

AI 多模态融合相关技术 专利分析报告

上海市知识产权服务中心
(上海市知识产权发展研究中心)

2025 年 7 月

声明

1. 本报告的分析基准日为 **2025 年 4 月 30 日**，一切分析结论均基于在此基准日之前能够获取的信息与资料而得出。对于基准日之后由于法律、技术、经济环境等发生变化，可能由此引发的不同分析结论。
2. 本报告形成过程中对于数据的检索、筛选、标引、分析等均依据本报告确定的分析“标的”或“事实”（分析“标的”或“事实”主要依据企业提供的特定技术信息以及市场信息为准）及当前所适用的专利法而做出，但并不意味分析服务机构能够保证上述分析“标的”或“事实”的准确性与全面性，也不意味着分析服务机构能够保证上述检索、筛选、标引、分析等处理过程的准确性与全面性。
3. 本报告在依据确定的分析“标的”或“事实”而进行创新启示建议时，不能保证也不能排除其运用于研发、生产与市场的有效性，且不能保证也不能排除与其他公开技术或者专利权利相冲突，在企业后续具体实施中，应当再次进行与技术、法律与市场层面的可行性分析。

摘要

AI 多模态融合技术作为人工智能领域的关键驱动力，正推动多模态应用向全局协同演进。本报告基于专利视角，深入分析了该技术在医疗健康、智慧交通、工业制造、消费电子及金融投资等领域的创新与应用趋势，旨在为相关企业和研究机构提供专利布局与技术研发的决策支持。

基于技术关键词、技术分类号和特定创新主体在全球范围进行多模态融合技术专利检索，检索得到原始专利数据 **11153** 件。经去重和人工筛选，剔除与技术不相关专利后，获得 **8685** 件与分析主题相关的专利申请。其中，技术地区布局方面，中国大陆地区，专利申请量已达到 **7861** 件，占比 **90%**；其中创新主体以院校/研究所为主，占比高达 **57%**，凸显了基础研究和前沿探索的活跃度。

专利主体方面，腾讯科技有 **167** 件专利，浙江大学为 **138** 件。核心主体如腾讯科技、百度、浙江大学、微软和谷歌等，报告分析了其专利布局特征、技术发展路线及市场竞争力。这些权利人通过构建多模态大模型、优化模型架构等手段，不断拓展技术的应用场景，如自动驾驶、医疗诊断、人机交互等，引领 AI 多模态融合技术的发展方向。

在技术发展方面，深度学习驱动的多模态融合技术已成为主流，专利布局主要集中在数据级、特征级和决策级融合等领域。中国地区在专利运营方面虽显不足，但在专利申请量上占据领先优势，凸显了国内在技术创新和知识产权保护方面的潜力。

根据国内 AI 多模态融合技术领域的市场及专利情况，针对中国地区提出了发展建议，包括加强产学研合作、提升专利运营能力、优化市场对接机制等，以促进技术创新与商业化应用的深度融合。

目录

声明	I
摘要	I
目录	I
第一章 AI 多模态融合技术研究概述	4
1.1 研究概述	4
1.1.1 研究背景	4
1.1.2 研究目的	5
1.1.3 研究意义	5
1.2 研究对象	6
1.2.1 多模态融合技术简介	6
1.2.2 多模态融合关键技术	6
1.2.3 多模态大模型行业现状	8
1.2.4 多模态大模型政策环境	14
1.3 研究思路	19
1.3.1 分析方法	19
1.3.2 分析内容	19
1.3.3 技术框架	20
1.4 检索范围及检索结果	21
1.4.1 检索策略	21
1.4.2 检索结果	21
第二章 AI 多模态融合技术宏观分析	23
2.1 专利整体分析	23
2.1.1 申请趋势情况	23
2.1.2 申请区域分布	25
2.2 专利主体分析	28
2.2.1 专利主体排名	28

2.2.2 专利主体类型	29
2.3 技术方向分布	30
2.3.1 技术主题分布	30
2.3.2 分类号分布	32
2.4 专利运营情况分析	33
2.5 中美技术对比分析	34
2.7 本章小结	35
第三章 AI 多模态融合技术主要权利人分析	37
3.1 国内主要权利人	37
3.1.1 腾讯科技	37
3.1.2 百度科技	46
3.1.3 浙江大学	55
3.2 国外主要权利人	65
3.2.1 微软	65
3.2.1 谷歌	73
3.3 主要权利人对比分析	82
3.4 本章小结	83
第四章 AI 多模态融合技术专题分析之模态融合	86
4.1 专利总量	86
4.2 专利趋势	87
4.3 法律状态	88
4.4 区域分布	89
第五章 AI 多模态融合技术专题分析之模态对齐	90
5.1 专利总量	90
5.2 专利趋势	91
5.3 法律状态	92
5.4 区域分布	93
5.5 本章小结	94

第六章 AI 多模态融合技术应用分析	96
6.1 医疗健康	96
6.2 智慧交通	100
6.3 工业制造	103
6.4 消费电子	108
6.5 金融投资	113
6.6 本章小结	117

第一章 AI 多模态融合技术研究概述

1.1 研究概述

1.1.1 研究背景

AI 多模态融合技术涉及视觉、语音、文本等多维度数据协同处理，通过联合分析和处理实现全面理解与推理，其专利布局直接反映技术交叉创新的核心方向。例如，浙江索思科技的异常行为检测系统专利通过融合视觉、声音及传感器数据构建分析模型，而捷通智慧则开发了基于视觉语言模型的少样本场景识别方法。专利导航可系统梳理跨模态数据融合、特征对齐等关键技术演进路径，帮助企业预判如“文本-3D 建模”等新兴融合场景的研发方向，避免重复投入。

多模态技术涉及复杂算法与硬件协同，海外巨头已构建专利壁垒。例如，亚马逊云科技、英伟达等企业通过分布式训练框架、异构计算架构专利主导生态。通过专利导航可识别关键封锁点（如跨模态注意力机制专利群），指导国内企业采用差异化技术路线。例如，北京鼎兴达的分布式训练方法结合改进的 Paxos 算法与 Nginx 负载均衡技术，形成自主技术方案。同时，专利数据揭示高校与企业的合作热点，例如多模态预训练模型、联邦学习等方向。通过分析清华大学、中科院等机构的专利许可动态，可精准对接技术转移需求。例如，中煤科工集团的爆炸风险决策模型专利可能衍生出与高校联合开发的边缘计算优化方案，形成“基础研究-产业应用”闭环。

多模态融合在自动驾驶、智能医疗、人机交互等领域具有不可替代的重要性。例如，医疗诊断中结合医学图像与病历文本可提升诊断准确性，智能家居通过融合语音与视觉数据实现自然交互。其核心价值在于增强系统鲁棒性（如

抗传感器干扰）、提供多维度上下文信息，并推动跨领域创新。随着少样本学习、深层语义推理等技术的发展，多模态融合将成为实现通用人工智能的关键驱动力，对提升复杂场景下的智能化水平具有战略意义。在此基础上对具体领域应用情况的专利数据研究可以催化跨领域创新，拓展应用边界。

1.1.2 研究目的

本报告旨在深入剖析 AI 多模态融合技术的专利布局与发展态势，为相关企业和研究机构提供决策支持。首先，报告目的在于梳理 AI 多模态融合技术的关键技术节点和专利分布情况，明确技术交叉创新的核心方向，助力企业把握技术发展趋势。其次，识别多模态技术领域的国外关键专利，指导国内企业采用差异化技术路线，提升我国在该领域的自主创新能力。同时，揭示高校与企业的合作热点和专利许可动态，促进产学研协同创新，加速技术转移转化，形成“基础研究 - 产业应用”闭环。最后，探讨多模态融合在自动驾驶、智能医疗、人机交互等具体领域的应用优势和潜在发展方向，为相关领域的创新发展提供参考，推动跨领域创新和应用拓展

1.1.3 研究意义

本报告研究意义在于明确 AI 多模态融合技术的关键专利和技术路线图，为国内科研人员和企业提供清晰的研究方向，提升企业在该领域的技术水平和创新能力。同时，促进多模态融合技术与其他领域的深度交叉融合，催生更多新兴应用场景和创新解决方案，推动人工智能技术向更广泛的应用领域拓展。这也有助于推动我国人工智能技术在关键领域的自主创新和应用突破，增强国家在人工智能领域的核心竞争力。

1.2 研究对象

1.2.1 多模态融合技术简介

AI 多模态融合技术是指通过整合文本、图像、音频、视频、传感器数据等多种模态信息，构建统一的语义空间，使人工智能系统能够更全面地理解、推理和生成跨模态内容。其核心在于打破单一数据类型的局限性，模拟人类多感官协同的认知方式，通过深度学习框架（如 Transformer、图神经网络）实现模态间的特征对齐与语义关联。例如，华为提出的中间层特征融合技术，通过交叉注意力机制动态整合图像和文本的中间层特征，显著提升了模型的鲁棒性和跨模态检索精度。从 AI 技术范式来看，多模态技术通过预训练+调参的方式颠覆了传统机器视觉小模型 CNN 高度定制化的业务模式，模型的泛用性大幅度提高。从 AI 商业 模式来看，产业的话语权逐渐由应用端走向研发端。

1.2.2 多模态融合关键技术

跨模态特征对齐是信息采集和整理的重要技术，是后续的特征融合分析训练的基础，其通过深度学习技术将文本、图像、音频等异构数据映射到共享的语义空间，实现多模态信息的语义级交互。其核心技术包括对比学习机制，如 CLIP 模型通过 4 亿图文对预训练，将图像编码器（ViT/ResNet）与文本编码器（Transformer）嵌入同一空间，利用对比损失函数最大化匹配对的相似度，最小化非匹配对关联性。这种技术使图像与文本的跨模态检索准确率提升 37 倍，在电商搜索场景中显著优化商品图文匹配效率；因果对齐框架也是跨模态特征对齐的重要技术点，如中山大学等团队提出的 CRA 框架引入因果干预，通过高斯平滑定位（GSG）、交叉模态对齐（CMA）和显式因果干预（ECI）三模块，消除视觉-语言数据中的虚假关联。在医疗影像分析中，该技术将病理特

征与诊断文本的匹配准确率提升 **18%**，并降低 **20%** 的误诊风险。同时，相关重要技术点还包括联合特征空间建模，OmniAlign-V 数据集通过模态无差异特征表示（MDF-FR）技术，将医学影像与电子病历的全局特征集成，建立可解释的跨模态词汇映射系统。该技术在肝脏肿瘤切除手术导航中，将多模态影像配准误差降低至 **0.3mm** 以内。

在信息整理后，需针对不同数据进行动态权重分配，通过实时评估数据质量与任务需求，实现模态重要性的动态调节。其主要技术有，可信度评估机制：基于观测支持度的动态加权算法，通过计算传感器数据的互支持度信息，实时调整融合权重。如在无人机导航中，使用该技术可将多传感器定位误差降低 **40%**，尤其在雾天环境下的红外-视觉数据融合表现突出。此外，场景自适应策略也十分重要，不同数据在不同的使用场景下会具备不同的策略。阿里 M2M 模型采用元网络（Meta Unit）生成动态权重矩阵，根据工业场景特征调整视觉与热成像数据的贡献比例。在面板缺陷检测中，误判率降低 **70%**，同时处理速度提升 **3** 倍。在权重的优化层面，除上述两种优化方法外，分层优化也是重要实现技术，如 APG 算法，其通过低秩参数化（Low-rank Parameterization）和分解前传（Decomposed Feedforwarding），将动态权重生成的计算复杂度从 $O(NMD)$ 降至 $O(N+M)$ ，支持边缘设备的实时多模态推理。

生成式融合是缺失数据的智能补全与增强技术，其增强了多模态数据的完整性，主要利用生成对抗网络（GAN）补全缺失模态数据，辅助检索与生成。其中，在跨模态生成技术中，BSAFusion 框架采用双向逐步变形场预测策略，通过 **5** 层变形场预测节点实现 CT 与 MRI 影像的缺失区域补全。在脑肿瘤分割任务中，生成影像的 Dice 系数达到 **0.92**，较传统方法提升 **15%**。多模态数

据增强技术以 **MuLan** 为例，其代理利用扩散模型渐进生成多对象场景，通过 **VLM** 反馈控制生成质量。在自动驾驶训练数据生成中，可批量创建包含 20 种天气条件与 100 类障碍物的合成场景，数据采集成本降低 80%。而语义引导生成的 **LLaVA** 模型通过 **GPT-4** 生成多模态指令数据，构建包含 108k 对话样本的跨模态数据集。在医疗教育领域，可根据病理描述自动生成 3D 解剖模型与交互式问答内容。

在数据处理完成后，融合技术只要体现在多任务联合训练中，通过共享表示与参数优化，实现知识迁移与计算效率提升：分层共享架构 **MM1** 模型采用视觉编码器（**ViT-H/14**）与语言解码器（**LLaMA-65B**）的联合训练框架，在图像描述生成任务中，**Bleu-4** 分数达到 0.42，同时支持少样本的思维链推理。

任务交互机制的 **ESMM** 模型通过 **CTR**、**CVR**、**CTCVR** 三任务联合训练，解决电商场景中的样本选择偏差问题。在淘宝推荐系统中，转化率预估误差降低 18%，**GMV** 提升 12%。动态资源分配的 **DiT-XL/2** 模型采用 **Transformer** 架构替代传统 **U-Net**，在 512×512 图像生成任务中，**FID** 指标从 3.85 优化至 3.04，推理速度提升 4 倍，同时支持文本到视频的跨模态生成。

1.2.3 多模态大模型行业现状

在市场需求的增长以及政策支持的背景下，我国大模型市场规模将不断增长。多模态大模型作为 **AI** 模型的发展方向，在各项相关技术愈发成熟下，其应用领域也将愈发广泛，比如说商业定制、游戏和影视等。根据观知海内的数据，预计到 2025 年市场规模将突破 40 亿人民币，具体如下：

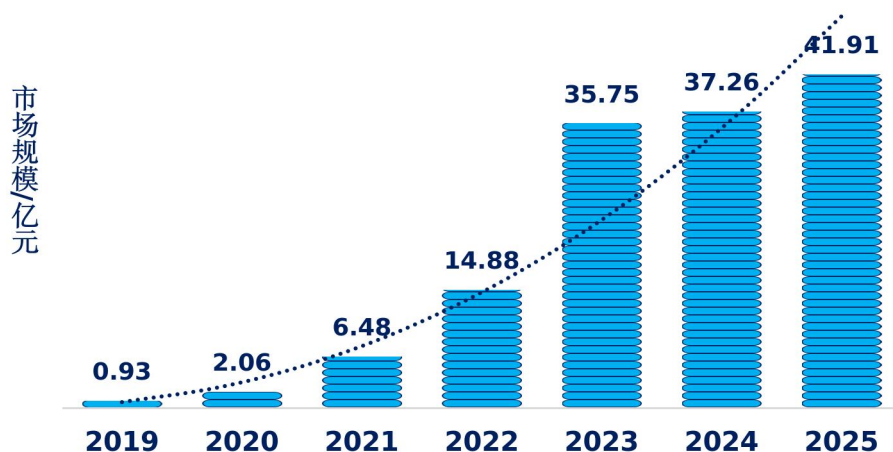


图 1-1 2019-2025 年中国多模态大模型行业市场规模

2023 年生成式人工智能概念兴起至今，国产生成式人工智能大模型如雨后春笋般涌现。截至目前，我国已初步构建了较为全面的人工智能产业体系，相关企业超过 4500 家，核心产业规模已接近 6000 亿元人民币，产业链覆盖芯片、算法、数据、平台、应用等上下游关键环节。

产品数量方面，生成式人工智能产品在我国百花齐放。截至 2024 年 7 月，我国完成备案并上线、能为公众提供服务的生成式人工智能服务大模型已达 190 多个，我国以大模型为代表的人工智能普及率达 16.4%。

产业融合方面，生成式人工智能与各行各业的融合正在我国加速落地。生成式人工智能与制造业、农业、医疗、教育等传统行业深度融合，推动产业转型升级，促进新业态、新模式的不断涌现。尤其在 2024 年，随着生成式人工智能技术的日趋成熟，各大科技企业的模型调用价格显著下降，从而明显降低了其他行业对生成式人工智能技术的应用成本。

截至 2025 年 1 月，我国共有 397 个生成式人工智能产品完成备案，如表 1-1 所示，上海、北京、广东三省的生成式人工智能备案产品数量占比分别达到 30.7%、27.2%和 9%。

表 1-1 《生成式人工智能已备案信息》的产品属地与数量分布

属地	数量	属地	数量
北京	107	上海	122
广东	36	浙江	28
江苏	31	四川	9
贵州	5	湖南	4
山东	9	天津	6
河北	5	重庆	4
海南	2	安徽	1
福建	1	河南	1
黑龙江	2	湖北	1
江西	3	辽宁	2
宁夏	4	陕西	1
云南	5	国资委	8

表 1-2 是中文大模型测评基准（SuperCLUE-V）于 2025 年 3 月提供的开源模型榜单，由表可知，在中文场景下，国内开源模型已具备较大优势。DeepSeek 系列开源模型，Qwen 系列开源模型，在 3 月 SuperCLUE 测评中表现优异，均有超过 Llama-3,3-70B-nstruct 的表现，引领全球开源生态。

表 1-2 SuperCLUE 开源榜单

排名	模型名称	机构	参数量	总分
1	DeepSeek-R1	深度求索	671B	70.34
2	QwQ-32B	阿里巴巴	32B	66.38
3	DeepSeek-R1-Distill-Qwen-32B	深度求索	32B	59.94
4	DeepSeek-V3	深度求索	671B	57.63
5	Qwen2.5-72B-Instruct	阿里巴巴	72B	51.9
6	DeepSeek-R1-Distill-Qwen-14B	深度求索	14B	49.67
-	Llama-3.3-70B-Instruct	Meta	70B	45.53
-	Phi-4	微软	14B	45.26
7	Qwen2.5-14B-Instruct	阿里巴巴	14B	40.7
8	DeepSeek-R1-Distill-Qwen-7B	深度求索	7B	39.07
9	Qwen2.5-7B-Instruct	阿里巴巴	7B	34.01
10	InternLM3-8B-Instruct	上海人工智能实验室	8B	32.02
11	GLM-4-9B-Chat	智谱 AI	9B	29.34
-	Gemma-2-9B-it	Google	9B	28.3
-	Llama-3.1-8B-Instruct	Meta	8B	25.42
12	Yi-1.5-34B-Chat-16K	零一万物	34B	23.29
13	Qwen2.5-3B-Instruct	阿里巴巴	3B	22.18
13	Yi-1.5-9B-Chat-16K	零一万物	9B	21.94
14	DeepSeek-R1-Distill-Qwen-1.5B	深度求索	1.5B	18.98
-	Llama-3.2-3B-Instruct	Meta	3B	18.15
-	Mistral-7B-Instruct-v0.3	Mistral AI	7B	11.78

目前，多模态是主流的迭代路径，互联网大厂利用生态优势将多模态能力融进具体使用场景。和海外相似，国内大模型的迭代方向也能捕捉到多模态的

趋势。热门中文大模型种类如图 1-2 所示。其中多模态方面，如百度文心一言的多模态体现在 ToB 平台“智能云千帆”，帮企业将大模型运用到需要文生图、文生视频的场景；讯飞星火则将多模态能力落地在教师助手、口语训练等教育场景。



图 1-2 热门的中文大模型种类图

1.2.4 多模态大模型政策环境

多模态融合技术已成为目前人工智能大模型跃迁的主要驱动力。以统一 Transformer 架构将视觉、语言、声学乃至时序信号映射到同一高维语义空间，不仅显著提升了模型在开放场景下的泛化能力，也为政务治理、工业质检、智慧交通、数字诊疗等复杂系统提供了端到端的决策基座。随着参数规模迈向千亿、万亿级别，多模态大模型对算力、数据、场景和资本的需求呈指数级放大，其技术迭代路径与产业落地节奏已超出单一企业或科研机构的可控范围，亟需政策层面的系统响应与精准干预，因此，对重要政策的研究也显得十分重要。以下依次展开国家顶层设计、地方实施工具箱及国外重点国家及地区产业政策进行相关整理分析。

