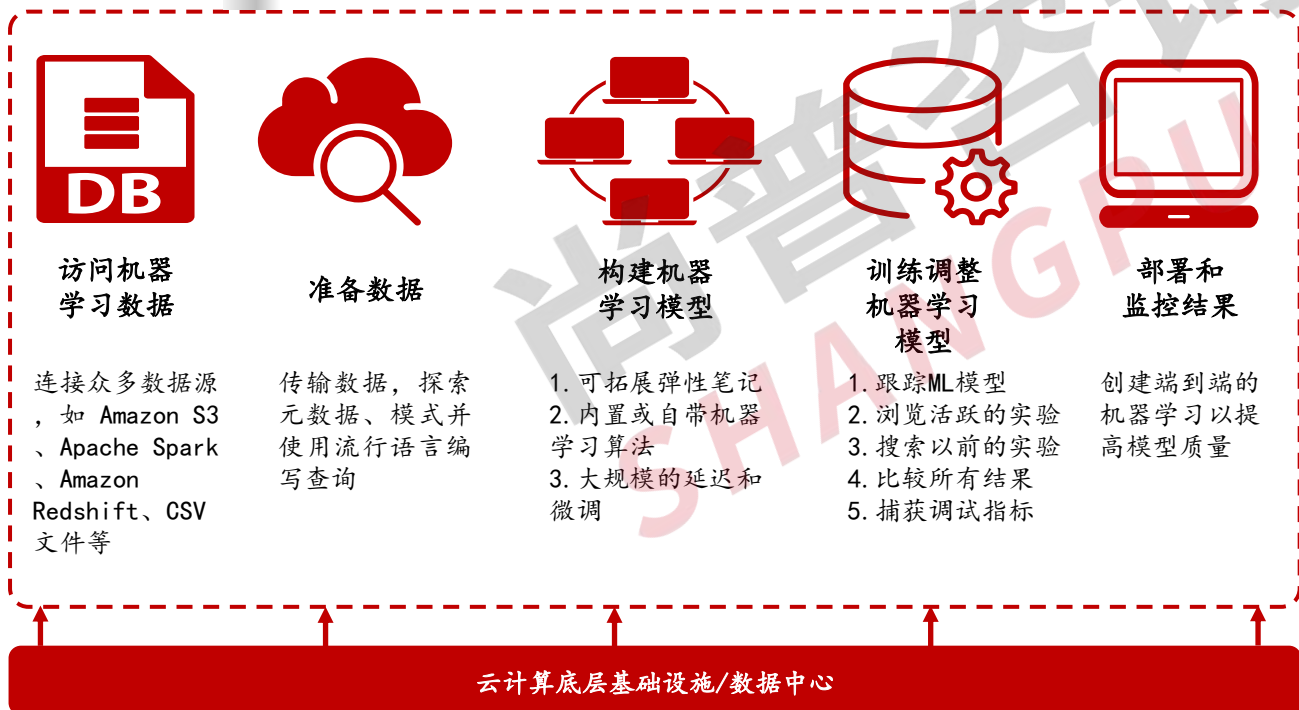
资料来源：IDC，尚普研究院

云计算服务

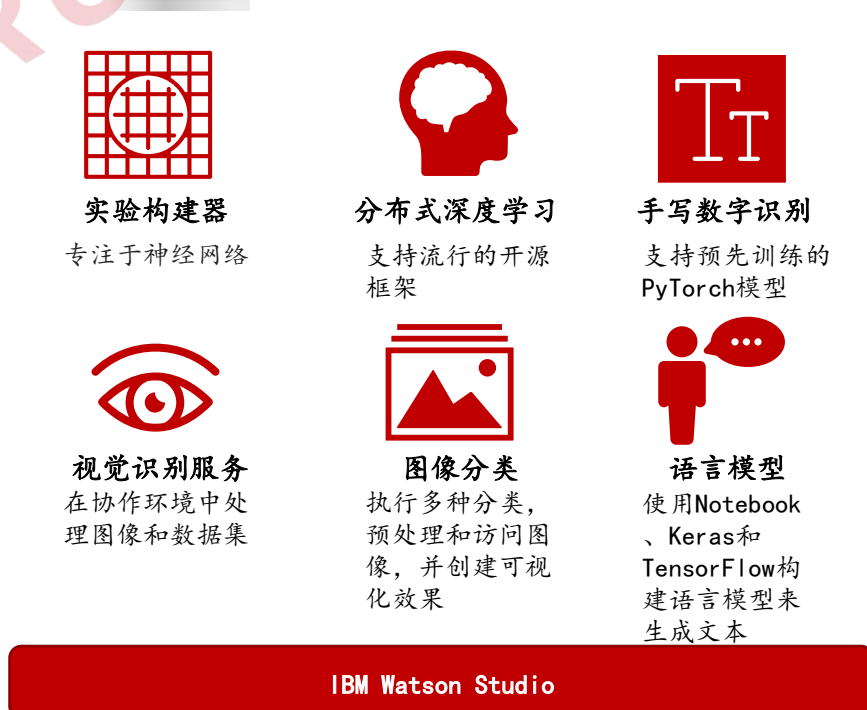
云计算服务显著降低人工智能算法开发成本，缩短产品开发周期

由于人工智能的开发和应用对于算力、数据有较大需求，云计算服务可以为开发者提供智能服务器集群等强大算力设施的租用，同时直接提供已经训练好的人工智能功能模块等产品，通过多元化的服务模式，降低开发者的开发成本和产品开发周期，为客户提供AI赋能。如亚马逊SageMaker，其可提供图片/图像分析、语音处理、自然语言理解等相关服务，使用者无需了解参数和算法即可实现功能的应用。随着底层技术的发展，IBM推出深度学习即服务（DLaaS），借助此项服务用户可以使用主流框架来训练神经网络，如TensorFlow、PyTorch及Caffe，无需购买和维护成本高昂的硬件，其每一个云计算处理单元都遵循简单易用的原则而设置，无需用户对基础设施进行管理，用户可以根据支持的深度学习框架、神经网络模型、训练数据、成本约束等条件进行挑选，然后DLaaS服务会帮助完成其余的事情，提供交互式、可迭代的训练体验。

亚马逊SageMaker机器学习流程



IBM深度学习即服务功能



资料来源：亚马逊AWS官网，尚普研究院结合公开资料整理绘制

资料来源：IBM官网，尚普研究院结合公开资料整理绘制

3

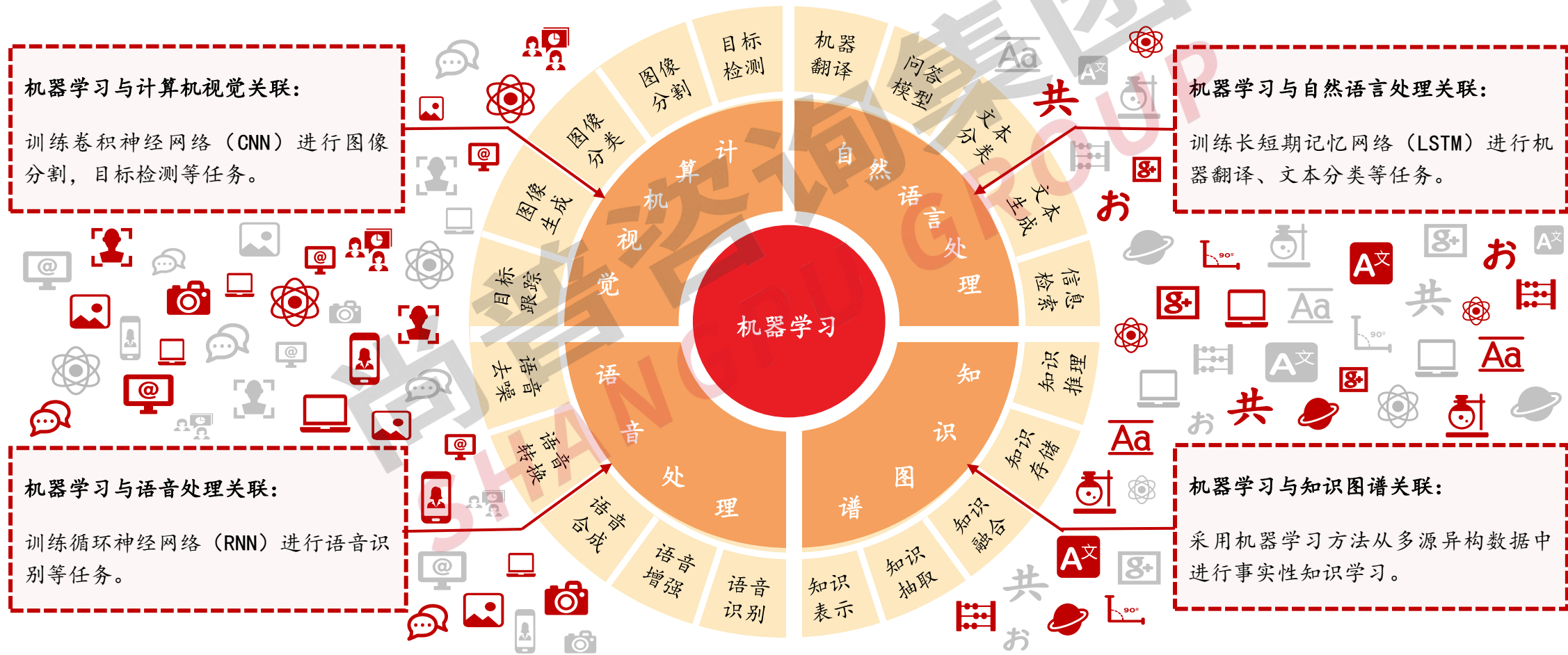
人工智能核心技术

Artificial Intelligence Core Technology

人工智能核心技术总览

机器学习作为人工智能的核心，与计算机视觉、自然语言处理、语音处理和知识图谱密切相关。机器学习作为人工智能技术的核心，近年来实现明显突破。机器学习与计算机视觉、自然语言处理、语音处理和知识图谱等关键技术紧密结合，相关机器学习算法主要应用于图像分类、语音识别、文本分类等相关场景中，从而提升人工智能技术的整体应用效果，使得人工智能技术在金融、医疗、交通等各领域实现广泛应用。

人工智能核心技术



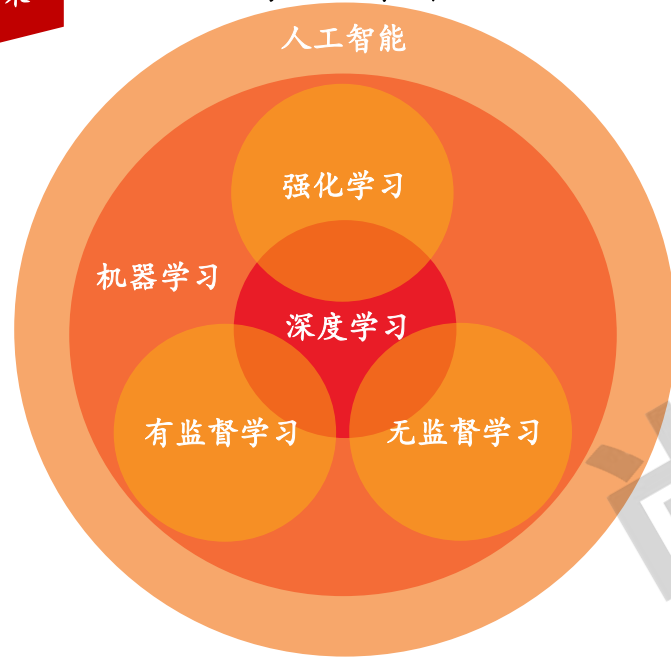
人工智能核心技术：机器学习总览

机器学习主要包括有监督学习、无监督学习和强化学习三类

机器学习是实现人工智能的核心方法，专门研究计算机如何模拟/实现生物体的学习行为，获取新的知识技能，利用经验来改善特定算法的性能。深度学习是机器学习算法的一种，深度学习算法具有多层神经网络结构，其在图像识别、语音处理等领域取得划时代的成果。根据学习范式的不同，机器学习可划分为有监督学习、无监督学习、强化学习。

人工智能、机器学习、深度学习关系图

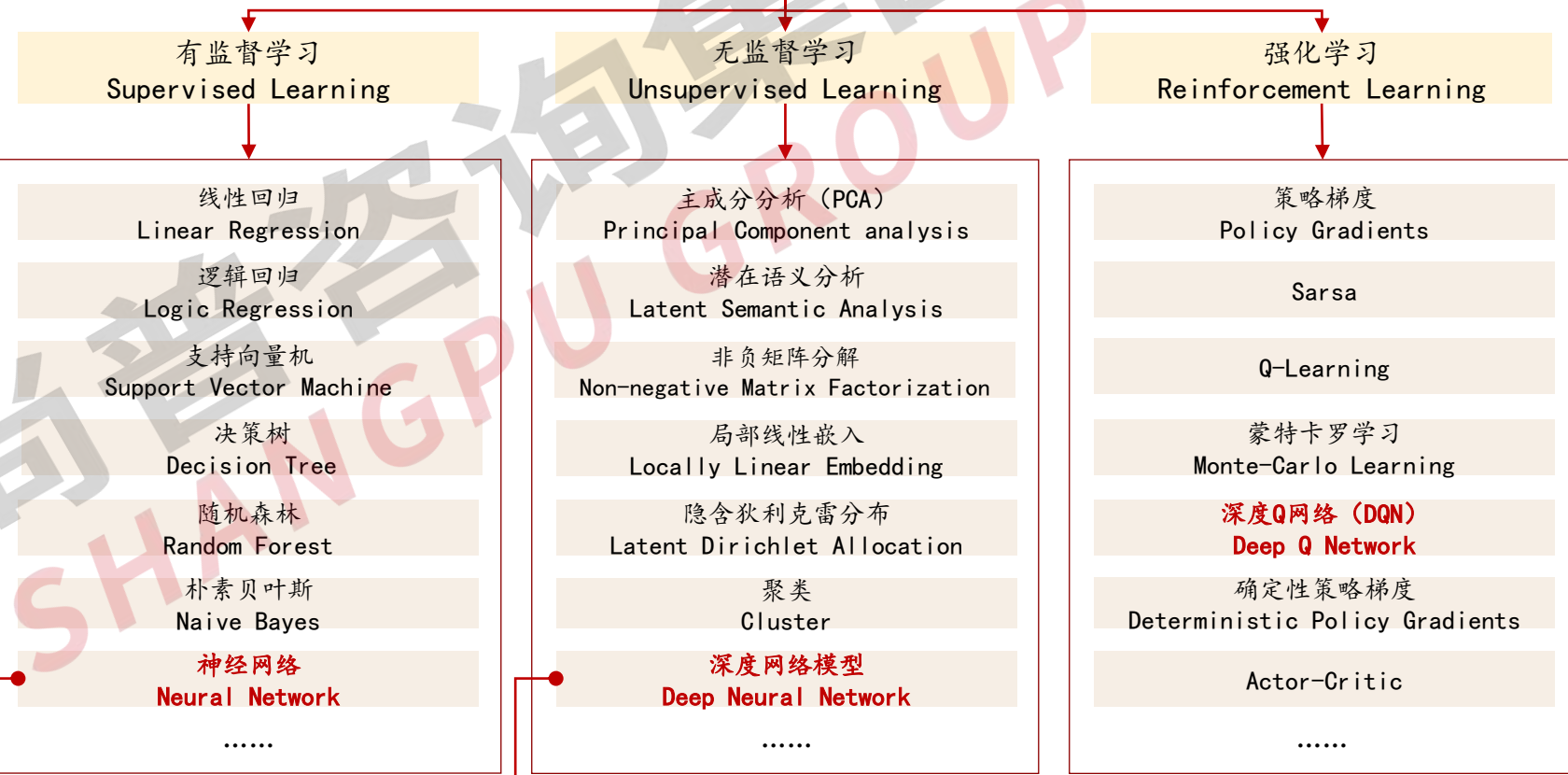
人工智能核心技术



DNN, CNN, RNN, GNN.....

AE, GAN.....

机器学习 Machine Learning



注：神经网络、深度网络模型、深度Q网络属于深度学习范畴

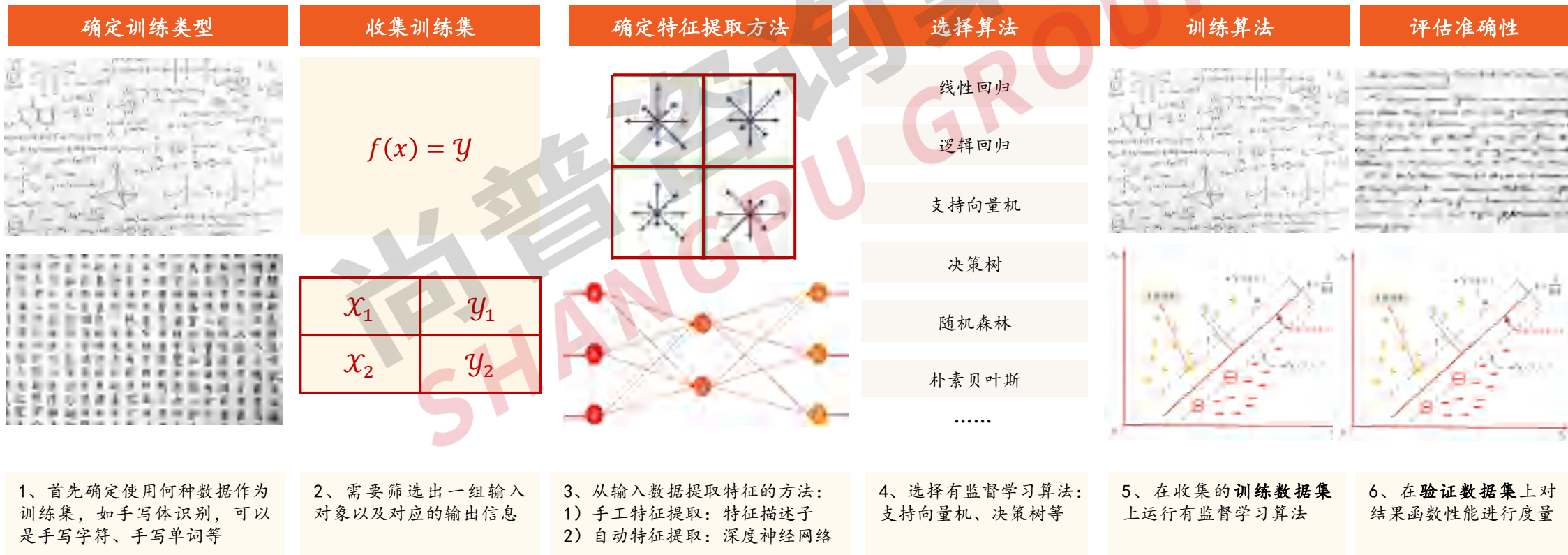
资料来源：尚普研究院结合公开资料整理绘制

人工智能核心技术：机器学习分类—有监督学习

有监督学习：从有标注训练数据中推导出预测函数

有监督学习 (Supervised Learning)：属于机器学习任务的一种类型，其主要是从给定的训练数据集中学习出一个函数（模型参数），当新的数据到来时，可以根据该函数预测结果。有监督学习的训练集要求包括输入和输出，或称为特征和目标，其中训练集中的目标是由人工标注的。有监督学习一般包括确定训练类型、收集训练集、确定特征提取方法、选择算法、训练算法、评估准确性六大环节，通过已有的训练样本（即已知数据及其对应的输出）去训练得到一个最优模型，再利用该模型将所有的输入映射为相应的输出，对输出进行简单的判断从而实现分类目的。

有监督学习流程 (Supervised Learning)



人工智能核心技术：有监督学习—逻辑回归

逻辑回归：对事件发生或不发生进行二元分类

逻辑回归 (Logistic Regression, LR)：是指学习某事件发生概率的算法，可对某个事件发生/不发生进行二元分类。逻辑回归使用 Sigmoid函数，输出结果范围在 [0, 1] 之间，逻辑回归的目标是发现特征与特定结果可能性之间的联系。如下示例：根据学习时长预测学生是否通过考试，响应变量为“通过和未通过考试”。

逻辑回归模型 (Logistic Regression, LR)

x : 学习时长 **3 5 7 9** 

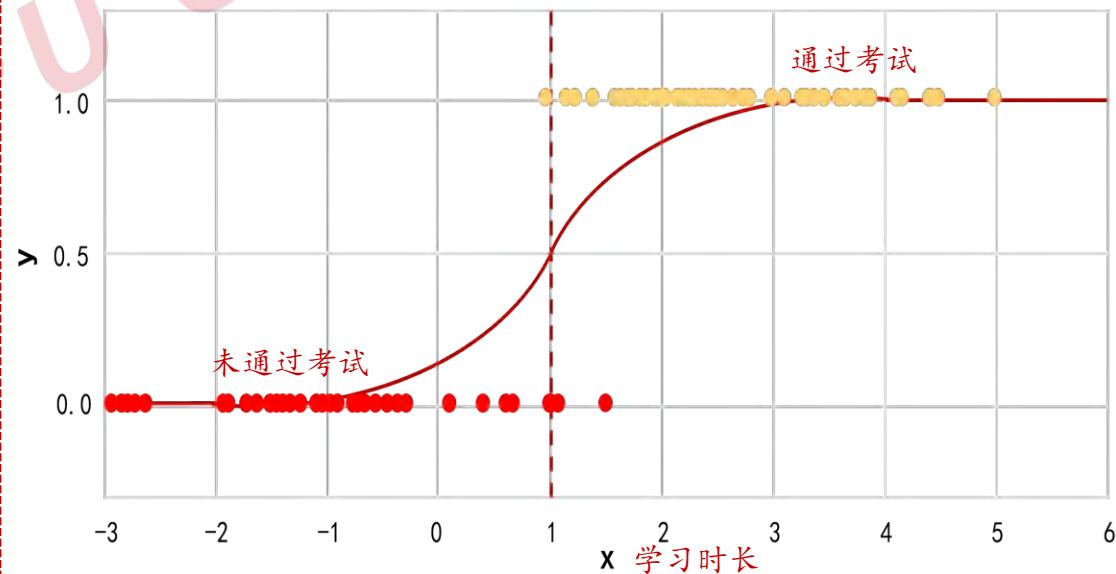
y = 0 未通过考试 (非预期结果) 

y = 1 通过考试 (预期结果) 

Sigmoid函数示例:
$$h(x) = \frac{1}{1 + e^{-x}}$$

二元分类通常将预测概率0.5作为阈值:

当概率 < 0.5 时, 将 y 的预测值分类为 0; 当概率 > 0.5 时, 将 y 的预测值分类为 1



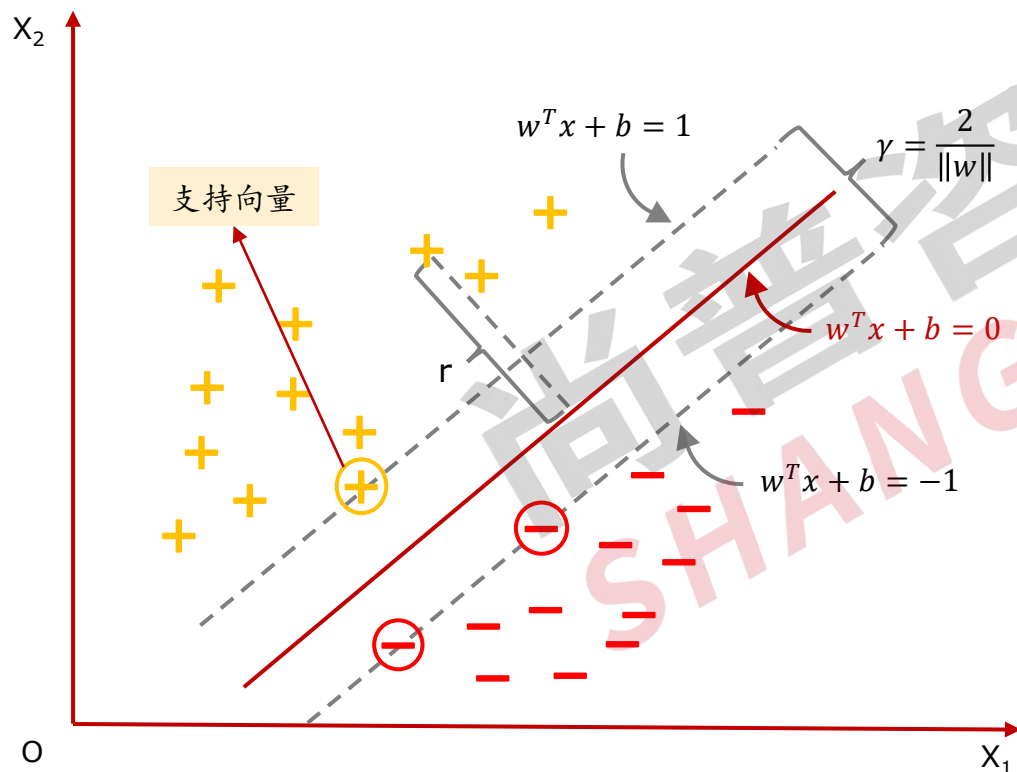
人工智能核心技术：有监督学习—支持向量机

支持向量机：以间隔最大化为基准学习远离数据的决策边界

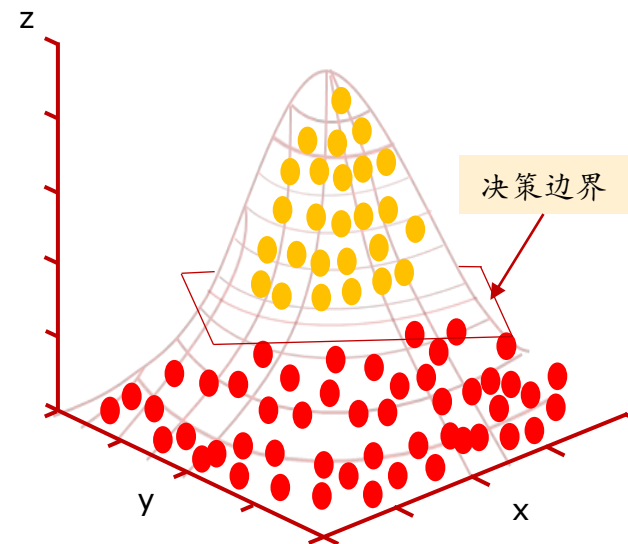
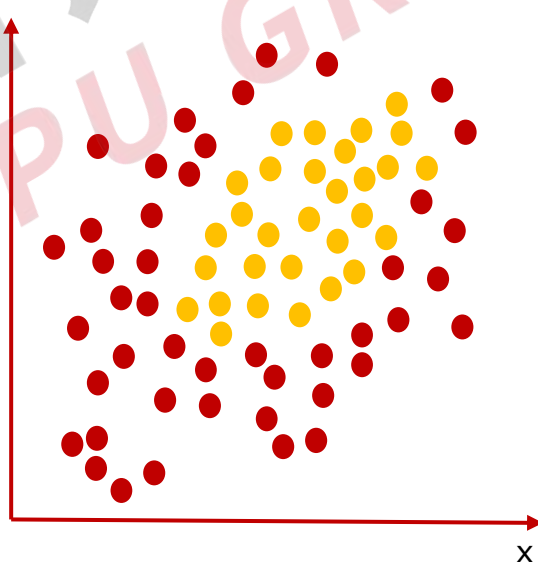
支持向量机 (Support Vector Machine, SVM)：主要是以间隔最大化为基准，学习得到尽可能远离数据的决策边界算法，支持向量是确定决策边界的重要数据。当训练样本线性可分时，通过硬间隔最大化，学习线性可分支持向量机；当训练样本近似线性可分时，通过软间隔最大化，学习线性支持向量机。在解决线性不可分问题时则会引入核函数，核方法主要将数据移动到另一个特征空间，然后进行线性回归，采用核方法的支持向量机使得原本线性不可分的数据在特征空间内变为线性可分。

人工智能核心技术

支持向量机 (Support Vector Machine, SVM)



核方法：模型可以在高维空间使用支持向量机学习决策边界，并将高维决策边界投影到原始特征形成的的向量空间，得到决策边界。



人工智能核心技术：有监督学习—决策树

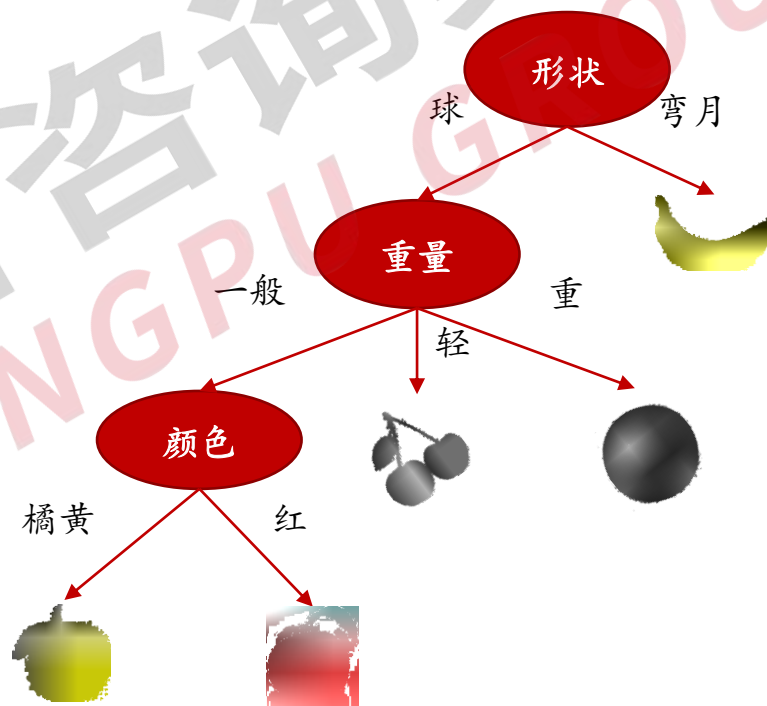
决策树：以树结构形式表达的预测分析模型

决策树 (Decision Tree) :是一种树状结构, 通过做出一系列决策(选择)来对数据进行划分, 这类似于针对一系列问题进行选择。一棵决策树一般包含一个根结点、若干个内部结点和若干个叶结点, 其中每个内部结点表示一个属性上的测试, 每个分支代表一个测试输出, 每个叶结点代表一种类别。**决策树生成是一个递归过程。**在决策树基本算法中, 有三种情形会导致递归返回: 1. 当前结点包含的样本全属于同一类别, 无须划分; 2. 当前属性集为空或是所有样本在所有属性上取值相同, 无法划分; 3. 当前结点包含的样本集合为空, 不能划分。

水果数据集

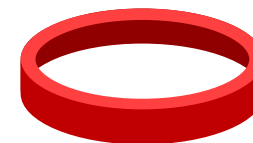
编号	形状	重量	颜色	类别
1	球	一般	红	苹果
2	椭球	重	绿	西瓜
3	弯月	一般	黄	香蕉
4	黄球	一般	桔	橘子
5	球	轻	红	车厘子

