

语义引擎产业专利导航摘要

一、产业概况

语义引擎技术是对输入数据进行语义理解计算的过程，其中高性能计算芯片是语义引擎技术的基石，高性能网络是数据传输通信的保证，语义引擎方法是计算的核心依据。

近年来，我国的语义引擎技术市场发展迅速，随之产品产出量也在持续不断的扩张。2021年，中国语音语义市场规模21.7亿美元，相比2020年增长超过30%。目前共有300多家公司从事语音语义等相关业务，目前尚未形成巨头垄断从市场格局端。商汤、旷视、海康威视、创新奇智、云从科技作为前5的厂商，构成了45.6%的市场份额，百度智能云、阿里云、华为云、腾讯云也贡献了一定的市场份额。

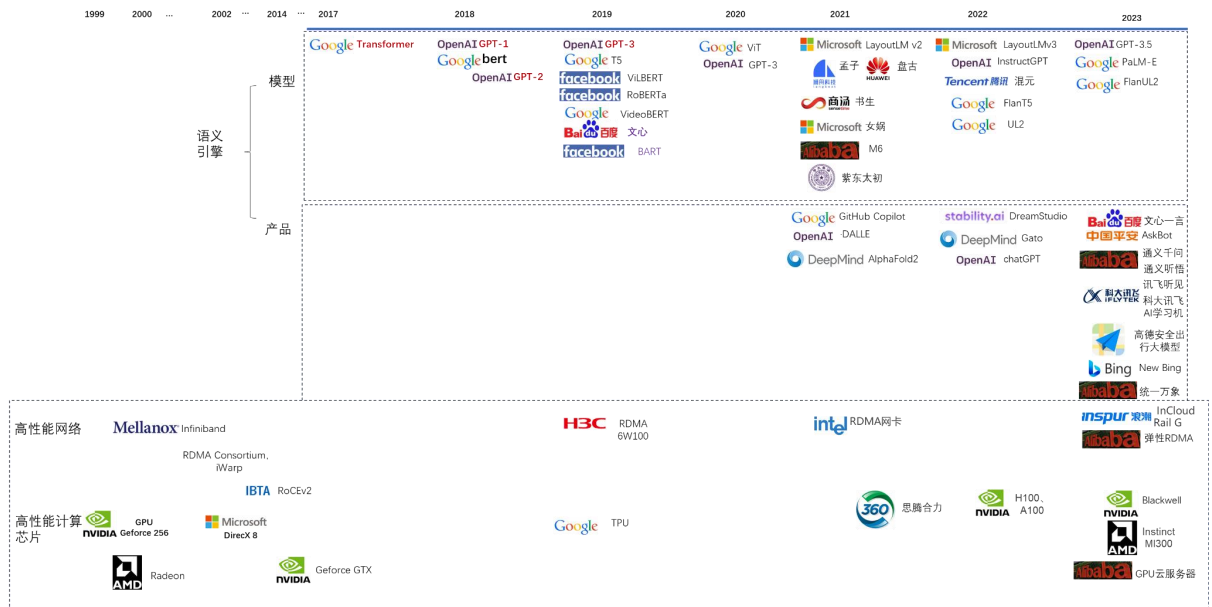


图1 语义引擎发展脉络

二、产业政策

随着信息时代的发展，人工智能的发展越来越迅速，各

个国家都在跟随时代的脚步，中国为了提高自身的人工智能水平，国家提出了一系列的相关政策，对该行业进行支持与推动其快速发展。

序号	文件名称	发布部门	发布时间	主要内容
1	《国民经济和社会发展规划第十四个五年规划和 2035 年远景目标纲要》	全国人大	2021 年	壮大经济发展引擎、大数据、人工智能等
2	《新时代崔进集成电路产业和软件产业高质量发展的若干政策》	国务院	2020 年	明确提出集成电路产业和软件产业是信息技术的核心，是引领新一代科技革命和产业的关键力量
3	《关于促进人工智能和实体经济深度融合的指导意见》	中央全面深化改革委员会	2019 年	强调市场导向与产业应用，打造智能经济形态
4	2018 年政府工作报告	国务院	2018 年	人工智能再次被列入政府工作报告
5	《人工智能标准化白皮书》	国家规划办 理委员会	2018 年	全面推荐人工智能标准化，促进人工智能产业发展
6	《国务院关于印发新一代人工智能发展规划的通知》	国务院	2017 年 7 月 8 日	人工智能，智能计算芯片与系统，自然语言处理技术
7	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》的通知	科技部 教育部 工业和信息化部 交通运输部 农业农村部 卫生健康委	2022 年 7 月 29 日	融合人工智能模型算法和领域数据知识
8	《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》	国家发改委	2021 年 3 月 11 日	理论突破、芯片研发、开源算法，自然语音识别
9	《国家新一代人工智能创新发展试验区建设工作指引》	科技部	2019 年 8 月 29 日	人工智能创新发展试验区建设
10	《高等学校人工智能创新行动计划》	教育部	2018 年 4 月 2 日	优化高校人工智能领域科技创新体系，把

				高校建成全球人工智能科技创新的重要策源地，自然语言理解和语音识别等
11	《促进新一代人工智能产业发展三年行动计划（2018-2020）》	工业和信息化部	2017年12月13日	支持建设高质量人工智能训练资源库、标准测试数据集并推动共享，鼓励建设提供知识图谱、算法训练、产品优化等共性服务的开放性云平台
12	《生成式人工智能服务管理办法（征求意见稿）》	国家互联网信息办公室	2023年4月11日	促进生成式人工智能健康发展和规范应用

三、专利分析

从专利申请的来源国来看，中国的专利申请数量占比远高于其他国家，占全球专利申请总量的 73%。其次为美国，占全球专利申请总量的 14%。

在语义引擎的按照模型结构，可分为基于 CNN、基于 RNN 和基于 transformer 三个分支。基于 transformer 的语义引擎技术起步时间晚，但增速极大，使得该分支的专利申请数量呈现后来居上的态势。