1) 国家相关引导政策

表 1-3 近年来国家关于人工智能发展的重点政策梳理

发布时间	文件/事件	发布机构
2017-07-20	《新一代人工智能发展规划》	国务院
2022-08-12	《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》	科技部等六部门
2023-07-10	《生成式人工智能服务管理暂行办法》	国家网信办等七部门
2023-12-31	《“数据要素×”三年行动计划（2024-2026年）》	国家数据局等十七部门
2024-01-18	《国家人工智能产业综合标准化体系建设指南》	工信部、网信办、发改委、国标委
2024-03-05	政府工作报告首提“人工智能+”行动	国务院
2025-03-05	政府工作报告再提“人工智能+”行动	国务院

国家层面，政府及各部委在人工智能发展方面做出系统性的总体部署，同时也逐步搭建并完善起配套的制度。2017年，国务院发布了《新一代人工智能发展规划》，该规划是国家首份系统性的人工智能国家战略文件，明确提出了我国人工智能发展的三步走战略。其核心是通过构建开放协同的人工智能科技

创新体系，推动人工智能产业的快速发展，进而为经济社会发展提供有力支撑。

2022 年，科技部等六部门发布了《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》，该指导意见聚焦于推动人工智能技术在不同场景下的应用创新，鼓励政产学研用各方合作，促进人工智能与经济社会的深度融合，提出加强场景供给，推动人工智能技术在更多领域的示范应用。这为多模态人工智能技术的应用提供了广阔的实践空间，加速了技术从实验室到实际应用的转化过程，促进了经济高质量发展。2023 年，国家数据局等十七部门发布了《“数据要素×”三年行动计划（2024-2026 年）》，该计划旨在推动数据要素与人工智能技术的深度融合，明确了数据采集、存储、处理、分析、应用等全链条的工作方向，特别是支持数据在人工智能模型训练中的应用，这为多模态人工智能模型的训练提供了丰富的数据资源，加速了数据要素价值在人工智能领域的释放，推动了人工智能技术的进一步发展。2024 年、2025 年，国务院政府工作报告中连续两年提及“人工智能+”行动：2024 年中首次提出并为战略锚定方向，以“适度超前”的基础设施为人工智能铺路，解决“有没有”的问题；在 2025 年报告中则深化落地，强调融合，以数据要素市场化推动技术价值转化，解决“用不用得好”的问题。

2) 地方相关引导政策

如表 1-4 所示，在地方层面，各地在多模态人工智能大模型领域的政策布局，充分体现了因地制宜的发展理念，具有重要意义。

北京凭借科研实力与政策资源优势，致力于打造全球领先的人工智能创新策源地，其全栈国产化路径为我国人工智能产业的技术自主可控提供了坚实保障。上海通过构建“模速空间”专业园区与工业大模型生态，聚焦人工智能在

工业等领域的深度应用，推动产业升级与经济高质量发展。深圳依托开源生态与深港合作，加速人工智能技术的创新与应用，为粤港澳大湾区的科技发展注入新动力。杭州以“1+N+X”生态为核心，通过提供丰富的政策工具，吸引企业集聚，助力当地人工智能产业的快速发展。苏州聚焦 AI+制造业场景示范，为制造业智能化转型提供有力支撑，提升产业竞争力。无锡以边缘智算与具身智能为特色，推动人工智能在物联网等领域的创新应用，拓展产业发展新空间。安徽通过“拎包入住”模式与全方位政策支持，吸引企业落地，促进产业集聚与创新发展。宁夏则利用其能源与算力优势，为东部地区的人工智能训练提供支持，实现区域间的协同发展。

总体而言，各地政策从资金、算力、数据、场景等多方面为多模态人工智能大模型产业提供了有力支撑，推动了技术突破与应用拓展，为我国在全球人工智能竞争中占据有利地位奠定了基础。

表 1-4 国内各地方人工智能发展重要规划

地区	量化目标	核心路径	政策工具	相关文件
北京	3-5 个自主可控基础大模型、100 个行业大模型、1000 个成功案例；公共算力 5000 PFLOPS	标杆应用牵引+全栈国产化	100 亿元 AI 产业基金；算力券首期≥4000 万元；模型沙盒监管	《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025 年）》
上海	2027 年集聚 100 家大模型生态企业；打造 100 个 AI 转型服务平台	“模速空间”专业园区+工业大模型	大模型最高 1 亿元综合支持；国家级测试验证中心；数据飞轮机制	《上海市推动人工智能大模型创新发展若干措施（2023-2025 年）》
深圳	2024 年商用通用大模型 3-5 个、标杆场景 10+	开源生态+深港合作	首版次目录产品最高 1000 万补贴；前海跨境研发通道	《深圳市加快推动人工智能高质量发展高水平应用行动方案(2023—2024 年)》
杭州	2025 年智算规模 50 EFLOPS；1 个国际一流通用大模型、10 个行业重大模型	“1+N+X”生态	每年 2 亿元券包（算力+模型+数据券）；重大项目 25%配套	《杭州市人民政府办公厅关于加快推进人工智能产业创新发展的实施意见》
苏州	千亿参数模型最高 2000 万；垂直模型场景≥5 个，300 万/年×3 年算力补贴	AI+制造业场景示范	场景采购双方最高各 200 万/100 万奖励；首台套/首版次补贴	《苏州市推进人工智能产业创新发展的若干措施》
无锡	2025 年智算集群 EUE ≤ 1.25；60%算力成本补贴	边缘智算+具身智能	三类创新券单企最高 1000 万；备案奖励 50 万起	《无锡市促进人工智能产业创新发展若干政策措施》
安徽	2027 年通用大模型性能全国前列；智算 30000P	“拎包入住”模式	揭榜挂帅项目最高 5000 万；算力/数据/场景/基金/场地“五个优先”	《安徽省打造通用人工智能产业创新和应用高地若干政策》
宁夏	百亿参数以上模型最高 1000 万	东西部算力走廊节点	研发补贴 30%最高 1000 万；绿电电价 ≤0.32 元/度	《宁夏回族自治区促进人工智能创新发展政策措施》

3) 国外重点地区政策梳理

其他重点国家或地区也在积极制定政策来引导人工智能产业有序发展。美

国，美国政府提供高质量大数据以促进人工智能发展，如特朗普政府提议向私营部门和学术界提供数据集以训练机器学习模型；《开放政府数据法》要求联邦机构以标准化格式发布信息，提高数据质量和可用性；2019 年签署《AI 政策法案》，推动政府机构更广泛地采用人工智能技术，包括人工智能培训和研究计划、人工智能伦理原则和人工智能安全措施等。日本，2017 年确定为人工智能元年，政府预算给予多方面支持；于 2017 年 3 月，AI 技术战略委员会发布《人工智能技术战略》报告，制定产业化发展路线图，将人工智能发展分为三个阶段，并将其应用到“社会 5.0”倡议的三个主要领域；2025 年 5 月通过《人工智能相关技术的研发及应用促进法》，旨在建立国家层面的 AI 发展框架，确立以技术自主、国际竞争力。韩国，2024 年 12 月通过《人工智能发展与信任基本法》，成为继欧盟之后第二个制定人工智能综合立法的国家；该法案聚焦于构建国家层面的人工智能协同治理体系，设立国家人工智能委员会等专门机构，统筹推进人工智能发展与安全。欧洲，2023 年 12 月，欧洲议会、欧盟成员国和欧盟委员会三方就《人工智能法案》达成协议，该法案于 2024 年 8 月正式生效，是全球首部全面监管人工智能的法规。此外，欧盟还针对特定数字场景制定了《数字服务法案》《数字市场法案》《人工智能责任指令》《产品责任指令》修订版等，致力于构建人工智能领域的全球标准。

由此可见，美国推动数据开放与立法保障，日本提供预算支持与战略规划，韩国构建协同治理体系，欧洲强化全面监管与标准制定。各国均积极部署政策，旨在引导人工智能大模型及相关产业有序发展，以期在全球竞争中占据有利地位。

1.3 研究思路

1.3.1 分析方法

本项目采用专利文献研究、产业研究相结合的方式进行研究。

专利文献研究通过综合量检测的技术特征关键字及其同位语、国际技术分类号、主要专利权利主体及其同位语、申请人所在区域等方式，通过构建检索式在专业专利数据库中获取全球、中国大陆的专利数据。主要采用聚类分析方法，开展趋势、区域、布局等分析。

同时结合市场及产业数据研究包括针对人工智能大模型等相关配套产业的综合研究，以配合专利信息利用梳理出专利分析工作内容。

1.3.2 分析内容

本报告分析内容具体而言，从宏观整体态势到中观关键参与者，再到微观技术分析及应用分析的脉络展开分析。

首先，通过分析全球范围内 AI 多模态融合相关技术的整体专利申请趋势、技术区域分布、专利主体类型、专利同族及引证等关键信息，揭示该技术的发展脉络与未来走向。

其次，报告聚焦核心权利人，进行具体分析，包括申请趋势、区域分布、法律状态、发明人团队、技术发展动态，来剖析其技术布局与创新实力，助力企业精准把握市场竞争格局。

同时，报告针对模态融合及对齐等核心技术进行专题探讨，解读关键专利的关键技术手段，了解不同技术问题背景下各创新主体是如何进行技术破题的，以此来推动技术深化与创新。结合医疗健康、智慧交通等领域的应用实例，彰显多模态融合技术的多元价值与实践成效，为行业应用拓展提供参考。

最后，基于详实的专利分析数据，提出针对性建议，助力强化知识产权管理、促进产学研协同、提升专利运营效能，加速多模态融合技术的成果转化与商业化进程，全方位赋能技术发展与产业升级。

1.3.3 技术框架

表 1-3 技术框架

技术主题	技术一级	技术二级	技术三级	技术四级
AI 多模态 融合技术	模态融合	数据级融合	数据预处理	归一化
			数据拼接	向量拼接
				矩阵拼接
		特征级融合	特征提取	CNN/RNN/LSTM
			特征拼接	/
			加权求和	/
		决策级融合	投票机制	多数投票
				加权投票
			概率平均	/
	模态对齐	显式对齐	无监督对齐	循环对抗网络
				CAA
				自编码器
			监督对齐	多模态嵌入
				多任务学习

		隐式对齐	注意力对齐	Transformer
				BERT-based
			语义对齐	GNN
				CLIP

1.4 检索范围及检索结果

1.4.1 检索策略

本项目研究对象为：AI 多模态融合技术相关专利

（1）检索地区的选择：根据本项目的实施方法和实施路径，确定本项目的专利数据分析范围为：全球。

（2）检索方式的选择：关键词、分类号等检索要素构建检索式，辅以主要权利人等权利人进行检索。

（3）分类号的确认：

a) 尝试性检索：利用 AI 多模态融合技术相关市场主体作为关键词，在专利网站和数据平台上进行检索，观察并记录检索结果，进而形成多模态融合技术相关 IPC 分类号表；

b) 修正检索：根据与多模态融合技术相关性对 IPC 分类号和关键词表进行修正，并进行梳理分类，进而形成最终 IPC 分类号表及最终关键词表。

c) 最终检索：使用最终确定的分类号确认表及关键词表，结合重点专利权人进行组合检索

1.4.2 检索结果

本报告检索范围：截至 2025 年 4 月 30 日在全球范围公开/公告之专利/专利申请。

本报告检索结果：检索得到原始专利数据 **11153** 件，去重去噪后得到 **8685** 件专利。

第二章 AI 多模态融合技术宏观分析

本章基于人工筛选所得的与项目技术密切相关的 8685 件专利进行统计分析，以宏观观察 AI 多模态融合相关技术专利情况，获取 AI 多模态融合技术的总体布局信息。具体地，本章重点研究了专利申请趋势分析、技术区域部署分析、专利权人排名和研发能力比较、主要专利权人分析、专利的同族数分析及被引证数分析等。

2.1 专利整体分析

2.1.1 申请趋势情况

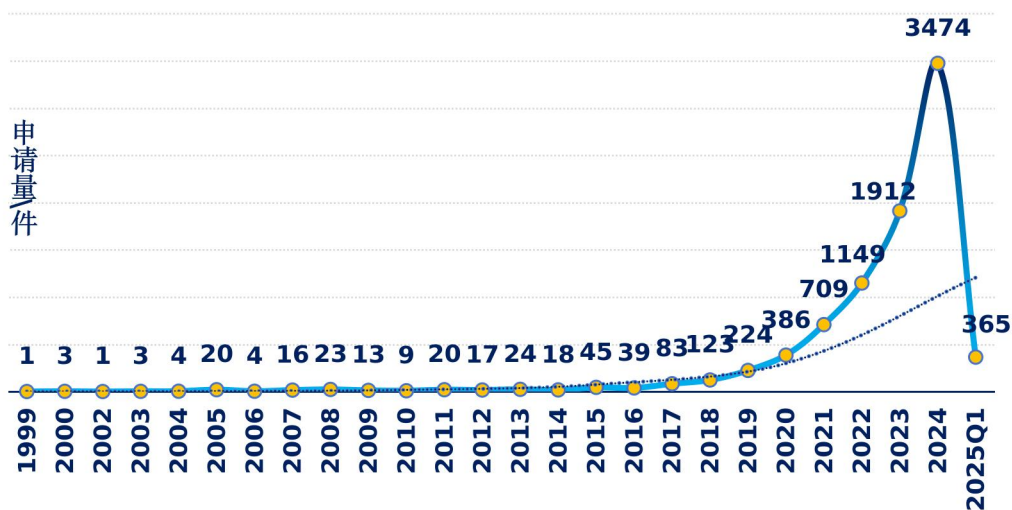


图 2-1 AI 多模态融合技术相关专利申请趋势

图 2-1 显示了多模态融合技术专利的发展趋势。从图中可以看出，存算一体技术的发展以 2010 年为界，分为技术起步期和深度学习驱动阶段。

（1）技术起步期（20 世纪 90 年代末-本世纪 10 年代）：

随着计算能力的提升和传感技术的发展，语音和图像的融合研究逐渐兴起，主要应用于多媒体信息检索和人机交互。这一时期的研究开始关注如何有效地将不同模态的信息进行融合，以解决实际应用中的问题，基于概率模型和机器

学习的方法开始应用于多模态融合。从专利申请来看，多模态融合技术专利申请，其申请时间开始于上世界 90 年代末期。其中，最早申请来自于微软公司的专利——公开号为 [US6502082B1](#)，其公开了一种进行对象跟踪的模态融合系统和方法。在最初的十年间，AI 多模态融合技术的专利申请数量一直维持低位，年均专利申请量均在 10 件以内，早期的融合技术暂未出现较大的进展，专利申请多涉及数据融合方法。而从权利人来看，微软、IBM 和摩托罗拉等主导了这一阶段的专利申请。

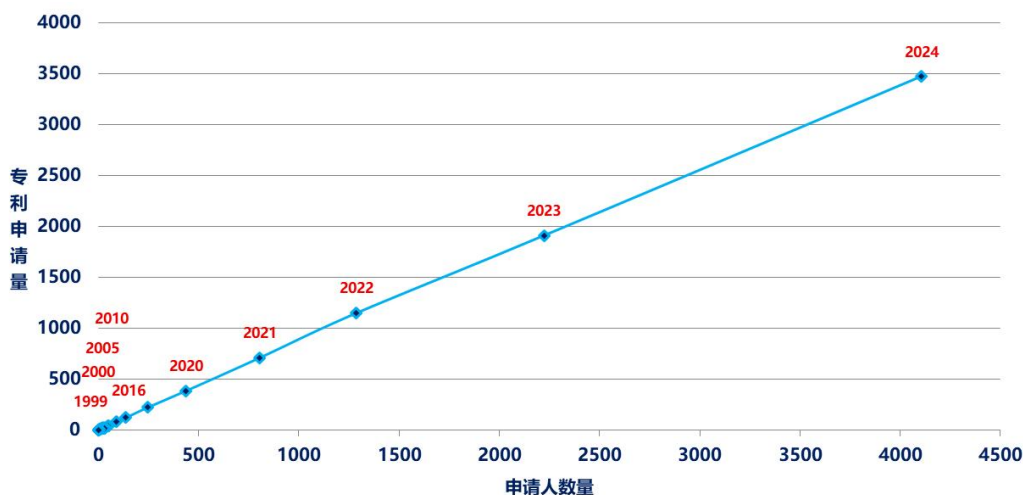


图 2-2 AI 多模态融合技术生命周期图

（2）深度学习驱动阶段（2013-至今）：

深度学习的兴起为多模态融合带来了新的机遇。2010 年左右，深度学习方法在计算机视觉和自然语言处理领域取得突破，Multimodal Deep Boltzmann Machines (DBMs) 和 Deep Belief Networks (DBNs) 等模型被提出用于多模态融合。2014 年，Vinyals et al. 提出了图像描述生成模型，结合卷积神经网络（CNN）和循环神经网络（RNN），实现了图像和文本的有效融合。2015 年，Google 的 Neural Machine Translation (NMT) 系统使用了多模态信息，提高了翻译的准确性和流畅性。

结合图 2-2 的多模态融合技术生命周期图可以看出：从 2010 年开始，多模态融合技术的申请人数量与各个年份的专利申请量都快速上升，二者近似形成

线性关系。这结合专利申请的整体趋势，表明了多模态融合技术呈现一个快速发展阶段。

进入 2020 年代，多模态融合技术在各个领域的应用得到了进一步扩展，从自动驾驶到医疗诊断，从虚拟现实到智能家居，多模态融合技术展示了其广泛的应用前景。2020 年，OpenAI 推出了 DALL-E，结合文本和图像生成技术，实现了基于文本描述生成图像的能力。2021 年，Meta 发布了 HoloLens，利用多模态融合技术实现了增强现实（AR）和虚拟现实（VR）的创新应用。2023 年，在自然语言处理和视觉理解的融合研究中，Transformer-based 模型（如 CLIP, DALL-E 2）展示了强大的多模态信息处理能力，推动了多模态融合技术的发展。这样有底层应用需求拉动、深度学习驱动的模式，在专利申请方面，今年来也一直保持高速增长势头。

2.1.2 申请区域分布

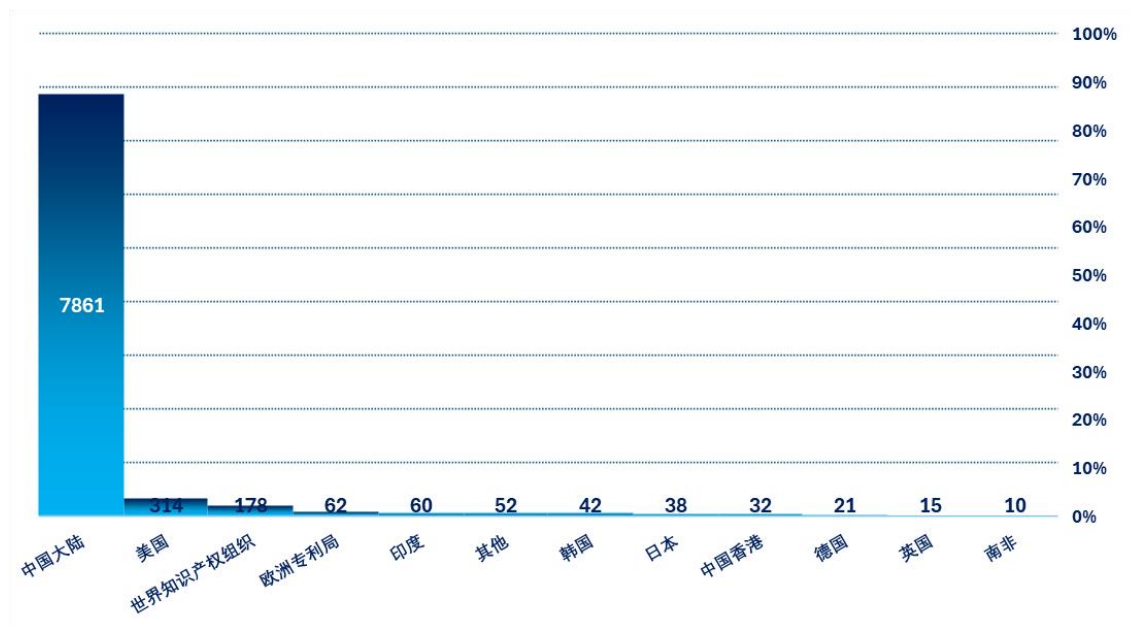


图 2-3 AI 多模态融合技术申请区域分布图

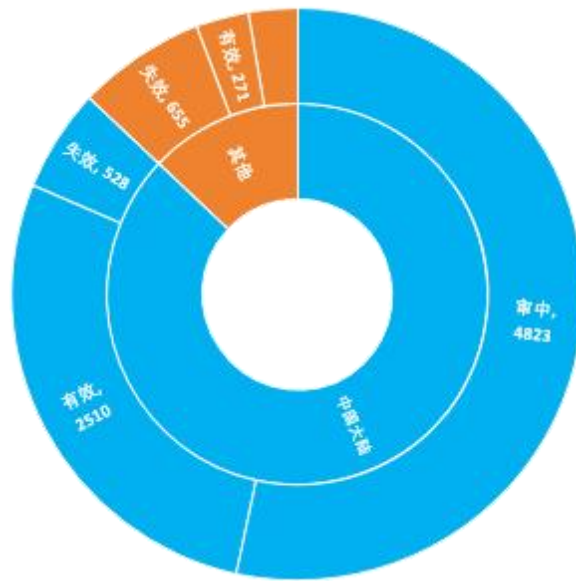


图 2-4 AI 多模态融合技术简单法律状态分布图

图 2-3 与 2-4 分别展示了多模态融合技术专利地区分布及法律状况。其中，中国大陆的专利申请量最多，为 7861 件，占比 90%；其次是美国，申请量为 314 件，占比接近 4%。整体来看，中国大陆和美国是 AI 多模态融合技术的主要布局区域。除这两个地区外，欧洲、印度和韩国等区域是多模态融合技术次一级布局区域。结合法律状态来看，中国大陆地区的有效专利为 2510 件，同时中国大陆地区的在审专利达到了 4823 件，将其他地区拉开了较大的差距。由此也可以看出对于 AI 多模态融合技术的研究方面，中国大陆地区的热情是远大于其他地区的。