决策树示例



决策树特点

优点



- ✓ 推理过程容易理解, 计算简单, 可解释性强。
- ✓ 比较适合处理有缺失属性的样本。

缺点



- ✓ 容易造成过拟合, 需要采用剪枝操作。
- ✓ 忽略了数据之间的相关性。

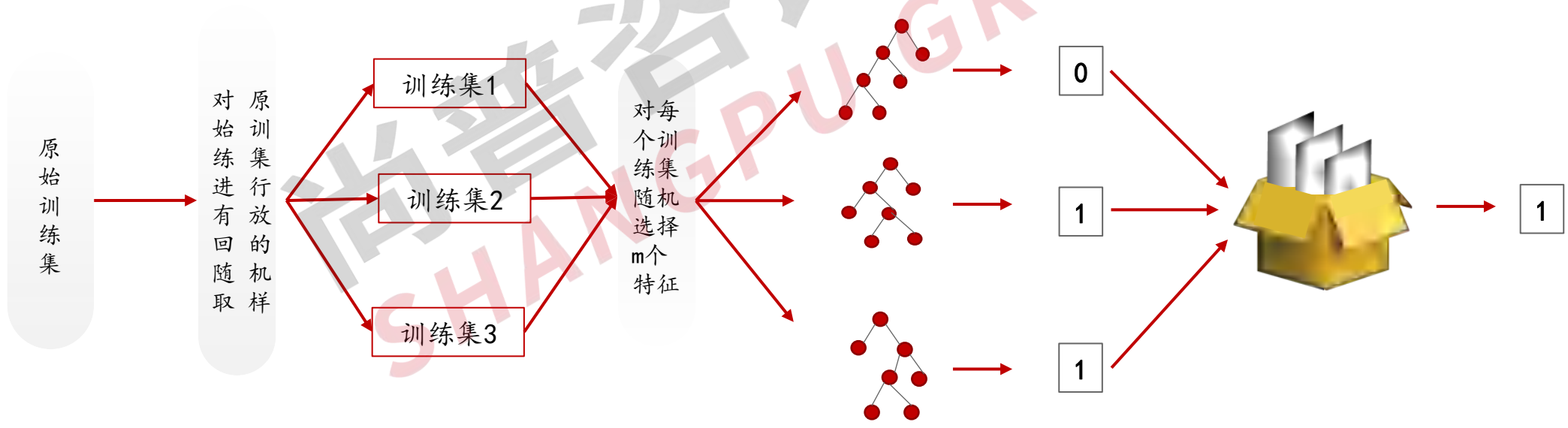
人工智能核心技术：有监督学习—随机森林

随机森林：利用多决策树模型，提高模型预测精度

随机森林 (Random Forest)：是指利用多棵决策树对样本进行训练并预测的一种分类器。**随机森林算法**：从每个决策树收集输出，通过多数表决得到最终的分类结果。随机森林的多数表决就像找人商量事情，不只听一个人的意见，而是在听取多人意见后综合判断。随机森林每棵树的生成规则如下图所示：1) 如果训练集大小为 N ，对于每棵树而言，随机且有放回地从训练中抽取 N 个训练样本，作为该树的训练集，重复 K 次，生成 K 组训练样本集；2) 如果每个特征的样本维度为 M ，指定一个常数 $m \ll M$ ，随机从 M 个特征中选取 m 个特征；3) 利用 m 个特征实现每棵树最大程度的生长，并且没有剪枝过程。

人工智能核心技术

随机森林 (Random Forest)

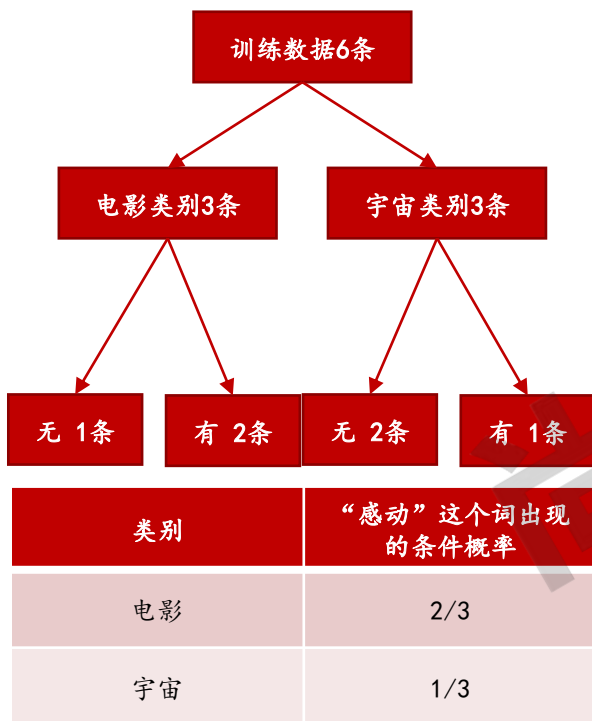


人工智能核心技术：有监督学习—朴素贝叶斯

朴素贝叶斯是常用于自然语言分类问题的算法

朴素贝叶斯是一个基于概率进行预测的算法，在实践中被用于分类问题。具体来说，就是计算数据为某个标签的概率，并将其分类为概率值最大的标签。朴素贝叶斯主要用于文本分类和垃圾邮件判定等自然语言处理中的分类问题。如训练数据类别为电影（那部让人感动的电影名作重映，华丽的动作电影首映，复映的名作感动了世界），训练数据类别为宇宙（沙尘暴笼罩着火星，火星探测终于重新开始，VR中看到的火星沙尘暴让人感动），验证数据为‘复映的动作电影名作让人感动’。以下为朴素贝叶斯算法过程：

第一步：条件概率的计算



依次计算出其他单词在电影和宇宙类别所占的条件概率

第二步：预处理阶段

训练数据	类别
感动、名作、电影	电影
华丽、动作、电影	电影
名作、世界、感动	电影
沙尘暴、火星	宇宙
火星、探测、重新开始	宇宙
VR、火星、感动	宇宙

在预处理阶段，我们将文本转换为BoW (Bag of Words, 词袋) 的形式，形成由特征值构成的向量和标签的组合。当所有单词的集合包含训练数据的单词时，将该单词列的值设为1，否则设为0。另外，当类别为“电影”时，将类别的值替换为1；当类别为“宇宙”时，将类别的值替换为0。

第三步：训练数据已成为特征值的向量

训练数据	名作	电影	华丽	动作	世界	感动	沙尘暴	火星	探测	重新开始	VR	类别
那部让人感动的电影名作重映	1	1	0	0	0	1	0	0	0	0	0	1
华丽的动作电影首映	0	1	1	1	0	0	0	0	0	0	0	1
复映的名作感动了世界	1	0	0	0	1	1	0	0	0	0	0	1
沙尘暴笼罩着火星	0	0	0	0	0	0	1	1	0	0	0	0
火星探测终于重新开始	0	0	0	0	0	0	0	1	1	1	0	0
VR中看到的火星沙尘暴让人感动	0	0	0	0	0	1	1	1	0	0	1	0

人工智能核心技术：有监督学习—朴素贝叶斯

朴素贝叶斯是常用于自然语言分类问题的算法

第四步：验证数据已成为特征值的向量

验证数据	名作	电影	华丽	动作	世界	感动	沙尘暴	火星	探测	重新开始	VR	类别
复映的动作电影名作让人感动	1	1	0	0	0	1	0	0	0	0	0	??

第六步：输出最大概率的标签作为结果

$$\frac{3}{6} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times (0.01)^8 = 1.48 \times 10^{-17}$$

第二步中，“电影”类别的文本出现的概率

每个单词的条件概率的乘积

第五步：概率的计算

参考第一步的方法，“名作”和“电影”的计算方式类似

	有“感动”	无“感动”	“感动”出现的条件概率
电影	2	1	2/3
宇宙	1	2	1/3

$$\frac{3}{6} \times \frac{2}{3} \times (0.01)^{10} = 3.33 \times 10^{-21}$$

第二步中，“宇宙”类别的文本出现的概率

每个单词的条件概率的乘积

类别 \ 条件概率	名作	电影	...	感动	...	VR
电影	2/3	2/3	...	2/3	...	0.01
宇宙	0.01	0.01	...	1/3	...	1/3

朴素贝叶斯使用训练数据来学习与各标签对应的单词的出现概率。在分类时求出每个标签对应的概率将概率最高的标签作为分类结果。朴素贝叶斯在训练时计算以下两种概率，一是每个标签出现的概率，二是在各标签下，每个单词出现的条件概率。图中只展示了部分特征值的列

通过计算，电影类别的概率为 1.48×10^{-17} 大于宇宙类别的概率 3.33×10^{-21} 。由于朴素贝叶斯算法是将概率最高的标签作为分类结果，因此，可以得出验证数据为电影的结果。

人工智能核心技术：机器学习分类—无监督学习

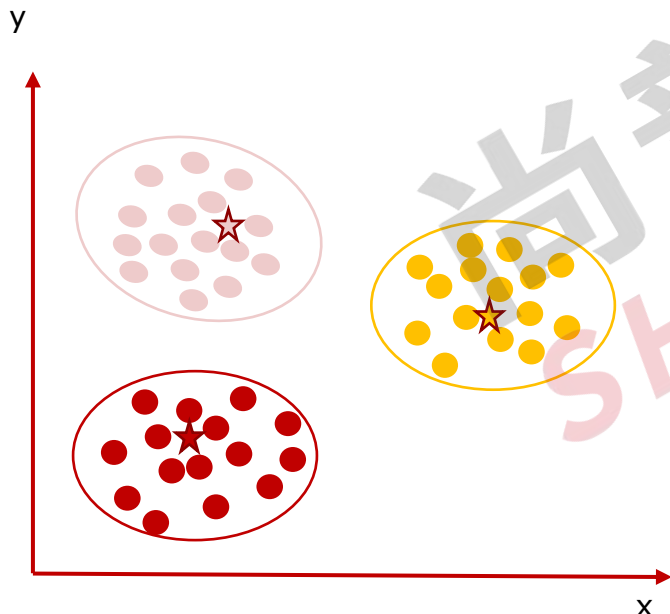
无监督学习：对无标签样本进行学习揭示数据内在规律

无监督学习 (Unsupervised Learning)：主要是指训练数据在不含标记的情况下生成模型（通常在缺乏足够先验知识难以人工标注类别，或进行人工类别标注的成本高等情况下），无监督学习的目标是通过对无标签样本的学习来揭示数据的内在特性及规律。无监督学习主要涉及聚类和降维问题，其中聚类问题涉及K-means聚类、概念聚类、模糊聚类等算法，聚类的目标是为数据点分组，使得不同聚类中的数据点不相似，同一聚类中的数据点则是相似的；降维问题主要是主成分分析、线性判别分析、多维尺度分析等算法，其中主成分分析将数据中存在的空间重映射成一个更加紧凑的空间，此种变换后的维度比原来维度更小。常见的无监督学习适用场景涵盖发现异常数据、用户类别划分、推荐系统等场景。

人工智能核心技术

无监督学习 (Unsupervised Learning)

聚类问题



K-means聚类

把数据划分为K个簇，再把每个点随机划分进相应簇，计算出每个簇的中心，然后计算每个数据点到中心点的距离（欧氏距离）；计算出数据点到k个中心点距离后，选取距离最近中心点所在簇作为该点所在的簇。反复执行相关操作，划分出k个簇，各个簇中数据点距离相近。

欧氏距离

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

降维问题

主成分分析

主成分分析主要构建由多个主成分组成的直角坐标系，可以用来降低数据维度。提取主成分时，可以按照贡献率高低顺序提取贡献率高的主成分，进行数据降维。

变量A	变量B	变量C	变量D

n维数据

降维到二维数据

降维到三维数据

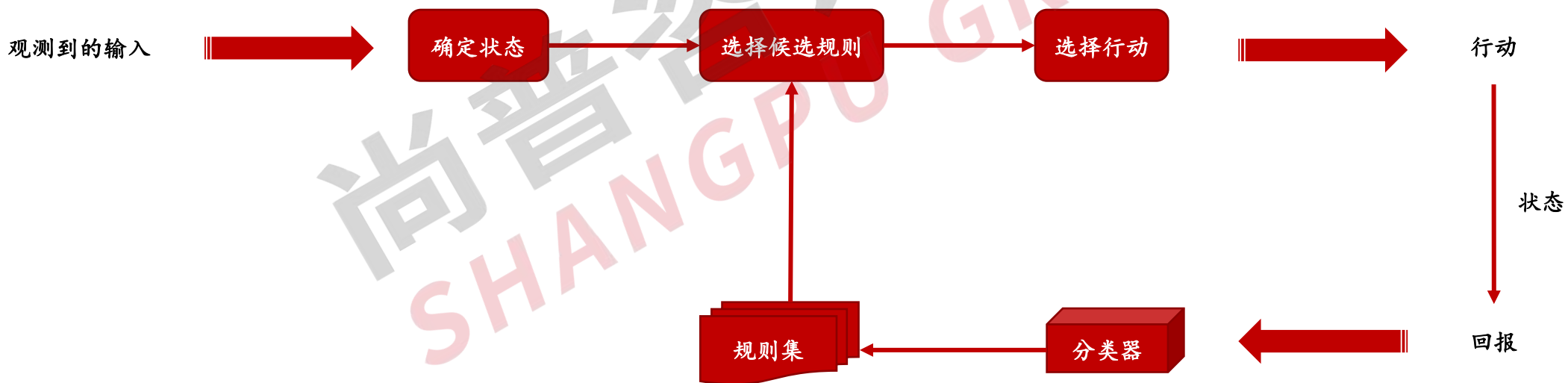
人工智能核心技术：机器学习分类—强化学习

强化学习：不依赖标注数据，有效解决序列行动优化问题

强化学习 (Reinforcement Learning, RL)：又称再励学习或评价学习，是系统从环境到行为映射的学习，以使激励信号（强化信号）函数值最大化。强化学习的关键要素有代理 (Agent)、环境 (Environment)、状态 (State)、环境回报 (Reward)。强化学习特点包括：①学习过程中没有监督者，只有激励信号；②反馈信号是延迟而非即时的；③学习过程具有时间序列性质；④系统的动作会影响到后续的数据。强化学习可分为策略迭代算法和价值迭代算法两类，典型算法包括策略梯度、Sarsa、Q-Learning、Actor-Critic等。强化学习在游戏、自动驾驶、推荐系统等领域有着广阔应用前景。

人工智能核心技术

强化学习 (Reinforcement Learning, RL)



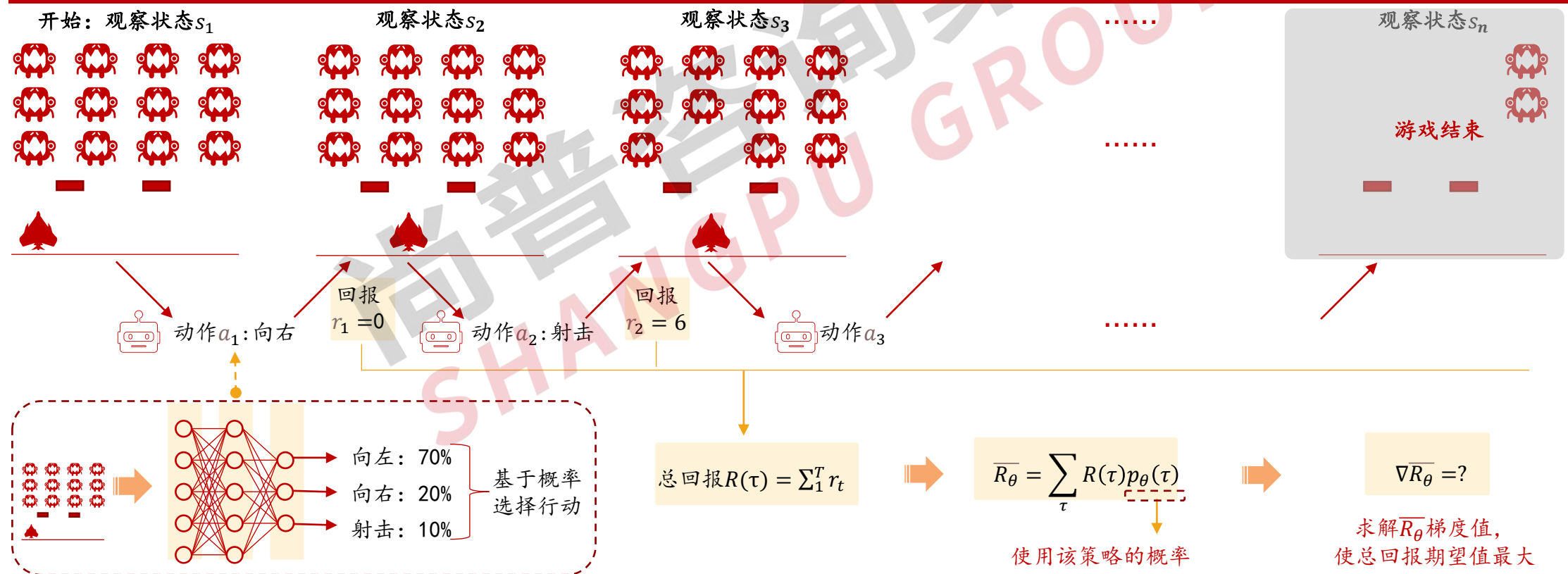
人工智能核心技术：强化学习—策略梯度

策略梯度：有效处理高维及连续动作空间问题

策略梯度 (Policy Gradient, PG)：是一种基于策略优化的强化学习算法。强化学习中，机器通过判别特定状态下采取不同动作所得环境回报大小来评价采取该动作的优劣。PG核心思想是：当一个动作环境回报较大，则增加其被选择概率；反之，减少该动作被选择概率。每个动作选择概率由神经网络决定，以参数化神经网络表示策略（记为 π_θ ），通过调整参数 θ ，使总体回报期望值最大化。而最大化问题可转化为求取回报函数关于参数 θ 的梯度，一般使用梯度上升算法，此即为策略梯度算法。在PG算法基础上，通常进一步采取引入基线 (baseline)、修改因果关系（当前策略不影响该时间点之前的回报）、引入回报权重等方式，来提高算法收敛性及收敛效率。

人工智能核心技术

策略梯度 (Policy Gradient, PG)

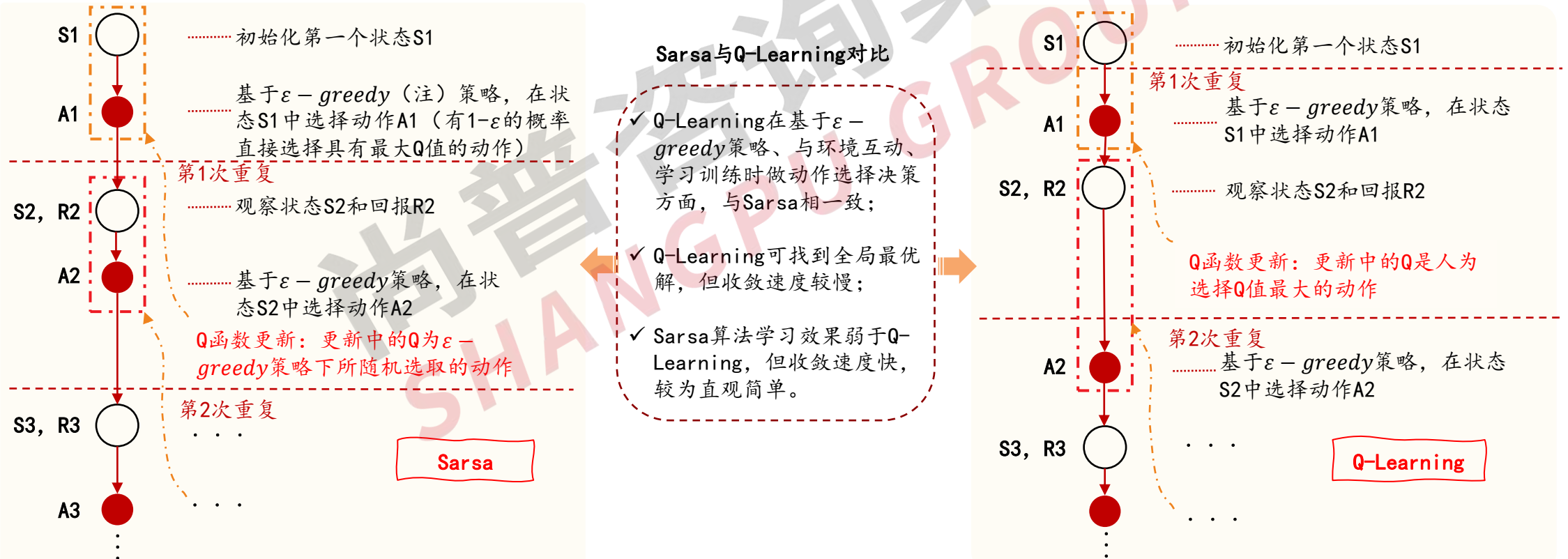


人工智能核心技术：强化学习—时序差分学习

时序差分学习：结合蒙特卡罗方法与动态规划特点，提升策略优化速度

时序差分学习 (Temporal Difference, TD)：与蒙特卡罗学习相似，不需要知晓具体环境模型，可直接从经验中学习。TD学习结合动态规划与蒙特卡罗方法，可在每执行一步策略后即进行Q函数更新，而不必在完成一个采样轨迹后才做更新。时序差分学习根据算法中采样策略和更新Q值策略是否相同，分为同策略算法 (on-policy) 和异策略算法 (off-policy)。同策略代表算法为Sarsa，其行为策略和目标策略均为贪心策略；异策略代表算法为Q-Learning (或称SarsaMax)，其行为策略为贪心策略，而目标策略为确定性策略。Sarsa收敛速度较快，但可能无法得到全局最优解；Q-Learning可找到全局最优解，但收敛速度较慢，因此需视具体问题选用。

时序差分学习 (Temporal Difference, TD)



注： ϵ -greedy：贪心策略，总是做出当前情况下的最优决策，而不考虑整体最优性。

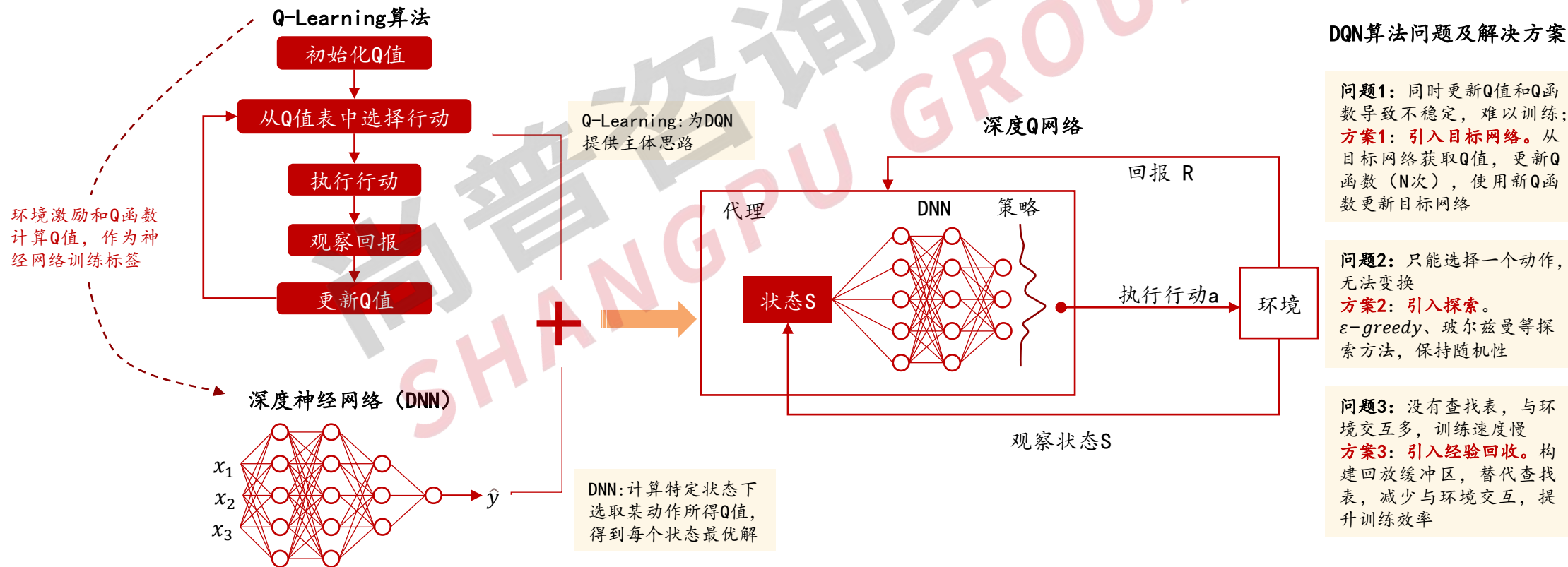
资料来源：清华大学出版社《机器学习》2016年1月第1版，尚普研究院结合公开资料整理绘制

人工智能核心技术：强化学习—深度Q网络

深度Q网络：神经网络与Q-Learning结合，有效应对复杂状态空间

深度Q网络（Deep Q-network, DQN）：由Google DeepMind团队于2013年提出，并在2015年进一步完善，是一种基于深度学习的Q-Learning算法。传统强化学习算法（如动态规划、蒙特卡罗方法、时序差分学习等），其状态均为离散、有限的，并使用查找表存储状态价值或Q函数，但现实生活中状态多为连续的，即使离散化后状态空间（系统全部可能状态的集合）仍十分庞大，因而产生内存不足、表格搜索耗时过长等问题。DQN则通过DNN、CNN或RNN等神经网络对价值函数做近似表示，利用经验回放方式替代查找表，令最终结果无限逼近全局最优解，有效解决Q-Learning状态空间有限问题。

深度Q网络（Deep Q-network, DQN）



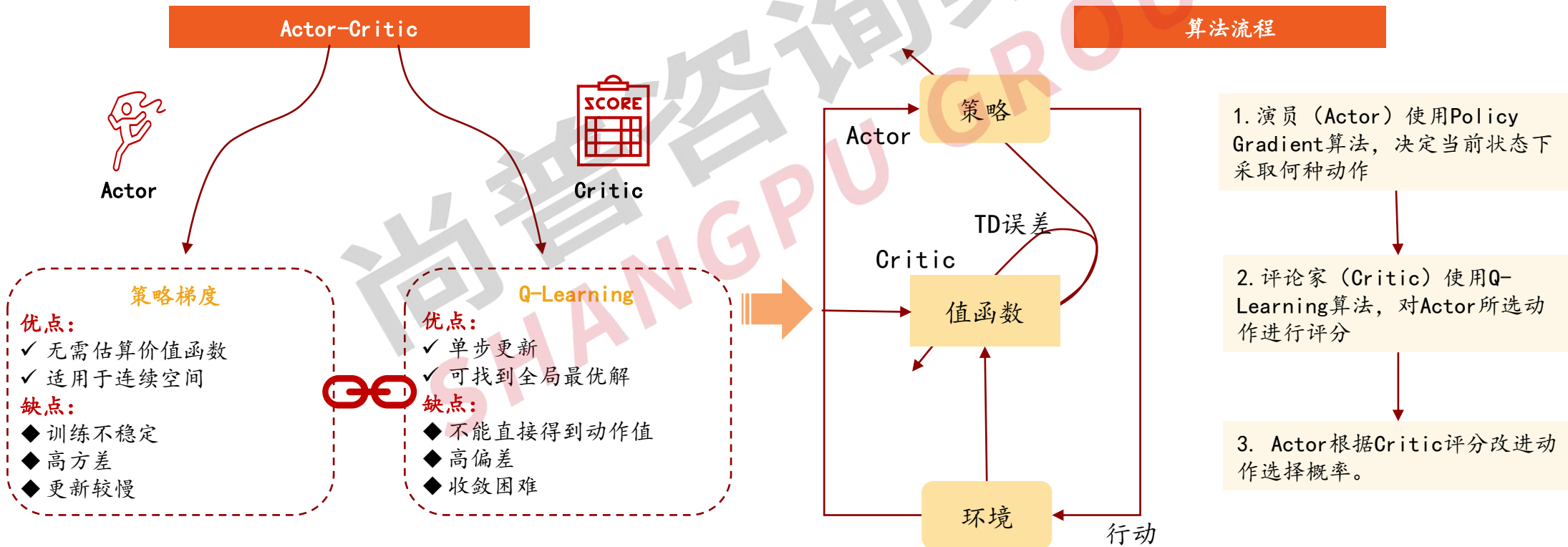
人工智能核心技术：强化学习—Actor-Critic算法

Actor-Critic算法：综合策略梯度与Q-Learning优势，提高学习效率和准确性

Actor-Critic算法 (AC)：又称“演员-评论家”算法，主要由策略网络Actor和评价网络Critic组成。其中，Actor使用策略梯度等基于策略优化的算法，决定当前状态下采取何种动作；Critic则使用Q-Learning等基于值优化的算法，对Actor所选动作进行评分，进而Actor根据Critic评分改进动作选择概率。Actor-Critic算法结合策略梯度与Q-Learning算法优点，既可做到单步更新，又可扩展至连续状态空间，提高策略更新效率同时解决高偏差问题。目前，AC算法已发展出Advantage Actor-Critic (A2C) 以及Asynchronous advantage Actor-Critic (A3C) 等改进型，进一步提高策略准确性、降低内存需求。

人工智能核心技术

Actor-Critic算法



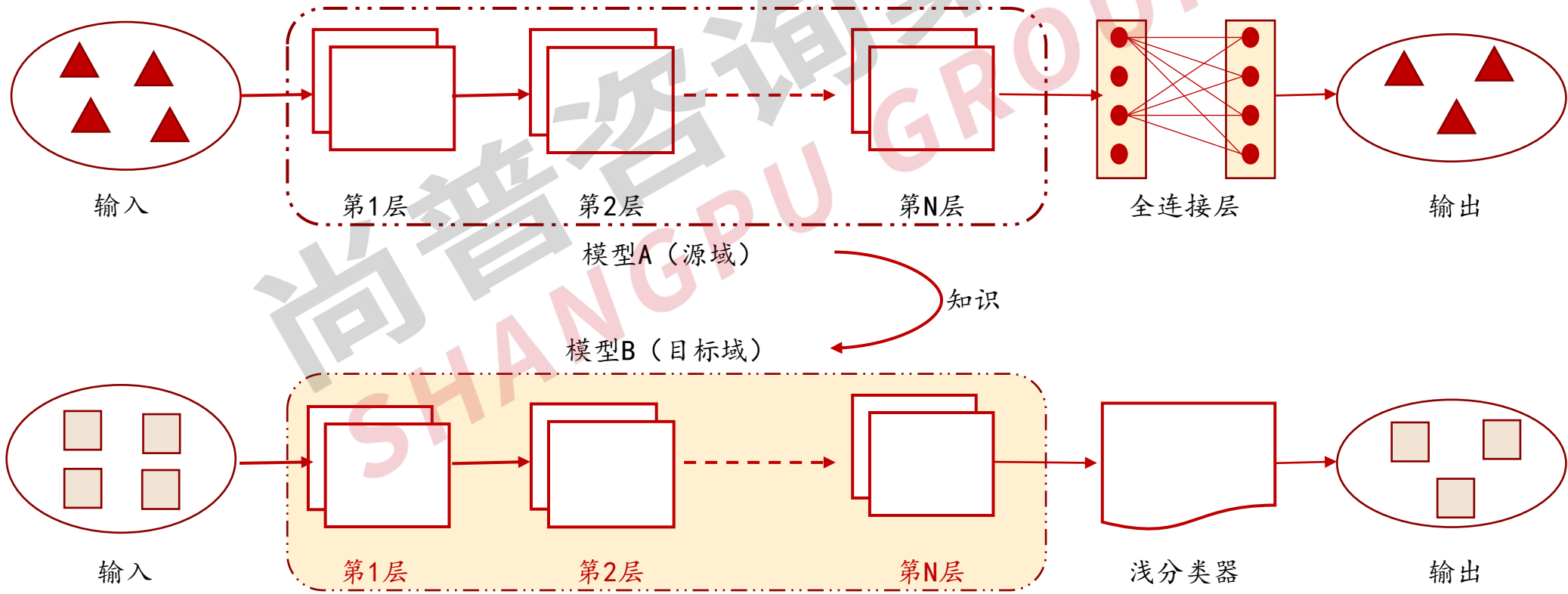
人工智能核心技术：机器学习—迁移学习

迁移学习：将知识由源域迁移至目标域，提高机器学习效率

迁移学习 (Transfer Learning, TL)：是一种机器学习方法,是把已训练好的模型参数迁移到新的模型来帮助新模型训练,其核心目标是将知识从源域迁移到目标域,让机器也可以做到“触类旁通”。迁移学习的主要优点是节省模型训练时间,且在目标域训练数据不足时,模型仍能取得较好的性能。迁移学习的训练框架可以概括为:1) 选择源模型,从可用模型中挑选出预训练模型;2) 重用模型,在目标域中使用源模型进行训练;3) 调整模型。模型可以在目标数据集中对输入-输出进行选择性的微调,以让其适应目标任务。实现迁移学习的方式主要包括样本迁移、特征迁移、模型迁移。目前,迁移学习主要应用在计算机视觉、自然语言处理等领域。