从语义引擎的专利布局情况来看，专利申请数量排名前五的申请人中，中国申请人占 73%，国外申请人占比 27%。

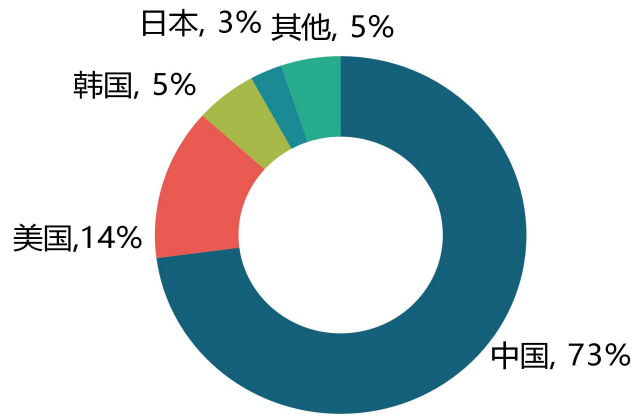


图 2 语义引擎全球专利申请来源国分布

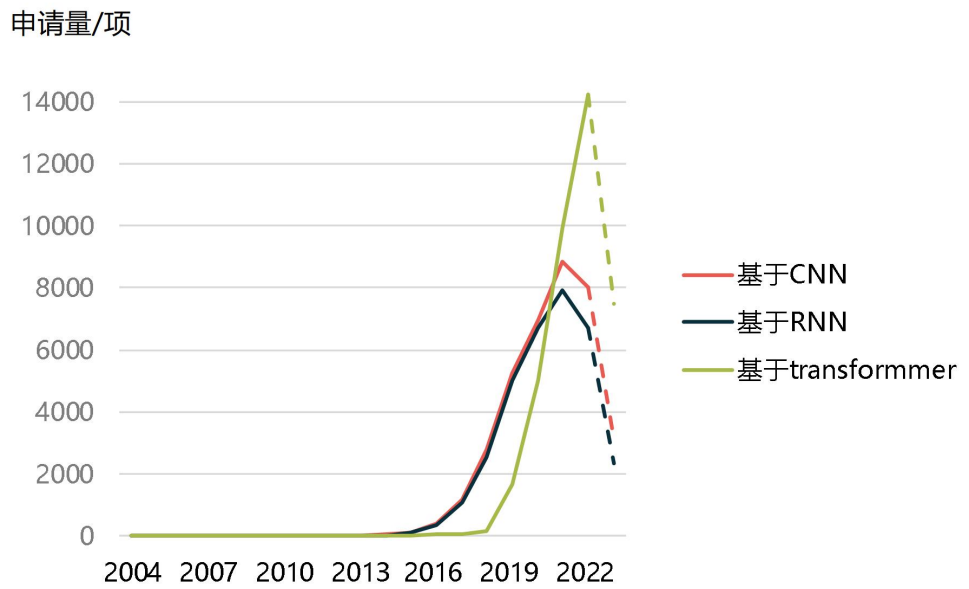


图 3 不同结构的语义引擎技术专利申请趋势

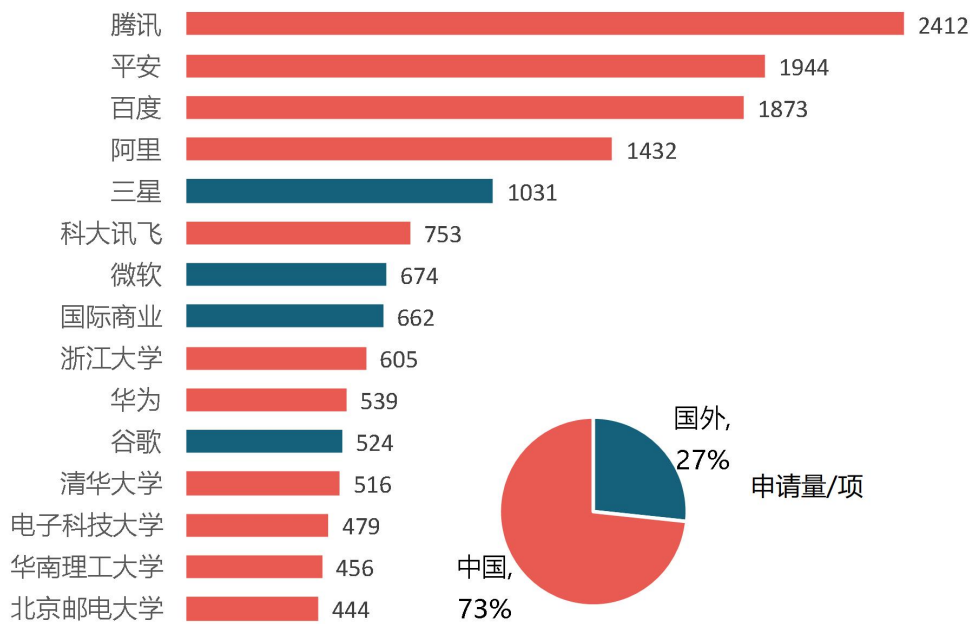


图 4 语义引擎全球专利申请申请人数量分布

四、重点技术

自然语言处理技术在近年来得到了快速发展，但仍存在语言的多样性、指代消解、语境依赖性、数据集和标注、模型的通用性和鲁棒性等一系列的问题。核心模型从卷积神经网络（CNN）、循环神经网络（RNN）到 Transformer 不断演进。