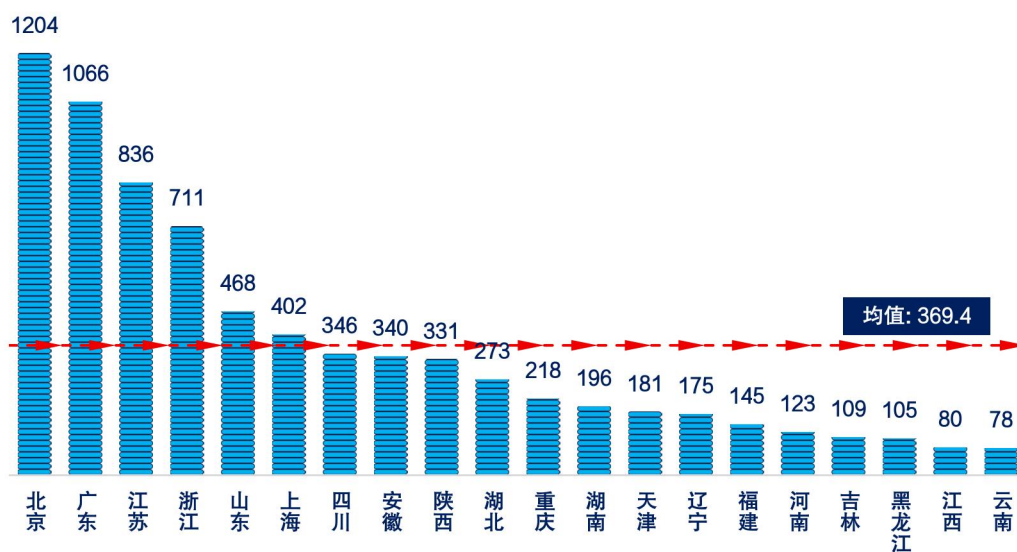


图 2-5 AI 多模态融合技术中国大陆各区域申请量

图 2-5 也进一步展示了截止分析基准日，中国大陆内部各个省市地区在 AI 多模态融合技术方面专利的申请数量情况。各个省份在 AI 多模态融合技术方面的专利平均申请量为 369 件。其中，北京的专利申请量最多，为 1204 件，占比 15.3%；其次是广东，其申请量为 1066 件，占比 13.5%；排名第三的地是江苏，其申请量为 836 件，占比为 10.6%。整体来看，超过申请平均量 369 件的地区有 6 个，多集中在华东、华南的沿海地区；中部地区各重要省份也有大量多模态融合技术相关专利申请，较为接近大陆地区平均申请量。这一分布特点，也与前序章节中我国各地人工智能大模型的备案数量大致情况相匹配。同时，这也反应了底层需求驱动技术发展的显著特点。

2.2 专利主体分析

2.2.1 专利主体排名

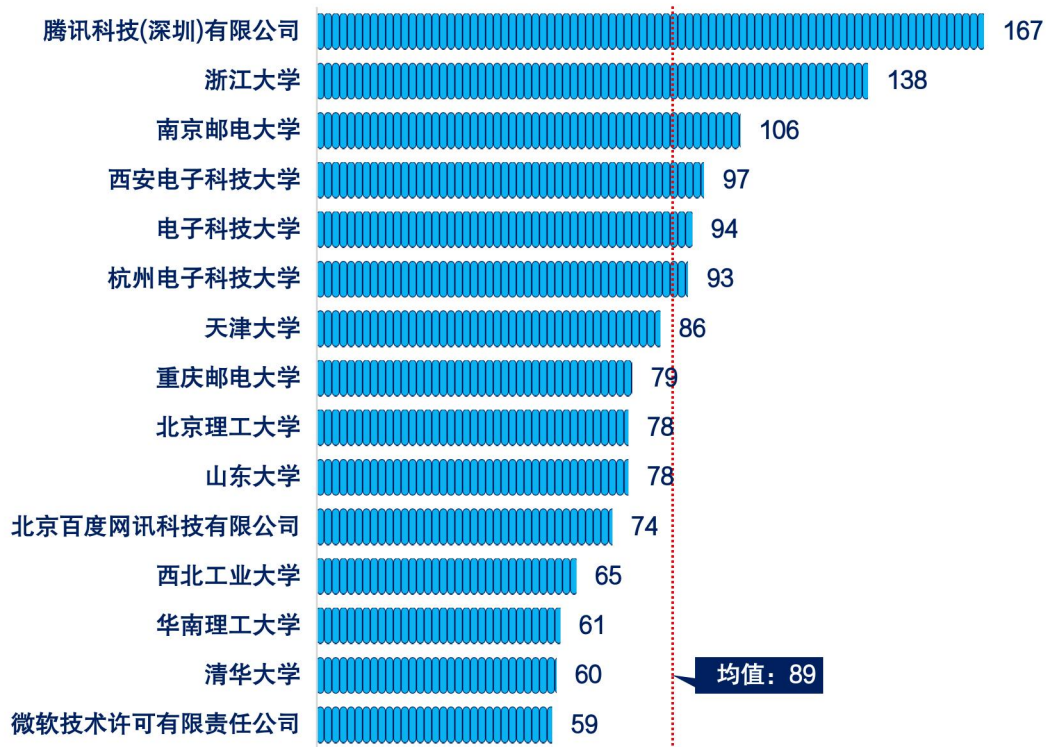


图 2-6 AI 多模态融合技术权利人申请排名

图 2-6 显示了 AI 多模态融合技术专利权利人申请量的排名情况。其中，腾讯科技(深圳)有限公司的专利申请数量最多，达到了 167 件；紧随其后的是浙江大学，其多模态融合技术的相关专利请量达到了 138 件，作为国内顶尖高校，其在科研创新方面也具有显著优势。

从整体排名来看，多模态融合技术相关专利的申请数量 TOP15 中，仅有三席申请人是公司性质，而且都是科技巨头；其余都是来自于中国的高校。由此可以看出，这突显出中国高校在 AI 多模态融合技术领域的科研热情高涨，为推动技术进步发挥了重要作用。同时，也说明大陆地区对于人工智能新兴人才的培养所做出的巨大努力。

从专利申请量的平均值来看，前十五名各个主体的平均申请量为 89 件，高出这一基线的创新主体有 6 家。这表明这些主体在多模态融合技术领域具有较强的研发实力 and 创新能力，对技术发展起到了引领作用。此外，北京百度网讯科技有限公司、微软技术许可有限责任公司，申请量也较为可观，分别为 74 件和 59 件，成为技术创新的重要力量之一。

2.2.2 专利主体类型

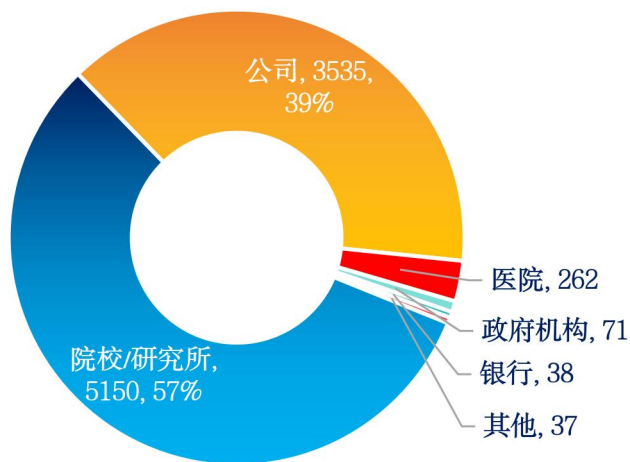


图 2-7 AI 多模态融合技术专利主体类型分布图

图 2-7 展示了 AI 多模态融合技术的申请专利的主体类型情况。显而易见的是，申请(权利)人为院校/研究所的占到绝大多数，其达到了 5150 家，占比 57%。这体现了在多模态融合技术领域，院校和研究所作为科研主力军，承担了大量的基础研究和前沿探索工作，是技术创新的核心动力。

而申请（权利）人为公司的有 3535 家，占比 39%，这一占比也相当可观。公司作为技术创新的主体，能够将科研成果快速转化为实际应用，推动技术的商业化和产业化发展。高校和企业 在专利申请量上的差异，从侧面反映出不同主体在技术研发和应用转化中的不同侧重。

此外，申请（权利）人为医院的有 262 家，银行的有 38 家，政府机构的有 71 家，其他主体的有 37 家。这些主体在多模态融合技术领域也有一定的参与度，从不同角度推动了技术的发展和 应用。例如，医院在医疗影像与多源

数据融合等场景下对相关技术有需求；银行等金融机构则可能在金融风控等领域探索应用；政府机构在智慧城市、公共服务等方面也有技术落地的需求，共同构成了多模态融合技术领域的多元创新生态。

综合来看，AI 多模态融合技术领域的专利申请主体呈现出多主体协同创新的格局，不同类型的主体在技术研发、应用转化和市场推广等方面各展所长，共同推动着该技术的快速发展和广泛应用。

2.3 技术方向分布

2.3.1 技术主题分布

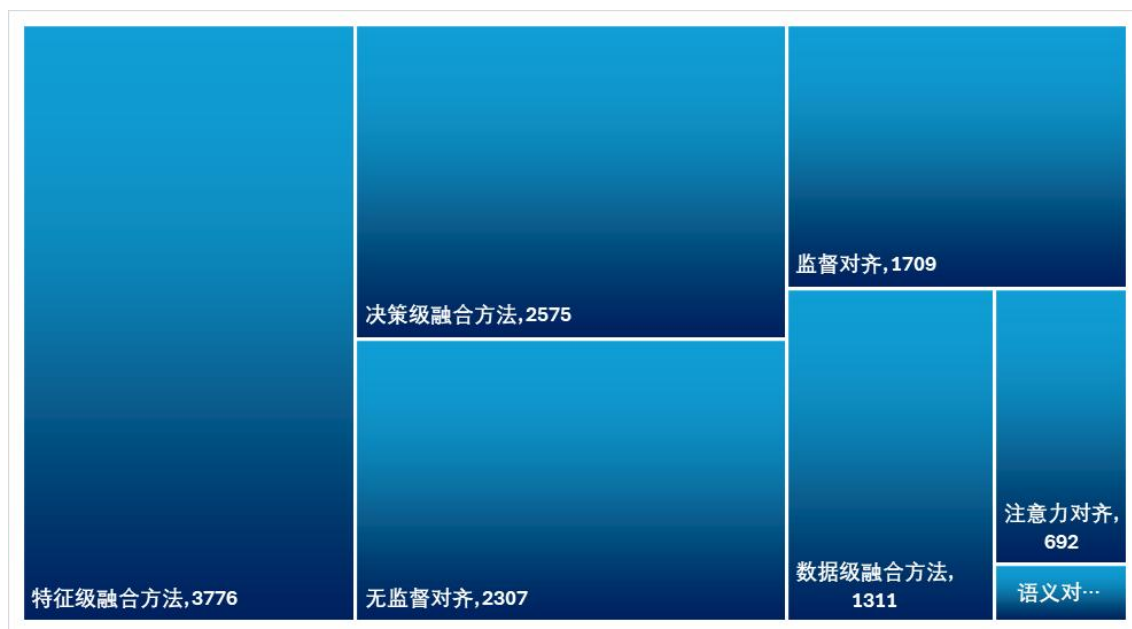


图 2-8 AI 多模态融合技术专利技术主题分布

图 2-8 显示了 AI 多模态融合技术的技术分布情况。。“特征级融合方法”专利数量最多，达到 3776 项，占据主导地位，说明它在多模态融合领域得到了广泛关注和应用，是当前研究和技术创新的一个重要热点方向，在各类多模态应用场景中，特征级融合能够有效整合不同模态的数据特征，为后续处理提供更丰富、更全面的信息基础。

“决策级融合方法”也有 2575 项专利，数量相对较多，表明在对多模态信息进行综合决策判断方面，业界投入了较多研发精力，旨在通过合理的决策融合

策略，提高多模态系统在复杂情况下的判断准确性和可靠性，如多模态智能安防系统中对各类危险行为的最终判定。

无监督对齐方向有 2307 项专利，无监督学习在多模态融合中具有重要意义，能够在没有大量标注数据的情况下实现模态间的对齐，这降低了对标注数据的依赖，拓展了多模态融合技术的应用范围，尤其在一些难以获取大量标注样本的领域，如部分医疗影像与文本描述的对齐等场景，具有较好的发展前景。

监督对齐专利数量为 1709 项，它借助标注数据来实现模态对齐，能有效保证对齐的准确性和一致性，广泛应用于对精度要求较高的多模态任务，像自动驾驶场景中，通过对车辆传感器数据（图像、雷达等）与标注的路况信息进行监督对齐，来精准地识别道路状况和周围物体。

数据级融合方法对应 1311 项专利，在多模态数据融合的基础环节发挥关键作用，不同模态的数据在数据级进行融合后，能够为后续的特征提取、决策判断等过程提供更完整、质量更高的数据输入，对于提升整个多模态系统的性能有着基础性的支撑。

注意力对齐方向有 692 项专利，注意力机制在多模态融合中可帮助模型聚焦于关键的模态信息和对应关系，提高融合效率，例如在多模态情感分析中，通过注意力对齐能更精准地抓住表情、语音语调等不同模态中反映情绪的关键特征。

语义对齐方向的专利数量虽然只有 145 件，不过语义对齐对于深入理解多模态信息的内在含义、建立不同模态间的语义关联等具有重要意义，也是未来多模态融合技术朝着更深层次发展需要攻克的关键方向之一。

整体来看，特征级融合方法、决策级融合方法、无监督对齐等占据多数专利数量，体现了多模态融合技术发展的阶段性和侧重点，即当前在数据融合的基础方式（如特征级、数据级）、比较直接的融合决策策略以及不依赖或少量依赖标注数据的对齐方法等方面的研究和应用更为成熟且活跃，而像语义对齐这类更偏向深层次语义关联构建的较新方向，可能还处于逐步探索和积累的阶段。

2.3.2 分类号分布

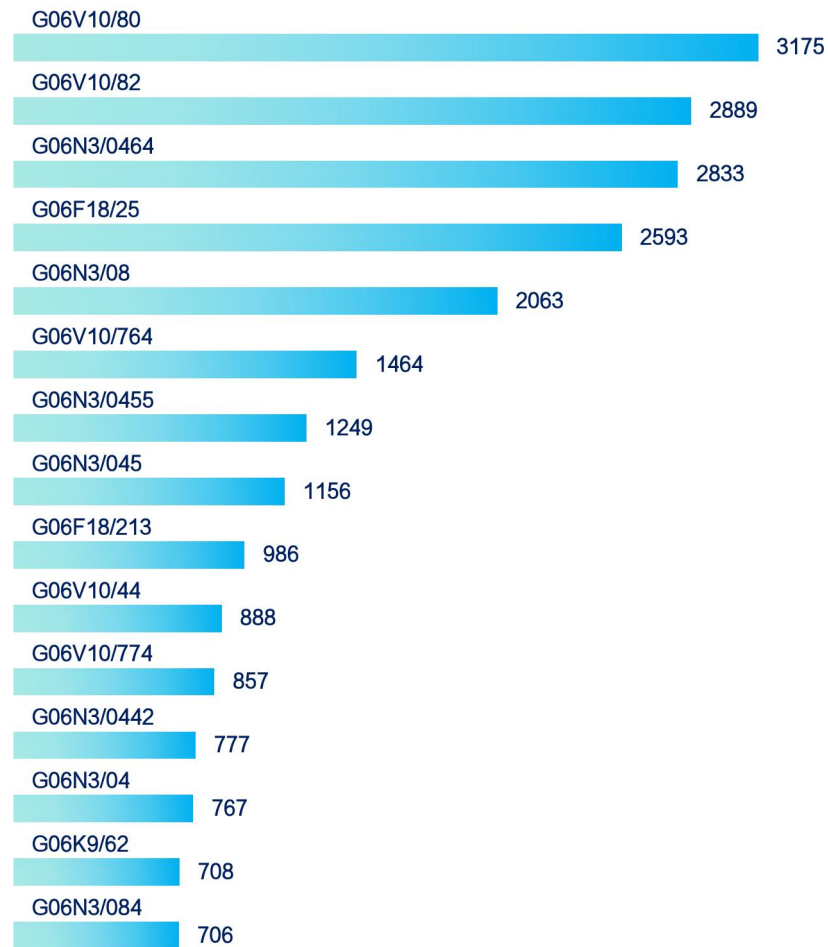


图 2-9 AI 多模态融合技术专利 IPC 分类号分布

图 2-9 展示了 AI 多模态融合技术专利的 IPC 分类号分布情况，从专利申请量来看，AI 多模态融合技术的专利申请数量较多，其中，分类号 G06V10/80 的专利申请量最多，达到了 3175 项，G06V10/82 次之，专利数量为 2889 项。G06N3/0464、G06F18/25、G06N3/08 等分类号的专利数量也都超过了 2000 项。这些数据表明，AI 多模态融合技术在相关技术领域的创新较为活跃，研发人员在这些方向上进行了大量的技术探索和尝试。