人工智能核心技术

迁移学习 (Transfer Learning, TL)



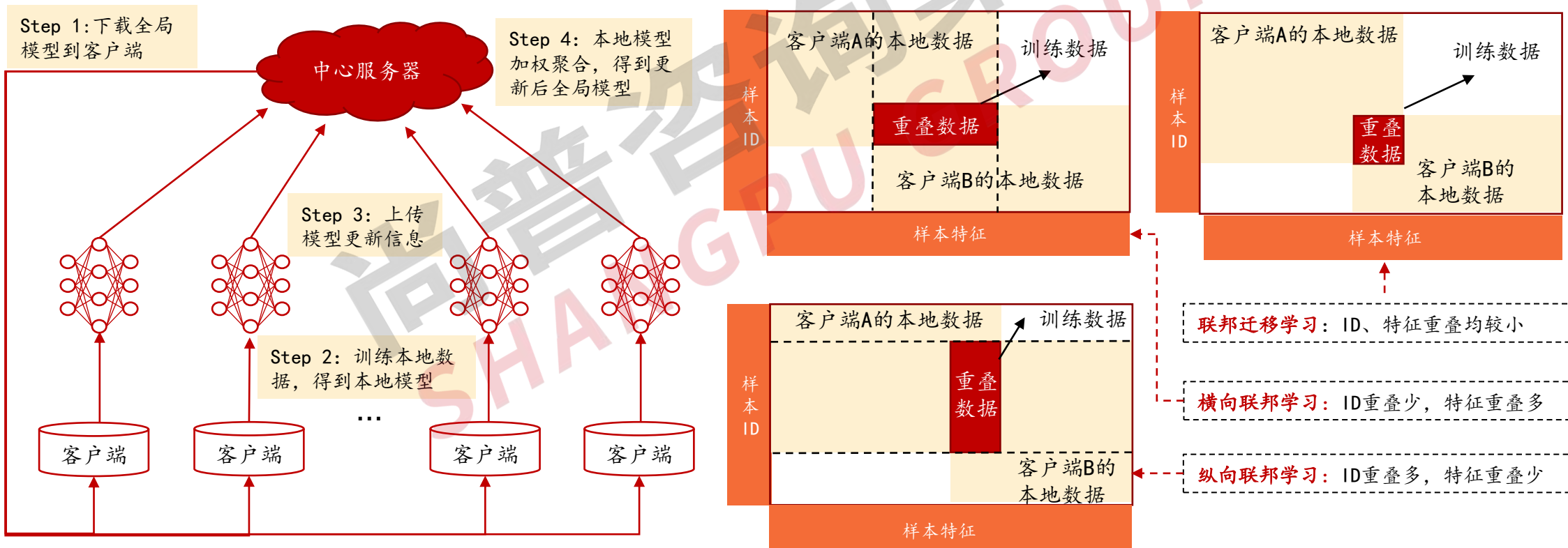
人工智能核心技术：机器学习—联邦学习

联邦学习：以分布式机器学习方式，有效化解信息孤岛问题

联邦学习 (Federated Learning, FL)：由Google公司首次提出，是一种由多个数据持有方协同训练模型，而不分享数据，仅在中间阶段交换训练参数的学习机制。根据其数据分布情况，可分为**横向联邦学习**、**纵向联邦学习**和**联邦迁移学习**。联邦学习的基本流程是：1) 客户端从服务器下载全局模型；2) 训练本地数据得到本地模型；3) 上传模型更新信息；4) 中心服务器加权聚合本地模型，得到更新后的全局模型。经过多次迭代，得到趋近于集中式机器学习结果的模型。作为一项新兴技术，联邦学习在物联网、智慧医疗、金融风控、智慧城市等领域均有广阔的应用前景。

人工智能核心技术

联邦学习 (Federated Learning, FL)



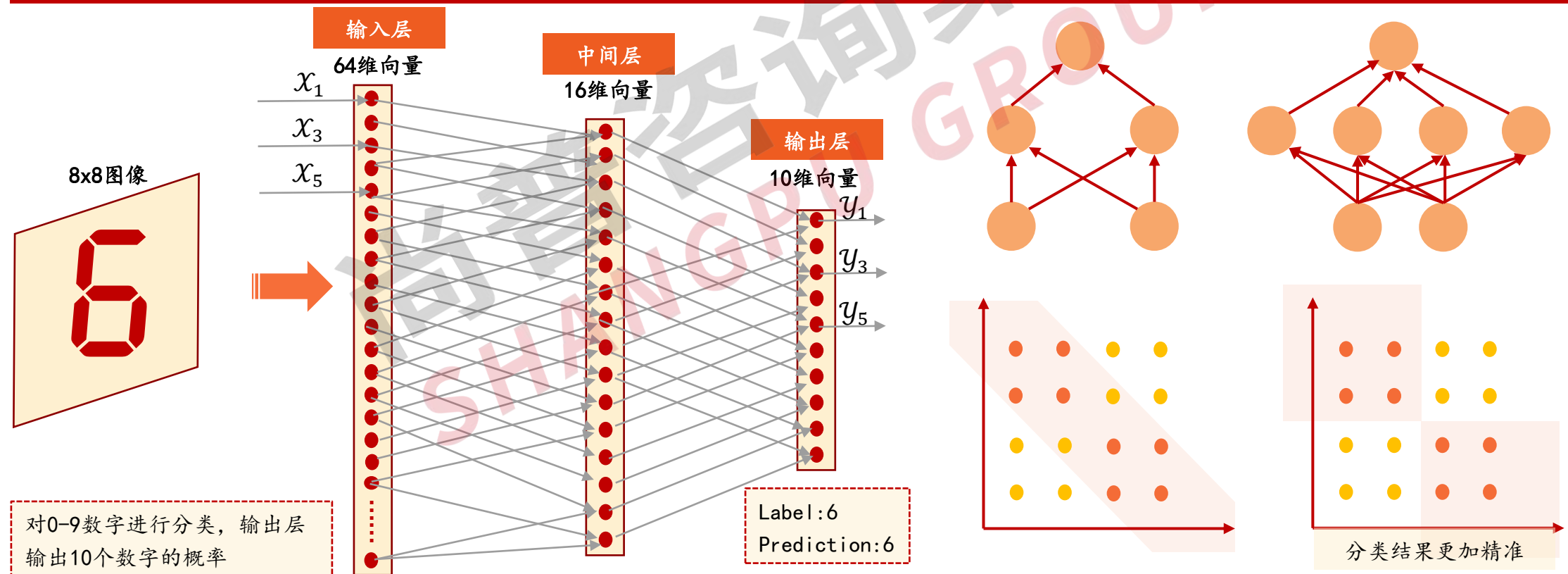
人工智能核心技术：深度学习—神经网络

神经网络：具有适应性简单单元组成的广泛并行互联网络

神经网络 (Neural Network)：由数千甚至数百万个紧密互连的简单处理节点组成，其主要包括输入层（输入数据）、中间层/隐藏层（学习复杂决策边界）和输出层（输出结果）。神经网络可以用于回归，但主要应用于分类问题。如下图所示：输入层表示输入图像（64维向量），中间层使用Sigmoid等非线性函数对于输入层数据进行计算，输出层使用非线性函数对于中间层数据进行计算。神经网络通过采取设置中间层的方式，利用单一算法学习各种决策边界，调节中间层数量以及层的深度，神经网络可学习更复杂的边界特征，从而得出更加准确的结果。

人工智能核心技术

神经网络 (Neural Network)



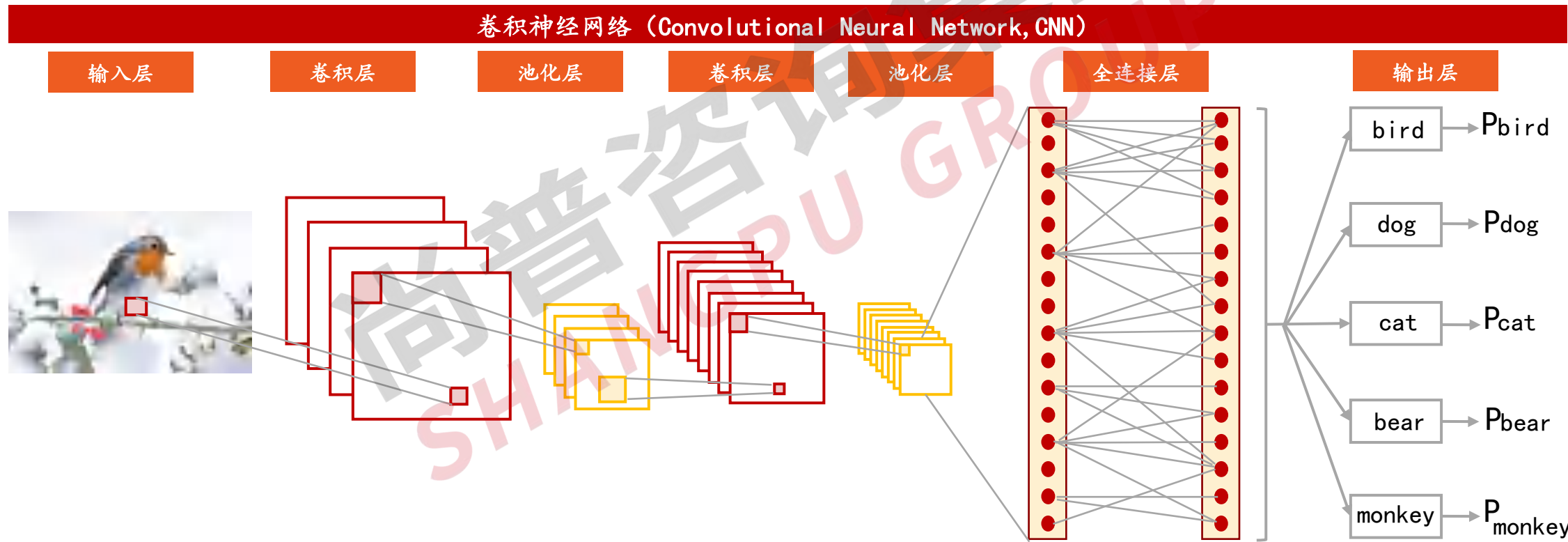
资料来源：清华大学出版社《机器学习》2016年1月第1版，人民邮电出版社《图解机器学习算法》2021年6月第1版，尚普研究院结合公开资料整理绘制

人工智能核心技术：深度学习—卷积神经网络

卷积神经网络：以图像识别为核心的深度学习算法

卷积神经网络 (Convolutional Neural Network, CNN)：由数千甚至数百万个紧密互连的简单处理节点组成，其主要包括输入层、卷积层、池化层、全连接层和输出层，适合处理图片、视频等类型数据。1980年，日本科学家福岛邦彦提出一个包含卷积层、池化层的神经网络结构。在此基础上，Yann Lecun将BP算法应用到该神经网络结构的训练上，形成当代卷积神经网络的雏形；1988年，Wei Zhang提出第一个二维卷积神经网络：平移不变人工神经网络 (SIANN)，并将其应用于检测医学影像；1998年Yann LeCun及其合作者构建了更加完备的卷积神经网络LeNet-5并在手写数字的识别问题中取得成功。

人工智能核心技术

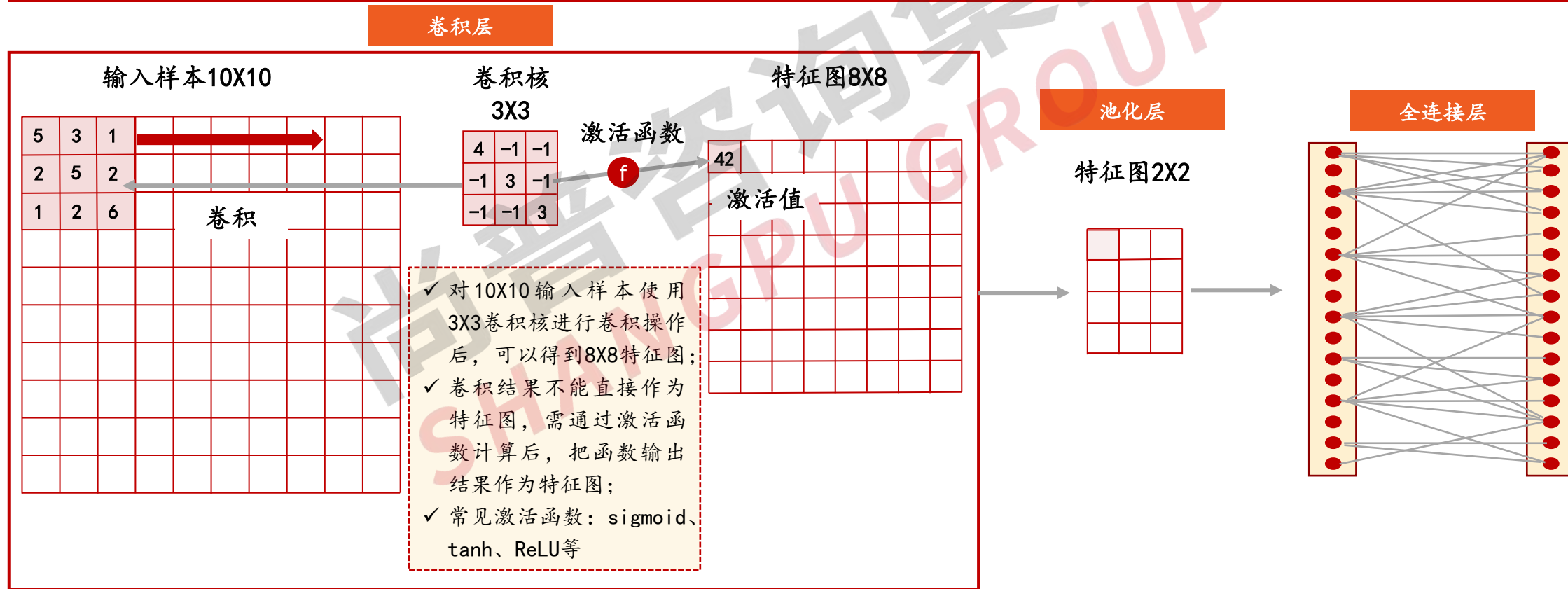


人工智能核心技术：深度学习—卷积神经网络

卷积神经网络：以图像识别为核心的深度学习算法

卷积层：图片输入转化成RGB对应的数字，然后通过卷积核做卷积，目的是提取输入中的主要特征，卷积层中使用同一卷积核对每个输入样本进行卷积操作；**池化层：**作用在于减小卷积层产生的特征图尺寸（压缩特征映射图尺寸有助于降低后续网络处理的负载）；**全连接层：**计算激活值然后通过激活函数计算各单元输出值（激活函数包括Sigmoid、tanh、ReLU等）；**输出层：**使用似然函数计算各类别似然概率。

卷积神经网络 (Convolutional Neural Network, CNN)

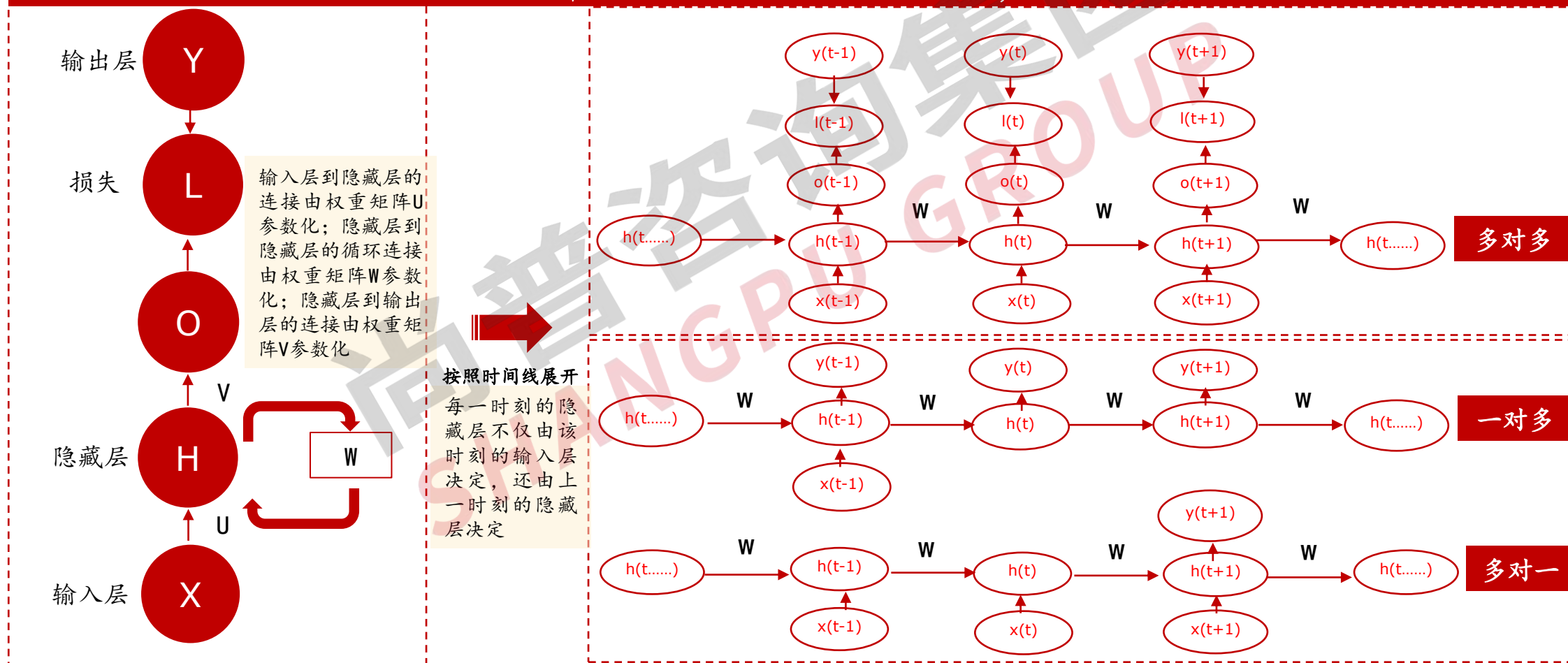


人工智能核心技术：深度学习—循环神经网络

循环神经网络：用于处理序列数据的神经网络

循环神经网络 (Recurrent Neural Network, RNN)：是一类以序列数据（指相互依赖的数据流，比如时间序列数据、信息性的字符串、对话等）为输入，在序列的演进方向进行递归且所有节点（循环单元）按链式连接的神经网络。目前，语言建模和文本生成、机器翻译、语音识别、生成图像描述、视频标记是RNN应用最多的领域。

循环神经网络 (Recurrent Neural Network, RNN)



资料来源：人民邮电出版社《深度学习》2018年5月第九版，尚普研究院结合公开资料整理绘制

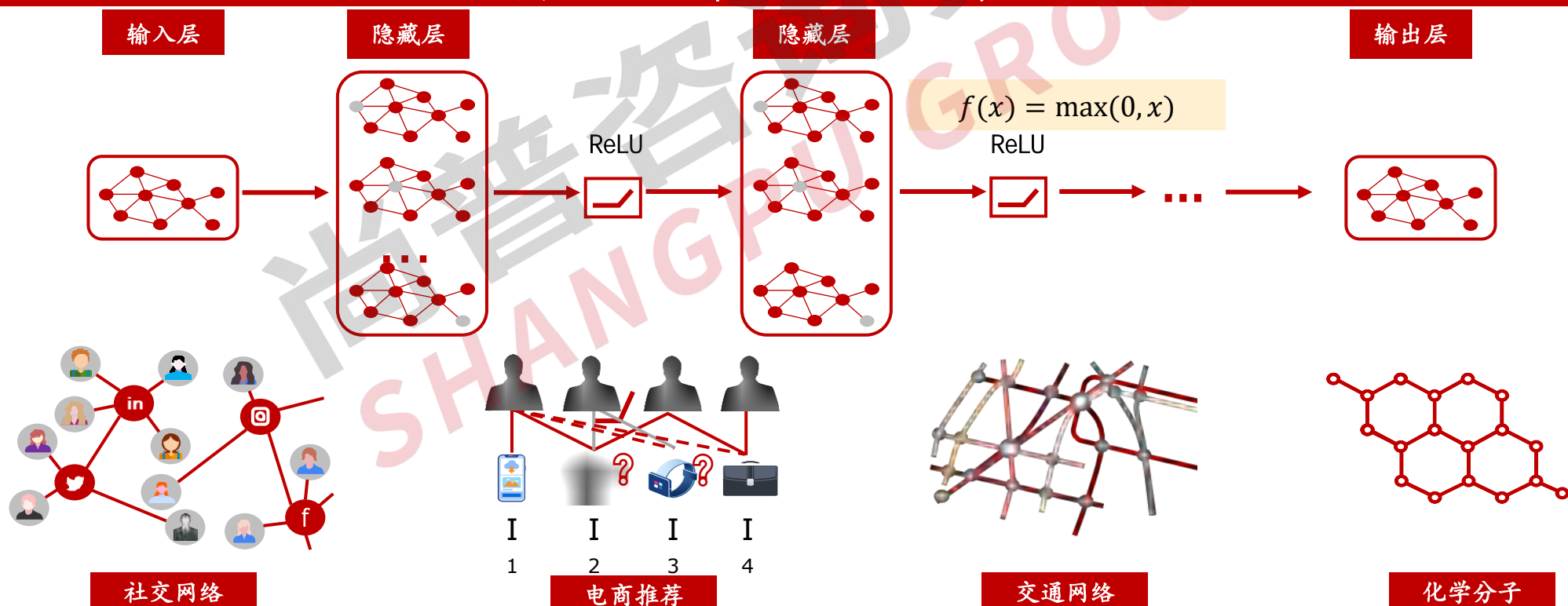
人工智能核心技术：深度学习—图神经网络

图神经网络：用于处理图结构数据的神经网络

图神经网络 (Graph Neural Networks, GNN)：将图数据和神经网络进行结合，在图数据上面进行端对端的计算，具备端对端学习、擅长推理、可解释性强的特点。图神经网络发展出多个分支，主要包括图卷积网络、图注意力网络、图自编码器、图生成网络和图时空网络等。图神经网络的训练框架如下：首先，每个节点获取其相邻节点的所有特征信息，将聚合函数（如求和或取平均）应用于这些信息。聚合函数的选择必须不受节点顺序和排列的影响。之后，将前一步得到的向量传入一个神经网络层（通常是乘以某个矩阵），然后使用非线性激活函数（如ReLU）来获得新的向量表示。目前，图神经网络在许多领域的实际应用中都展现出强大的表达能力和预测能力，如物理仿真、科学研究、生物医药、金融风控等。

人工智能核心技术

图神经网络 (Graph Neural Networks, GNN)



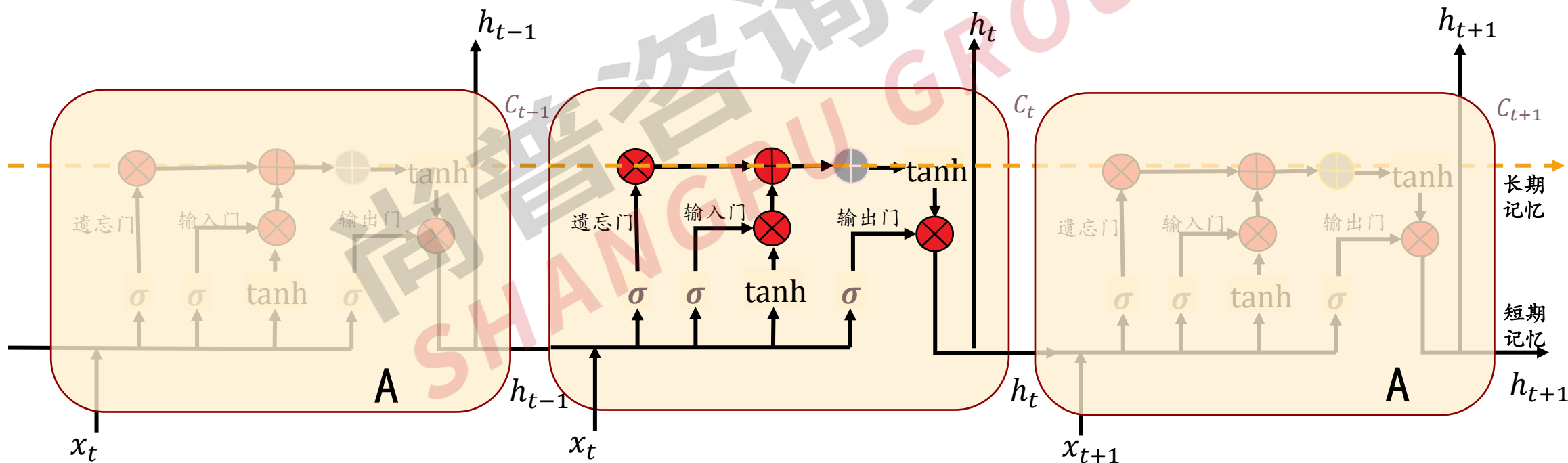
人工智能核心技术：深度学习—长短期记忆神经网络

长短期记忆神经网络：在RNN中加入门控机制，解决梯度消失问题

长短期记忆神经网络 (Long Short-Term Memory, LSTM)：LSTM是一种特殊的循环神经网络 (RNN)。传统RNN在训练中，随着训练时间的加长和层数的增多，很容易出现梯度爆炸或梯度消失问题，导致无法处理长序列数据，LSTM可有效解决传统RNN“长期依赖”问题。LSTM由状态单元、输入门 (决定当前时刻网络的输入数据有多少需要保存到单元状态)、遗忘门 (决定上一时刻的单元状态有多少需要保留到当前时刻)、输出门 (控制当前单元状态有多少需要输出到当前输出值) 组成，以此令长期记忆与短期记忆相结合，达到序列学习的目的。LSTM应用领域主要包括文本生成、机器翻译、语音识别、生成图像描述和视频标记等。

人工智能核心技术

长短期记忆神经网络 (Long Short-Term Memory, LSTM)

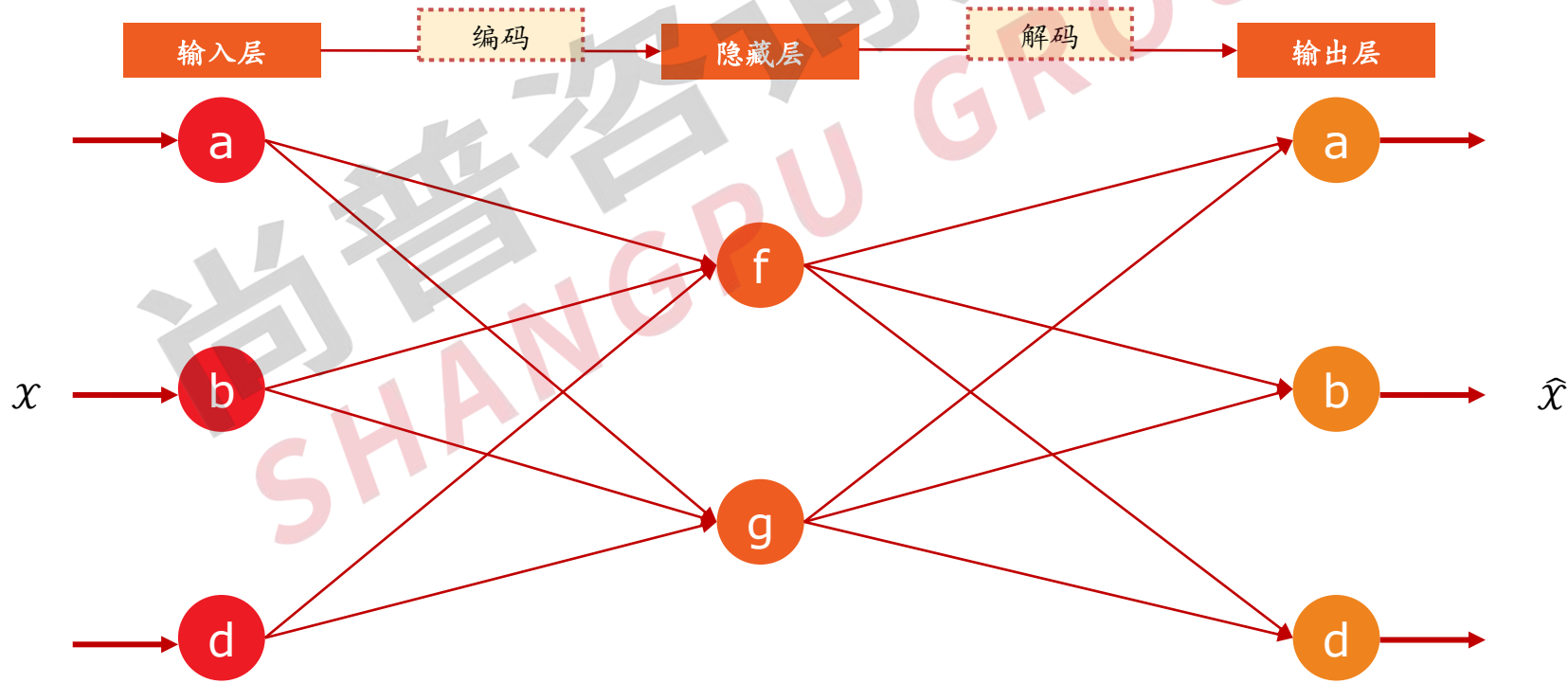


人工智能核心技术：深度学习—自编码器

自编码器：通过期望输出等同于输入样本的过程，实现对输入样本抽象特征学习

典型深度无监督学习模型包括自编码器、受限波尔兹曼机与生成对抗网络。自编码器 (Autoencoder, AE)：包括编码器和解码器两部分，其中编码器将高维输入样本映射到低维抽象表示，实现样本压缩与降维；解码器将抽象表示转换为期望输出，实现输入样本的复现。自编码器的输入与期望输出均为无标签样本，隐藏层输出则作为样本的抽象特征表示。自编码器仅通过最小化输入样本与重构样本之间的误差来获取输入样本的抽象特征表示，无法保证自编码器提取到样本的本质特征。为避免上述问题，需要对自编码器添加约束或修改网络结构，进而产生稀疏自编码器、去噪自编码器、收缩自编码器等改进算法。自编码器凭借其优异的特征提取能力，主要应用于目标识别、文本分类、图像重建等诸多领域。