Transformer 模型引入了注意力机制，使得模型能够将不同位置的信息进行加权聚合，从而更好地表示输入序列和输出序列之间的关系。其具有并行计算能力强、适用于长序列处理、无记忆性、支持多语言处理、适合高维度数据处理等多重优点。

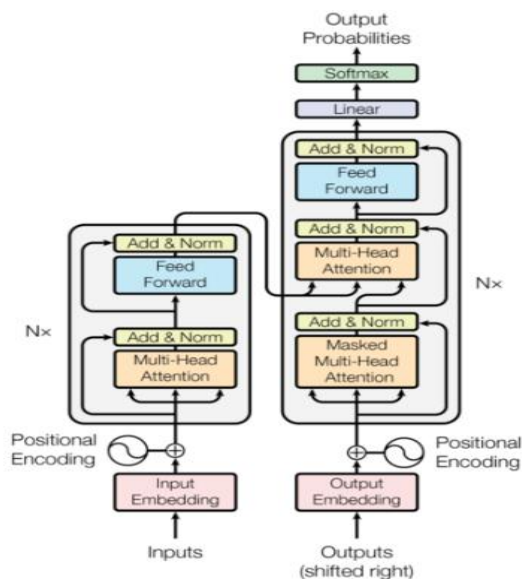


图 5 Transformer 的总体框架结构图

五、重点企业

自 2017 年 transformer 架构提出以来，国内外各类创新主体纷纷加入预训练模型全球竞技赛中，从自身技术优势出发，提出适用于各类任务的预训练模型。由预训练模型申请人统计信息排名可知，百度、平安、谷歌、腾讯、清华大学等均为该领域的主要创新主体。