整体而言，这些 IPC 分类号的分布反映出 AI 多模态融合技术在数据融合、神经网络应用、特征提取与学习、编码器网络等关键技术方向上的研发热度和

创新趋势，各分类号下的专利数量差异也体现了不同技术路径在当前发展阶段
的关注程度和应用潜力。

2.4 专利运营情况分析

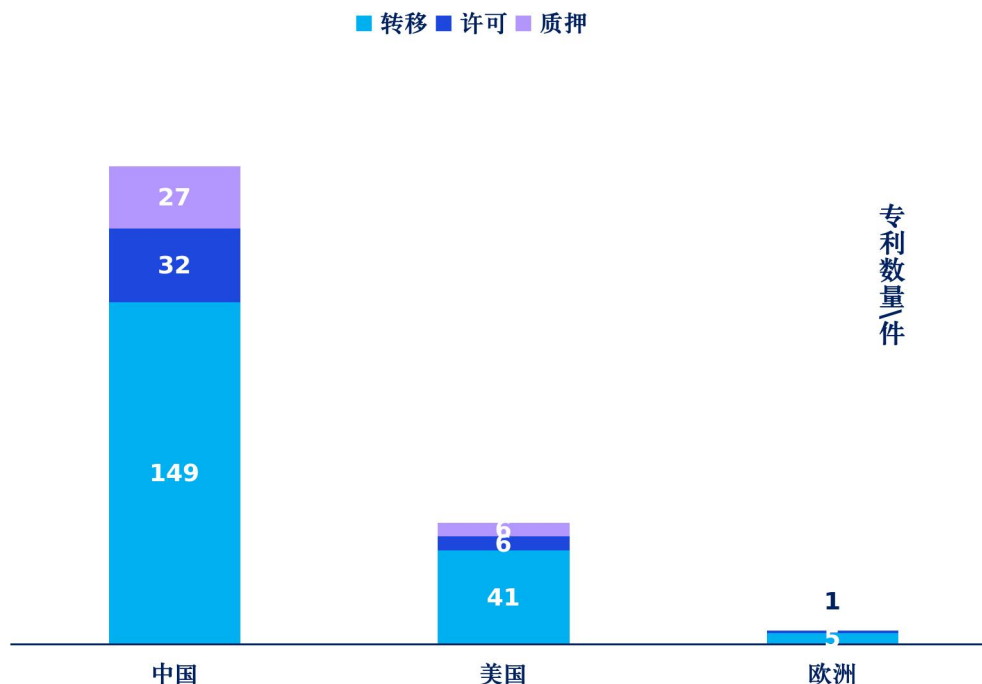


图 2-10 AI 多模态融合技术主要地区专利运营情况

如上图 2-10 所示，展示了 AI 多模态融合技术相关专利市场化运营情况。
从类别上看，主要涉及三种，分别是专利质押、转让和许可。

在中国，这三种运营方式的专利数量分别为 149 件、32 件和 27 件，表明中国在该领域的专利运营活动较为活跃，尤其是专利质押融资活动较为频繁，这可能与中国企业对知识产权的重视以及对资金需求有关。

美国的专利运营数量相对较少，其中转让有 41 件，许可和质押分别为 6 件和 1 件，反映出美国企业在专利运营上更倾向于转让。欧洲的专利运营活动较少，仅有 5 件转让和 1 件许可，可能由于其专利制度和市场环境的差异导致运营热度较低。

总体而言，全球对该技术的专利运营处于发展阶段，中国在数量上占据优势，未来还需进一步优化运营模式，促进技术的市场转化，而美国则在转让方面表现相对突出，这可能与其成熟的知识产权交易市场有关，欧洲则相对滞后，可能需要加强专利运营的相关政策支持和市场培育。

2.5 中美技术对比分析

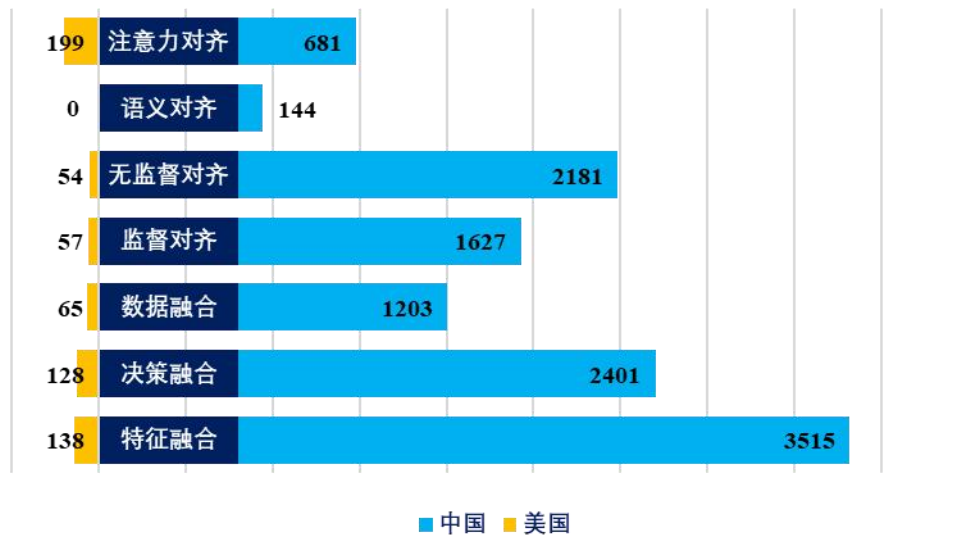


图 2-11 AI 多模态融合技术中美权利人技术分布对比

针对分别来在于中国大陆地区、美国地区的原始申请人进行主要的技术分布情况。通过对比可以发现：

首先，AI 多模态的模式融合方面，特征融合、决策融合、数据融合三个技术分支，均呈现中国专利数量远超美国专利数量的情况。这表明中国在特征融合领域具有绝对优势，可能在技术研发和应用方面投入了大量资源，同时中国庞大的市场需求和科研环境也促进了该领域的专利产出。

其次从多模态的模式对齐来看，监督对齐方面，中国拥有 1627 项专利，美国仅有 57 项。监督对齐依赖于标注数据，中国在这一领域的优势可能得益于丰富的数据资源和标注能力，以及在相关应用场景中的深入研究。无监督对

齐方面，中国有 **2181** 项专利，美国仅有 **54** 项。无监督对齐技术因其在数据标注依赖少的特点，在一些难以获取大量标注数据的领域具有重要作用，中国的专利数量优势表明在这一方向上投入了较多研发精力。注意力对齐方面，中国有 **681** 项专利，美国仅有 **199** 项。注意力对齐通过聚焦关键信息提高融合效率，中国在该领域的专利数量较多，反映了在这一技术方向上的积极布局。语义对齐方面，中国有 **144** 项专利，语义对齐涉及更深层次的语义理解，中国在这一领域的专利数量也处于领先地位，表明在探索多模态语义关联方面取得了一定进展。

总的来说，在模态融合和模态对齐技术方向上，中国的专利数量均大幅领先于美国，尤其在特征融合、决策融合、无监督对齐等领域，优势明显。这反映了中国在 **AI 多模态融合技术** 领域的快速发展和强大创新能力。

美国在各方向上的专利数量相对较少，但在注意力对齐领域相对较强，这可能表明美国在这一方向上有一些特定的技术优势或应用场景。中国的专利分布较为均衡，涵盖多个技术方向，而美国的专利数量相对集中在少数方向。

2.7 本章小结

本章从宏观视角对 **AI 多模态融合技术** 相关专利进行了分析。首先，在专利申请趋势方面，清晰呈现出该技术自 **2010** 年起的快速发展态势，申请量逐年攀升，尤其在 **2024** 年达到峰值，这不仅反映出技术的蓬勃生命力，更凸显其在众多领域巨大的应用潜力和市场需求。从地域分布来看，中国以 **7861** 件专利申请量占据首位，美国紧随其后，两国在该领域的技术优势和市场重视程度显而易见。进一步分析发现，中国大陆的高校和科研机构成为技术创新的主力

军，其中浙江大学表现尤为突出，这充分说明高校在前沿技术探索中扮演着不可或缺的角色。

技术主题分布上，“特征级融合”相关专利申请量高居榜首，达到 3776 件，占据主导地位，说明它在多模态融合领域得到了广泛关注和应用，是当前研究和科技创新的一个重要热点方向，，在各类多模态应用场景中，特征级融合能够有效整合不同模态的数据特征，为后续处理提供更丰富、更全面的信息基础。同时，IPC 分类号分布也有力地印证了多模态融合技术在数据融合、神经网络应用等关键技术方向上的创新活跃度，为后续的深入分析提供了详实的数据支撑和明确的方向指引。

但是，从专利主体类型和中国大陆地区专利运营情况的数量对比来看，可以发现以下问题：尽管中国大陆的院校和科研机构在 AI 多模态融合技术领域贡献了大量的专利申请，但专利运营数量相对较低，表明这些专利大多停留在理论研究或实验室阶段，未能有效转化为实际的产品或服务，反映出在专利成果转化方面存在一定的障碍；企业作为市场运营的主体，与院校/科研机构之间的合作机制可能不够完善，导致专利技术难以快速、有效地进入市场；院校和科研机构可能在这方面的投入和重视程度不够，缺乏专业的专利运营人才和有效的运营策略。

中美在 AI 多模态融合技术专利数量上差异显著：中国在各方向上专利数量全面领先，尤其在特征融合、决策融合、无监督对齐等领域优势明显；美国专利数量相对较少，但在注意力对齐方向相对较强。这体现了中国在该领域的快速发展和强大创新能力，以及美国在特定技术方向上的一定优势。

第三章 AI 多模态融合技术主要权利人分析

权利人是专利申请的主体，也是技术发展的主要推动力量，重要权利人更是研发和创新的领先力量和领先指标，是市场竞争的主体。通过对重点权利人的研究和分析，可以掌握整个市场竞争格局，获取相关公司发展战略、专利布局特征、专利战略和研发方向等。

基于前一章对国内外重要权利人专利申请量的排名情况。特分别选取两位国内外重要权利人，具体分析包括公司概况、专利申请趋势分析、专利申请区域分析、法律状态分析和技术发展路线分析等，以期获得其在 AI 多模态融合技术领域的整体发展态势和技术发展路径。其中，国内主要权利人选取的是腾讯科技和百度这两家专利申请总量位列前两位的企业，同时选取了专利申请量排名第二的浙江大学进行研究。国外方面，结合产业情况和专利申请情况，选择微软科技和谷歌科技作为研究对象。

3.1 国内主要权利人

3.1.1 腾讯科技

1) 公司简介

腾讯科技（深圳）有限公司成立于 1998 年，总部位于中国深圳，是中国领先的互联网科技企业之一。公司以社交平台微信和 QQ 为核心，业务广泛布局于游戏、金融科技、云计算、人工智能等关键领域，在全球拥有庞大用户群体。腾讯致力于通过技术创新推动行业发展，积极参与 AI 等前沿技术研究与应用，旨在打造智能便捷的产品和服务。在多模态融合技术领域，腾讯的研发投入大，专利申请量国内领先，技术应用覆盖内容创作、医疗影像分析、智能驾驶等多个场景，持续探索新兴应用方向。

腾讯的混元大模型是其在多模态融合技术领域的重要成果，支持文本、图像、视频和 3D 等多种模态内容的理解与生成。模型采用混合专家模型（MoE）结构，其总体性能相比上一代提升了 50%，部分中文能力已追平 GPT-4。例如，混元大模型的图像到文本模型包含视觉变换器（ViT）用于视觉编码、视觉 - 语言适配器，以及采用 MoE 架构的大型语言模型。



图 3-1 腾讯科技的多模态模型框架图

腾讯不断探索和优化多模态融合的架构和方法。例如，通过视觉大语言模型的文本 - 图像融合模块，实现文字描述和图像内容的精确匹配和融合。开发了基于视觉变换器（ViT）和 Swin Transformer 的多模态编码器，用于视觉特征提取，并通过多模态对齐和融合技术，将不同模态的特征进行有效整合。

腾讯的多模态技术在内容创作领域得到了广泛应用，如影视创作、社交内容生产、游戏资产生成和智能广告投放等。通过多模态模型的混合生成模式，实现了图生图、图生视频等创新功能，降低了创作门槛，提高了创作效率。此外，混元大模型还为腾讯视频、QQ 浏览器、微信等内部产品提供智能剪辑、智能创作等能力支持，助力内容的高效生产和推荐。

2) 申请趋势

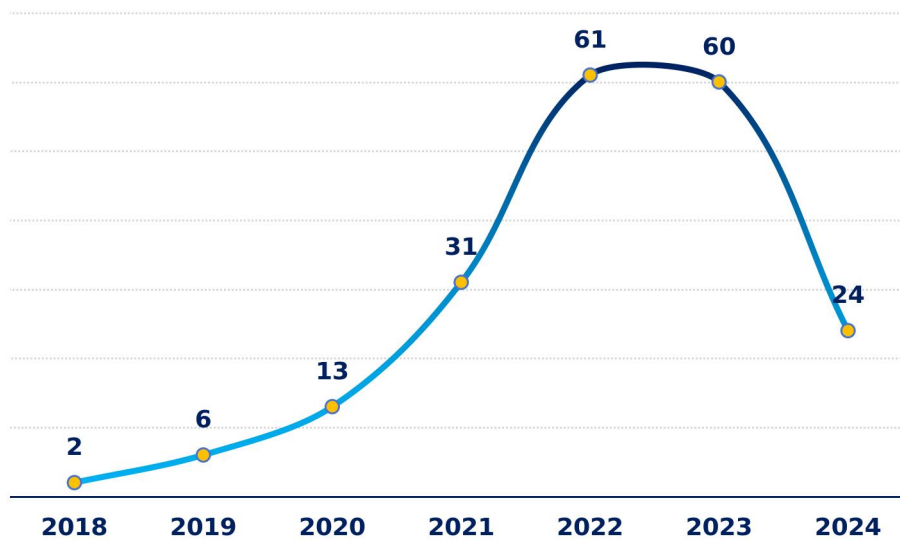


图 3-2 腾讯科技 AI 多模态融合技术专利申请趋势

从图 3-2 中可以看出,腾讯科技在 AI 多模态融合技术领域的专利申请量从 2018 到 2024 年经历了以下趋势: 2018-2021 年增长阶段: 专利申请量逐年增长, 从 2018 年的 2 件增加到 2021 年的 31 件, 显示出腾讯对该技术的持续投入和研发力度的加大。2022 年达到峰值: 2022 年申请量达到最高点 61 件, 表明腾讯在这一年对多模态融合技术的研发和专利布局达到了一个高潮。2023-2024 年下降阶段: 2023 年申请量有所下降至 60 件, 2024 年进一步降至 24 件, 一部分是由于公开信息延迟的缘故, 另外也有可能表明腾讯在技术储备逐渐成熟后, 研发重点或策略有所调整, 或者是市场需求变化导致专利申请量减少。

3) 布局情况

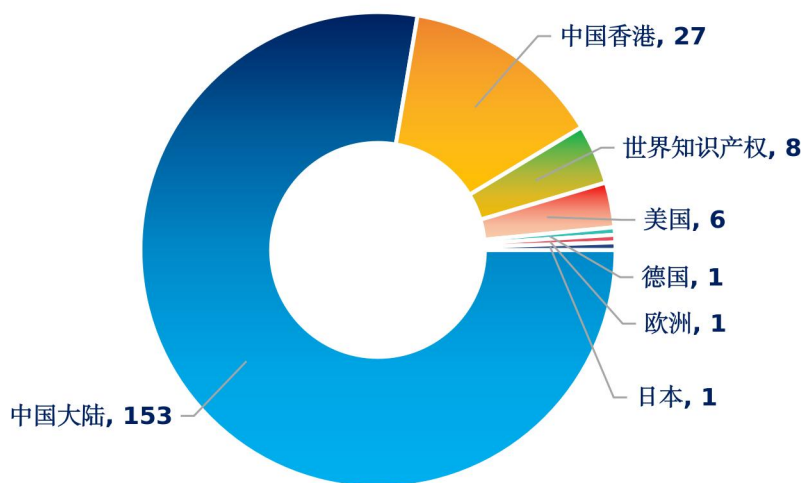


图 3-3 腾讯科技 AI 多模态融合技术各地区专利申请量

在中国大陆，腾讯的专利申请量高达 153 件，显示该地区是其核心市场与创新基地，腾讯在此投入大量资源进行研发，以专利保护推动技术应用和市场拓展。腾讯在中国香港申请了 27 件专利，借助香港的国际化地位和知识产权保护体系，作为拓展国际业务的窗口，为进入海外市场积累经验并提升国际影响力。通过世界知识产权组织（WIPO），腾讯提交了 8 件专利申请，利用其国际专利体系高效布局全球市场，扩大技术创新的国际覆盖范围，增强全球竞争力，并与国际企业开展合作交流。在美国，腾讯拥有 6 件专利，凸显美国市场在腾讯全球战略中的重要地位，通过专利保护创新，提升国际竞争力，与美国科技企业竞争合作。腾讯在欧洲、德国、日本各有 1 件专利申请，虽数量较少，但体现了其积极的全球市场拓展策略，借助专利布局突破市场壁垒，寻求新的业务增长点，同时加强与当地企业和科研机构交流与合作。

4) 技术分布

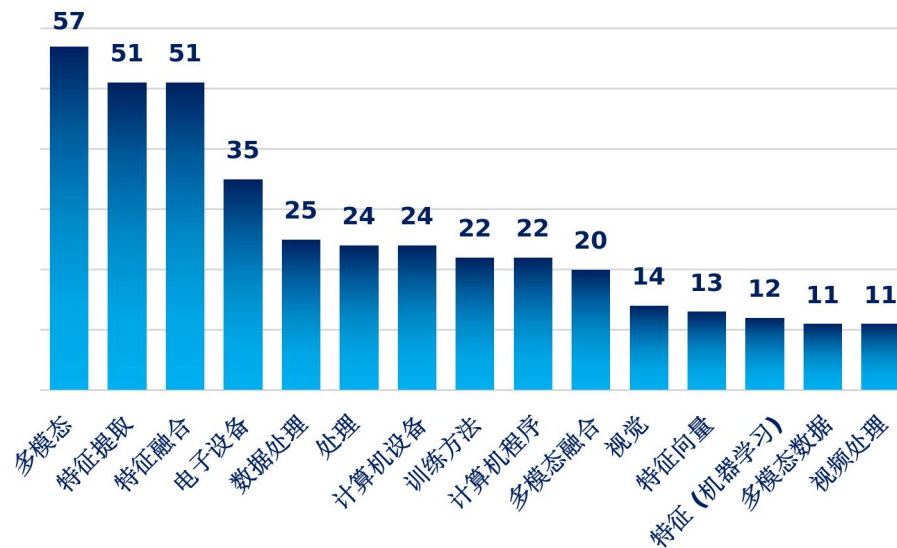


图 3-4 腾讯科技 AI 多模态融合技术主题申请量

腾讯科技在多模态融合技术的专利申请呈现多元布局,涵盖多模态(57 件)、特征提取(51 件)、特征融合(51 件)等关键领域,表明其在整合多源数据、提取与融合特征方面投入巨大,致力于提升数据处理效能。电子设备相关专利(51 件)凸显其在智能硬件与多模态交互结合上的研发深度。数据处理(35 件)及相关技术主题的专利申请则进一步展示腾讯在优化多模态数据全周期流程上的全面探索,整体彰显其推动多模态技术应用和发展的决心。

5) 发明团队

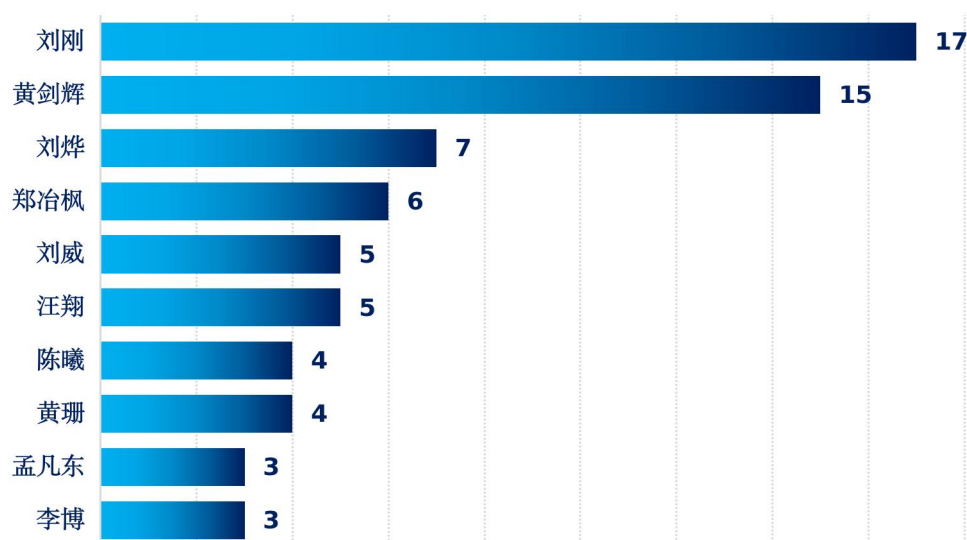


图 3-5 腾讯科技 AI 多模态融合技术发明人申请量

从图中可以看出，腾讯科技在 AI 多模态融合技术领域的发明团队中，刘刚以 17 件专利申请量位居首位，表明他在该领域的研发贡献最为突出。黄剑辉和刘烨分别以 15 件和 7 件专利申请量位列其后，显示出他们在团队中的重要贡献。郑治枫、刘威、汪翔等团队成员也有一定数量的专利申请，分别贡献了 6 到 5 件专利，这表明团队整体的研发实力较为均衡，各个成员都积极参与技术创新。陈曦、黄珊等成员的专利申请数量相对较少，但他们的贡献也为团队的整体研发工作提供了支持和补充。总体来看，腾讯科技的发明团队在多模态融合技术领域的创新动力十足，团队成员各展所长，共同推动了相关技术的发展和专利申请的积累，为腾讯在该领域的技术优势奠定了坚实的基础。