自编码器 (Autoencoder, AE)



人工智能核心技术：深度学习—生成对抗网络

生成对抗网络：通过对抗训练机制使得生成器生成以假乱真的样本

生成对抗网络 (Generative Adversarial Network, GAN)：通过使用对抗训练机制对两个神经网络进行训练，避免反复应用马尔可夫链学习机制带来的配分函数计算，明显提高应用效率。生成对抗网络包含一组相互对抗模型—判别器和生成器，判别器目的是正确区分真实数据和生成数据，使得判别准确率最大化，生成器是尽可能逼近真实数据的潜在分布。生成器类似于造假钞的人，其制造出以假乱真的假钞，判别器类似于警察，尽可能鉴别出假钞，最终造假钞的人和警察双方在博弈中不断提升各自能力。

生成对抗网络 (Generative Adversarial Network, GAN)



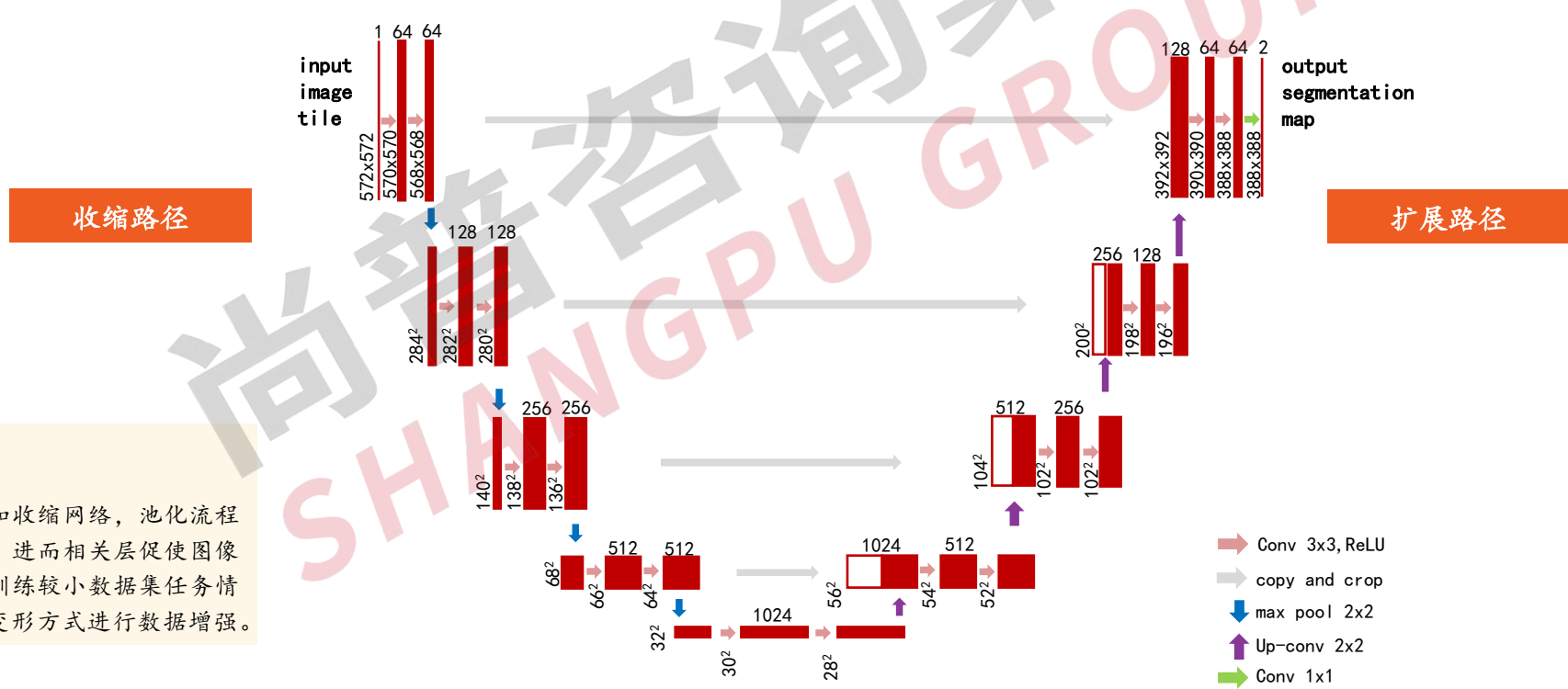
人工智能核心技术：计算机视觉—U-Net

U-Net：在有限数据集图像分割方面独具优势

U-Net：由Olaf Ronneberger、Philipp Fischer等人于2015年首次提出，起初应用在医学图像分割领域。U-Net网络结构可视化的结果形如字母“U”，U-Net由左侧的收缩路径和右侧的扩展路径两部分组成。相较于传统模型，U-Net在架构和基于像素的图像分割方面更有优势，且在有限数据集图像上更加有效。在生物医学图像领域，U-Net通过弹性变形方法帮助扩充数据集，从而提高学习效果。从实际应用来看，图像分割不仅限于医学图像领域，其在卫星图像遥感和无人驾驶系统等相关领域也有应用。

人工智能核心技术

图像分割算法：U-Net结构



U-Net主要特点：

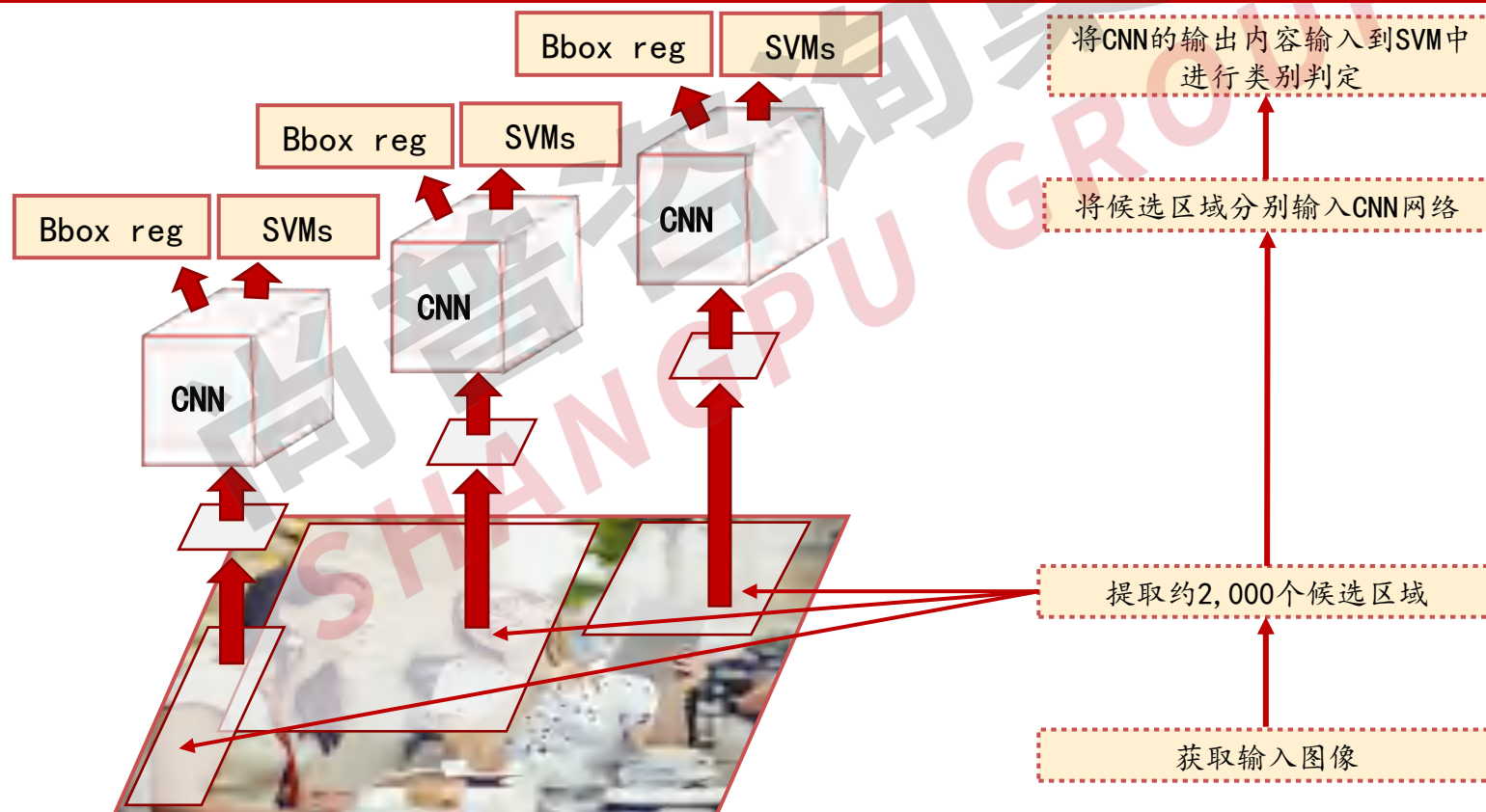
构建连续层的方式增加收缩网络，池化流程被上采样流程所取代，进而相关层促使图像输出分辨率提高。在训练较小数据集任务情况下，通过采取弹性变形方式进行数据增强。

人工智能核心技术：计算机视觉—R-CNN

R-CNN：基于深度学习模型的目标检测算法

目标检测算法主要分为R-CNN和Yolo两类，R-CNN准确度高但速度慢；Yolo算法速度快但准确性相对低。R-CNN (Region-CNN) 最早将深度学习应用在目标检测任务中，目标检测任务是计算机视觉中的重要组成部分，其在人脸识别、自动驾驶等领域有着广泛应用。传统目标检测方法大多以图像识别为基础，在图片上使用穷举法选出所有物体可能出现的区域框，对这些区域框提取特征并使用图像识别方法分类，得到所有分类成功的区域后，通过非极大值抑制算法得到输出结果。R-CNN遵循传统目标检测的思路，但在提取特征这一步，将传统特征换成深度卷积网络提取的特征。

目标检测算法：R-CNN结构



R-CNN目标检测过程一般包括以下步骤：
1) 输入一张图像，使用无监督算法提取约2,000个物体可能位置；
2) 将所有候选区域取出并缩放为相同大小，输入卷积神经网络中提取特征；
3) 使用支持向量机对每个区域的特征进行分类。

注：Bbox reg是指边界框回归，即对边界框进行平移回归

资料来源：中国科学技术出版社《人工智能导论》2018年8月第1版，尚普研究院结合公开资料整理绘制

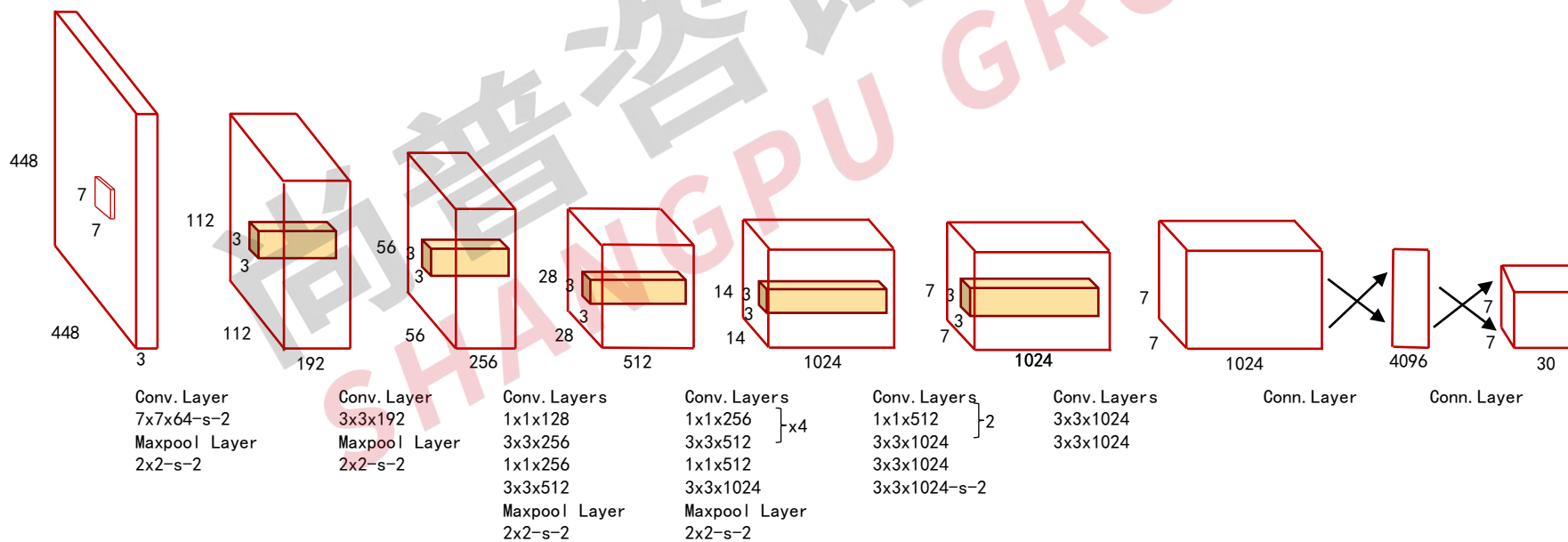
人工智能核心技术：计算机视觉—Yolo

Yolo: 采用CNN模型实现端到端的目标检测算法

以往采用滑动窗口的目标检测算法将检测问题转化为图像分类问题存在计算量过大等弊端。Yolo算法很好的解决此类问题，其直接将原始图片分割成互不重合的小方块，然后通过卷积产出特征图。具体而言，Yolo算法采用单独的CNN模型实现端到端的目标检测，其主要包含24个卷积层和2个全连接层。卷积层主要使用1x1卷积来做通道缩减，然后紧跟3x3卷积。卷积层和全连接层采用Leaky ReLU激活函数，最后一层采用线性激活函数。Yolo算法的优点在于通过采用CNN实现检测，且训练与预测环节都是端到端过程，因而该算法比较简洁且速度快。同时，由于Yolo是对整张图片做卷积，因而其在检测目标时拥有更大的视野且不容易对背景误判。Yolo算法的缺点是其应用在物体的宽高比方面泛化率相对较低，无法定位不寻常比例的物体。

人工智能核心技术

目标检测算法：Yolo结构



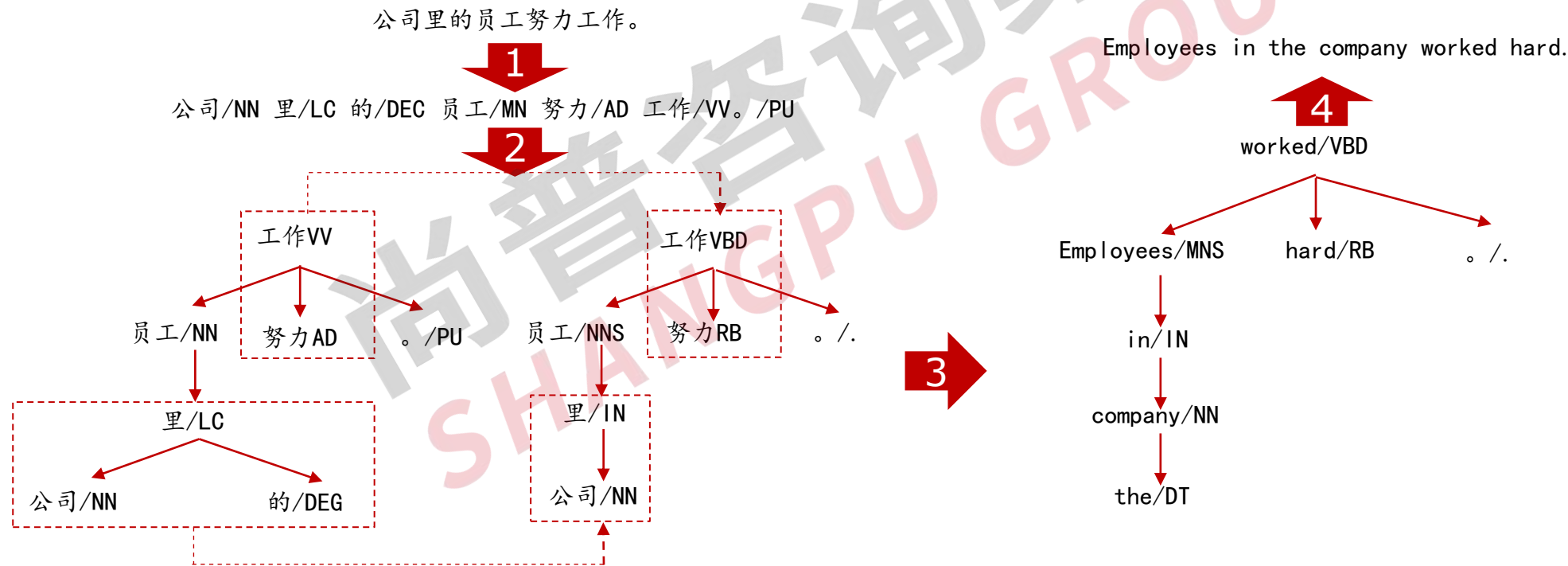
人工智能核心技术：自然语言处理

自然语言处理：用于分析、理解和生成自然语言，便于人机间交流

自然语言处理 (Natural Language Processing, NLP) 通过对词、句子、篇章进行分析，对内容里面的人物、时间、地点等进行理解，并在此基础上支持一系列核心技术（如跨语言的翻译、问答系统、阅读理解、知识图谱等）。基于这些技术，NLP可以应用于诸多领域，包括搜索引擎、客服、金融、新闻等。以机器翻译为例，给定输入的源语言句子经过词法和句法分析得到句法树，然后通过转换规则将源语言句子句法树进行转换，调整词序、插入词或者删除词并将句法树中的源语言词用对应的目标语言词替换，生成目标语言的句法树，最后基于目标语言的句法树遍历叶子节点，得到目标语言句子。

人工智能核心技术

自然语言处理 (Natural Language Processing, NLP) — 机器翻译流程图

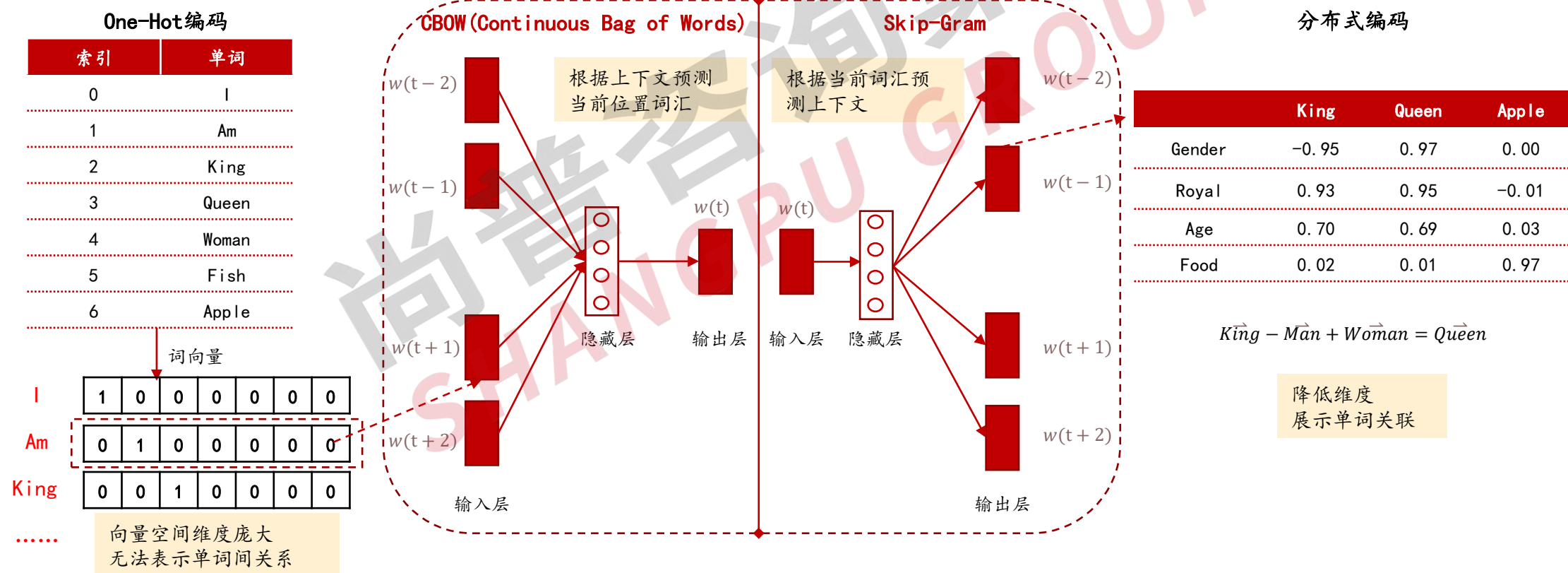


人工智能核心技术：自然语言处理—Word2Vec

Word2Vec：以分布式方法表示词向量，有效降低维度，丰富词语间关系

Word2Vec (Word to Vector)：由Google团队Tomas Mikolov等人于2013年提出，是一种高效词向量训练模型。在NLP任务中，需要将单词或词语转化为数值向量（词向量）供计算机识别，这一过程称为词嵌入。早期One-Hot编码方法简单易懂、有较好鲁棒性，但存在维度爆炸问题，且无法表示词语相关关系。Word2Vec使用分布式词向量表示方法，有效降低词向量维度，丰富单词间关系。Word2Vec主要使用CBOW（连续词袋）模型和Skip-Gram（跳词）模型两种预测方法，前者根据上下文预测当前词，适用于小语库；后者根据当前词预测上下文，在大型语料库中表现更好。

Word2Vec (Word to Vector) 主要预测模型

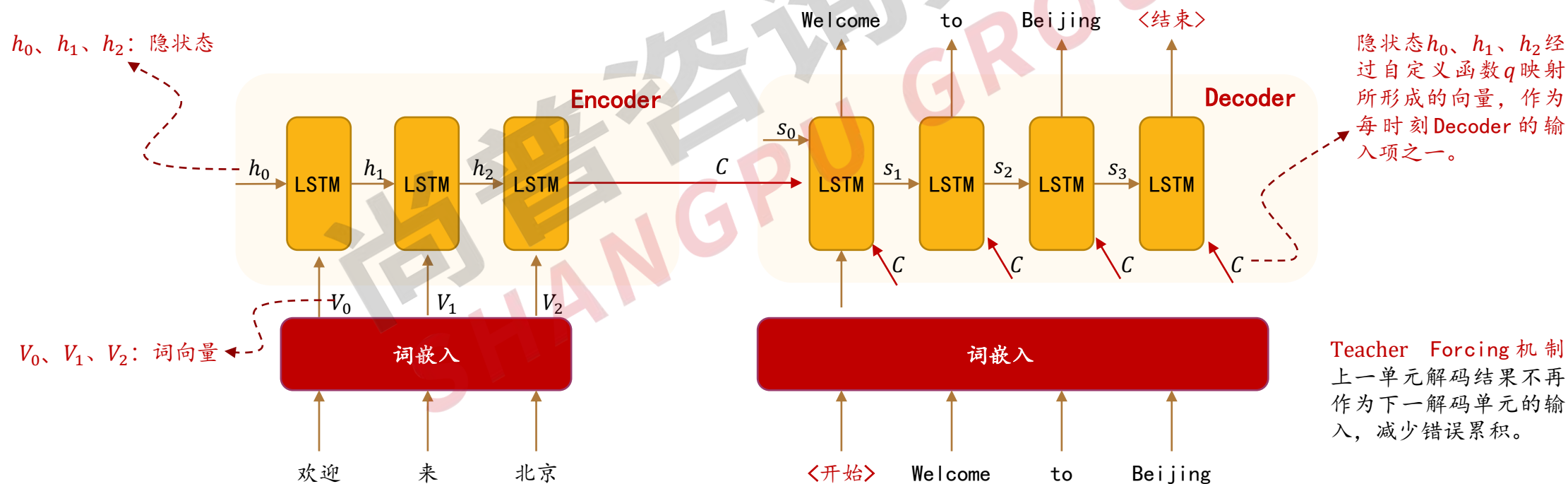


人工智能核心技术：自然语言处理—Seq2Seq

Seq2Seq：基于循环神经网络输出非固定长度语句

Seq2Seq (Sequence to Sequence)：由Google Brain和Yoshua Bengio团队在2014年分别独立提出，主要用于解决机器翻译任务中输出语句长度不确定问题。Seq2Seq模型基于RNN或LSTM等神经网络算法，主要由编码器（Encoder）和解码器（Decoder）两部分构成。其中，编码器负责将输入序列压缩为指定长度语义向量；解码器负责根据语义向量生成指定序列。因而，Seq2Seq模型可以实现序列信号转化为无固定长度序列输出（RNN只能将序列信号转化为固定长度序列输出）。除机器翻译外，Seq2Seq也被应用于文本摘要、图像描述文本生成、语音识别等场景。

引入Teacher Forcing机制的Seq2Seq模型框架



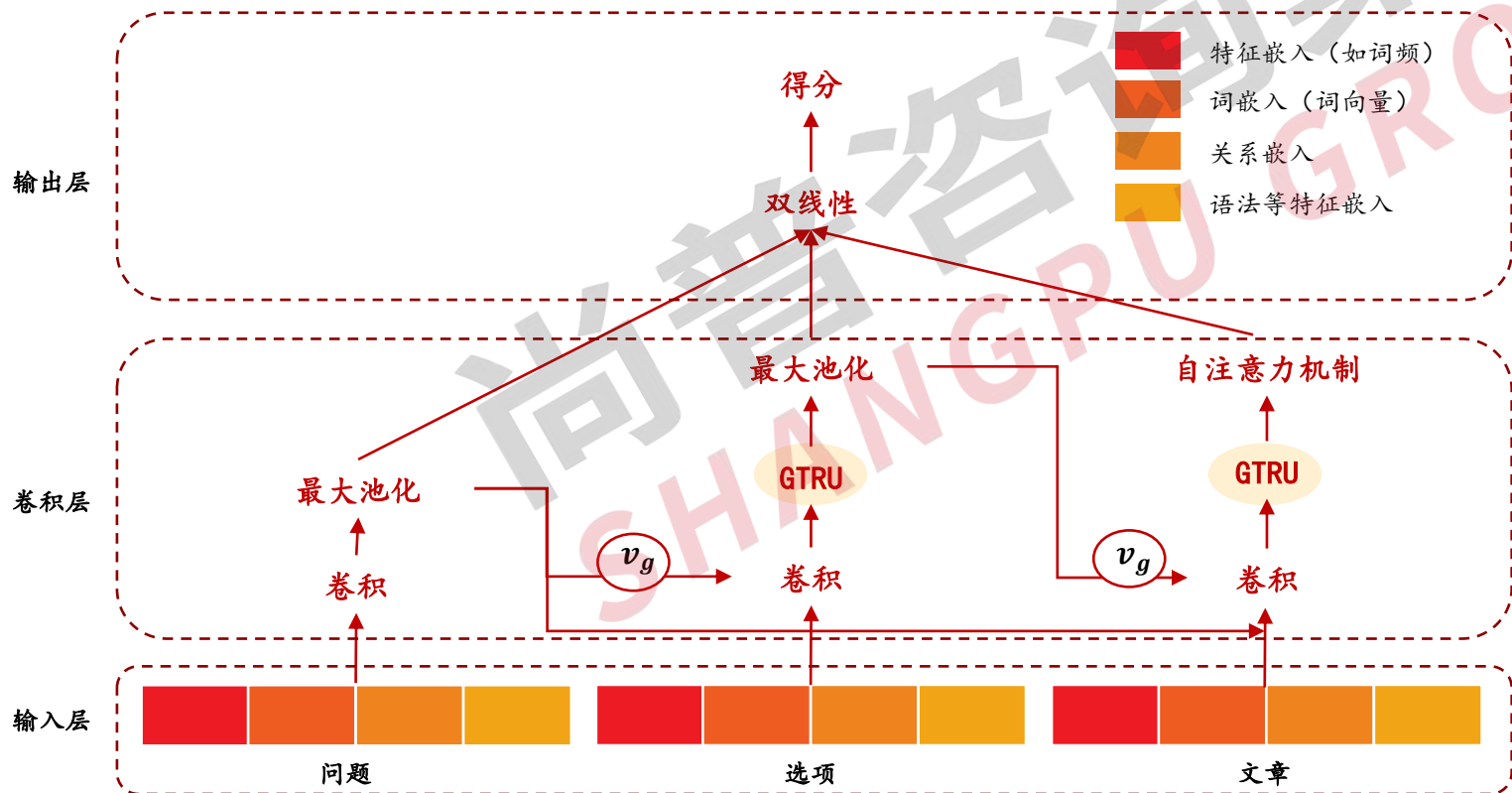
人工智能核心技术：自然语言处理——门控卷积神经网络

门控卷积神经网络：应用于语言建模，以并行计算方式加速学习过程

门控卷积神经网络（Gated Convolutional Neural Networks）：由Facebook团队Yann等人在2016年提出，使用基于门控机制的卷积神经网络处理语言建模问题。与循环神经网络（RNN）相比，门控卷积神经网络可实现并行计算，有效缩短运行时长。由于卷积神经网络不存在梯度消失问题，因而门控卷积神经网络中的门控机制更加简化，仅需保留输出门用于判定信息是否传递至下一层，可实现更高精度及更快收敛速度。目前，门控卷积神经网络模型已被应用于光学字符识别（OCR）、机器阅读理解等多个领域。