2021	GitHub Copilot	DALL-E	AlphaFold2		
2022	DreamStudio	Gato	chatGPT		
2023	文心一言 通义千问 通义听悟 高德安全出行大模型	讯飞听见 科大讯飞AI学习机 New Bing 360智脑全家桶 序列猴子	百度搜索 Claude Google Bard 飞书“My AI” DriveGPT	腾讯会议AI小助手 Alexa语音助手 一站式行业大模型精 选商店腾讯云 Maas 天工 魔力写作、MOSS、曹植	AskBot 统一万象 钉钉 图查查 知海图AI

图 6 预训练模型专利申请的发展趋势

六、浙江省专利分析

从整体趋势来看，浙江省年申请量分布趋势整体呈现持

持续增长状态。从专利申请的地域分布来看，浙江省在预训练模型领域的专利申请主要来自于杭州，占浙江省全部申请总量的 91.8%。从专利申请的主体来看，浙江大学和阿里是浙江省最主要的创新主体。模型方面浙江大学提出了多模态的 OmModel 模型，阿里巴巴提出了通已、M6 等模型。

	申请量/项				
	通用	文本	图像	语音	多模态
谷歌	44	17	21	26	16
百度	8	103	30	15	23
清华大学	6	25	7	5	11
浙江大学	4	38	21	2	12
阿里	5	70	10	9	15

图 7 预训练模型浙江省专利申请数据类型分布

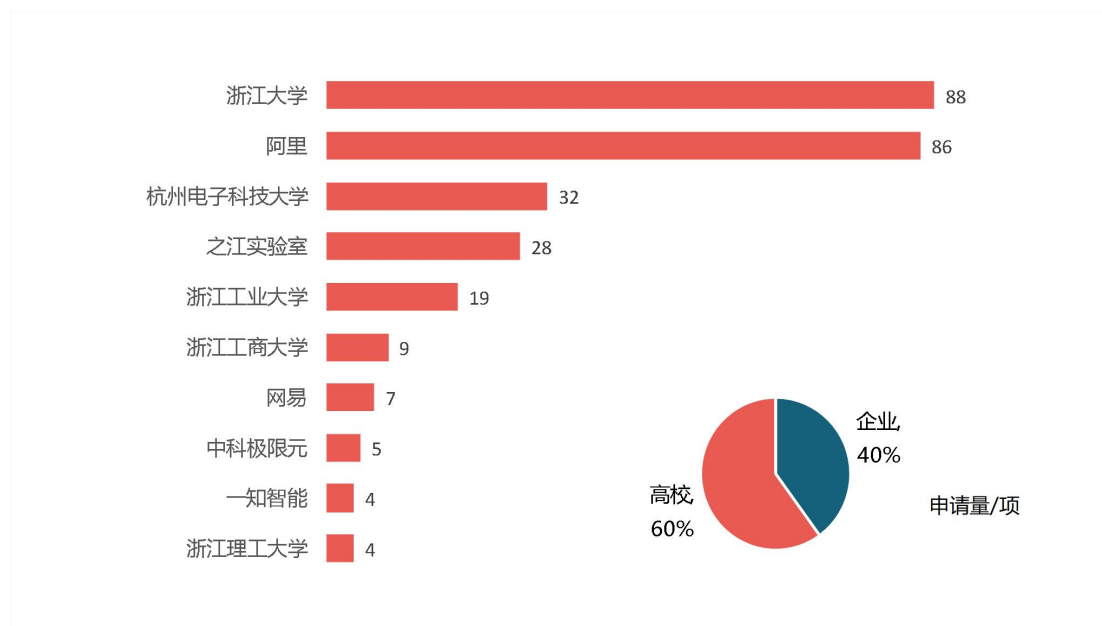


图 8 预训练模型浙江省专利申请申请人分布