6) 技术发展动向

综合考虑腾讯科技专利相关度、简单家族专利数、引用量及所记载的技术方案情况，选取腾讯科技近年来多模态融合技术相关专利，据此构建其技术主要技术发展动向，如图 3-6 所示。

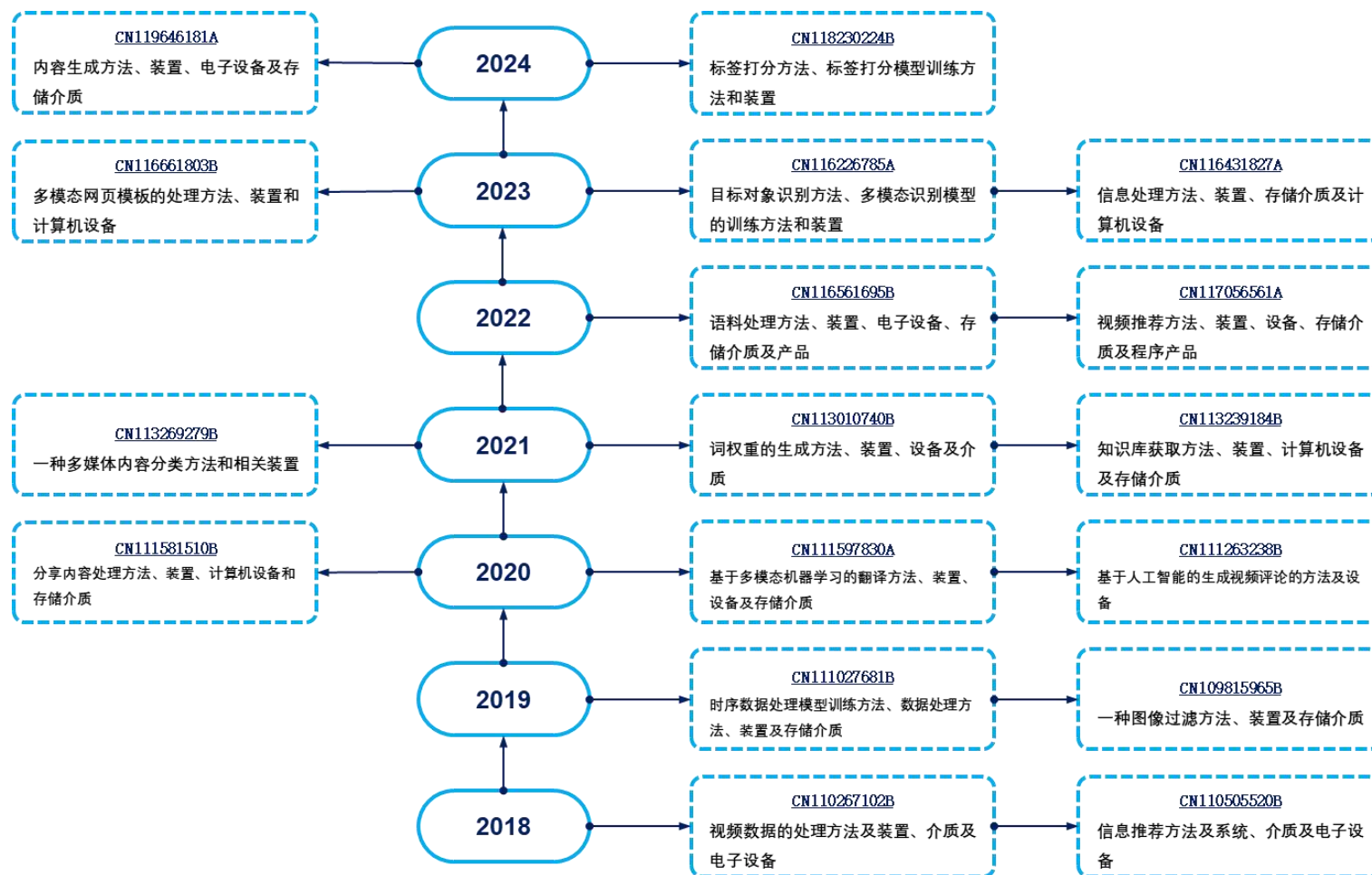


图 3-6 腾讯科技 AI 多模态融合技术动向

图 3-6 展示了近年来,腾讯科技在 AI 多模态融合技术领域主要的技术发展动向,从数据可以看出腾讯科技的模态融合技术专利申请比较频繁。

在 2018 年,腾讯科技通过 CN119646181A 公开了通过将不同模态的输入内容编码融合,并在多模态大语言模型中采用双向注意力机制进行去噪处理,解决了因果注意力掩码导致的内容生成准确性问题,实现了更高效的多模态信息处理和内容生成。同年, CN118230224B 公开了通过模态特征注入模块和聚合模块的融合技术,解决了视频标签准确率低的问题,提高了标签的准确性和召回率,适用于多模态学习和信息检索等领域。

2020 年,腾讯科技公开了 CN111597830A,该专利通过多模态机器学习方法对多模态源语言进行语义关联和融合,解决了现有技术中图片与英文注释语义关联不足的问题,提高了翻译结果的贴近性和准确性; CN111581510B 则公开了通过提取分享内容的文本和图像特征,形成多模态向量预测点击率,解决了优质内容推送难题,提高了推荐质量和效率。

2021 年,腾讯科技关于模态融合的专利申请主要集中在以下方向: CN113010740B 通过在视频搜索系统中采用多模态特征融合方法,结合视频和文本特征生成词语权重,解决了现有技术中词语权重生成不准确的问题,提高了关键词语与混淆词语的区分度和权重准确性; CN113239184B 公开了通过对信息资源的多模态数据进行特征提取和融合,解决了现有技术中知识库构建效率低下的问题,实现了更为精准和高效的知识库获取; CN113269279B 通过在多媒体内容分类模型中引入卷积参数进行卷积融合,解决了不同模态语义特征交互性差的问题,显著提高了多媒体内容的分类准确性。

2022 年,腾讯科技通过 CN116561695B,公开了通过对多个图文对样

本进行机器学习训练，调整预设模型参数，获得目标模态融合模型，解决了现有技术无法提取反映模态间特征关系的内容特征的问题，实现了更高效的内容特征提取和处理准确度。CN117056561A 通过融合被分享视频和用户的多维特征向量与候选视频的特征向量，计算相似度值，解决了现有技术中热点变化导致的视频推荐不准确问题，提高了视频推荐的准确性。

2023 年，腾讯科技公开专利 CN116226785A，通过多模态识别模型对文本和图像数据进行特征提取和融合，解决了单模态识别模型在识别准确性和效率上的不足，实现了更高的识别准确性和效率；公开专利 CN116431827A，通过在知识图谱中使用实体对齐模型进行多模态信息的对比学习和特征嵌入，解决了多模态实体对齐准确性差的问题，提高了知识图谱的表示能力和对齐效果。

2024 年，腾讯科技公开了 CN119646181A，其通过将不同模态的输入内容编码融合，并在多模态大语言模型中采用双向注意力机制进行去噪处理，解决了因果注意力掩码导致的内容生成准确性问题，实现了更高效的多模态信息处理和内容生成。而专利 CN118230224B，腾讯科技通过模态特征注入模块和聚合模块的融合技术，解决了视频标签准确率低的问题，提高了标签的准确性和召回率，适用于多模态学习和信息检索等领域。

由此可见，腾讯科技近年来在多模态融合技术领域重点布局了专利，主要集中在模态特征提取与融合、语义关联与内容生成、信息检索与推荐系统、知识图谱与对比学习、模型训练与优化等方向。通过这些专利，腾讯致力于提升多模态信息处理的准确性、效率和应用效果，推动多模态技术在内容生成、推荐系统、知识图谱构建等领域的创新发展。

3.1.2 百度科技

1) 公司简介

百度科技（全称北京百度网讯科技有限公司），成立于 2000 年，位于北京市，是一家以从事网络服务为主的 AI 公司。其创始人李彦宏拥有“超链分析”技术专利，使中国成为全球仅有的几个拥有搜索引擎核心技术的国家之一。百度业务涵盖搜索引擎、自动驾驶、智能云、AI 芯片等领域，是全球为数不多在“芯片-框架-模型-应用”四层全栈布局的人工智能公司。



图 3-7 百度 AI 多模态大模型框架图

百度的文心大模型是其在多模态融合技术领域的重要成果。文心大模型 4.5 是百度首个原生多模态大模型，实现了文本、图像和视频的混合训练。文心大模型 4.5 Turbo 进一步提升了模型的速度和性价比，多模态能力与 GPT 4.1 持平。模型采用多模态异构专家建模、自适应分辨率视觉编码等技术，大幅提升跨模态学习效率和多模态融合效果。

2) 申请趋势

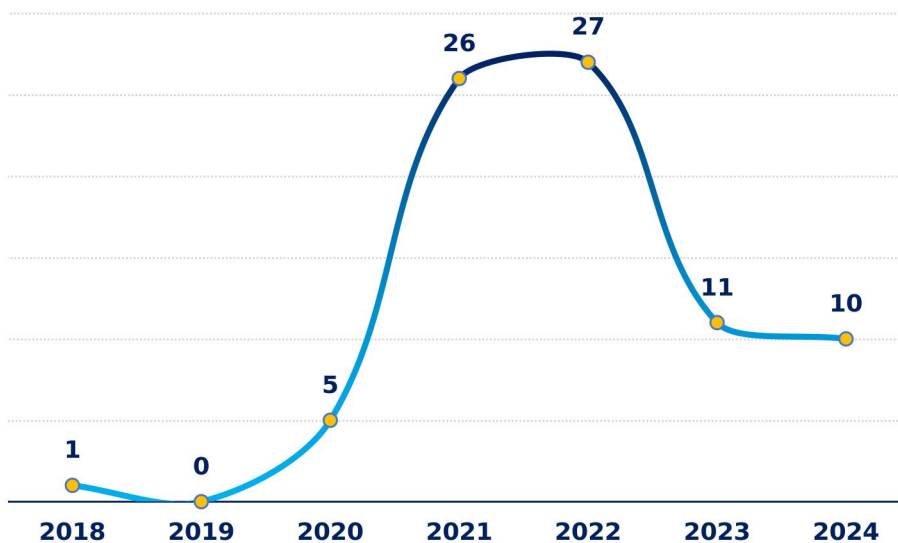


图 3-8 百度 AI 多模态融合技术专利申请趋势

结合上图 3-8，2018 年至 2020 年期间，百度的专利申请量相对较少，2018 年仅有 1 件申请，2020 年增加至 5 件。这表明在这一阶段，百度开始逐步加强对多模态技术的研发和专利布局，但整体投入相对谨慎。2021 年是百度在这个领域的专利申请量快速增长的一年，申请量达到 26 件，较 2020 年翻了 5 倍以上，这一显著增长反映了百度在多模态技术上加大了研发力度，可能与 AI 技术的整体发展以及市场需求的增加有关。2022 年百度的专利申请量达到了 27 件，维持了较高的研发活跃度，表明百度在多模态技术领域持续投入，巩固了其在该领域的技术优势。然而，到了 2023 年，申请量下降至 11 件，2024 年进一步减少至 10 件。这种下降趋势可能表明百度在多模态技术的研发重点有所调整，导致专利申请的增长速度放缓。此外，市场环境的变化、技术研发方向的转移以及公司战略的调整等因素也可能对专利申请趋势产生影响。

总体来看，百度在 AI 多模态融合技术领域的专利申请趋势显示出早期的稳步增长，中期的快速增长，以及后期的逐渐放缓。这一趋势与百度在该领域的

技术发展和产品推出节奏相吻合，也反映了其对多模态技术发展的重视和持续投入。

3) 布局情况

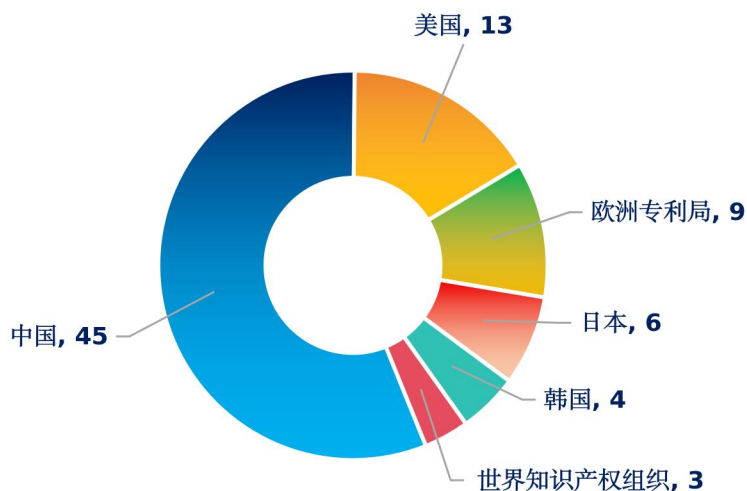


图 3-9 百度 AI 多模态融合技术各地区布局数量

在中国，百度拥有 45 件专利，占其专利布局的绝大部分。这与百度的主要市场和业务集中在中国密切相关，其在国内拥有庞大的用户基础和广泛的应用场景，如搜索引擎、Apollo 自动驾驶平台、小度智能设备等，这些业务的发展推动了多模态技术在国内的研发和应用。在美国，百度申请了 13 件专利。美国作为全球科技竞争的关键市场，百度通过专利布局保护其技术创新，为拓展北美市场及与美国科技企业竞争提供保障。百度在欧洲专利局申请了 9 件专利，显示其积极开拓欧洲市场，利用统一专利体系保护创新，为技术和产品进入欧洲多国奠定基础。

总体来看，百度的专利布局以中国为核心，同时积极拓展美国、欧洲、日本、韩国等海外市场，并借助国际组织进行全球布局。这种布局策略不仅保护

了百度的技术创新，也为百度在全球范围内的业务拓展和市场竞争提供了有力支持。

4) 技术分布

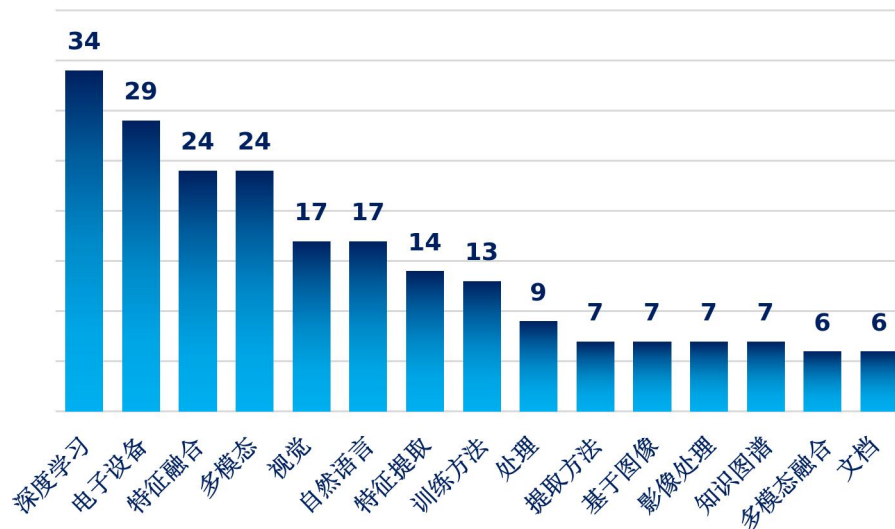


图 3-10 百度 AI 多模态融合技术各技术主题专利数量

百度在 AI 多模态融合技术领域的专利技术分布广泛，涵盖多个关键主题。其中，深度学习以 34 件专利申请量居首，显示其在基础技术研发上的重点投入，为多模态融合提供核心支撑。电子设备相关专利达 29 件，凸显百度在智能硬件领域的创新应用。特征融合与多模态主题各占 24 件，表明百度在多模态数据处理和特征整合方面研发成果显著。视觉与自然语言处理各有 17 件专利，体现百度在计算机视觉和自然语言理解领域的均衡发展。特征提取、训练方法等主题也有一定数量的专利，显示百度在多模态技术的研发和应用上全面布局，致力于推动相关技术在各领域的落地与发展。

5) 发明团队

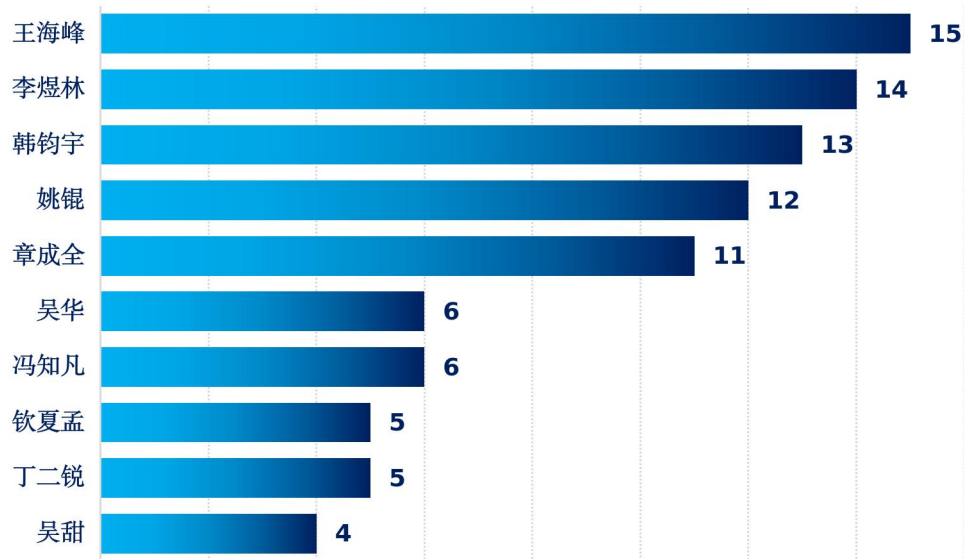


图 3-11 百度 AI 多模态融合技术发明人专利数量

图 3-11 可以看出，王海峰以 15 件专利居首，李煜林和韩钧宇分别以 14 件和 13 件专利紧随其后，三人构成核心发明力量。姚锴、章成全分别贡献 12 件和 11 件专利，是团队中的中坚力量。吴华、冯知凡各贡献 6 件专利，钦夏孟、丁二锐各有 5 件，吴甜也有 4 件。整体来看，团队成员贡献存在差异，但合作紧密，保障了技术创新活力与多样性。

6) 技术发展动向

综合考虑百度专利相关度、简单家族专利数、引用量及所记载的技术方案情况，选取百度近年来多模态融合技术相关专利，据此构建其技术主要技术发展动向，如图 3-12 所示。