人工智能核心技术

门控卷积神经网络（Gated Convolutional Neural Networks）模型架构



Gated Tanh-ReLU Units：其中Tanh-ReLU表示激活函数。GTRU即在卷积神经网络中的门结构部分，有助于文字表征建模。

v_g ：Reference Vector，来自其他部分的参照向量，强化各部分间联系。

人工智能核心技术：语音处理—语音识别

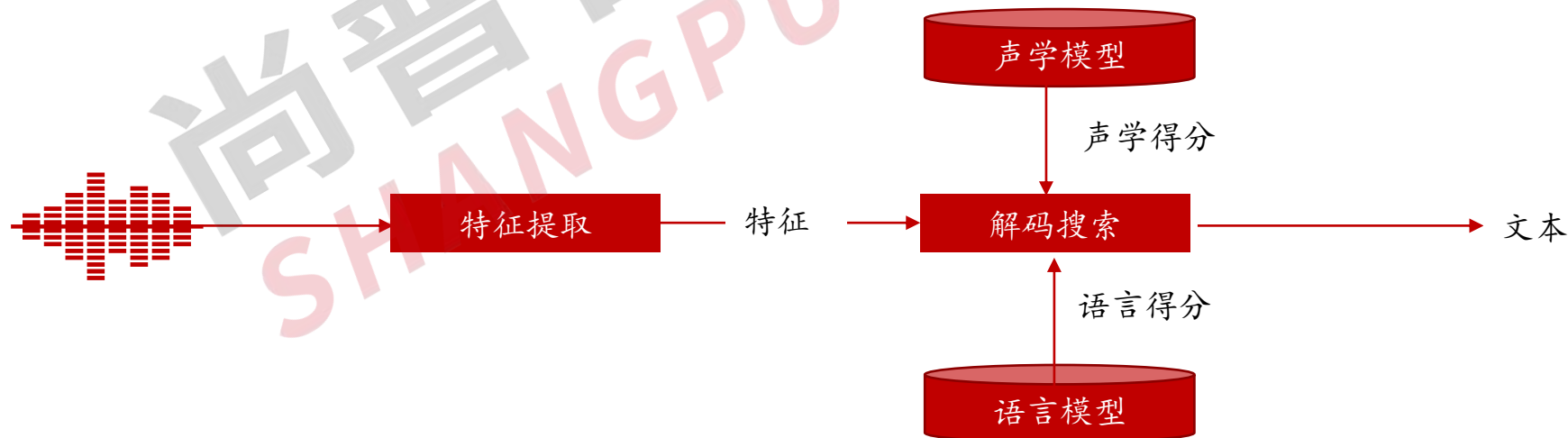
语音识别：将语音自动转换为文字的过程

语音识别 (Speech Recognition, SR)：是将语音自动转换为文字的过程，目的是将语音信号转变为计算机可读的文本字符或命令。20世纪60年代，最具代表性的研究成果是动态规划匹配方法，该方法能够将输入语音与样本语音的各自特征，按时间轴进行伸缩、匹配。20世纪80年代以后，基于隐马尔可夫模型的统计建模方法逐渐取代了基于模版匹配的方法，基于高斯混合模型-隐马尔可夫模型的混合声学建模技术推动语音识别技术的蓬勃发展。到21世纪初，深度学习技术在语音识别中发挥重要作用，使语音识别的研究获得突破性进展。至此，语音识别技术开始走向大规模商业化阶段。**语音识别技术主要包括：**

- ①**特征提取**，将语音信号转化为计算机能够处理的语音特征向量。常用的特征参数有基音周期、共振峰、短时平均能量或幅度等；
- ②**模式匹配**，根据一定的准则，使未知模式与模型库中的某一个模型获得最佳匹配的过程；
- ③**模型训练**，提取语音库中语音样本的特征参数作为训练数据，合理设置模型参数初始值，对模型参数进行重估，使识别系统具有更佳的识别效果。

人工智能核心技术

语音识别 (Speech Recognition, SR) 系统框架



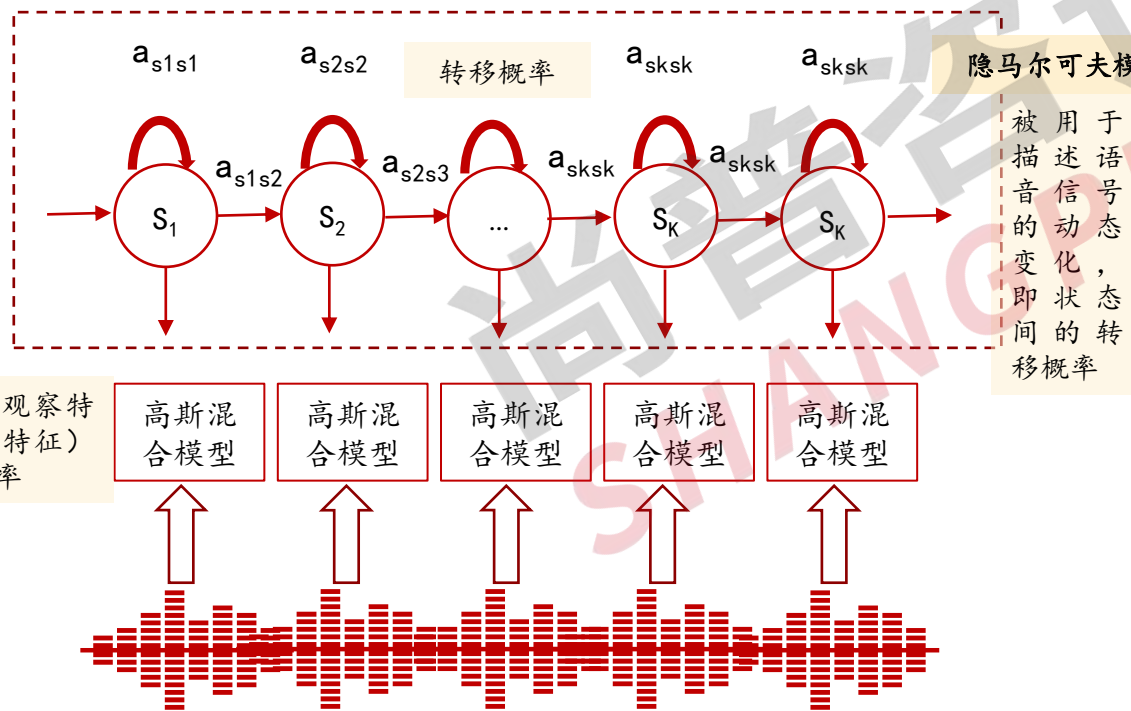
人工智能核心技术：语音处理—语音识别

基于深度神经网络-隐马尔可夫模型的声学模型成为主流声学建模技术

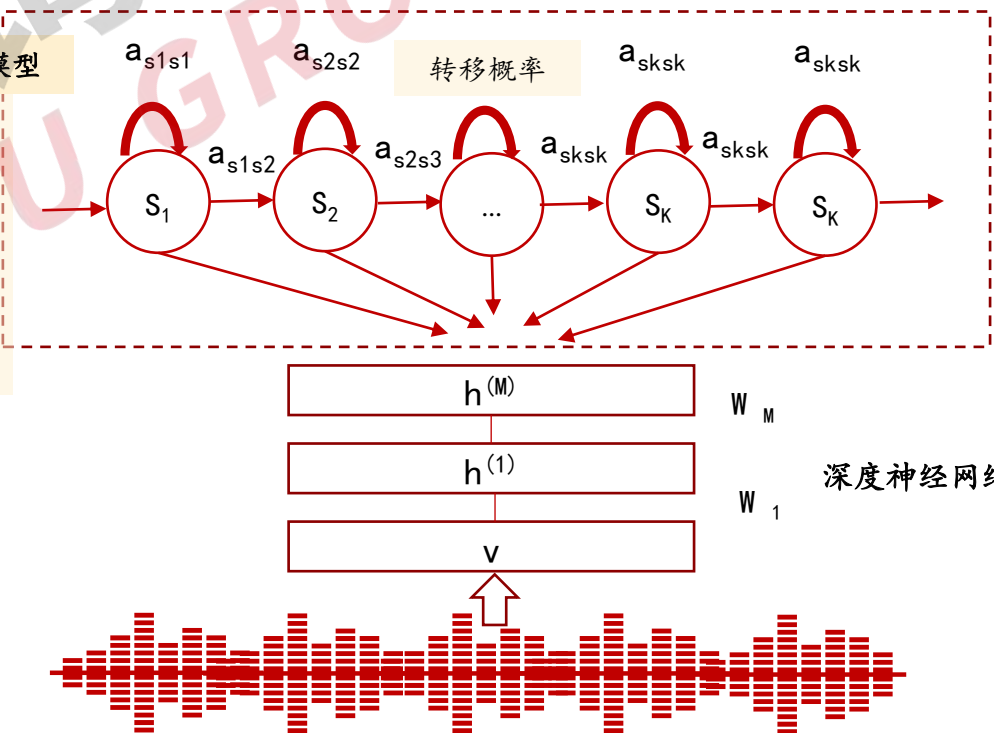
声学模型承载声学特征与建模单元之间的映射关系，建模单元包括音素、音节、词语等，其单元粒度依次增加。若采用词语作为建模单元，每个词语的长度不等，导致声学建模缺少灵活性，很难充分训练基于词语的模型。相比之下，词语中包含的音素是确定且有限的，利用大量的训练数据可以充分训练基于音素的模型，因此目前大多数声学模型一般采用音素作为建模单元。比较经典的声学模型是混合声学模型，通常可以分为两种：基于高斯混合模型-隐马尔可夫模型的声学模型和基于深度神经网络-隐马尔可夫模型的声学模型。后者具有两方面优势：一是深度神经网络能利用语音特征的上下文信息；二是深度神经网络能学习非线性的更高层次特征表达。因此，基于深度神经网络-隐马尔可夫模型的声学模型已成为目前主流的声学建模技术。

人工智能核心技术

基于高斯混合模型-隐马尔可夫模型的声学模型



基于深度神经网络-隐马尔可夫模型的声学模型



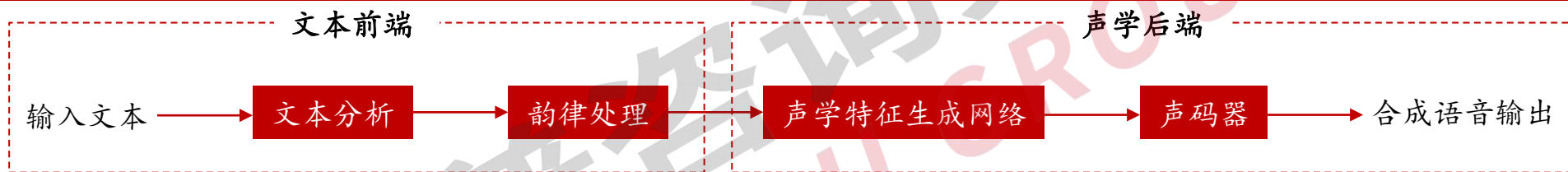
资料来源：中国科学技术出版社《人工智能导论》2018年8月第1版，人民邮电出版社《图解语音识别》2020年4月第1版，尚普研究院结合公开资料整理绘制

人工智能核心技术：语音处理—语音合成

语音合成：将任意输入的文本转换成自然流畅的语音输出

语音合成 (Text To Speech, TTS) 是一种将文本转换为语音的技术，赋予机器像人一样自如说话的能力，是人机语音交互中重要的一环。语音合成系统分为文本前端和声学后端。文本前端主要负责在语言层、语法层、语义层对输入文本进行预处理，如为文本添加韵律信息，并将文本词面转化为语言学特征序列；声学后端可以分为声学特征生成网络和声码器，声学特征生成网络根据文本前端输出的信息产生声学特征。声码器利用频谱等声学特征，生成语音样本点并重建时域波形。近年来出现的端到端语音合成系统则将声学后端合并为一个整体，可直接将语言学特征序列，甚至字符直接合成语音。目前，语音合成技术在银行、医院信息播报系统、自动应答呼叫中心等领域实现广泛应用。

语音合成 (Text To Speech, TTS) 系统框架



- #### 文本分析流程
- ✓ 文本预处理 (删除无效符号)
 - ✓ 文本规范 (识别特殊字符)
 - ✓ 分词 (以词为单位划分单元序列)
 - ✓ 词性标注
 - ✓ 多音字消歧
 - ✓ 节奏预测

- #### 语音合成模型
- ✓ Tacotron (端到端, 引入注意力机制)
 - ✓ DeepVoice (大量采集卷积结构, 提升模型推断速度)
 - ✓ Transformer (全连接层组成预处理网络, 提高训练速度)
 - ✓ FastSpeech (借鉴Transformer, 更快速)
 - ✓ LightTTS (应用于缺少平行语料的情况下)

- #### 四类声码器
- ✓ 纯信号处理 (Griffin-Lim算法)
 - ✓ 基于自回归神经网络模型 (WaveNet, WaveRNN)
 - ✓ 基于非自回归神经网络模型 (WaveGlow)
 - ✓ 基于生成对抗网络 (MelGAN)

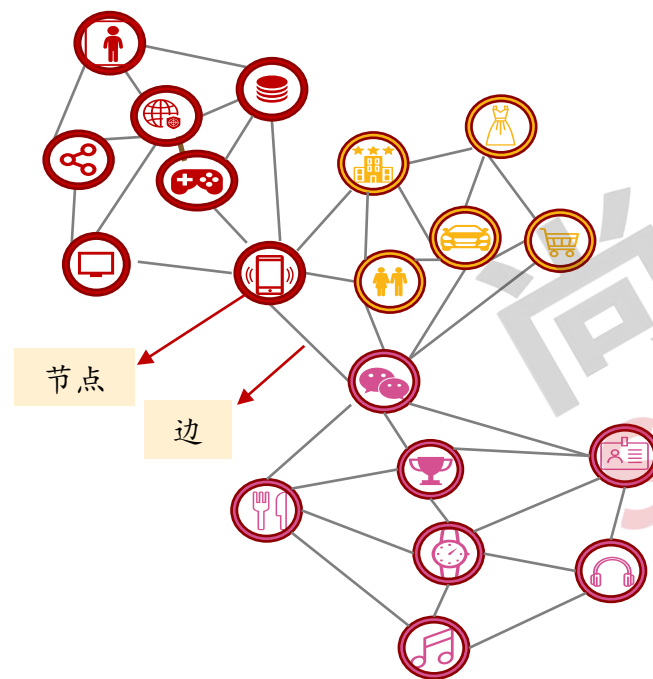
人工智能核心技术：知识图谱

知识图谱：利用图模型来描述知识和建模万物关系的语义网络

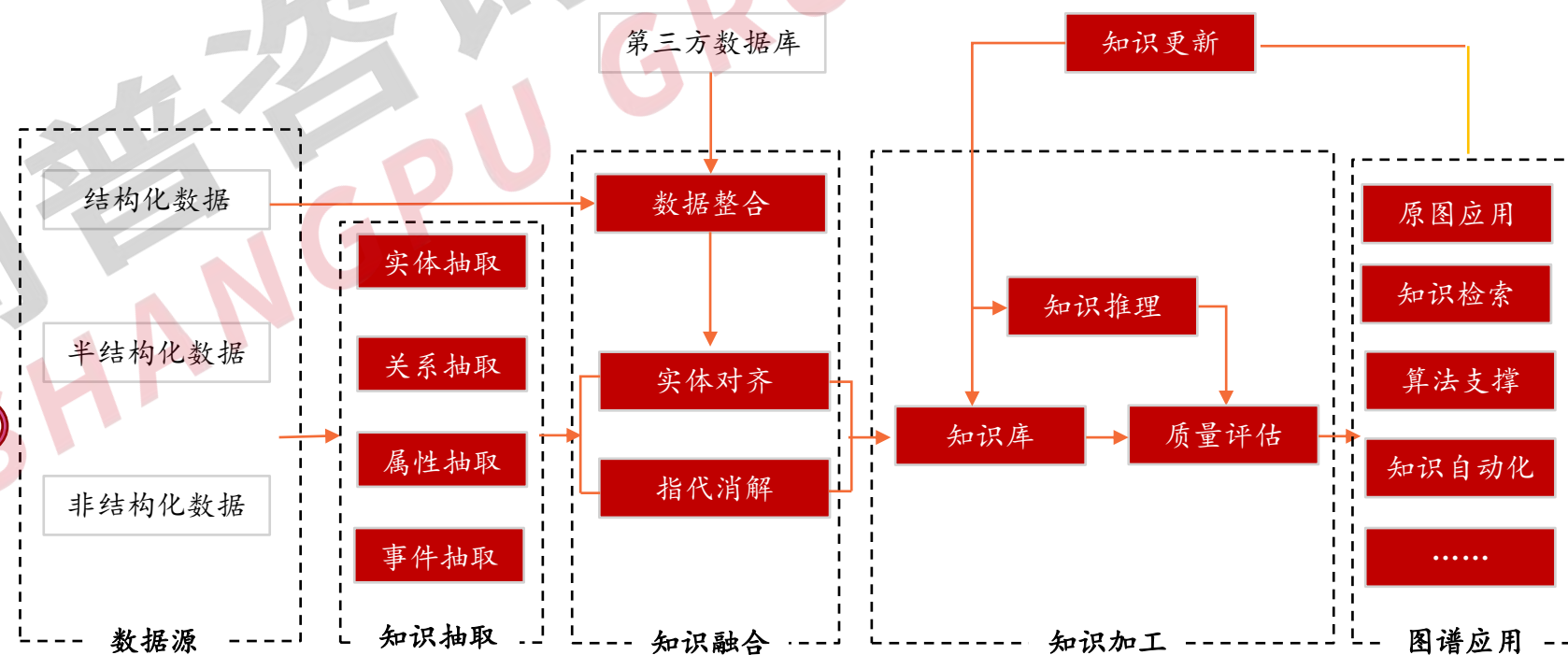
知识图谱 (Knowledge Graph)：于2012年由Google提出，其目标在于提高搜索引擎能力，提高用户搜索质量，改善用户搜索体验。知识图谱是以结构化方式描述客观世界中概念、实体之间的关系，最常见的表示形式是RDF（三元组），即“实体x关系x另一实体”或“实体x属性x属性值”集合，其节点代表实体 (entity) 或者概念 (concept)，边代表实体/概念之间的各种语义关系。知识图谱的构建过程是让计算机更好理解各行业领域信息，使其具备人的认知能力。**知识图谱基本构建流程**：知识抽取（如实体抽取、关系抽取）、知识融合（如数据整合、实体对齐）、知识加工（如知识推理）等。由于图数据相对于传统的关系型数据具有更强大的表达能力，善于处理大量、复杂、互联、多变的网状数据，因此图数据的计算与推理逐渐成为知识图谱的重要研究任务之一。

知识图谱 (Knowledge Graph)

知识图谱示例



知识图谱基本构建流程



4

人工智能应用领域

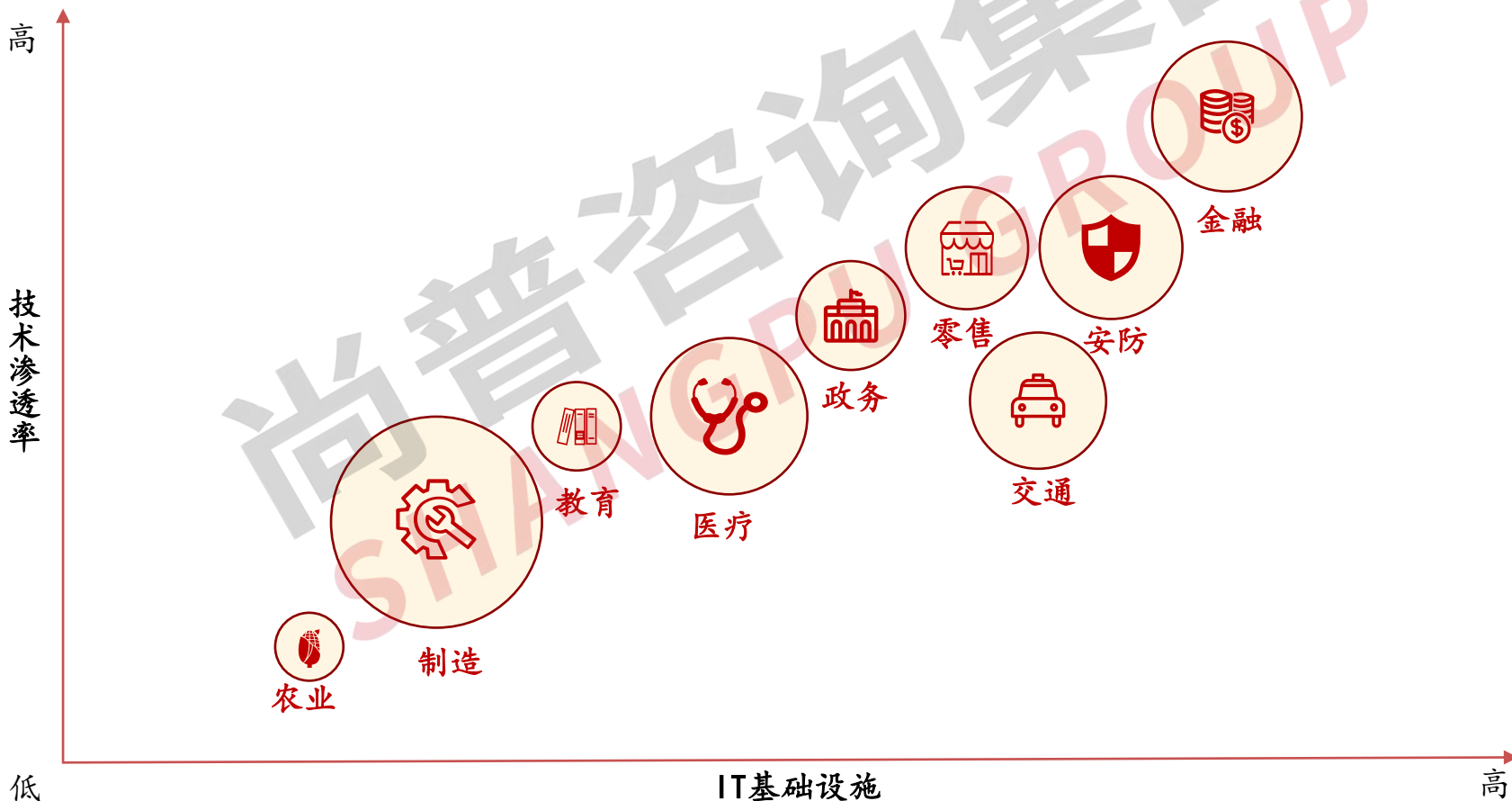
Artificial Intelligence Application Fields

人工智能应用领域总览

金融、安防等领域人工智能技术渗透率高，下游应用市场前景广阔

人工智能作为驱动全球产业数字化转型的核心技术，被广泛应用在金融、安防、制造、交通、医疗、教育、零售等多个领域。尚普研究院按照IT基础设施、技术渗透率和全球市场规模三大维度对人工智能技术的底层基础以及应用成熟度进行分析。综合来看，人工智能在金融、安防和零售领域技术融合程度高，医疗、制造领域未来发展空间较大。伴随技术的快速发展，未来人工智能在多个领域也会呈现技术渗透程度不断加深的趋势。

人工智能下游应用成熟度模型



注：圆圈大小表示市场规模 资料来源：Gartner，尚普研究院结合公开资料整理绘制 87

人工智能应用领域：安防

AI技术与安防领域深度融合，前端/边缘端实现智能决策成为主要趋势

伴随人工智能、5G通信、物联网等新兴技术的快速发展，安防产业正从传统的视频监控向智能安防方向转变，并逐步渗透到千行百业。从产业构成来看，AI+安防产业涵盖前端感知层（智能摄像机、传感设备等）、网络层（5G传输网络等）、后端服务平台层（智能计算设备、存储设备等），其中AI相关算法（人脸识别、机器学习等）主要涉及前端摄像机等感知设备以及后端智能服务器相关领域。未来，拥有AI芯片和算法的智能摄像机将得到广泛应用，从原有的后端决策转移到前端/边缘端决策，减少后端数据处理压力，降低网络带宽时延，进一步提升安防场景决策的时效性和准确性，由此带动全球智能安防产业规模不断扩大。

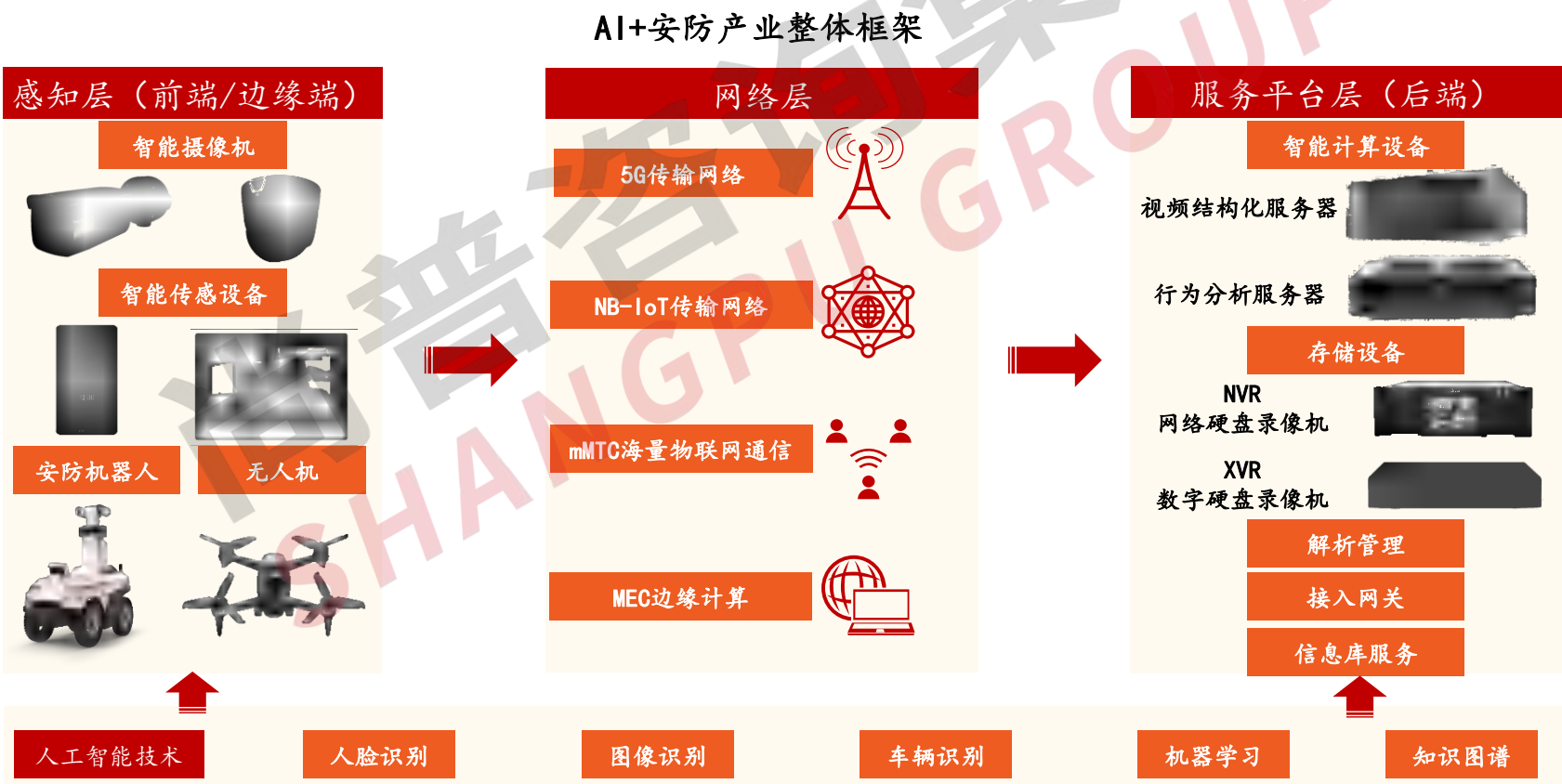
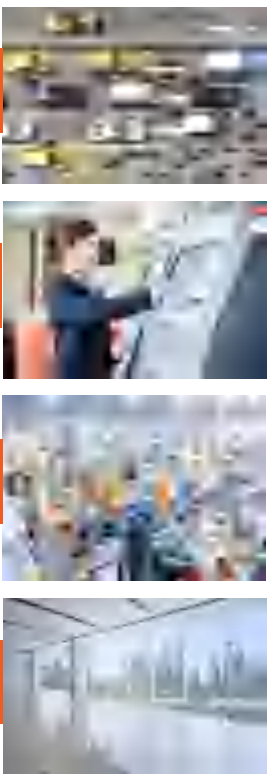
人工智能应用领域

交通

金融

工业

楼宇

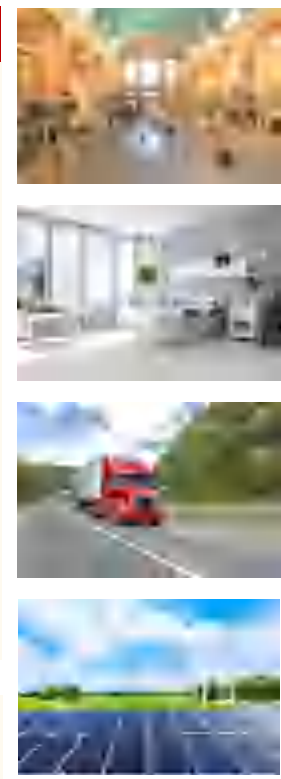


公共安全

医疗

物流

能源



人工智能应用领域：金融

人工智能推动金融领域开启全方位、多层次数字化转型

人工智能技术成为继互联网技术之后变革金融行业的全新力量，以其在机器学习、语音处理、图像识别、文字提取等方面的强大能力，逐渐渗透金融业务全价值链，从而不断拓展金融服务的广度和深度，大幅提高金融行业的经营效率。具体来看，人工智能与金融行业的融合体现在产品创新、精准营销、智慧运营、风险控制等方面，在智能投研、智能投顾、智能客服、智能营销、智能运营、智能合规、智能理赔、智能身份识别、智能风控等细分领域全面赋能金融业创新发展。



银行



保险



证券



支付



征信

产品研发		营销获客		运营管理			风险控制	
智能投研	智能投顾	智能客服	智能营销	智能运营	智能合规	智能理赔	智能风控	智能身份识别
深	深	深	深	深	深	浅	深	深
浅	浅	深	深	深	深	深	深	深
浅	深	深	深	深	深	浅	深	深
浅	浅	浅	深	深	深	浅	深	深
浅	浅	浅	深	深	深	浅	深	浅

- ✓ 机器学习
- ✓ 知识图谱
- ✓ 自然语言处理

- ✓ 机器学习
- ✓ 知识图谱
- ✓ 自然语言处理
- ✓ 智能语音

- ✓ 机器学习
- ✓ 知识图谱
- ✓ 自然语言处理
- ✓ 智能语音
- ✓ RPA
- ✓ 计算机视觉
- ✓ 生物特征识别

- ✓ 机器学习
- ✓ 知识图谱
- ✓ 自然语言处理
- ✓ RPA
- ✓ 计算机视觉
- ✓ 生物特征识别

图例 ■ 应用程度较深 ■ 应用程度一般 ■ 无应用

人工智能应用领域：医疗

人工智能技术在医疗影像、药物研发等细分领域实现广泛应用，应用程度逐步加深

当前，机器学习、深度学习、计算机视觉、自然语言处理等AI技术在医疗多个领域实现应用。在医疗影像方面，利用计算机视觉、深度学习技术实现了心血管、肺结节、眼底等多种疾病影像筛查；在药物研发方面，AI技术已覆盖到化合物研究、临床前研究和临床研究领域，提升新药研发效率；在医疗机器人方面，产品主要以康复机器人和手术机器人为核心，但机器人智能化水平和应用程度仍需提升；在健康管理方面，借助可穿戴设备等多样化手段，AI为人们身体健康提供精细化指导。未来，伴随技术的持续进步，人工智能将会在医疗领域发挥更大价值，持续提升人类的健康水平和幸福指数。

人工智能应用领域

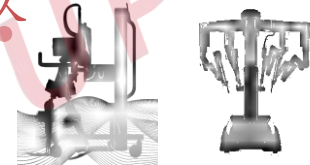
医疗影像

通过CV技术对医学影像进行快速阅片，辅助医师解读医学影像；通过DL对模型进行训练，实现分类、检测等功能。主要适用于心血管、肺炎、骨折等疾病。



主要AI技术：计算机视觉、深度学习

医疗机器人



主要AI技术：机器学习

康复机器人：辅助人体完成肢体动作，用于损伤后康复。
手术机器人：医生控制操作杆，手部动作传达到机械臂尖端完成手术操作，增加操作的精确性和平稳性。

化合物研究

利用NLP高效分析海量文献，增加新靶点发现几率，利用ML或CV实现化合物筛选，缩短研发周期降低研发成本

临床前研究

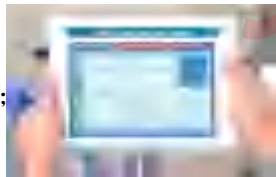
利用ML技术将药物与疾病进行匹配提升研发效率

临床研究将ML技术应用到统计分析等方面，提升临床试验效率

药物研发

主要AI技术：机器学习、计算机视觉、自然语言处理

语音输入自动转化为电子病历；利用ML和NLP抓取病历数据形成数据库；基于脱敏的电子病历数据，使用深度学习模型对患者病情进行预测。



主要AI技术：语音处理、自然语言处理、深度学习

电子病历

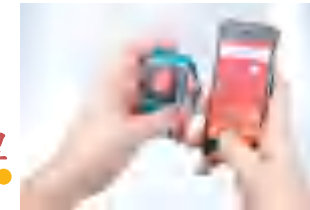
基因测序



主要AI技术：深度学习

目前高通量测序技术较难实现基因解读，从基因序列中挖掘有效信息有限；利用DL技术重建真实的基因组序列，结果准确性明显提升。

健康管理



主要AI技术：机器学习、知识图谱

通过智能可穿戴设备、家庭智能健康检测设备，实时动态监测健康数据，在血糖管理、血压管理、用药提醒、健康要素监测等方面，人工智能提供常态化、精细化指导。

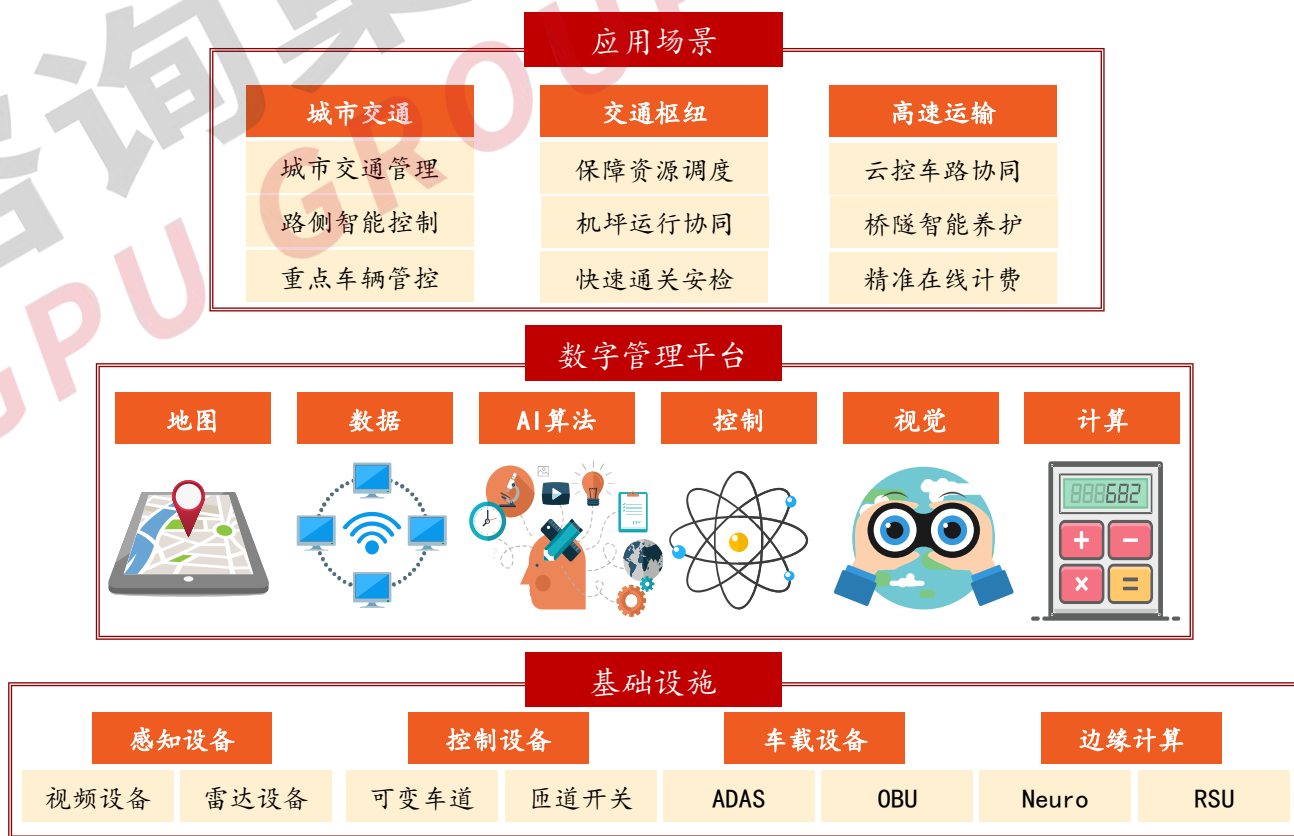
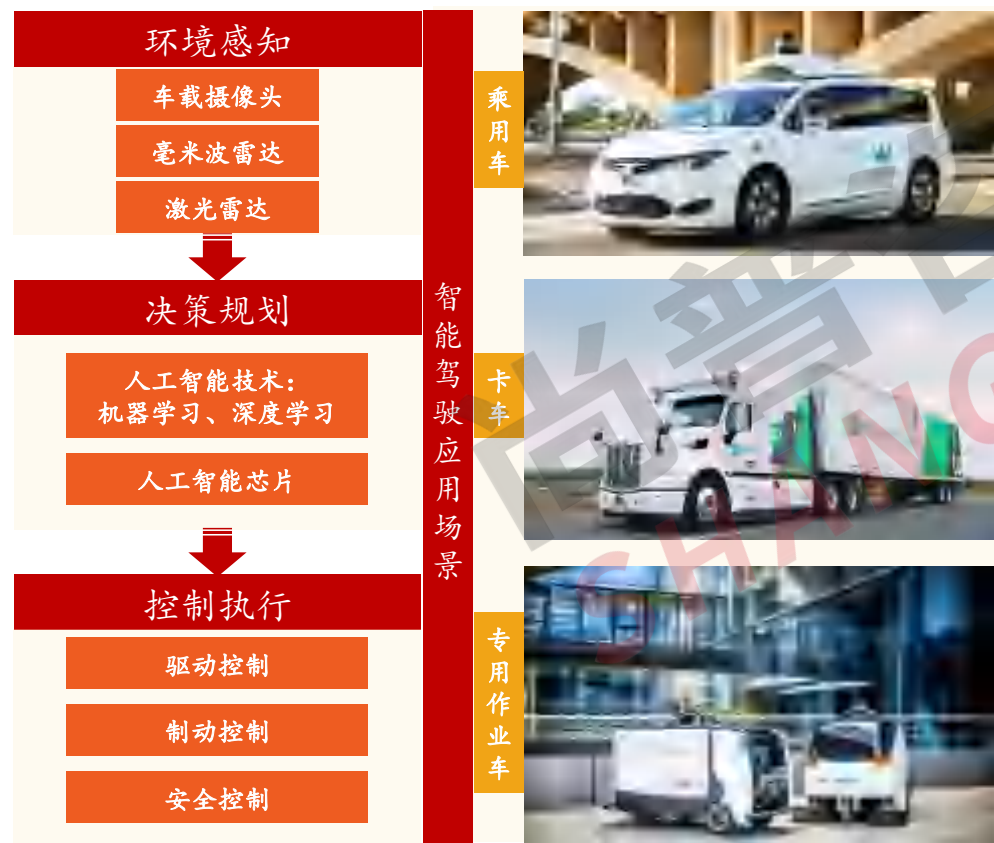
人工智能应用领域：交通

智能驾驶和智能交通管理平台为主要应用领域

人工智能技术在交通领域的应用主要涵盖智能车辆以及智能交通管理平台等领域。典型的智能驾驶系统主要由环境感知、决策规划和控制执行三部分构成，其中决策规划主要被AI算法和AI芯片所主导，结合环境感知所获取的数据信息，实现车辆轨迹预测和行驶路线规划。智能交通管理平台的核心为数字管理平台，其包含地图、数据、AI算法、控制、视觉、计算六大能力，为城市交通、交通枢纽、高速运输等各场景赋能。此外，感知设备、边缘计算等路侧、边缘侧基础设施也为智慧交通的实现提供持续动力。

智能驾驶技术及应用框架

智能交通管理平台



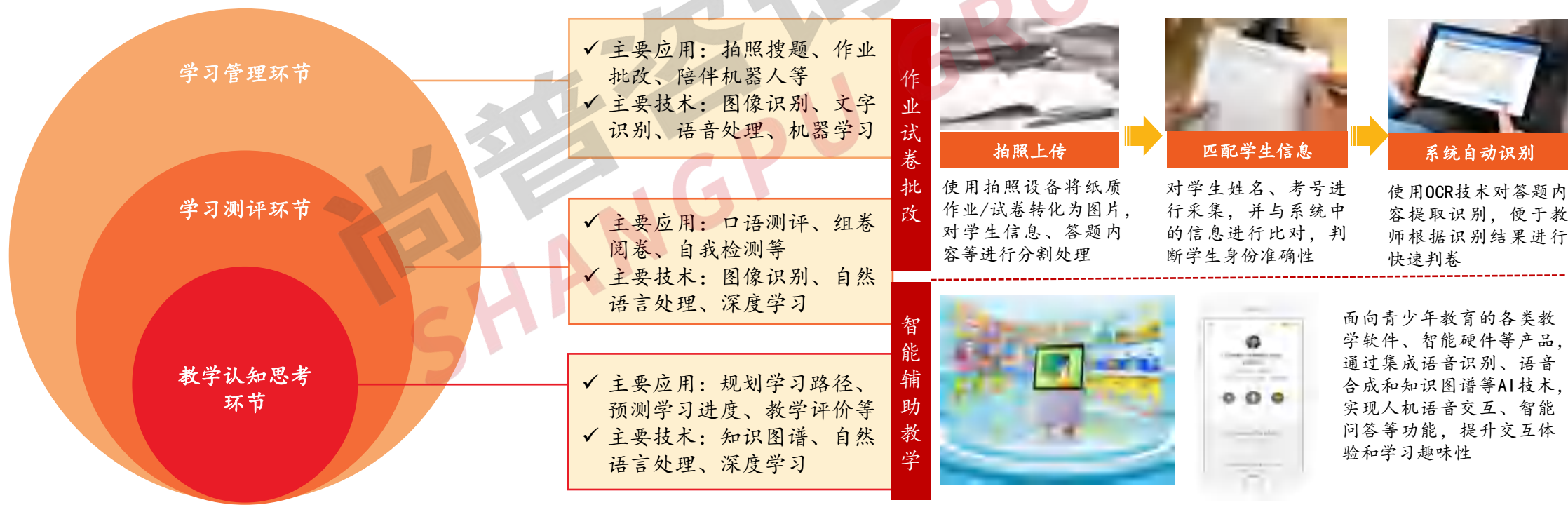
资料来源：部分图片来源于Waymo、图森未来、智行者，尚普研究院结合公开资料整理绘制 91

人工智能应用领域：教育

人工智能与教育三大环节紧密结合，持续提升教育智能化水平

人工智能为K12教育、高等教育、职业教育等多个教育细分领域实现赋能，并逐步覆盖外围“学习管理”、次外围“学习测评”以及核心“教学认知思考”三大环节。1) **学习管理环节**：采用图像识别、文字识别等技术实现拍照搜题、作业批改；2) **学习测评环节**：利用自然语言处理、深度学习等技术实现口语测评、组卷阅卷；3) **教学认知思考环节**：通过知识图谱、自然语言处理等技术开展规划学习路径、预测学习进度、教学评价等工作。通过人工智能技术与教育的融合，为教学活动主要参与者持续赋能。对于学生而言，通过智能设备进行数据分析，实现对学习情况的个性化分析，不断提升学习的效率与质量；对于教师而言，通过引入AI技术实现教学手段的多样化与智能化，为教师教学减负增效；对于教育管理者而言，人工智能为教学评价提供辅助决策与建议，为教学管理提供科学支撑。

智能教育主要应用环节



人工智能应用领域：制造

人工智能加速制造业由传统制造向智能制造转型升级

全球新冠疫情的蔓延导致供应链断裂、劳动力短缺等问题出现，一定程度上制约了全球制造企业的正常运营，人工智能在制造业中的应用能够有效抵御相关风险，推动制造业走向数字化、智能化的道路。借助AI技术，制造业在对海量工业数据的深度集成和分析的基础上，实现柔性化、自动化和网络化生产，大幅提高制造业生产效率。2010年以后，德国、中国、美国等国家相继提出智能制造战略，从顶层设计角度谋划制造业未来发展，抢占智能制造发展制高点。目前，人工智能与制造业的深度融合体现在设计、生产、运维、检测、物流等多个环节，在预测性维护、产品质量检测等多个场景均有所应用。



主要国家战略	主要领域	典型案例
 德国工业4.0	物理信息系统	西门子公司研发用于数控机床的控制器平台Sinumerik One，其具有创建相应“数字孪生模型”的多功能软件，可实现虚拟和现实的无缝交互，帮助机器制造商和操作员在纯虚拟环境中开发产品。
 中国制造2025	高端制造 智能工厂 核心装备	蔚来汽车选用达索系统3DEXPERIENCE作为全球研发平台，工程师可以随时访问完整的车辆数据，实现多国工程师在开发设计时的协作，快速实现产品设计迭代。
 美国先进制造 领先战略	先进传感 与控制 下一代机器人	波音公司通过模拟空气的流速对机翼产生的压强、发动机喷气产生的推力等，对飞机起飞条件进行测试，缩短飞机的研制周期，并降低研发测试所需的能耗，降低碳排放。

人工智能应用领域：政务

人工智能持续提升政务服务的数字化、智能化水平

人工智能作为新一代信息技术，将为数字政务的发展注入新动力，开启“人工智能+政务”治理新范式。人工智能与政务服务的融合主要体现在服务性、管理性与决策性的应用场景，具体如下：**1) 简化服务流程。**政务服务涉及审批材料等程序，人工智能以“用户画像”的方式向公众提供订制化、自动化和精准化服务；**2) 提高管理效率。**人工智能可以通过智能外呼等途径对政务服务过程进行有效的监督、检查；**3) 赋能科学决策。**依托人工智能可以打破部门间行政数据壁垒，实现跨层级、跨系统、跨部门、跨业务的信任和协同，助力政府决策的自动化、精准化，实现经济社会的和谐治理。

人工智能应用领域

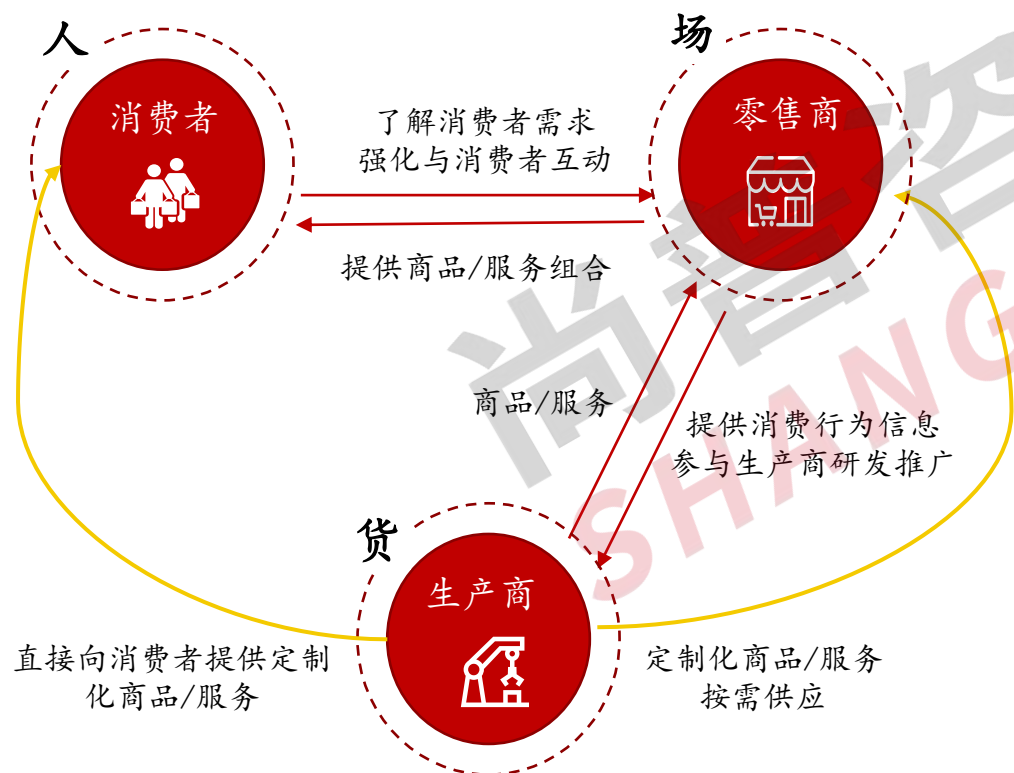


人工智能应用领域：零售

人工智能促进零售企业降本增效、提升消费者购物体验

计算机视觉、自然语言处理等AI技术在零售场景中的落地应用，为零售行业的参与主体、不同业务环节赋能，进而实现对零售行业的整体升级和改造。人工智能推动零售由线性结构向环状结构转变，实现用户体验最优、生产商成本最低、零售商收益最大。根据Global Market Insights数据，到2024年全球人工智能在零售领域的市场规模将达到80亿美元，2018—2024年复合增速超40%。人工智能在零售领域应用场景主要包括：吸引消费者参与、智能货品管理、重新定义门店、智慧供应链。未来，以大型零售企业为主导的人工智能技术中台（通用化AI智能服务综合设施）建设将成为“人工智能+零售”的主流技术趋势。此外，人工智能技术也将逐渐覆盖零售的全产业链条和线上线下多零售业态。

AI+零售商业模式图



吸引消费者参与
用户画像、多层次线上购物体验、智能客服



智慧供应链
智能定价、智慧物流、仓储管理



智能货品管理
商品定价设置、自主补货提醒、促销活动设计



重新定义门店
智能门店选址、智慧门店、无人零售

人工智能应用领域：农业

人工智能驱动农业向精准化、智慧化方向发展

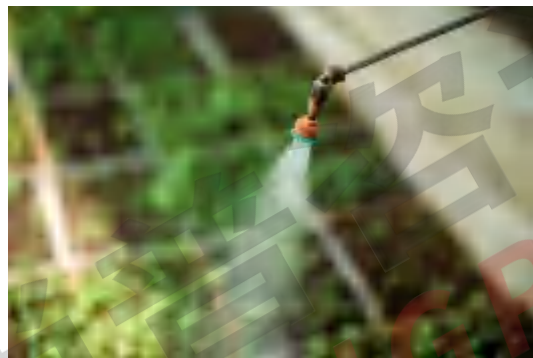
高盛预计，到2025年新一代人工智能技术在农业领域中的应用将达到200亿美元市场规模。人工智能技术和可提供实时数据的物联网传感器可以提高作物产量，降低粮食生产成本。目前，人工智能与农业领域的结合已经逐步体现在播种、施肥、除草、灌溉、病虫害防治等农业生产的多个环节。未来，随着人工智能技术与农业场景的深度融合，人工智能赋能农业将逐步由作业管理走向农业预测，农业生产的集约化、精准化、智慧化程度将进一步提高。

病虫害管理



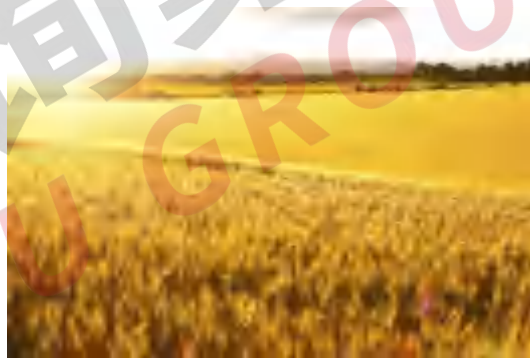
- ✓ 利用无人机的红外相机数据与田间的传感器相结合，可以监测植物的相对健康水平，可以在虫害发生之前预测和识别虫害。
- ✓ 通过使用智能传感器结合无人机的视觉数据流，可以定义农药的最佳组合，以减少害虫的威胁进一步扩散。

优化农业资源



- ✓ 基于人工智能技术的智能灌溉控制系统，可以根据当地气候数据，对灌溉用水进行分析，选择最佳灌溉策略。
- ✓ 跟踪和追溯系统依靠先进的传感器获得对每个农作物状况的更多了解，有助于通过在供应链上提供更大的可视性和可控性来安排农业生产。

产量预测



- ✓ 使用机器学习技术分析3D地图、基于无人机的土壤颜色数据，预测特定作物的潜在土壤产量。
- ✓ 根据产量对作物进行价格预测，确定特定作物的价格弹性曲线，从而预测生产总量。

农业机器人



- ✓ 农业机器人是一种可由不同程序软件控制，以适应各种作业，能感觉并适应作物种类或环境变化的新一代无人自动操作机械。对自走式机器人机械进行编程，在每行作物上分配肥料，有助于降低运营成本，进一步提高田间产量。



5 人工智能趋势展望

Artificial Intelligence Trend Outlook

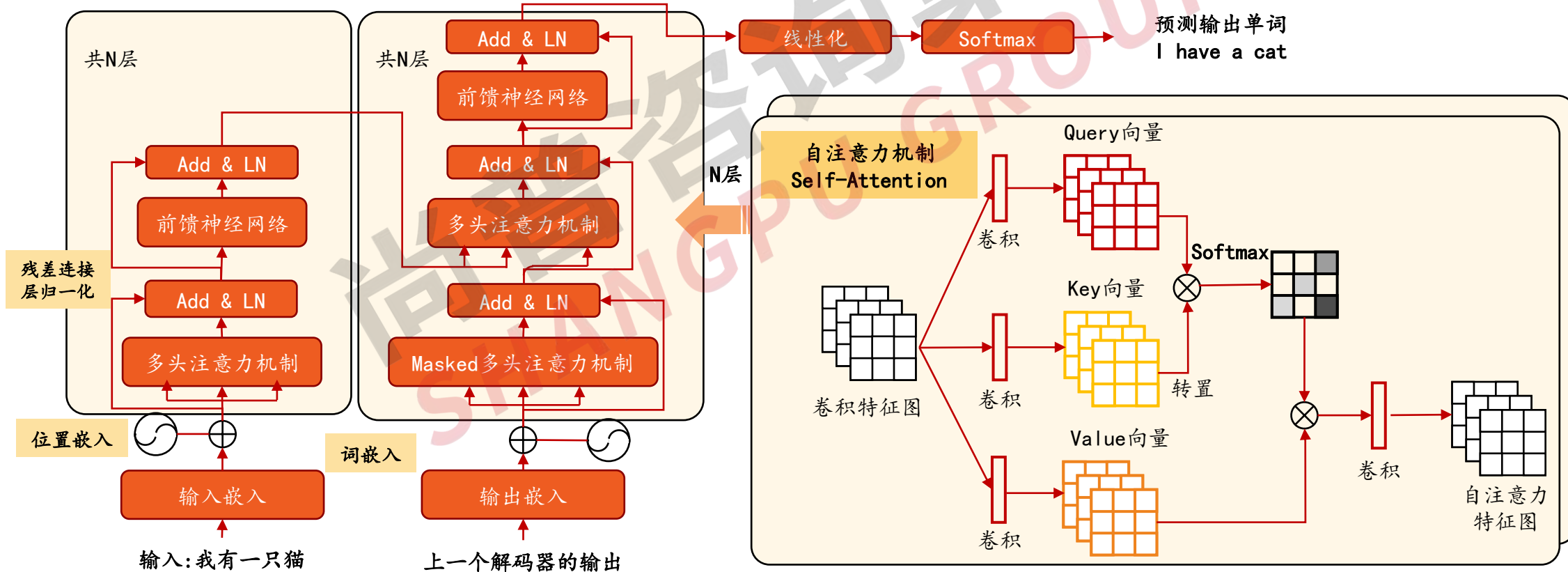
前沿技术：Transformer模型

Transformer模型：基于自注意力机制，有效提高模型训练效率

Transformer模型：由Google的Ashish Vaswani等人和多伦多大学的Aidan N. Gomez于2017年首次提出，是一种基于自注意力机制（在Transformer模型中起基础作用，可减少对外部信息的依赖，更擅长捕捉数据或特征的内部关系，优化模型训练结果）的深度学习模型，该模型主要由编码器和解码器构成，模型本身并行度较高，在精度和性能上均要优于传统的循环神经网络（RNN）和卷积神经网络（CNN）。Transformer模型在简单语言问答和语言建模任务上有着较好表现。Transformer模型仍存在一些不足，如对局部信息的获取能力弱于RNN和CNN、不能很好表征单词的位置信息、顶层梯度消失等问题。

人工智能趋势展望

Transformer模型



资料来源：Vaswani A et al., Attention Is All You Need, 尚普研究院结合公开资料整理绘制 98

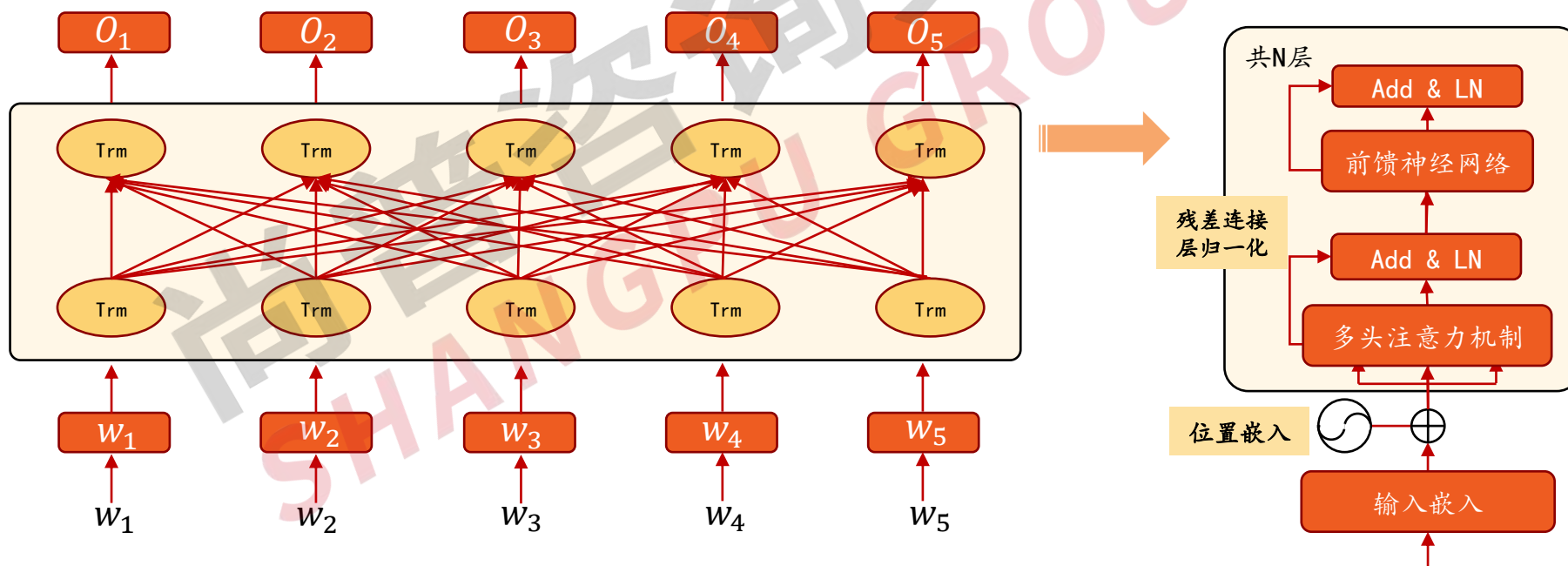
前沿技术：BERT模型

BERT模型：基于Transformer Encoder构建的预测模型

BERT模型：由Google于2018年提出，是基于Transformer Encoder构建的一种模型。**模型基本思想**：给定上下文来预测下一个词。BERT模型架构是由多接口组成的Transformer编码器层，即全连接神经网络增加自注意力机制。对于序列中的每个输入标记，每个接口计算键、值和查询向量，相关向量用于创建加权表示，合并同一层中所有接口输出并通过全连接层运行。每个层使用跳跃连接进行包装，之后将层归一化处理。BERT模型传统 workflow 主要包含预训练和模型微调两部分，其中预训练环节涉及MLM和NSP两个半监督任务；模型微调环节包含一个或更多全连接层，通常增加到最终编码器层的顶部。BERT模型已集成在谷歌搜索中，并提升10%的查询准确率。

人工智能趋势展望

BERT模型 (Bidirectional Encoder Representation from Transformer)



前沿技术：ViT模型

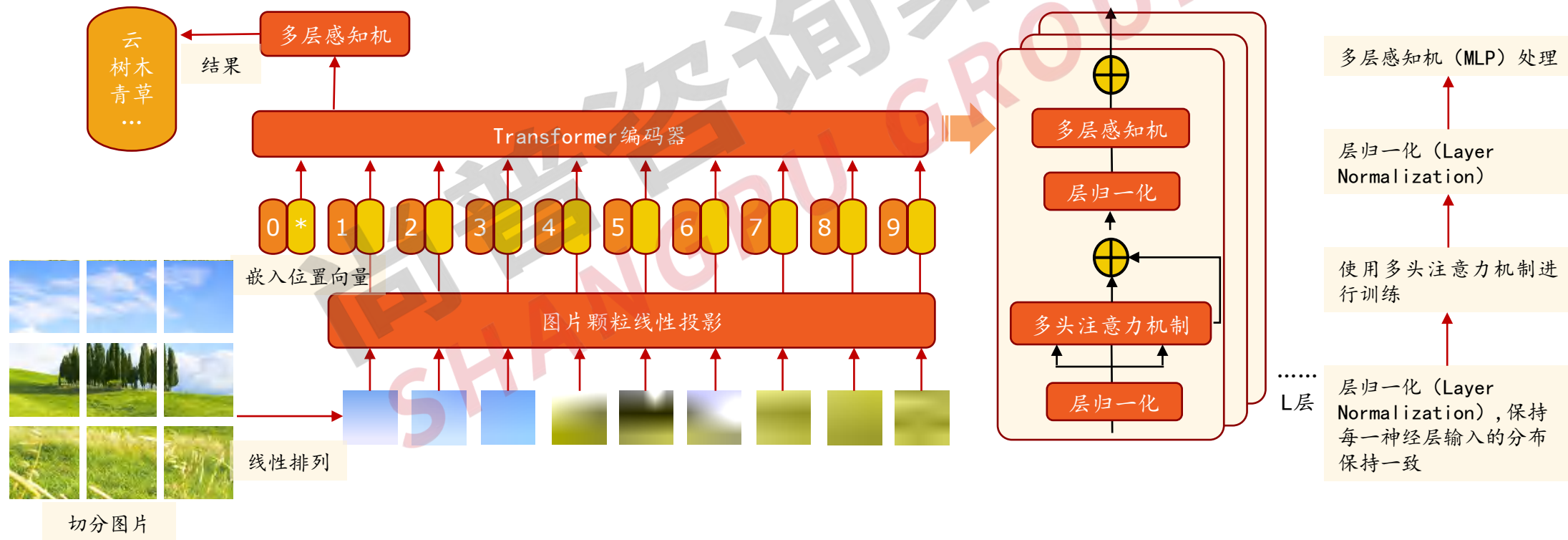
ViT模型：将Transformer模型引入计算机视觉领域，大幅提高图片识别率

ViT模型 (Vision Transformer): 由Google公司Alexey Dosovitskiy等人在2021年提出，其原理是将原本用于NLP领域的基于自注意力机制的Transformer模型应用于计算机视觉领域。相比于传统CNN算法，ViT模型在大数据集上识别率更高、成本更低。**ViT模型的基本原理:**

- 1) 将图片切分为相同大小的块，每个块可以看做一个“单词”；
- 2) 每个块经过线性投射层（全连接层）的操作得到位置编码（图像分类问题转化为NLP问题）；
- 3) 将线性投射后的图像块输入由L层Transformer模型构成的编码器；
- 4) 输出结果经过多层感知机（MLP）得到最终分类结果。目前，ViT已发展出多种改进模型，在计算机视觉物体检测等方面具有很大应用潜力。

人工智能趋势展望

ViT模型 (Vision Transformer)



前沿技术：自监督学习

自监督学习：将无监督问题转化为有监督问题的方法

自监督学习 (Self-supervised Learning): 旨在对于无标签数据, 通过设计辅助任务来挖掘数据自身的表征特性作为监督信息, 来提升模型的特征提取能力, 将无监督问题转化为有监督问题的方法。辅助任务是自监督学习最关键内容, 目前主要包括自然语言处理NLP和计算机视觉CV两大类任务。其中, 自然语言处理包括单词预测、句子序列预测、词序列预测; 计算机视觉包括图像任务、视频任务等。著名AI科学家Yann Lecun曾经提出, 如果将人工智能比作一块蛋糕, 蛋糕的大部分是自监督学习, 蛋糕上的糖衣是监督学习, 蛋糕上的樱桃是强化学习, 自监督学习在人工智能领域占据重要地位。

人工智能趋势展望

自监督学习 (Self-supervised Learning)

自监督学习中辅助任务的构造—自然语言处理 (NLP)

单词预测	原始信息	A quick rabbit jumps over the lazy dog.
	随机遮挡一个单词	A quick [MASK] jumps over the [MASK] dog.
	预测	A quick rabbit jumps over the lazy dog.