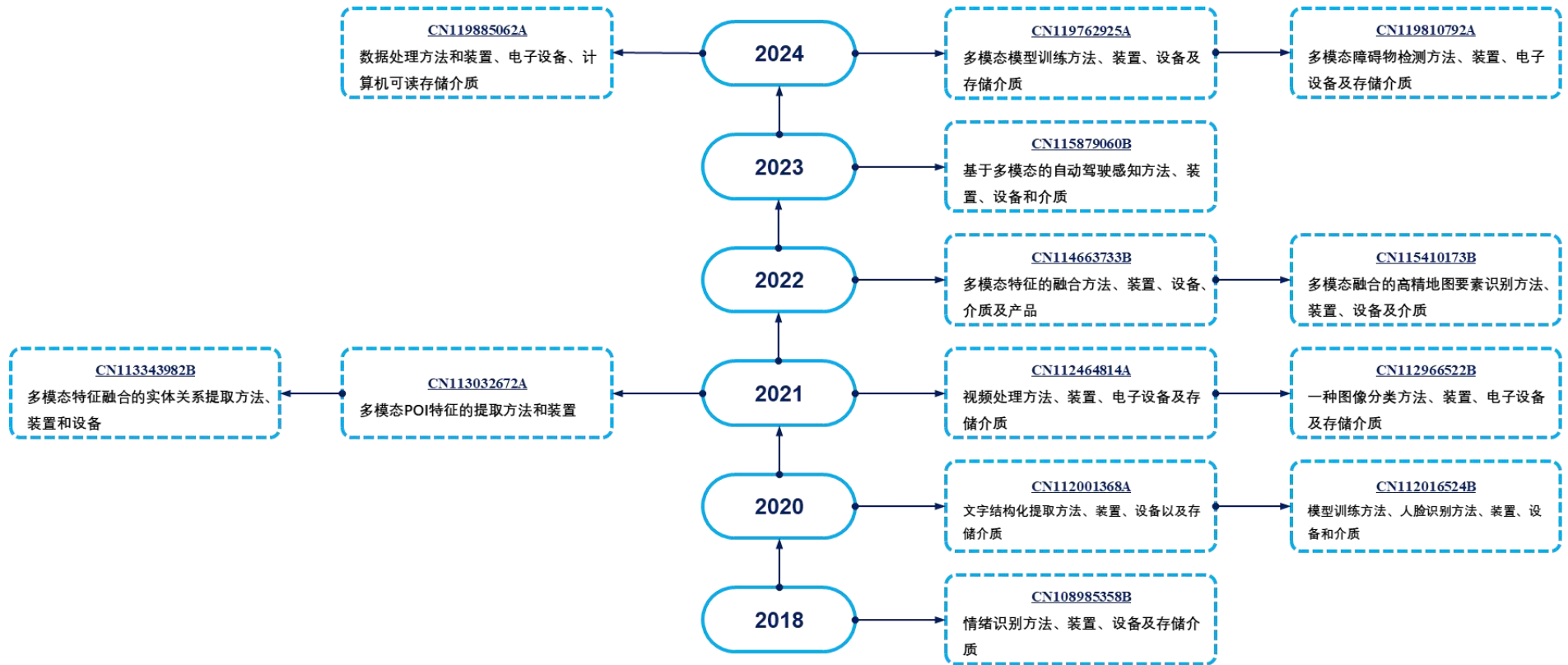


图 3-12 百度 AI 多模态融合技术发展动向

图 3-12 展示了近年来，百度在 AI 多模态融合技术领域主要的技术发展动向。

2018 年，百度公开了 CN108985358B 专利，其通过将多模态会话信息的会话特征融合并输入统一的多模态情绪识别模型，解决了现有技术中情绪识别方法通用性差和成本高的问题，提高了情绪识别的准确性和可控性。

2020 年，通过专利 CN112001368A，百度通过对实体文本图像进行文字检测、多元信息提取和特征融合，结合图神经网络进行类别关系推理，解决了现有技术中实体文本图像结构化信息提取效率低的问题，实现了高效自动化处理和广泛应用；CN112016524B 专利则通过对多模态人脸识别技术中特征融合和监督训练的方法，解决了现有 RGBD 人脸识别准确率不足的问题，实现了更高的识别准确率。

2021 年，百度申请了 CN112464814A 专利，该专利通过视频帧的细粒度拆分和多模态信息特征编码与相似度匹配，结合深度学习技术，解决了现有技术中视频处理在高时效性场景下的准确性和效率问题，实现了更高效、更准确的视频处理效果；CN112966522B 通过多模态特征融合模型融合文档图像的特征子图、语义特征和位置特征，解决了现有技术中忽略语义和位置特征的问题，提升了文档图像分类的精度并简化了处理流程；CN113032672A 通过提取和融合 POI 的多模态特征表示，解决了现有技术中 POI 相似度计算复杂度高的问题，实现了更快速和高效的相似度计算；CN113343982B 通过多模态特征融合方法，解决了文本识别准确率低的问题，通过识别图像中的视觉和文本特征，计算区域关联度，实现了高准确性的实体关系提取，适用于多种文档类型。

2022 年，百度公开了 CN114663733B，通过多头注意力机制和多层感知机对多模态特征进行交叉融合和再次融合，解决了现有技术中多模态特征拼接导致的冗余信息和低判别性问题，实现了更高效的特征融合和召回率提升；CN115410173B 通过多模态融合高精地图要素识别方法，结合点云数据与多种图像数据，解决了单一数据类型融合时识别准确性低的问题，提高了地图要素识别的准确性和自动驾驶系统的感知能力，降低了人工修正成本。

2023 年，百度的 CN115879060B 通过在自动驾驶感知系统中采用多模态感知方法，将车端和路侧的多模态信息进行全局 3D 空间的特征融合，解决了独立感知模块的信息不足问题，提升了感知精度和安全性。

2024 年，CN119762925A 披露了百度通过对多模态模型进行细化训练，利用不同分辨率的文档图像和目标掩膜区域的特征，解决了多模态模型在理解文档图像时的难题，实现了更强的感知理解能力和用户体验提升；CN119810792A 通过多模态障碍物检测方法，结合激光点云、视觉图像和毫米波点云的特征，生成障碍物检测框，解决了现有技术难以精确检测异常动态障碍物的问题，提高了检测精度和安全性；CN119885062A 通过对用户输入的多模态数据进行编码和融合，生成多模态嵌入向量，并使用 NTP 方法进行预测，解决了大型语言模型在多模态数据处理方面的 ICL 能力不足和鲁棒性差的问题，实现了更准确和符合上下文的回答数据生成。

百度在多模态融合技术领域的专利布局主要集中在情绪识别、信息提取、特征融合、视频处理、文本识别、多模态特征编码、自动驾驶感知等多个方向。通过这些专利，百度致力于提升多模态信息处理的准确性、效率和应用效果，推动多模态技术在情绪识别、信息提取、视频处理、自动驾驶等领域的创新发展。

展。

3.1.3 浙江大学

1) 学校简介

浙江大学是中国顶尖的高等学府，位于浙江省杭州市，是教育部直属的全国重点大学，也是国家“双一流”建设高校。学校前身为 1897 年创立的求是书院，1998 年由浙江大学、杭州大学、浙江农业大学、浙江医科大学合并组建新的浙江大学，学科门类齐全，在多个领域具有深厚的研究基础和强大的科研实力。

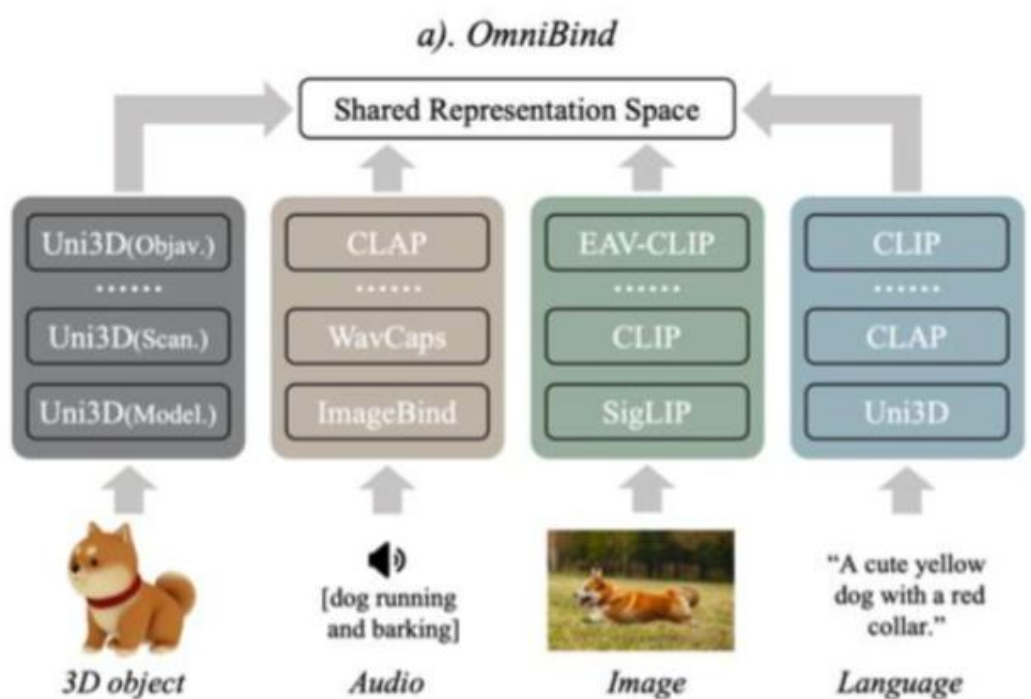


图 3-13 浙江大学 AI 多模态大模型框图

浙江大学发布了名为 OmniBind 的全能多模态大模型，支持 3D、音频、图像和文本等多种模态输入，并在 13 个主要评测基准上取得领先成绩。该模型创新性地采用“空间绑定”方法，将 14 个专业模型知识整合，参数规模达 70 亿到 300 亿不等，还设计了独特权重路由策略以融合不同来源知识，大幅提高模型性能并降低训练成本。研究团队的主要贡献者包括陈帅等，该研究

成果体现了浙大在多模态大模型开发方面的强大实力和创新能力。

2025 年 5 月，浙江大学和蚂蚁集团的研究团队在 arXiv 上发布了题为“ACTIVE-O3: Empowering Multimodal Large Language Models with Active Perception via GRPO”的研究论文，由浙江大学的申春华教授领衔，朱沐之、钟浩、赵灿宇等人共同完成。该研究为多模态大语言模型赋予了类似人类的“主动感知”能力，让 AI 系统能够主动决定“看哪里”和“怎么看”，提升了模型在复杂场景中的感知和理解能力。

浙江大学庄越挺教授团队从若干新视角对多模态表示学习进行了改进，提出了一种基于层级复合语义的深度多模态嵌入方法，以及以点击图上的随机游走为基础的多模态表示学习方法，将用户点击信息引入到多媒体信息检索中，提高了跨模态检索的质量。

2) 申请趋势

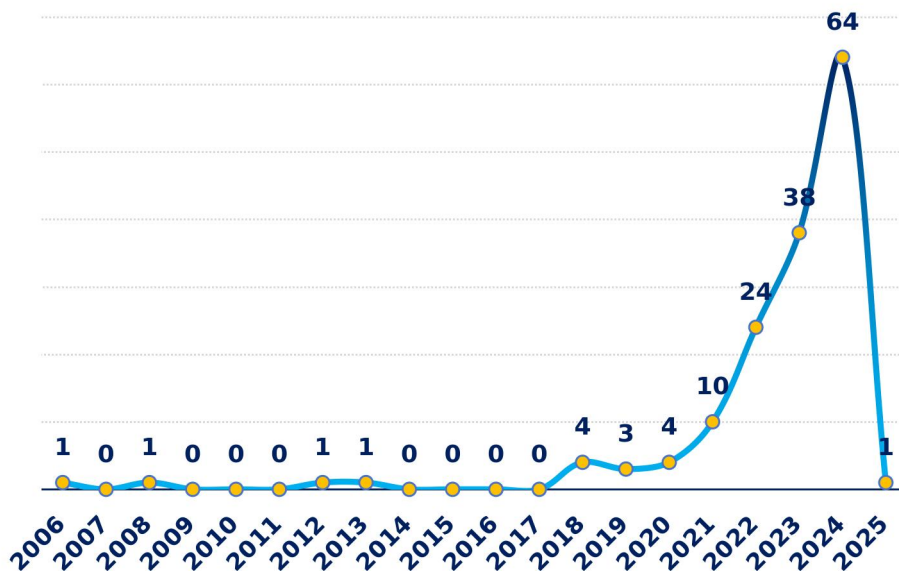


图 3-14 浙江大学 AI 多模态融合技术专利申请趋势

结合图 3-14，浙江大学在 AI 多模态融合技术领域的专利申请趋势呈现出以下特点：

2006 年至 2017 年期间，浙江大学的专利申请量相对较低，年申请量多为个位数，反映出该阶段其多模态融合技术处于研究初期，专利申请活动较少。2018 年至 2024 年期间，专利申请量显著增长，尤其在 2022 年达到峰值 64 件，这表明随着因为 AI 技术快速发展，浙江大学在该阶段加大了研发力度，多模态融合技术在各领域应用前景广阔。2023 年申请量有所回落至 33 件，2024 年降至 24 件。

总体来看，浙江大学的专利申请趋势与 AI 多模态技术的发展阶段一致，早期积累，中期爆发，逐步研究并进行创新保护。

3) 布局情况

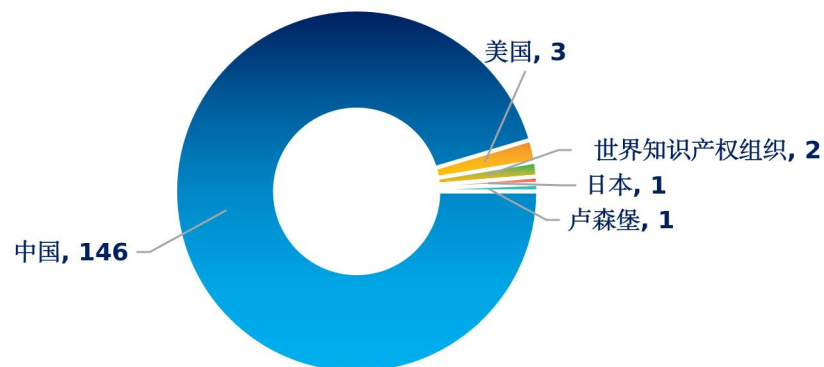


图 3-15 浙江大学 AI 多模态融合技术各地区申请量

浙江大学在 AI 多模态融合技术领域的专利布局以中国为核心，共申请 146 件专利，凸显其在国内的科研优势与市场专注度。同时，积极拓展国际布局，在美国申请 3 件、通过世界知识产权组织（WIPO）申请 2 件，彰显其国际影响力的提升。此外，日本和卢森堡各有 1 件专利申请，进一步完善全球布局。这种策略不仅保护了创新成果，更为未来的技术共研合作奠定了坚实基础。

4) 技术分布

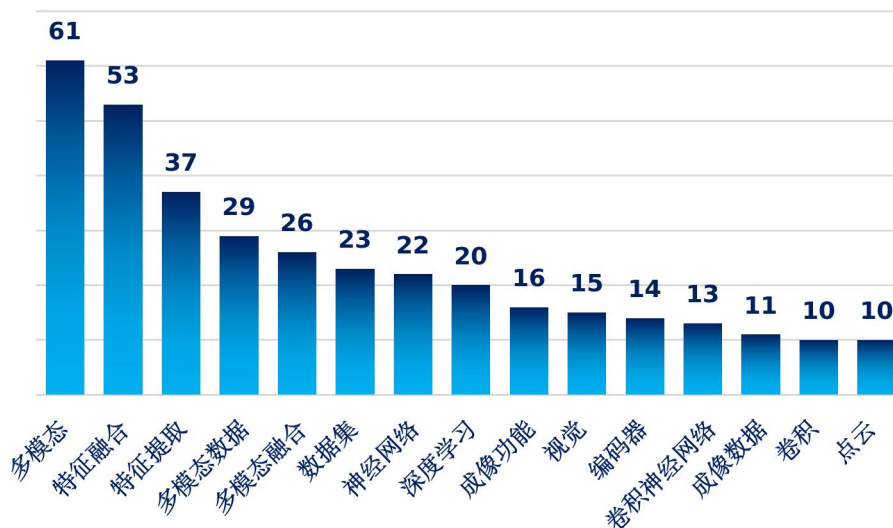


图 3-16 浙江大学 AI 多模态融合技术各技术主题申请数量

浙江大学在多模态融合技术领域的专利技术分布广泛，涵盖了多个关键主题。其中，多模态相关专利申请量最多，达到 61 件，显示了其在整合多种模态数据方面的强大研发能力。特征提取和特征融合分别以 53 件和 37 件的专利申请量紧随其后，表明浙江大学在从多模态数据中提取和融合关键特征方面进行了大量研发，以提高数据处理的准确性和效率。多模态数据和多模态融合等主题也有较高数量的专利申请，分别为 29 件和 26 件，进一步凸显了浙江大学在多模态技术领域的全面布局和深入研究。数据集和神经网络分别有 23 件和 22 件专利申请，显示了其在数据处理和模型构建方面的技术创新。此外，浙江大学在深度学习、成像功能、视觉、编码器等技术主题上也进行了专利布局，虽然申请量相对较少，但这些专利进一步丰富了其多模态技术的研发体系。总体来看，浙江大学的专利技术分布多模态融合领域的各个环节，从基础技术到具体应用场景，体现了其在该领域的广泛探索和深度投入。

5) 发明团队

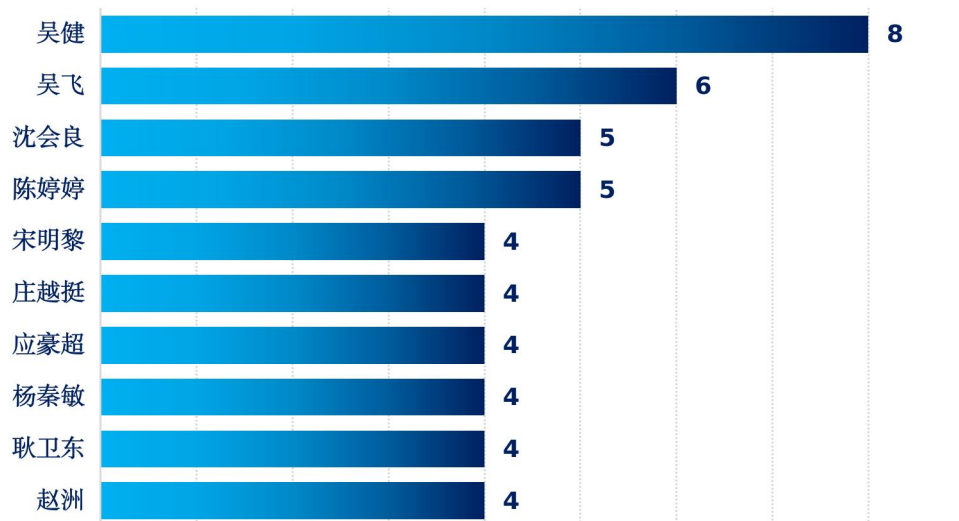


图 3-17 浙江大学 AI 多模态融合技术发明人申请量

浙江大学在 AI 多模态融合技术领域的发明团队表现活跃, 专利申请量突出。吴健以其 8 件专利申请量在团队中占据首位, 显示其在该领域的深入研究和卓越贡献。吴飞以 6 件专利申请量位居其次, 其贡献也不容小觑。沈会良、陈婷婷等团队成员分别贡献了 5 件专利申请, 显示他们在多模态技术中的积极探索与创新实践。宋明黎、庄越挺、应豪超等成员各贡献了 4 件专利, 不同课题组在多模态技术研究中广泛参与。在多模态技术的各个方向均有深入研究, 涵盖医学、自动驾驶等多个领域。这种广泛的研究布局与发明团队的多样化学科背景相呼应, 凸显了浙江大学在推动多模态技术发展上的综合实力。

整体而言, 浙江大学的发明团队凭借其多样化的研究方向和深厚的学术积累, 不断驱动多模态技术的创新与发展, 展现出该团队在科研前沿的引领地位及推动技术落地转化的潜力。