句子序列预测	原始信息	I completed high school. Then I did my undergrad.
	句子1 ✓	I completed high school.
	句子2 ✗	Then I did my undergrad.

词序列预测	打乱正常语句中的单词顺序, 让模型学习组句, 标签信息为原来正确的词序。
	I did X. Then I did Y. Finally I did Z.

自监督学习中辅助任务的构造—计算机视觉 (CV)

图像任务

将图像分为不同块, 比如九宫格, 然后让网络预测不同块相对位置信息。该过程可提高模型局部特征提取能力以及全局空间信息提取能力。

视频任务

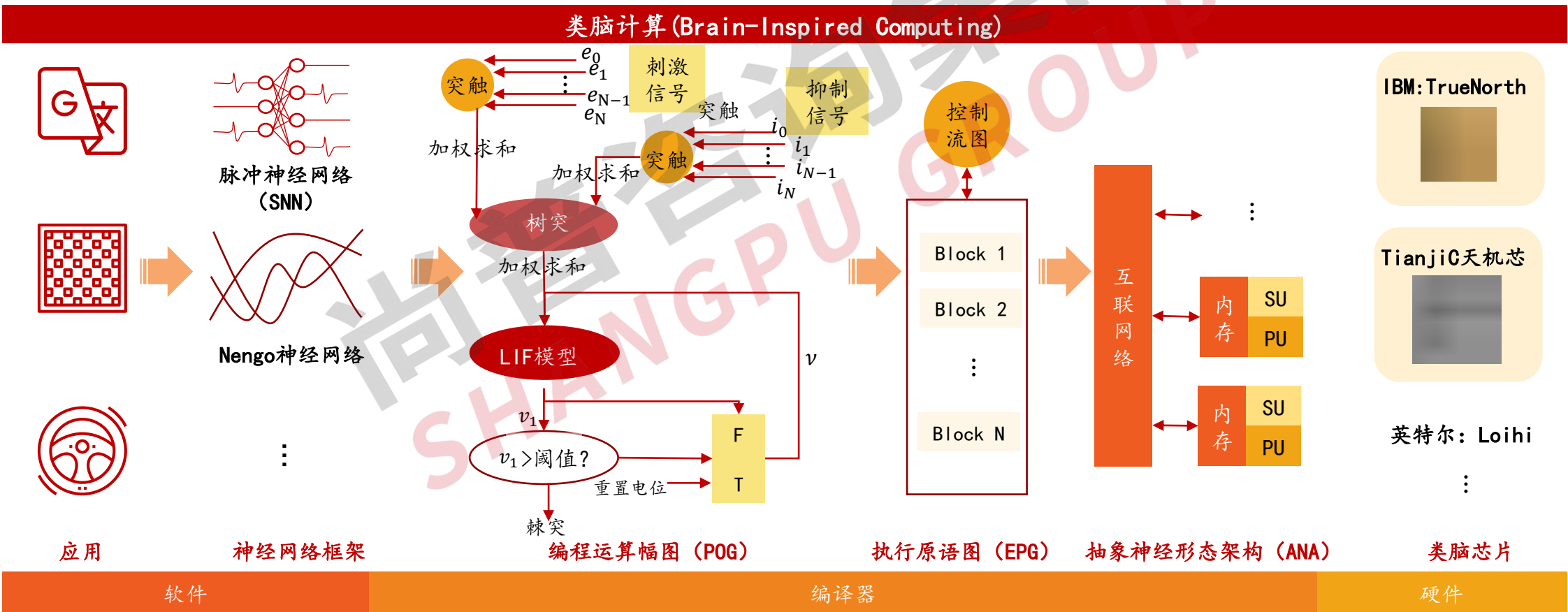
对于视频中的每一帧, 存在特征相似的概念。视频相邻帧特征是相似的, 而相隔较远的视频帧是不相似的, 通过构建正负样本来进行自监督约束。如图所示, positive range是在一个区域内随机选择, negative range在margin range之外随机选择。

前沿技术：类脑计算

类脑计算：模拟大脑结构和信息加工过程，提高机器认知能力、降低运行功耗

类脑计算(Brain-Inspired Computing): 又称神经形态计算，是借鉴生物神经系统信息处理模式和结构的计算理论、体系结构、芯片设计以及应用模型与算法的总称。类脑计算可模拟人类大脑信息处理方式，以极低的功耗对信息进行异步、并行、高速和分布式处理，并具备自主感知、识别和学习等多种能力，是实现通用人工智能的途径之一。2020年10月，清华大学张悠慧等人在Nature杂志发表论文，首次提出“类脑计算完备性”，填补类脑研究完备性理论与相应系统层次结构方面的空白。类脑计算技术的发展将推动图像识别、语音识别、自然语言处理等前沿技术的突破，有望推动新一轮技术革命。

人工智能趋势展望



资料来源：张悠慧等著, A system hierarchy for brain-inspired computing, Nature, 尚普研究院结合公开资料整理绘制 102

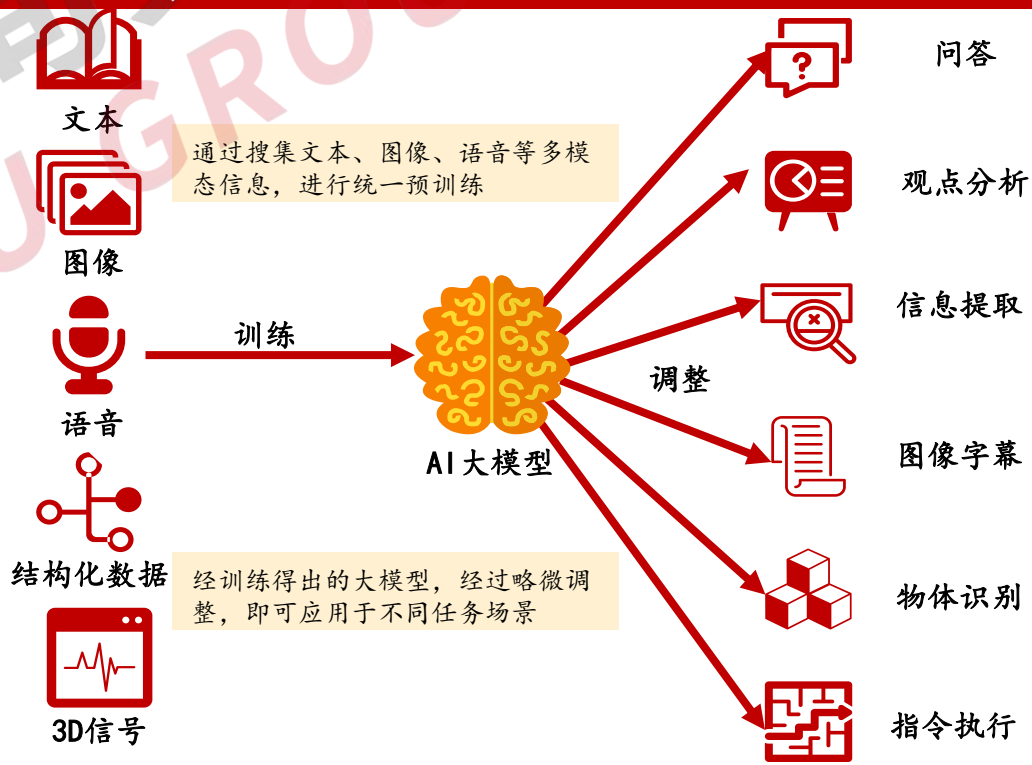
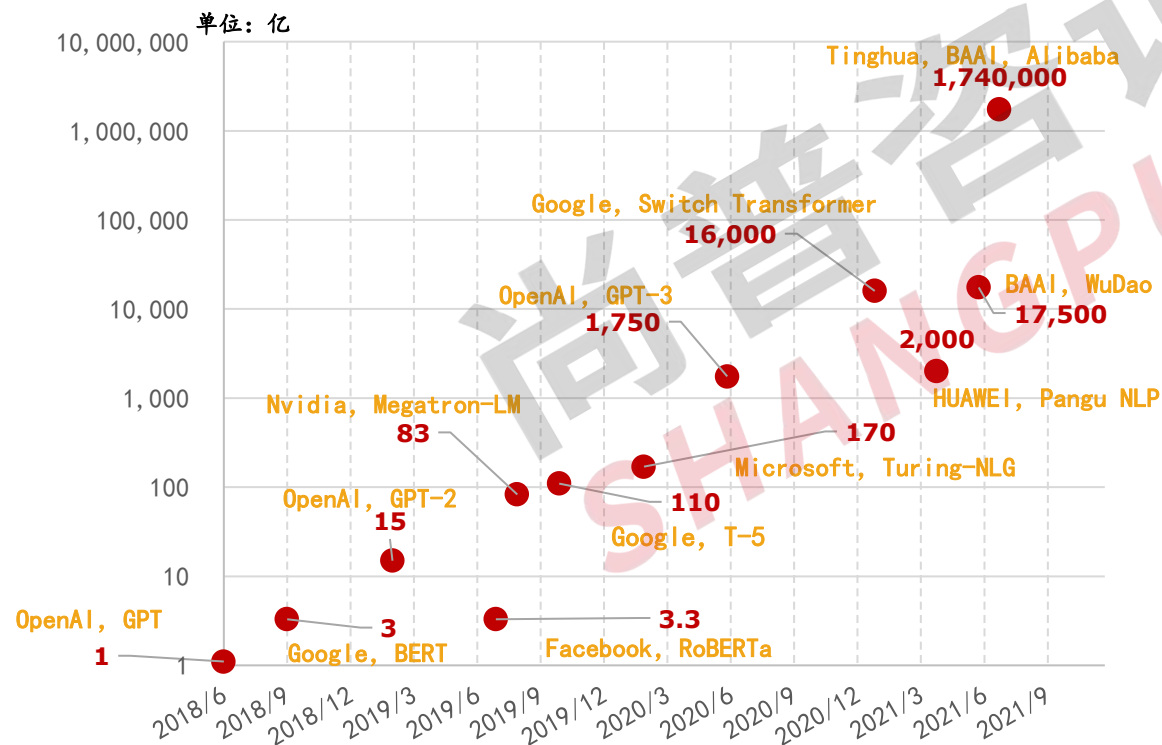
前沿技术：AI大模型

AI大模型：包含万亿量级参数的预训练模型，显著降低模型训练成本

AI大模型 (Foundation Models):是指经过大规模数据训练且在经微调后即可适应广泛下游任务的模型。随着参数规模不断扩大，AI大模型在语言、视觉、推理、人机交互等领域涌现出新能力。由于各类AI大模型可有效应用于多种任务，各类AI大模型同质化特征愈发明显。伴随2017年Transformer模型的提出，深度学习模型参数数量突破1亿个。此后，模型参数数量迅速增长，其中BAGUALU模型于2021年7月发布，参数规模已达到174万亿个。模型参数规模的增大，有助于进一步提高模型的运算精度。AI大模型的应用向多个应用场景提供预训练方案，其使用的自监督学习方法也可减少标注数据的使用，降低训练研发成本。具体而言，AI大模型在医疗和生物医药、法律和教育等领域拥有广阔应用前景。

人工智能趋势展望

AI大模型 (Foundation Models)



资料来源: Li Fei-Eei et al., On the Opportunities and Risks of Foundation Models, 尚普研究院结合公开资料整理绘制 103

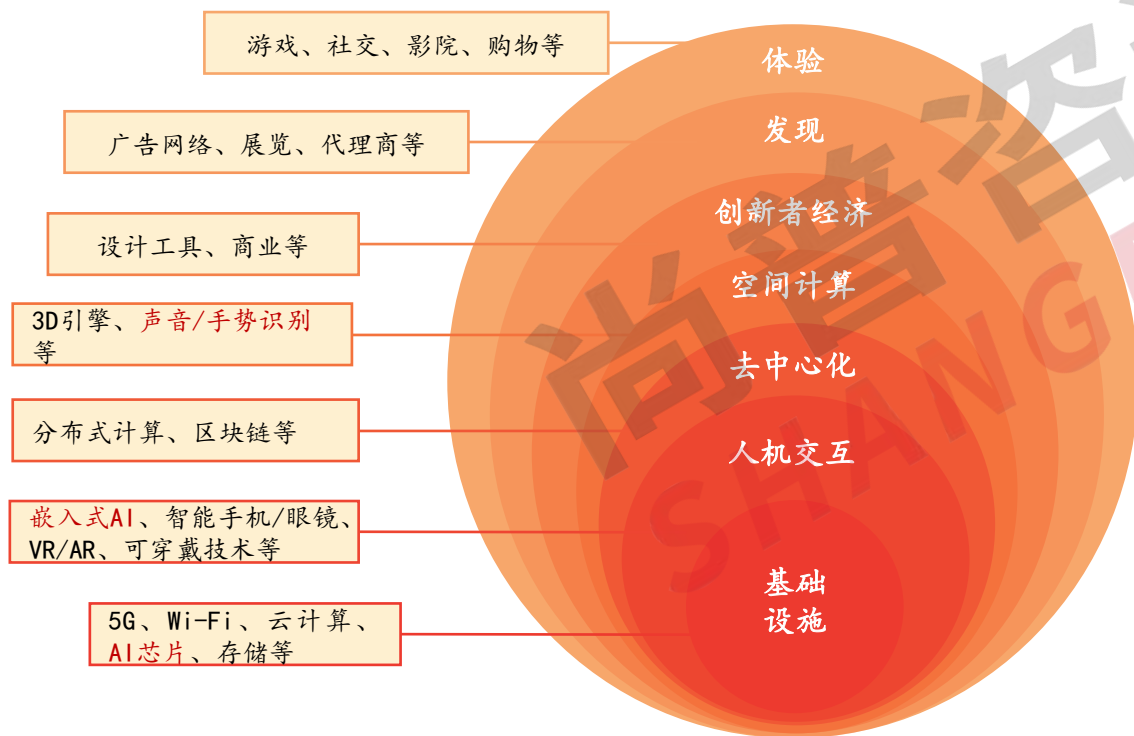
产业融合：人工智能与元宇宙

人工智能作为元宇宙时代的核心生产要素，加速元宇宙商业化落地

元宇宙 (Metaverse)：本质上是对现实世界的虚拟化、数字化过程，其主要包括基础设施、人机交互、空间计算等七层架构，其中计算机视觉、AI芯片和嵌入式AI等人工智能技术及基础设施共同助力元宇宙加速落地。元宇宙涵盖芯片、云计算、技术平台、通信、智能设备、内容服务等庞大生态系统。当前全球科技企业纷纷投入到元宇宙建设热潮中，如Nvidia、Meta、Microsoft等科技巨头及国内企业推出元宇宙解决方案，应用于办公、娱乐、设计等领域。

人工智能趋势展望

元宇宙七层经典架构 (The Seven Layers of the Metaverse)



元宇宙典型案例

 <p>NVIDIA Omniverse Avatar AI助手利用语音、计算机视觉、自然语言理解等技术，可以处理日常客户服务互动，比如餐厅订单、银行交易、个人预约等。</p>	 <p>NVIDIA Omniverse Create 用于加速高级场景合成的应用，支持用户依照 Pixar USD，以实时、交互方式组装、照明、模拟、渲染环境并与环境互动。</p>	 <p>Meta Horizon Workrooms 虚拟会议室成为来自任何地点同事更好工作的共同空间。可以使用VR以虚拟人形式或者通过视频通话的方式加入会议。</p>
 <p>NVIDIA Omniverse Replicator 合成数据生成引擎，能够生成用于训练深度神经网络的物理模拟合成数据。涵盖各类场景，如在现实世界中无法经常经历或安全体验的罕见和危险情况。</p>	 <p>NVIDIA Omniverse View 建筑、工程和设计、制造和产品开发以及媒体和娱乐行业的项目评审人员能以令人惊叹的逼真度协同评审3D设计项目。</p>	 <p>Meta Horizon Worlds 为用户带来虚拟现实世界的更多娱乐操控体验，通过设定个人边界防止虚拟人之间产生过度交互。</p>

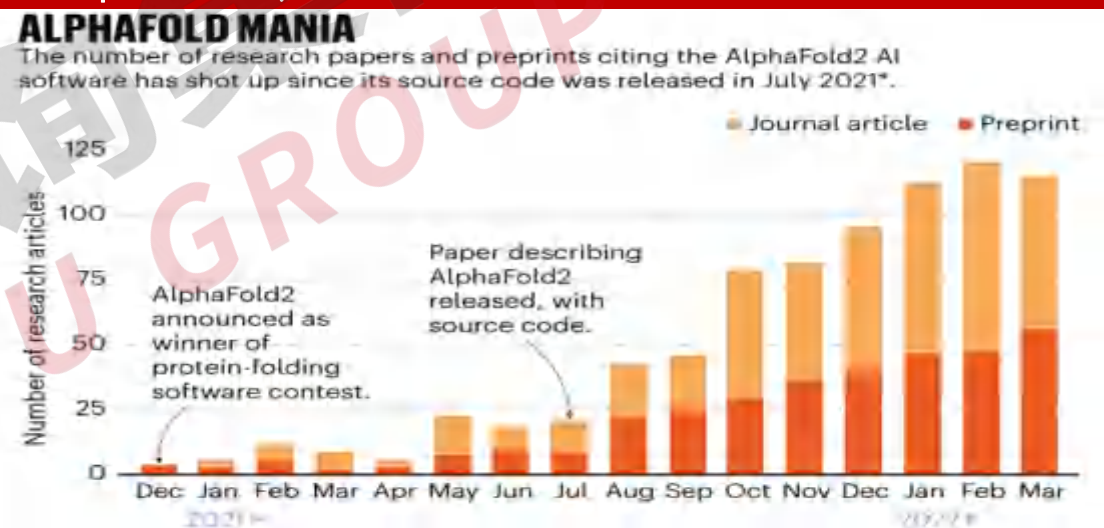
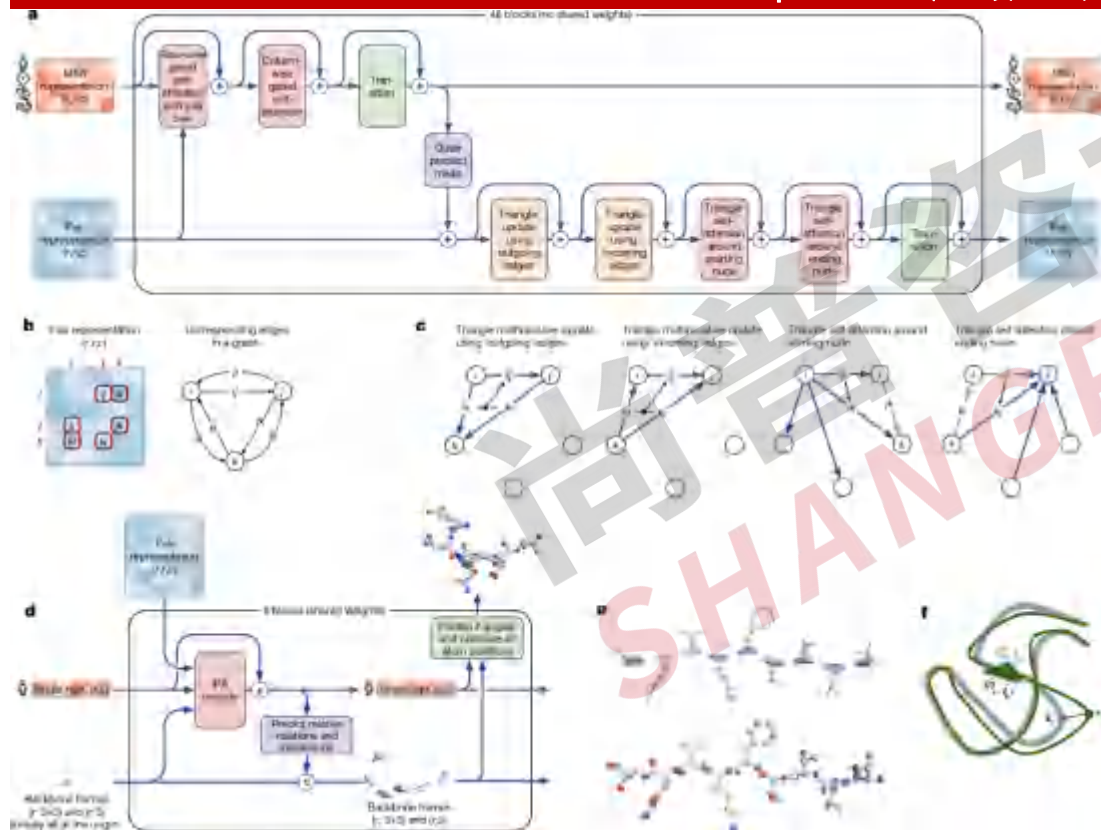
产业融合：人工智能与生命科学

AlphaFold驱动人工智能在生命科学领域实现突破

AlphaFold是由谷歌旗下DeepMind团队基于深度学习算法的蛋白质结构预测的人工智能系统，其被视作人工智能深入到生物领域的一大突破。目前AlphaFold已对98.5%的人类蛋白质结构做出预测，此外还对于大肠杆菌、果蝇、斑马鱼、小鼠等研究时常用生物的蛋白质结构进行预测。DeepMind与欧洲生物信息研究所（EMBL-EBI）联合发布由AlphaFold预测的蛋白质结构数据库—AlphaFold Protein Structure Database，该数据库已包含约35万个蛋白质结构。

人工智能趋势展望

AlphaFold蛋白质结构预测及AlphaFold发展历程



AlphaFold由两个主要阶段组成：

第1阶段：以氨基酸序列和多序列对齐（MSA）作为输入。目标是学习一个丰富的「成对表示」。该表示可提供「哪些残基对在3D空间中接近」的信息。

第2阶段：使用第一阶段的表示法直接产生原子坐标；将每个残基作为一个单独的物体，预测放置每个残基所需的旋转和平移；最终组装成一个完整的结构链，可以生成一个基于网络中间层表示的三维结构。

资料来源：Nature官网，尚普研究院结合公开资料整理绘制 105

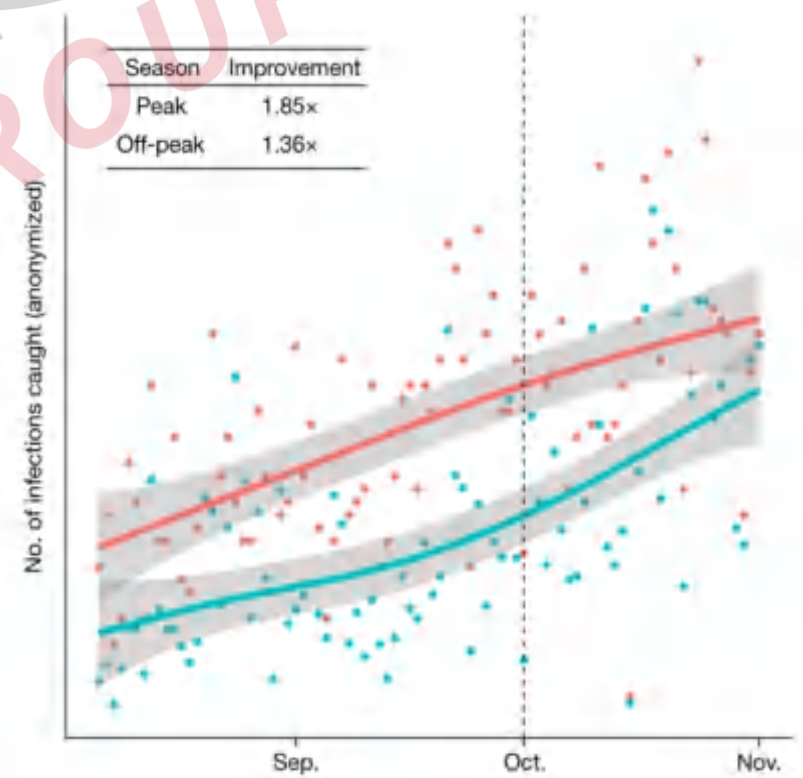
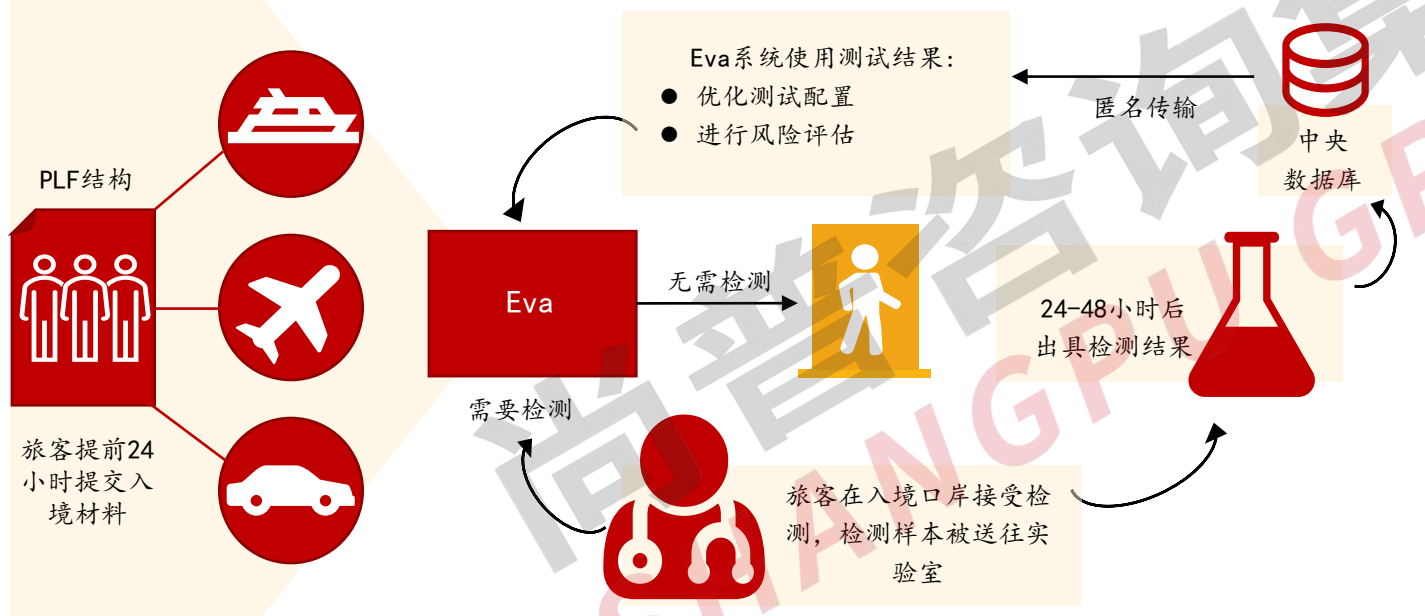
产业融合：人工智能与新冠疫情

Eva强化学习系统提升入境旅客新冠病毒检测效能

Eva是用于检测入境旅客新冠病毒的强化学习系统，其由美国南加州大学、美国宾夕法尼亚大学、AgentRisk以及希腊相关专家合作开发。2020年，Eva系统被部署到希腊所有入境口岸（机场、港口、车站等），用于识别限制新冠无症状旅客入境。借助Eva系统，希腊每天对抵达或途经该国大约41,000户家庭中约17%人员进行检测。通过比较，Eva发现感染的旅行者比原先严格按照其国籍检测的方式多1.25-1.45倍。与随机检测相比，Eva在旅游旺季发现感染的旅行者是前者的4倍，非旅游旺季的结果是随机检测的1.85倍，取得良好检测效果。

人工智能趋势展望

Eva强化学习系统：新冠无症状入境旅客检测流程



新冠无症状入境旅客检测流程：

- 1) 旅客在到达目的地24小时之前提交旅行信息；
- 2) 根据旅客的相关数据，Eva选择部分乘客进行测试；
- 3) 选定的旅客需要进行1-2天的自我隔离；
- 4) 旅客检测结果呈阳性需要进一步隔离并进行接触者追踪；
- 5) 检测结果可以用来更新Eva，持续提升检测效果。

产业融合：人工智能与半导体

AI与EDA紧密融合，促使芯片PPA结果更加稳定

功耗、性能和面积（PPA）是芯片设计中的三个重要优化指标。为使PPA优化结果更佳，同时为应对芯片安全性需求提升、设计规模攀升及工艺节点微缩等趋势，EDA厂商开始利用AI技术解决半导体芯片设计问题。在EDA中，数据快速提取模型、布局和布线、电路仿真模型、PPA优化决策等环节均有AI技术参与。AI应用于EDA主要有两种形式：1) 应用于EDA工具以优化单一芯片设计环节，如Google、西门子EDA；2) 应用于整体芯片设计流程优化，如Cadence、Synopsys。此外，华大九天、台积电等公司亦将AI纳入芯片生产各个环节。将AI与芯片设计工作结合，不仅有助于释放人力成本、提升工作效率，还将进一步弱化人在其中的作用，甚至可能改变EDA产业格局。

人工智能趋势展望

EDA企业AI技术应用案例

Synopsys

DSO.ai应用程序

使用AI算法在超大方案空间中搜索优化，自动执行影响较小的决策
与瑞萨电子合作，将DSO.ai设计系统引入先进汽车芯片设计



Cadence

Cerebrus

采用知识图谱和强化学习方式
工程生产力提高10倍，最多可将功耗、性能和面积(PPA)结果改善20%
5nm工艺手机CPU做示范，Cerebrus可在10天内完成全部设计



西门子EDA

AI/ML开发套件、Catapult HLS AI工具包

开发人员使用基于AI/ML的加速器快速找到PPA最佳参数
以3倍速度、纳米级准确度预测OPC输出



Ansys

KNN (K-Nearest-Neighbor) 机器学习算法

用于EM (Electro Migration) 签核工具
自动检查电流是否平均，判断是否会引发电流过载



NVIDIA

NVCell自动标准单元版图生成器

使用强化学习、遗传算法
增强其先进技术节点中标准单元布局



其他公司

台积电——机器学习设计支持平台

XILINX——基于ML的FPGA EDA工具套件

.....