6) 技术发展动向

综合考虑浙江大学专利相关度、简单家族专利数、引用量及所记载的技术方案情况, 选取浙江大学近年来多模态融合技术相关专利, 据此构建其技术主

要技术发展动向，如图 3-18 所示。

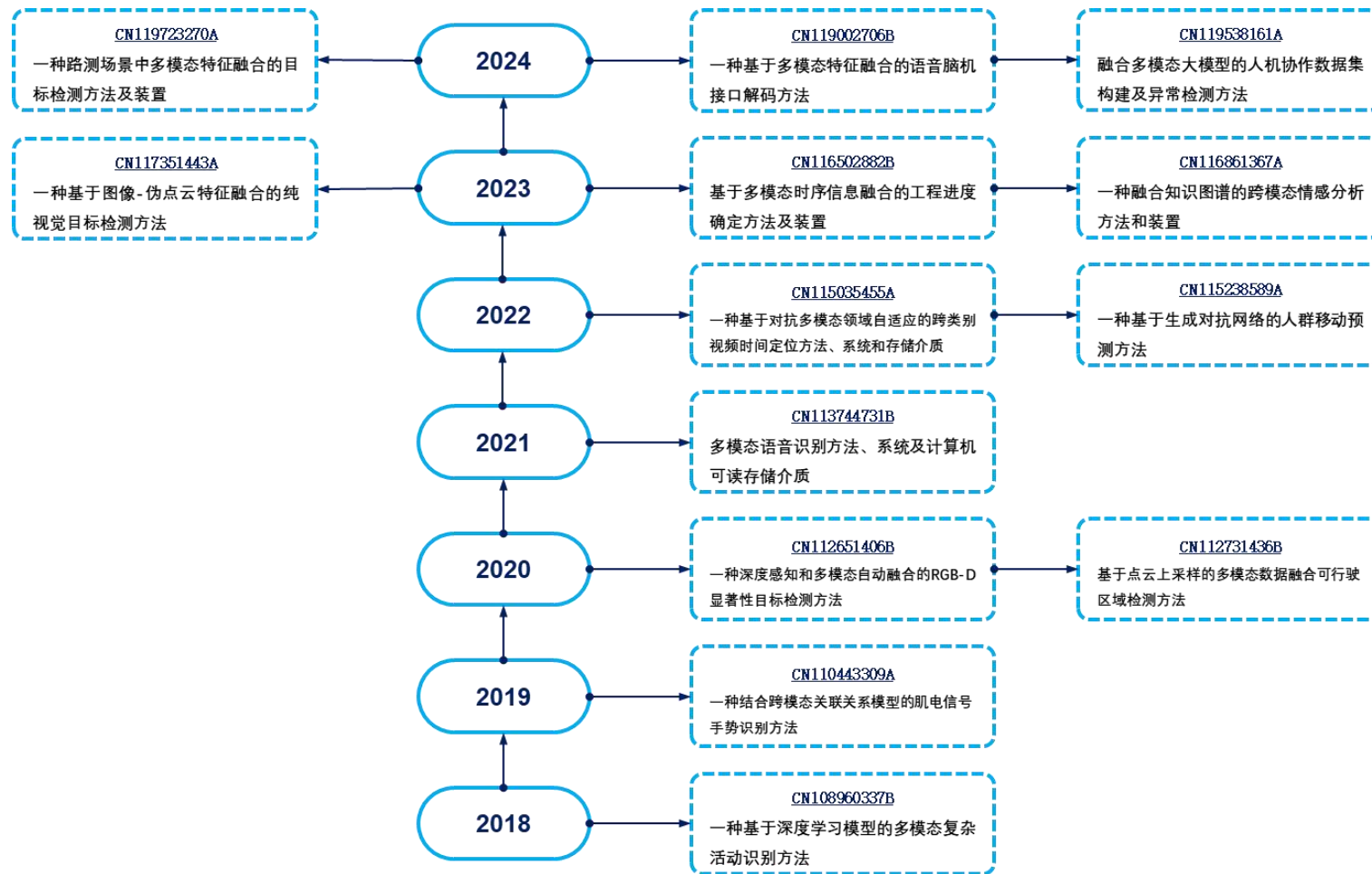


图 3-18 浙江大学 AI 多模态融合技术发展动向

图 3-18 展示了浙江大学 2018 年以来在 AI 多模态融合技术发展动向。

2018 年，通过申请专利 CN108960337B，浙江大学公开了通过深度学习模型的多模态复杂活动识别方法，利用卷积子网络和长短时记忆网络提取和融合时序数据特征，解决了现有技术中复杂活动识别准确性不足的问题，提高了识别的准确性和特征表征能力。2019 年，CN110443309A 通过构建跨模态关联关系模型和多模态融合网络，利用肌电信号与手部姿态数据的关联关系，提高了肌电信号手势识别的准确率，解决了现有技术中识别准确率不足的问题，并保持了用户体验。

2020 年，浙江大学公开了 CN112651406B，其通过深度神经网络拆解深度图像辅助彩色图像特征提取，并利用多模态自动融合的超网络结构，解决了现有技术中未能有效利用深度图像信息的问题，显著提高了 RGB-D 显著性目标检测的准确度和分割精度；同年公开了 CN112731436B，通过基于边缘强度自适应的点云上采样和多层卷积网络的多模态数据融合方法，解决了现有技术中单一传感器检测局限性和边缘恢复模糊的问题，实现了对结构化程度较差道路场景的准确检测和细节信息的保留。2021 年，浙江大学申请专利 CN113744731B，通过在语音识别技术中结合毫米波信号和音频信号的特征校准和融合，解决了噪音环境下语音识别准确性的问题，实现了在噪音和远距离条件下的高准确率识别。

2022 年，CN115035455A，浙江大学的研究团队通过对抗多模态领域自适应的方法，提取和融合视频和文本特征，解决了现有技术中视频时间定位方法耗时长和泛化性差的问题，实现了在跨类别视频数据上的良好泛化能力和性能提升；同年，CN115238589A 通过基于生成对抗网络的人群移动预测方

法，解决了疫情期间人群移动模式不可预测的问题，实现了对人群移动变化的精准预测和政策模拟，帮助政府制定科学合理的防疫措施，避免了经济和健康影响。

2023 年，CN116502882B 通过获取和融合施工现场的多模态时序信息，解决了现有技术中施工进度确定的准确性和灵活性不足的问题，实现了更准确和灵活的施工进度预测；CN116861367A 通过构建多模态知识图谱和使用预训练模型进行情感分析，解决了多模态情感分类中图像和文本特征融合困难的问题，提高了情感分类的准确性和表现能力，降低了成本；CN117351443A 通过图像-伪点云特征融合技术，解决了纯视觉模型在三维目标检测中性能较差的问题，通过伪点云特征融合技术使得纯视觉模型能够有效利用激光雷达提供的信息，显著提升了检测性能。

2024 年，浙江大学团队公开了专利 CN119002706B，其通过在神经信号解码中引入多模态特征融合方法，结合神经信号、面部肌电信号和文本序列的特征，构建多模态解码模型，解决了神经信号不稳定性和噪声干扰导致的解码性能不足问题，实现了更高效和鲁棒的文本解码效果；同年，CN119538161A 通过在虚拟人机协作场景中构建多模态数据集并利用多模态大模型进行微调训练，解决了人机协作数据集构建中数据采集效率低和潜在碰撞风险的问题，实现了高效的多模态数据集构建和虚拟人机协作场景中的异常检测；而 CN119723270A 通过在路测场景中对毫米波雷达点云进行数据增强，并结合图片和激光雷达点云的高层语义特征进行融合，解决了毫米波雷达点云稀疏和多传感器数据融合鲁棒性问题，实现了高效的 3D 目标检测。

浙江大学在多模态融合技术领域的专利布局覆盖了复杂活动识别、肌电信

号手势识别、目标检测、语音识别、视频处理和情感分析等多个方向。通过深度学习模型和创新融合方法，这些专利有效提升了多模态信息处理的准确性、效率及应用效果。

3.2 国外主要权利人

3.2.1 微软

1) 公司简介

微软（Microsoft）是全球领先的科技公司，成立于 1975 年，总部位于美国华盛顿州雷德蒙德市，主要从事计算机软件、硬件及相关服务的研发与销售。其产品涵盖操作系统（如 Windows）、办公软件（如 Microsoft Office）、云服务（如 Azure）及游戏（如 Xbox）等多个领域，业务遍布全球，拥有庞大的用户群体和强大的技术研发能力。

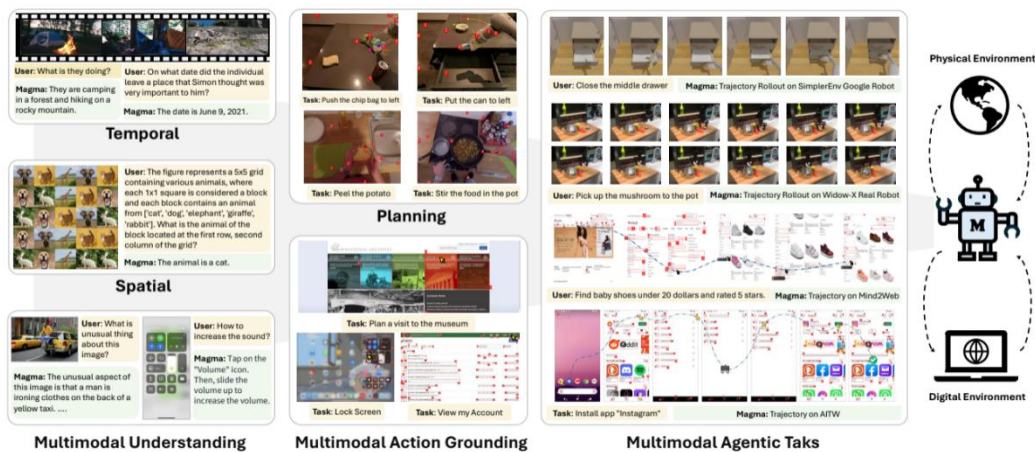


图 3-19 微软 AI 多模态大模型示意图

微软推出了全球首个多模态 AI 基础模型 Magma。该模型整合了自然语言处理、计算机视觉、语音识别等技术，能够高效处理和融合多模态数据。例如在图像和文本融合任务中，Magma 可以同时理解图像内容和相关文本描述，生成更准确和丰富的输出结果，适用于智能客服、教育、医疗等多个领域。2023 年 3 月，微软推出了多模式大型语言模型 Kosmos-1，支持处理文本、图像、音频和视频等内容。2025 年，微软发布了最新的视觉语言模型（VLM），代表了多模态技术的最新发展方向。其 Florence-2 系列模型采用多阶段、分层

式的视觉编码器架构，结合多尺度 patch 处理策略，提升了模型在复杂场景下的表现能力。微软的多模态模型采用跨模态特征编码与融合技术，有效解决了信息多源整合中的异构性问题，实现了多维度信息的互补与增强。同时，微软采用先进的 Transformer 架构和自注意力机制，优化了模型的推理效率。例如，其提出的多阶段 Transformer 结构在捕获局部细节与全局语义方面实现了平衡。

2) 申请趋势

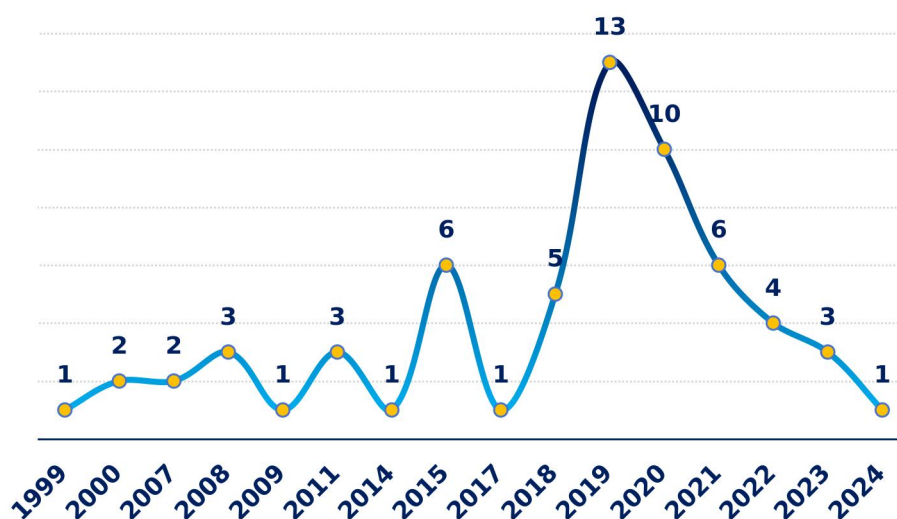


图 3-20 微软 AI 多模态融合技术专利申请趋势

微软的专利申请趋势从 1999 年至 2024 年经历了早期的稳步增长、中期的波动、快速上升期、高峰期以及后期的下降与稳定。早期（1999-2010 年）专利申请量从 1 件逐步增至 2 件；中期（2011-2015 年）波动中增长，申请量在 1-3 件间波动；2016-2018 年进入快速上升期，2016 年达 6 件，2018 年略降至 1 件；2019 年迎来高峰期，申请量达 13 件；后期（2020-2024 年）逐渐下降并趋于稳定。这一历程反映了微软在 AI 多模态融合技术领域的持续创新与战略调整，体现了其适应技术发展和市场需求的灵活性。

3) 布局情况

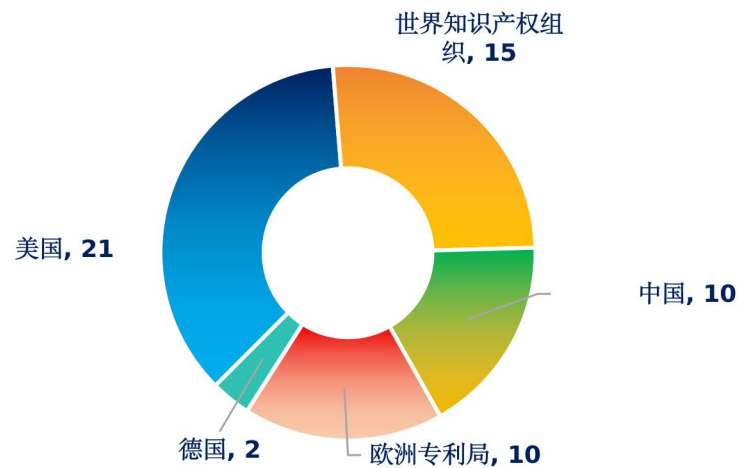


图 3-21 微软 AI 多模态融合技术各地区申请量

微软在 AI 多模态融合技术领域的专利布局以美国为核心，拥有 21 件专利申请，凸显其本土市场的研发优势。同时，积极通过世界知识产权组织（WIPO）布局 15 件专利，拓展全球市场。在中国和欧洲专利局各有 10 件专利申请，显示其对亚欧两大市场的重视。此外，德国和日本也有少量专利布局，进一步完善全球保护体系。这种全方位的专利布局策略，不仅有力保护了微软的技术创新，更为其在全球范围内的市场拓展和竞争提供了坚实支撑。

4) 技术分布

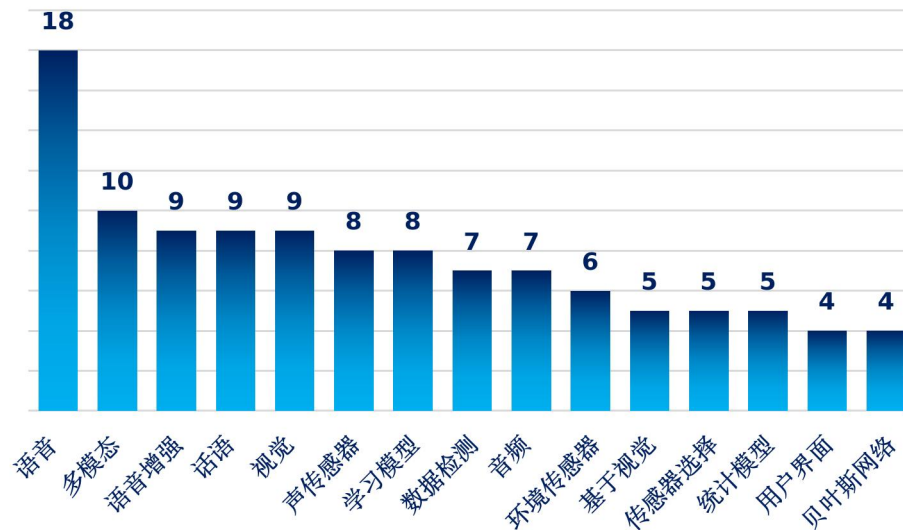


图 3-22 微软 AI 多模态融合技术各技术主题申请量

从图中可以看出，微软在 AI 多模态融合技术领域的专利技术分布广泛，涵盖多个关键主题。其中，“多模态”相关的专利申请量达到 18 件，表明微软在整合多种模态数据方面进行了重点布局。此外，“语音增强”和“语音”相关专利各有 10 件和 9 件，显示了微软在语音处理技术上的深厚积累。其他技术主题如“声传感器”“声学习模型”“数据检索”等也有较多专利申请，表明微软在多模态技术的各个方向均有深入研究和创新。总体来看，微软的专利技术分布有多模态融合领域的各个环节，从基础技术到具体应用场景，体现了其在该领域的广泛探索和深度投入。

5) 发明团队



图 3-23 微软 AI 多模态融合技术发明人申请量

从图中可以看出，微软的 AI 多模态融合技术发明团队中，Zhou Ming 以 6 件专利申请量位居首位，表明其在该领域的突出贡献和影响力。Lovitt, Andrew William 以 4 件专利申请量位列第二。Abavisani, Mahdi、Cui, Lei、Duan, Nan、Hesketh, John Benjamin、Horvitz, Eric J.、Huang, Shaohan、Iuzzolino, Michael、Ji, Lei 等团队成员均贡献了 3 件专利申请，显示了团队整体在多模态技术领域的广泛参与和积极创新。这种较为均衡的贡献模式表明微软在多模态技术的研发中注重团队协作，充分发挥各成员的专业优势，形成强大的创新合力。

6) 技术发展动向

综合考虑微软专利相关度、简单家族专利数、引用量及所记载的技术方案情况，选取微软近年来多模态融合技术相关专利，据此构建其技术主要发展动向。

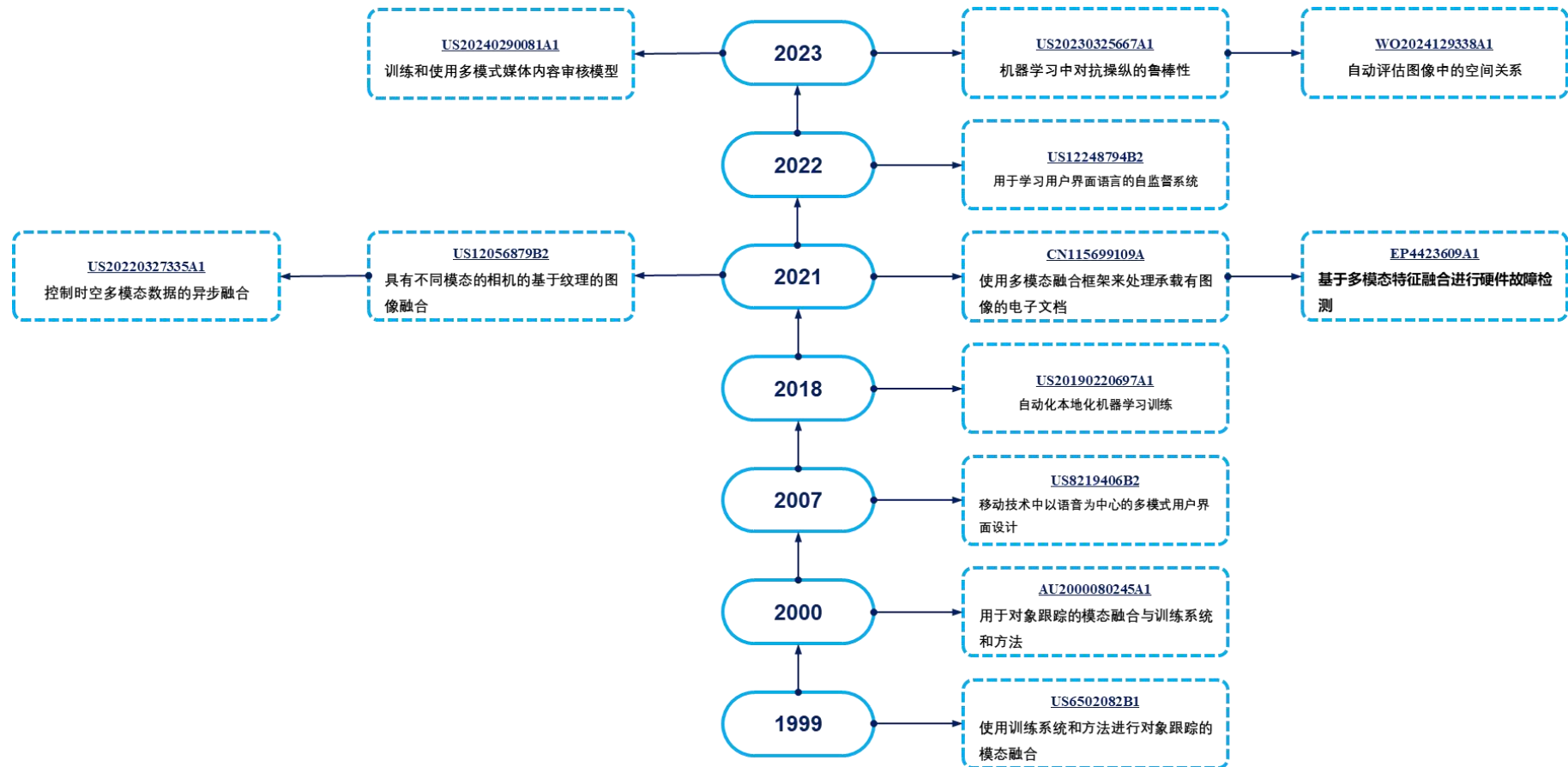


图 3-24 微软 AI 多模态融合技术发展动向

图 3-24, 可以发现, 微软在多模态融合技术的专利技术布局并未呈现一定的连续性, 除早年的先创研究布局外, 近年来并没有呈现爆发式布局。

微软在多模态融合技术最早的专利布局是在 1999 年, US6502082B1 通过系统训练贝叶斯网络自动推断和适应对象状态和跟踪模态之间的概率依赖性, 通过贝叶斯学习和模态融合增强视觉跟踪的准确性和鲁棒性, 从而解决当前跟踪系统的局限性; 2000 年, 专利申请 AU2000080245A1 采用系统训练贝叶斯网络以融合多种视觉处理方式, 通过自适应模态融合和可靠性适应来提高对象跟踪准确性和鲁棒性, 从而解决当前跟踪系统的局限性, 从而改善了头部跟踪性能。