产业融合：人工智能与碳中和

人工智能在预测、监测、优化三大环节赋能碳中和

自2015年第21届联合国气候变化大会后，碳中和已成为全球共识。当前，碳中和已获得全球超过40个国家和地区承诺，其中大部分国家宣布将于2050年左右实现碳中和目标。从整体来看，人工智能将从预测、监测、优化三大环节助力碳中和，如预测未来碳排放量、实时监测碳足迹、优化工作流程等。根据BCG数据，到2030年人工智能的应用有望减少26-53亿吨二氧化碳排放量，占全球减排总量5-10%。从行业来看，人工智能在不同领域及不同环节发挥重要作用，其主要在城市、制造、汽车、电力四大领域助力“碳中和”。

人工智能趋势展望

AI赋能碳中和应用场景

- ✓ AI技术快速发展，推动智能驾驶高级别迭代
- ✓ 实现车辆编组运行，提高行车密度和道路容量，减小风阻，减缓交通拥堵，降低油耗
- ✓ 通过AI优化运输路线和取送货顺序
- ✓ 优化驾驶员个体差异和行为波动，降低单位里程油耗

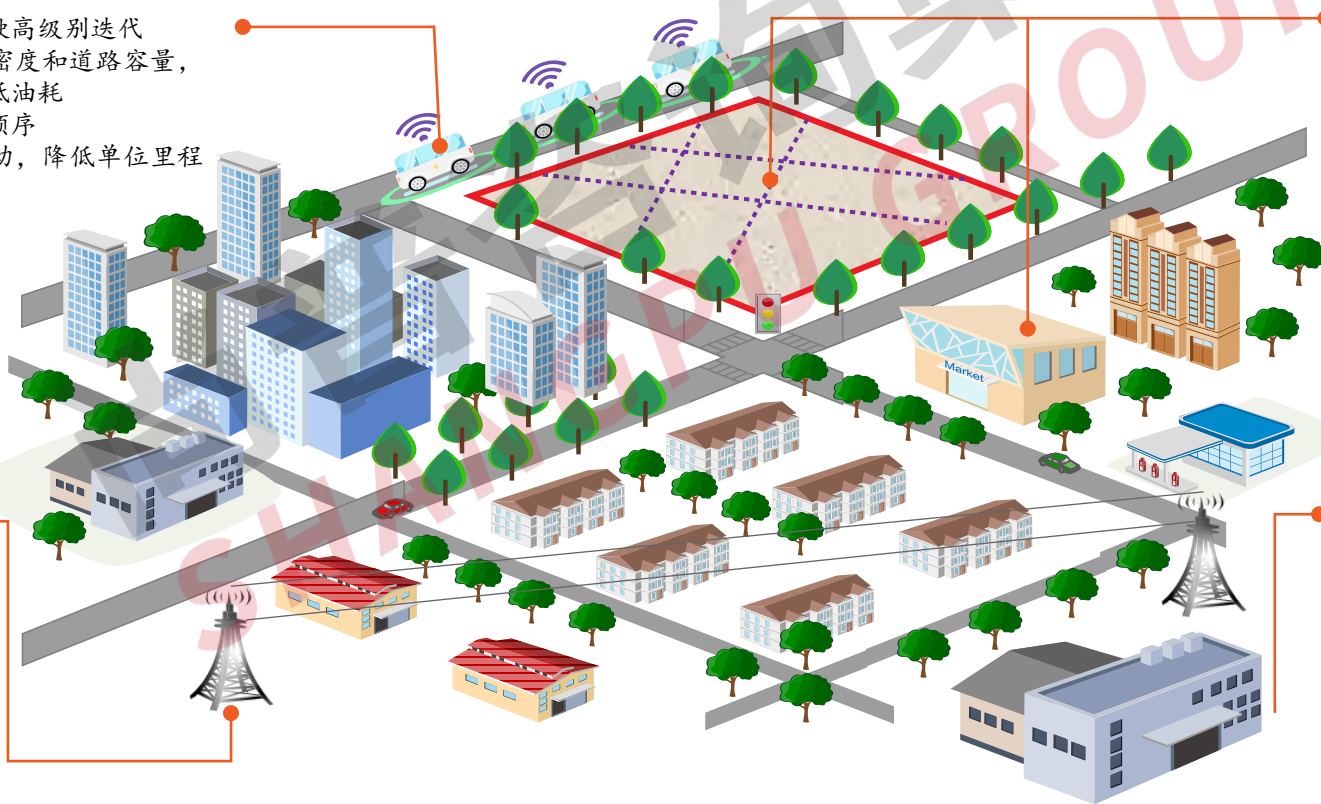


AI+汽车

- ✓ **发电端：**
AI助力提升传统火力发电效率；消纳新能源及提高电网稳定性
- ✓ **巡检端：**
AI帮助降低电网各环节能源损耗



AI+电力



- ✓ 城市规划借力AI实现合理布局，减少大拆大建、缩短通勤距离
- ✓ AI提升城市管理效率，缓解交通拥堵
- ✓ AI与安防结合，从被动监控到主动分析，减少出警次数及重大事故



AI+城市

- ✓ 集成AI模块，实现高效模拟仿真研发
- ✓ 智能制造新功能提升生产力
- ✓ 工业大数据+AI算法智能预测设备异常，降低维护成本
- ✓ AI+机器视觉检测减少人力成本，提高检测效率
- ✓ AI+供应链：车货匹配、仓库管理、托盘标准化、智能分拣、无人车、智能快递柜



AI+制造

产业融合：人工智能与冬奥会

人工智能在开幕式、比赛项目等环节助力实现科技冬奥

2022年2月，第24届冬季奥林匹克运动会成功在北京举办。人工智能技术在冬奥会开幕式、比赛项目、运动员训练等多个场景实现应用，助力科技冬奥目标实现。Intel打造的3DAT技术，一方面可以帮助教练员提出科学训练计划，有效提升运动员训练效率，同时还可以在开幕式中实现与演员互动效果；商汤科技为冬奥会冰壶比赛打造的“冰壶运动轨迹捕捉”技术，实现对冰壶检测追踪和轨迹捕捉。未来，人工智能与体育赛事、体育运动的融合程度将逐步加深。

人工智能趋势展望



3DAT技术在冬奥会中的应用



英特尔3DAT技术具备很高的通用性，面向不同的项目只需要简单的适配过程，大概需要5万张图片即可训练出针对性的AI模型，可对于采集和分析空间定位和三维姿态进行参考。2022北京冬奥会开幕式，英特尔联合第三方专业团队定制开发「基于人工智能技术的演出实时特效系统」，将3DAT技术应用在开幕式相关环节中，仅通过4台摄像机覆盖全场，实现演员与现场大屏实时互动。



冰壶运动轨迹捕捉技术



基于SenseMARS火星混合现实平台，商汤科技为冬奥会冰壶比赛打造“冰壶运动轨迹捕捉”技术，从而增强线上线下观赛体验。“虚实结合”的创新应用依托商汤科技视觉AI感知、深度学习、三维建模和空间定位等技术，克服在超大空间环境下基于远视距对较小物体运动轨迹精准捕捉的难题，并通过对真实场景的数字化映射，实现对运动冰壶的检测跟踪和轨迹捕捉。

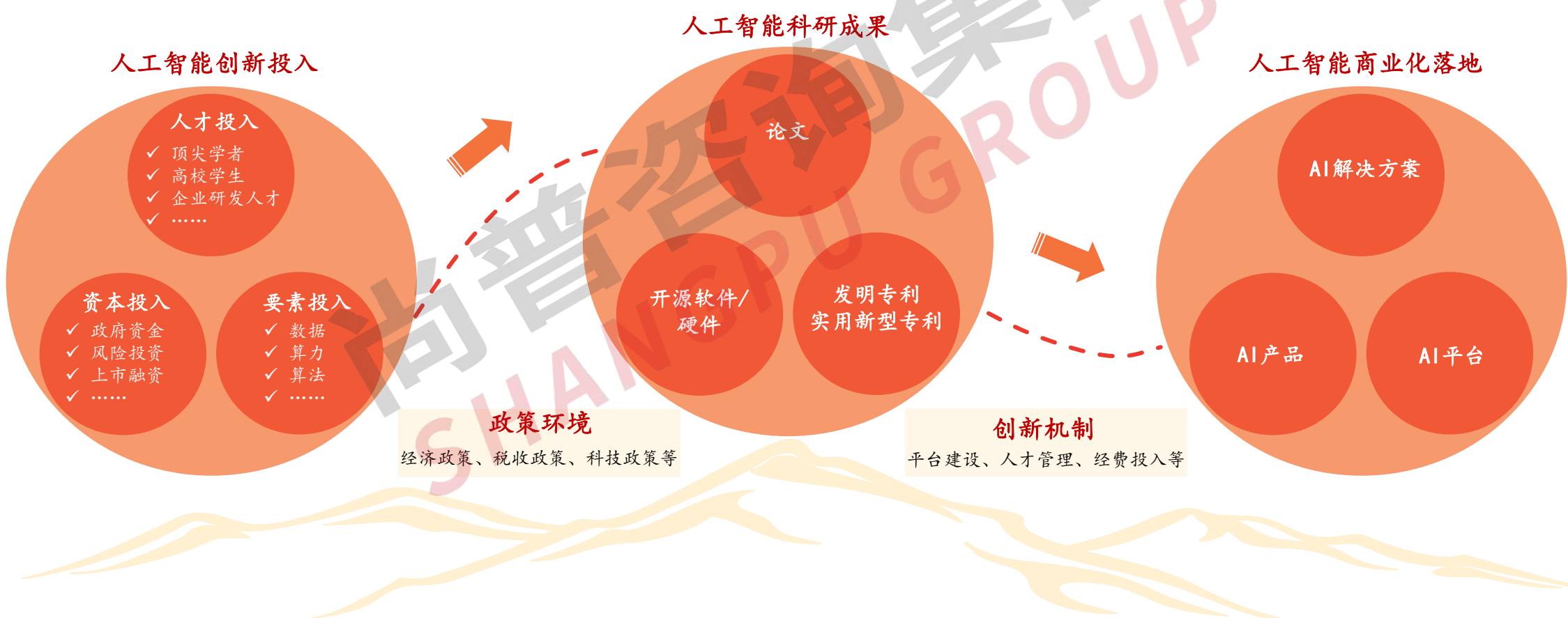
资料来源：部分图片来源于网络，新华社，尚普研究院结合公开资料整理绘制109

热点问题：人工智能产业发展的路径探究

人工智能在“科研成果—商业化落地”过程中依然存在诸多挑战

尚普研究院结合人工智能产业特点，梳理出“创新投入—科研成果—商业化落地”的发展路径。其中，创新投入主要涉及人才投入、资本投入和要素投入；科研成果涵盖论文、开源软硬件和专利；商业化落地包含AI产品、AI解决方案和AI平台。当前人工智能产业从投入端到科研成果端发展较快，人工智能在学术研究领域不断取得突破。但另一方面，人工智能从技术成果到商业化落地环节仍面临诸多挑战，如算法训练的通用性不强、泛化能力不强、商业化成本较高、相关产业及企业对于人工智能技术接受程度不高等问题，需要不断从政策机制、技术发展以及管理层观念等方面不断完善，才能最终推动人工智能在商业化方面实现快速发展。

人工智能趋势展望



热点问题：伦理与安全

人工智能发展面临隐私保护与算法合规使用等方面挑战

随着人工智能技术的高速发展与普及应用，由其产生的伦理与安全问题日益受到关注。人工智能不但延续信息技术的伦理问题，又因深度学习算法具有不透明、难解释、自适应、运用广泛等特征而在基本人权、社会秩序、国家安全等方面产生新问题。人工智能典型伦理问题包括：威胁公民生命安全、算法歧视、威胁隐私权、影响公民知情权和监督权、虚假信息泛滥、破坏市场竞争环境、引起权力结构变化、影响劳动者权益、AI武器威胁世界和平等问题。目前，全球人工智能治理尚处于初期探索阶段，各国正加大人工智能领域的治理力度，出台人工智能伦理相关制度，以预防AI在应用过程中产生的风险。未来，全球人工智能治理将由准则向操作指南、行业标准等层面逐步深入，加快构建人工智能国际治理体系。

人工智能典型伦理问题

2018年，Uber自动驾驶汽车在美国亚利桑那州与行人发生碰撞，造成行人死亡。原因是车辆物体识别发生混乱，未能及时刹车。



信息精准推送、自动化假新闻撰写和智能化定向传播、深度伪造等AI技术的滥用和误用，导致虚假信息泛滥问题，可能煽动民意，引起政局动荡。

人工智能武器在世界范围内加剧区域间不平等，威胁人类生命和世界和平。



人工智能的滥用可能威胁公民隐私权、个人信息权；AI决策不透明产生算法黑箱，影响公民知情权、监督权。

AI算法在算法歧视、训练数据中的偏失，可能产生或扩大社会歧视，侵害公民平等权利。



AI算法在不易被察觉和证明情况下，利用价格算法合谋和算法歧视，以横向垄断或轴辐协议等形式，破坏市场竞争环境。

2019年11月，美国国防部下属国防创新委员会《人工智能伦理道德标准》

人工智能五大伦理原则，即负责、公平、可追踪、可靠和可控。

2019年4月，欧盟委员会《可信赖AI的道德准则》

实现可信赖人工智能的七个要素：人的管理和监督、鲁棒性和安全性、隐私和数据管理、透明度、多样性、非歧视性和公平性、社会和环境福祉、问责制。

2019年6月，科技部《新一代人工智能治理原则——发展负责任的人工智能》

协调发展与治理的关系，确保人工智能安全可靠可控。

2022年3月，科技部《关于加强科技伦理治理的意见》

加强人工智能等领域的科技伦理立法研究等。

2021年11月，联合国教科文组织《人工智能伦理问题建议书》

定义人工智能建设必要基础性项目，确保人工智能健康发展的共同价值观和基本原则，关于人工智能主题的首份全球性规范框架。

2019年10月，德国数据伦理委员会《针对数据和算法的建议》

围绕数据和算法系统展开，包括一般伦理与法律原则、数据、“算法系统”、“欧洲路径”四部分内容。

2018年12月，日本内阁府《以人类为中心的人工智能社会原则》

主张在推进人工智能技术研发时，综合考虑其对人类、社会系统、产业构造、创新系统、政府等带来的影响，构建能够使人工智能有效且安全应用的“AI-Ready社会”。

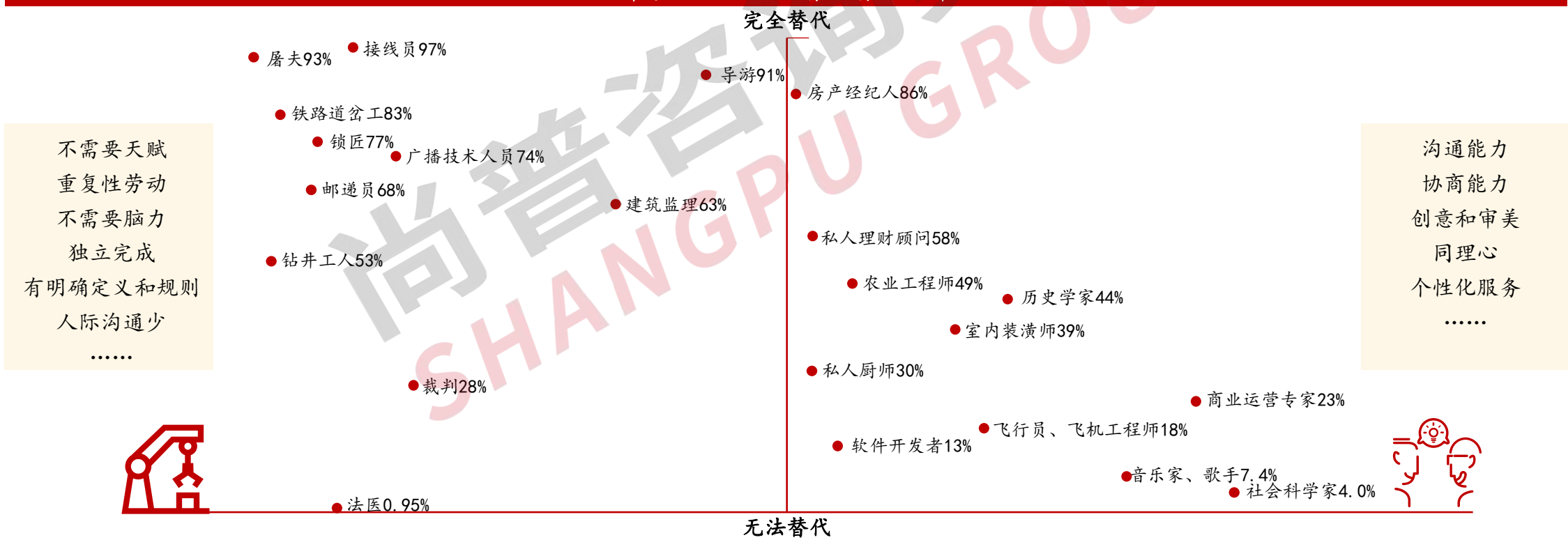
热点问题：就业

人工智能将替代大量重复性劳动，而难以取代创造性和沟通类工作

人工智能将通过改变劳动分工与人力资本价值结构深刻影响就业市场。AI与劳动力就业关系包含三个方面：1) 当AI成本低于劳动力工资水平、且产品附加值又不足以弥补用工成本时，AI应用将直接替代相应劳动岗位；2) AI应用填补劳动者无法胜任的岗位，既可降低错误率，提高产品质量，也可保护人身安全和健康；3) AI应用催生新工作岗位，AI带来生产生活方式的变革与社会效率的提高，全社会产能实现跃升，进一步产生新工作岗位。世界经济论坛发布的《2020未来就业报告》预计，到2025年，机器可能会取代8,500万个工作岗位，在AI推动下经济增长会产生9,700万个新岗位。随着AI技术发展，工作岗位、员工技能和任务类型将重塑，就业替代呈现行业性特征，总体就业岗位数量仍将上升。

人工智能趋势展望

部分职业被人工智能替代概率



热点问题：国家间技术限制

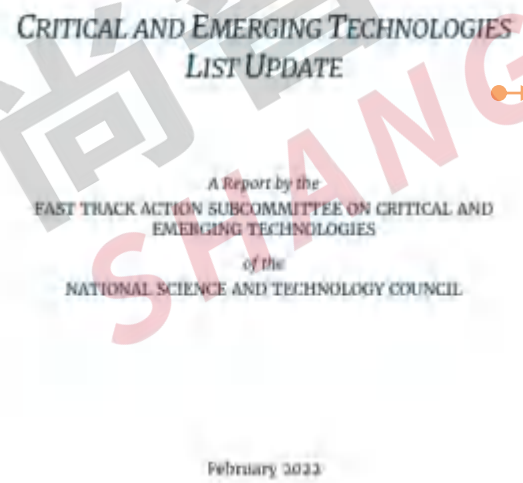
国家间技术限制阻碍人工智能技术进步

当前，开源深度学习框架、开源工具集、开源应用软件快速发展，国际间AI技术交流不断深入，但部分国家和政府间组织为保持自身AI技术优势，限制AI技术交流。如美国在2021年6月发布《创新与竞争法案》，在AI、无人机、芯片等多个领域限制与中国合作；美国商务部于2019年10月和2020年5月将商汤科技、科大讯飞等多家中国AI公司加入其实体清单，实施投资限制；2022年白宫修订“关键和新兴技术（GET）清单”，对AI技术具体分类并实行技术封锁。欧盟则于2021年9月通过最新出口管制法规，内容涵盖人脸识别等AI技术。上述相关政策与未来人工智能发展趋势背道而驰，不利于各国开展技术合作。

人工智能趋势展望



2019年10月美国商务部将28家中国组织纳入实体清单
2020年5月美国商务部以新疆人权为由将9家中国实体纳入实体清单
包含科大讯飞、旷视科技、商汤科技、云从科技4家AI公司。



2022年2月美国白宫总统行政办公室发布最新修订“关键和新兴技术（GET）清单”

对AI技术具体分类，包括机器学习、深度学习、强化学习、感官知觉和识别等，便于实施技术封锁。



2021年6月美国国会发布《创新与竞争法案》

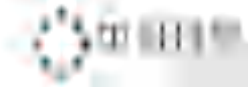
在AI技术、AI芯片、机器人、量子计算、脑机接口、先进材料等新兴和基础技术领域限制与中国合作。



2021年9月欧盟议会实行《建立欧盟控制两用物项的出口、中介、技术援助、过境和转让的管制制度》

限制出口类别，包含人脸识别等网络监控工具。

合作机构



中科创星介绍

中科创星是中国首个专注于硬科技创业投资与孵化的专业平台。作为“硬科技”理念的缔造者和“硬科技”投资的先行者，中科创星致力于打造以“研究机构+早期投资+创业平台+科创服务”为一体的硬科技创业生态，为科技创业者提供专业、深度、全面的投资、孵化及融资解决方案。中科创星围绕光电芯片、人工智能、航空航天、生物技术、信息技术、新材料、新能源、智能制造等重点投资领域，主要投向具有成长潜力、拥有自主创新能力的初创期、早期科技型中小企业，通过发掘和培养科技创业领军人才，孵化科技企业，使高新技术从“实验室”走出来，使科技成果得到市场的验证，推动科技成果产业化，实现科研院所科技资源优势与社会资本深度结合。截至目前，基金总规模61亿元。已投资364家硬科技企业，投资金额50亿元。

中科创星网址：www.casstar.com.cn



北京智谱华章科技有限公司介绍

北京智谱华章科技有限公司（简称智谱AI）依托于清华大学团队十余年的技术积累及人才优势，专注于从海量数据资源中，挖掘隐含语义信息、构建隐含关联关系以及具有价值的情报信息，服务于政府、企业、科研机构及广大科研人员。项目团队拥有自主知识产权的科技情报大数据挖掘与服务平台，收集文献2.7亿、学者1.3亿、专利1亿，以及学术会议、科研项目、科研机构、科技资讯等各类科技信息资源总计超过10亿；构建超800万概念，上亿连边的科技知识图谱。智谱AI致力于打造数据与知识双轮驱动的认知智能，让机器像人一样思考。公司拥有下一代认知智能创新平台悟道++，并在此基础上打造和提供综合科研信息服务平台AMiner 2.0、科技情报知因分析、智慧人才和智能数字人等业务。

智谱AI网址：www.zhipu.ai

合作机构

荆州教育经济研究院 紫荆数字经济研究院介绍

清控紫荆教育旗下“紫荆数字经济研究院”是依托紫荆教育多年的行业积累，以新兴技术为基础，聚焦数字经济领域，整合智库、行业机构、科研院校、龙头企业、媒体等优质资源，打造数字经济研究学术高地，建立服务地方经济发展和企业数字化转型、应用推广的创新发展咨询平台。目前围绕学术研究、政策与案例分析、教学研究三大方向展开，汇聚国内外知名学者、优秀学青代表，组织学术交流，发表期刊论文、开展科研课题研究，发现和激励优秀青年人才，形成更多高水平研究成果；汇聚政府专家、知名企业家，整合最新政策和行业案例，形成分析报告、案例库，引领数字经济领域的探索实践，推动优势产业发展；汇聚教育专家、高校名师、产业资源，建立教学资源库、形成教学成果，将前沿技术与人才培养模式有机融合，建立良性发展的人才生态圈。旨在通过产学研紧密协作培养顶尖数字人才，创新技术研发、企业培育、校企合作培养专业人才等方面发挥重要作用，为数字经济人才提供学习支撑和提升机会，助推国家数字经济人才发展战略。

中国风险投资有限公司 中国风险投资有限公司介绍

中国风险投资有限公司（中国风投）是于2000年由民建中央发起、民建会员参股设立的以风险投资和基金管理为主营业务的专业投资机构，是中国本土风投行业的重要开拓者之一。公司作为民建中央落实全国政协九届一次会议《关于加快发展我国风险投资事业的提案》（即一号提案）的重要载体，肩负着推动我国风险投资事业发展的使命和责任，致力于风险投资领域的理论研究和实践探索。公司业务范围覆盖天使投资、创业投资、股权投资、产业投资、资本运营、并购整合等。公司重点投资领域包括新材料、新能源、集成电路、高端制造、轨道交通、军工航天、清洁技术、医药医疗等。目前已投资项目超过250个，并助力其中70余家企业登陆国内外资本市场，同时管理了近40只各类基金。

中国风险投资有限公司网址：<http://www.c-vc.com.cn/>

合作机构

中电海康科创基金介绍

中电海康（杭州）股权投资管理有限公司是海康系基金管理人。海康系科创基金以助力海康创新联合体生态构建为使命，以基金投资作为海康创新联合体生态纽带，通过海康创新生态，从战略管理、创新赋能、技术合作、渠道共享、资本运作及研发合作等业务与资源导入，开展多维度、体系化精准赋能，助力快速提升被投资企业基本面与核心价值，提升被投资企业产业化与市场化水平，形成与被投资企业多赢格局。海康科创基金聚焦智能物联网硬核科技，助力推进我国新计算、新存储领域前沿科技，如自旋电子、新型存储、类脑计算等硬核科技领域的股权投资。截至目前，已成功投资于多家高科技企业并实现对被投资企业产业赋能，在智能物联网硬核科技领域积累了丰富的基金投资和运作管理经验。

容亿投资介绍

容亿投资源自产业，拥有丰富的通讯、智能终端和汽车等高端产业资源，是一家快速成长的产业投资机构。依托核心团队深厚的产业背景和产业经验，坚持用经营思维做投资，重点布局硬科技和数字科技两大领域。目前在上海、杭州、苏州、北京、深圳、嘉兴六地均设有投资平台，投资资金主要来源于知名机构、上市公司和政府产业引导基金。

容亿投资，让创业更容易。集丰富的产业、人才和政府等资源，助力创业企业快速成长为行业龙头。

容亿投资网址：<http://www.winrealvc.com>

合作机构



创汇高合（简称高合咨询）专注于帮助企业通过利益分配模式激活个体动力，达成创新增长。通过理清不同类型的合伙人以及各类合伙人之间的关系，从顶层设计、业务模式及个体激励三个维度出发，根据企业不同阶段的不同需求，设计相匹配的内外部合伙人机制，明确合伙人选拔路径、发展路径、利益分配路径、激励路径，整合内外部资源，最大限度的发挥合伙人的作用。其所运营的合伙人机制研究院拥有强大的咨询研究能力，拥有知识产权60余项，理论指导实践，实践升华理论，承接国家研究课题并培育中国本土的管理类咨询人才，为中国管理科学的发展贡献力量。每年为企业家、创业者、商界精英、人力资源高管等提供事业合伙人领域的辅导和咨询，包括合伙人制度、合伙人梯队、股权激励、虚股激励、薪酬激励、平台型组织、城市合伙人等，在事业合伙人的领域中引领前行！

尚普咨询集团介绍

尚普咨询集团有限公司（简称“尚普咨询”），注册资本5000万，前身为“北京尚普信息咨询有限公司”，成立于2008年初。作为2009年A股创业板开板便投身IPO上市咨询的国内首批专业机构之一，尚普上市咨询业务常年处于市场领先地位，稳居行业头部，并曾被《人民日报》旗下《国际金融报》评选为“IPO咨询行业先锋机构”，成为资本市场公认的中国最大IPO咨询服务机构之一。

尚普咨询服务内容包括IPO及再融资募投项目可行性研究、发行人行业定位与所处行业细分市场研究、申报底稿整理与咨询、债券可研报告编制、企业社会责任&ESG报告、港股上市行业顾问、上市辅导与规划咨询、企业发展战略咨询、融资并购财务顾问等专业服务。

尚普咨询作为北京中关村高新技术企业，是行业中资质证书最为齐全的机构之一，具备中华人民共和国涉外调查许可证，为中国科学技术情报学会、北京工程咨询协会、SEMI国际半导体产业协会、CSIA中国半导体行业协会及CAAI中国人工智能学会会员单位，是全国首家获得募投报告编制甲级资质的专业IPO咨询机构。

尚普研究院作为尚普咨询旗下的专业研究机构，依托尚普咨询主业开展深度行业研究，专注于半导体、人工智能、5G通信、云计算、新能源、医疗大健康等领域，持续为政府机构、社会团体及企业提供专业咨询服务。

法律声明

版权声明

本报告为尚普咨询集团旗下尚普研究院联合中科创星、北京智谱华章科技有限公司、紫荆数字经济研究院、中国风险投资有限公司、中电海康科创基金、容亿投资等六家机构参与撰写,其版权归属尚普咨询集团有限公司所有,没有经过相关机构的书面许可,任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境内外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

免责条款

本报告中部分文字和数据采集于公开信息,行业数据及相关市场预测主要为尚普研究院分析师采用桌面研究、行业访谈、市场调查及其他研究方法,并通过尚普统计预测模型估算获得;企业数据主要为访谈获得。尚普研究院尽最大努力确保该等信息的准确性、完整性及可靠性,但不作任何保证。在任何情况下,本报告中的信息或所表述的观点均不构成任何建议。本报告只提供给用户作为市场参考资料,尚普研究院对该报告的数据和观点不承担法律责任。

行业市场研究/募投可研报告/工作底稿咨询/IPO一体化咨询

北京总公司

地址：海淀区北四环中路229号海泰大厦1118、1120、1127、1129室

电话：15033256189

上海分公司

地址：黄浦区河南南路33号新上海城市广场1809室

电话：13611322396

深圳分公司

地址：福田区金田路3038号现代国际大厦1栋11层1101B室

电话：18148563314

网址：IPO.shangpu-china.com

邮箱：IPO@shangpu-china.com



欢迎关注尚普IPO咨询