2007 年, 微软公开专利 US8219406B2, 其采用的多模态人机交互系统通过结合多种输入模态来增强用户意图识别, 从而解决传统 GUI 和语音识别错误的局限性, 提高移动设备的交互效率和准确性; 到了 2018 年, US20190220697A1 采用双传感器设置来生成特定位置的训练数据, 增强机器学习模型在物联网设备中事件检测的准确性和适应性, 通过根据性能自动更新模型来解决低成本传感器和环境变化的局限性指标。

2021 年, CN115699109A 通过多模态融合方法, 利用神经网络和光学字符识别组件分析图像中的区域特征和外部文本, 解决了现有技术中图像对象检测不准确的问题, 实现了更高效和准确的对象名称识别; US12056879B2 通过使用基于边缘检测权重的 α 图合并来自不同相机模式的纹理, 系统生成增强的直通图像, 解决当前 MR 系统的局限性并改善用户交互和环境感知; US20220327335A1 公开了一种异步融合方法, 通过对齐空间和时间嵌入来解决数据融合中时间异步的挑战, 从而实现时空多模态数据的高效、准确的数据

融合,从而增强决策系统。

2022 年,US12248794B2 通过使用配对的 visio 语言数据训练视觉和语言模型,该方法解决了教学机器以理解和操作用户界面的挑战,从而实现了有效的 UI 表示和操作,而无需依赖人标记的数据。2023 年,微软公开了 US20230325667A1,介绍了一种方法,通过将特征向量、分类和操作向量之间的因果关系纳入机器学习模型中,解决了对抗性训练的局限性,增强了针对看不见的操作的鲁棒性,并提高了跨不同扰动的性能。US20240290081A1 则公开了多模态融合变压器模型通过生成和组合语言和视觉模态的嵌入来解决检测有害跨模态内容的挑战,通过对比损失数据训练来提高内容审核效率和准确性。

总体而言,微软的多模态融合技术专利布局呈现阶段性特点。早期专利集中于贝叶斯网络在视觉跟踪和对象识别中的应用,提升系统准确性和鲁棒性。后续专利拓展至人机交互、物联网事件检测等领域。近年来,专利涉及多模态对象检测、异步数据融合、UI 操作学习、对抗性训练增强及跨模态内容审核等方向,解决各领域挑战,提升系统性能和效率。微软的专利布局涵盖多模态融合的基础研究和应用拓展。

3.2.1 谷歌

1) 公司简介

谷歌（Google）是美国一家全球知名的跨国科技公司，由拉里·佩奇和谢尔盖·布林于 1998 年 9 月 4 日在加利福尼亚州门洛帕克创立。最初以搜索引擎项目“BackRub”起家，后凭借独特的 PageRank 算法迅速崛起，成为互联网搜索领域的领头羊。2015 年，谷歌重组为 Alphabet Inc.，成为其子公司，专注于互联网相关业务。谷歌的业务涵盖搜索引擎、广告、云计算、软件、硬件技术等多个领域，旗下拥有 Google Search、YouTube、Android 操作系统等世界级产品。



图 3-25 谷歌 AI 多模态大模型 Gemini

谷歌在人工智能大模型，尤其是多模态融合技术方面取得了显著的创新和进展。其最新的多模态大模型 Gemini 系列在多项国际权威基准测试中刷新纪录，展现出卓越的性能。谷歌还推出了多模态模型 Veo3 视频生成模型、Imagen4 图像生成模型以及 Lyria2 音乐生成模型，分别在视频、图像和音乐生成领域展现了强大的能力。此外，谷歌在搜索引擎中引入了“AI 模式”，结合

深度搜索、实时视觉识别与智能交互技术，提升了搜索体验。在产业应用方面，谷歌不断将 AI 技术融入到实际产品中，如 Project Starline 升级为 Google Beam，利用多摄像头和 AI 融合技术实现逼真的 3D 视频通话体验。

2) 申请趋势

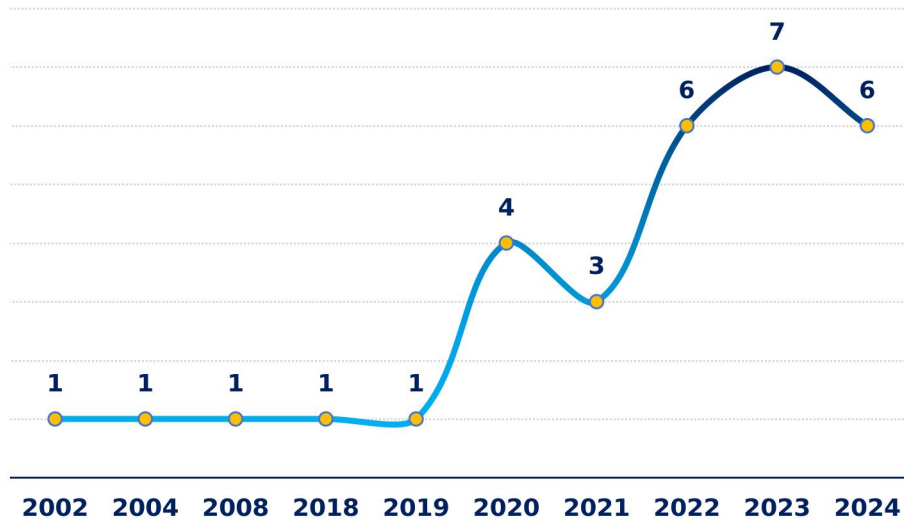


图 3-26 谷歌 AI 多模态融合技术专利申请趋势

谷歌在 2002-2018 年期间专利申请量低，年申请量多为 1 件，显示其在多模态融合技术的初步探索阶段。2019-2021 年申请量逐渐增加，2019 年 3 件，2020 年 4 件，2021 年 6 件，这与谷歌在 AI 领域的加速布局有关，如 2020 年发布的多模态模型 Veo，支持文生视频功能。2022 年申请量达 7 件，为高峰期，这与 Gemini 等重要产品的发布密切相关。2023 年申请量稳定在 6 件，2024 年保持，表明谷歌在技术应用和优化上进入新阶段。总体来看，谷歌的专利申请趋势与多模态技术产品发布的节奏相吻合，反映了其在技术研发和产品落地上的持续投入与战略调整。

3) 布局情况

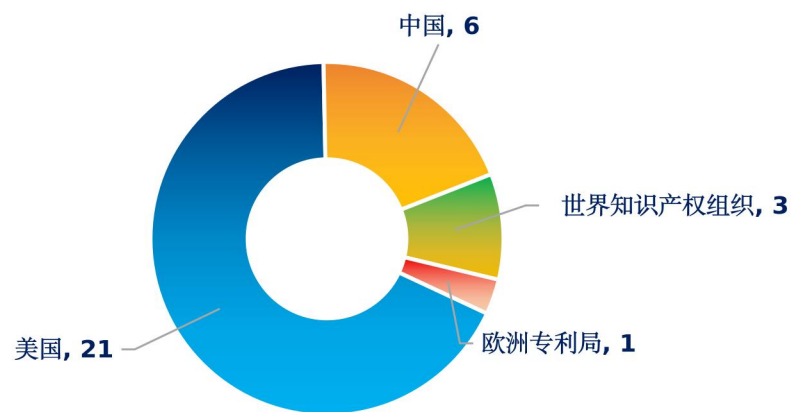


图 3-27 谷歌 AI 多模态融合技术各地区申请量

谷歌在 AI 多模态融合技术领域的专利布局以美国为核心，共申请 21 件专利，凸显其在本土市场的强劲研发实力与技术保护力度。在中国申请 6 件专利，显示对中国市场的重视及推动技术落地的决心。通过世界知识产权组织（WIPO）申请 3 件专利，借助国际专利体系拓展全球市场，提升国际竞争力。此外，向欧洲专利局申请 1 件专利，初步涉足欧洲市场。总体而言，谷歌的专利布局聚焦美国，同时积极开拓中国及全球市场，有力支撑其全球技术竞争力与市场拓展战略。

4) 技术分布

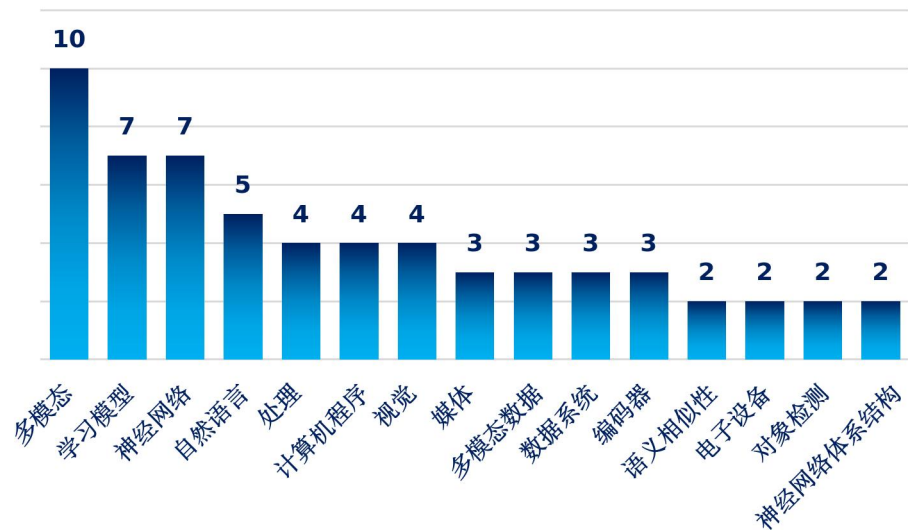


图 3-28 谷歌 AI 多模态融合技术各技术主题申请量

谷歌在 AI 多模态融合技术领域的专利布局呈现明显的层次分布，重点突出且覆盖广泛。多模态以 10 项专利居首，是核心方向，表明谷歌注重构建通用技术框架，实现多模态信息融合处理，助力智能助手、无人驾驶等多场景应用。学习模型和神经网络各有 7 项专利，前者关乎设计优化模型架构以适应多模态数据特点，提升学习效率准确性，后者侧重探索高效稳定神经网络架构处理复杂多模态数据，如在图像语音识别中的应用。处理计算机程序、视觉、媒体各 4 项专利，分别聚焦多模态数据处理交互的算法架构、图像目标检测与视频分析等视觉信息提取关联融合、以及多媒体内容生成编辑管理，将多模态技术应用于媒体产品创新。多模态数据、数据系统、编解码、语义相似性各 3 项专利，分别涉及数据特点采集存储预处理、数据管理检索服务、编解码方案优化传输存储效率，以及挖掘不同模态数据语义关联，为融合语义理解奠基。电子设备、对象检测、神经网络体系结构各 2 项专利，分别关注多模态技术在智能设备中的应用拓展、复杂场景下精准对象检测，以及新型神经网络架构探索以提升多模态融合任务性能。

总体而言，谷歌的专利布局在核心多模态技术及相关学习模型、神经网络等方面重点投入研发，同时积极拓展多个具体技术方向和应用场景，有力巩固其在该领域的技术领先地位。

5) 发明团队



图 3-29 谷歌 AI 多模态融合技术发明人申请量

谷歌在 AI 多模态融合技术领域拥有一支多产且多元的发明团队。其中，MIAO, HUI、PIERGIOVANNI, ANTHONY J. 和 ANASTASAKOS, TASOS 等三位发明人各有 3 项专利，是该领域的核心贡献者，表明他们在技术研发和创新方面发挥了重要作用。同时，还有多位发明人如 ANGELOVA, ANELIA、ARNAB, ANURAG 等各有 2 项专利，反映出该领域研发力量的广泛参与和团队协作的高效性。

整体来看，谷歌的多模态融合技术能够取得领先地位，这些发明人的创新成果为谷歌在该领域的技术进步和产品应用提供了有力支持。

6) 技术发展动向

综合考虑 Google 专利相关度、简单家族专利数、引用量及所记载的技术

方案情况，选取近年来多模态融合技术相关专利，据此构建其技术主要发展动向。

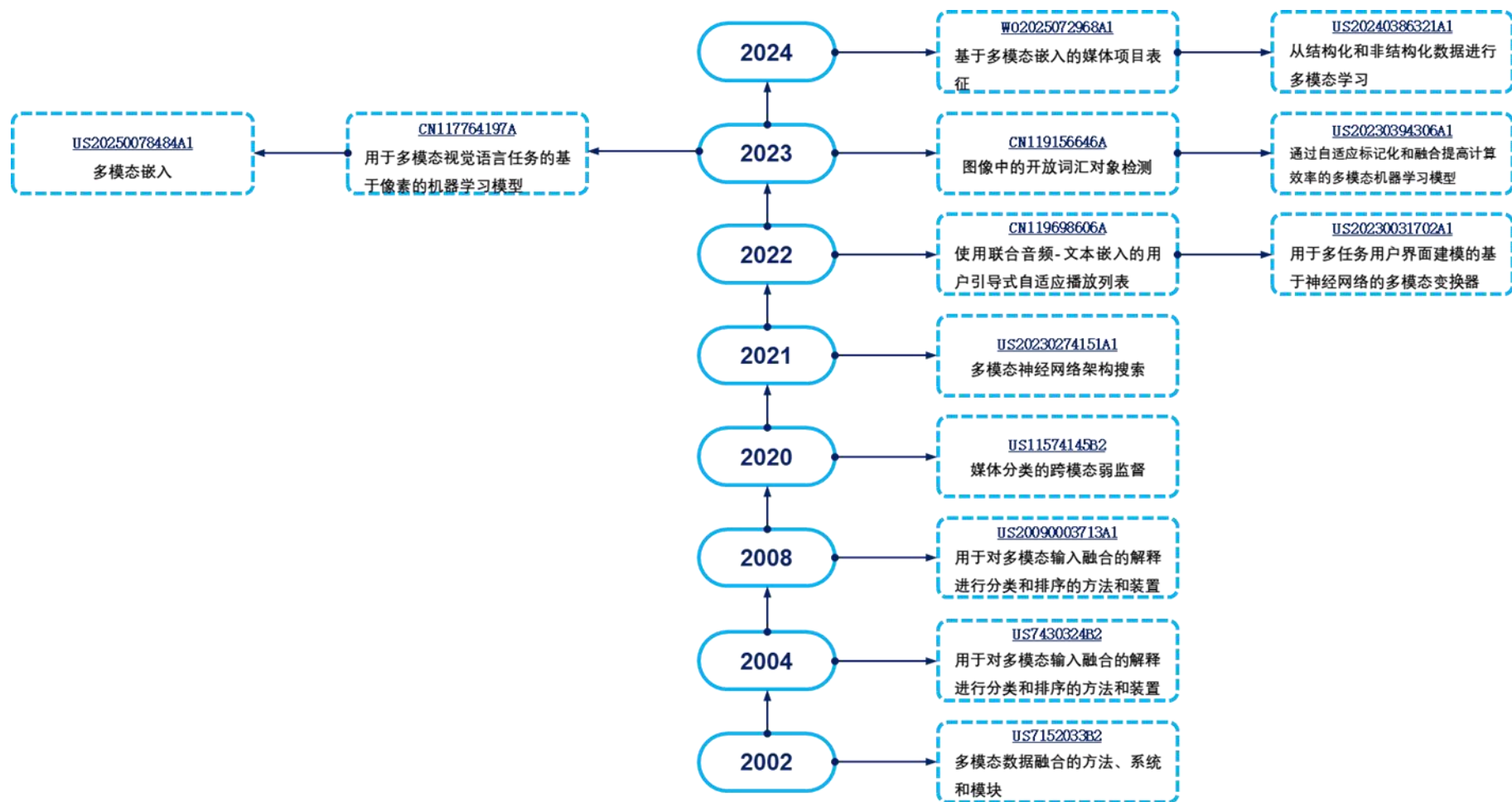


图 3-30 谷歌 AI 多模态融合技术发展动向

图 3-30 展示了谷歌 AI 多模态融合技术发展动向。其中：

2002 年，谷歌公开了 US7152033B2，采用了基于统计数据分析，在多模式数据融合系统中实现动态可变等待期，解决了静态等待期的低效率问题，减少了延迟并提高了不同输入模式的响应准确性。2004 年，US7430324B2 采用了使用语义分类器和交互管理器生成具有多级置信度得分的联合解释来解决处理同时多模态输入的挑战，从而提高多模态系统中输入处理的可靠性和效率。

2008 年，US20090003713A1 公开了该系统通过使用语义分类器和交互管理器生成具有多级置信度得分的联合解释，解决了处理多模态输入的低效率问题，从而提高了多模态系统中输入处理的可靠性和速度。

2020 年，谷歌披露了 US11574145B2，其弱监督机器学习技术利用现有模型自动对媒体格式的内容进行分类，减少了手动标记的需要并提高了内容分类的效率。US20230274151A1 公开了一种方法，该方法通过联合优化多模态融合和模型架构，优化多模态任务的神经网络架构，解决现有技术的局限性，并在各种多模态任务(特别是电子健康记录数据)上实现卓越的性能。

2022 年，CN119698606A 通过在音乐推荐系统中分析用户的收听/跳过行为和自然语言输入，使用双塔模型生成用户偏好的表示，训练机器学习模型生成更新的播放列表，解决了现有系统难以即时自适应用户心情的问题，实现了更高效的音乐推荐；US20230031702A1 公开了 VUT 模型通过集成图像结构和语言转换器来解决执行多个图形用户界面任务的资源密集型问题，从而实现高效、准确的多模态处理，从而增强移动用户体验。

2023 年，CN119156646A 通过对象检测神经网络的图像编码子网络和文本编码子网络的联合训练，生成对象嵌入和查询嵌入，解决了现有系统难以

检测开放词汇对象类别的问题，实现了高效的对象检测和推理能力；

US20230394306A1

公开了多模态处理模型中的自适应标记化层有效地减少了特征数据大小并提高了可扩展性,通过实现更快、更经济高效的多模态数据处理来解决当前模型中的低效率问题,特别是对于视觉问答等任务; CN117764197A 通过渲染文本内容为图像并使用基于像素的机器学习编码模型处理图像和文本内容，解决了现有技术中多模态视觉语言任务中计算资源高耗和文本处理复杂的问题，实现了更高效和准确的多模态任务处理;US20250078484A1 通过使用 LLMS 和 VLMS 来处理和生成视觉和音频数据中的多模式嵌入,可以解决现有的自动化助手将视觉数据纳入对话中的局限性,从而改善参与度和隐私,同时减少资源使用情况。

2024 年，谷歌公开了 WO2025072968A1，其通过使用视频和音频编码器生成的嵌入来表征媒体项目,该平台有效地检测了媒体趋势并评估媒体质量，从而解决了现有方法的资源密集型挑战。同年，US20240386321A1 通过多模式处理系统通过使用基于模态和基于相似性的掩盖目标来解决跨越缺失模式的挑战,从而提高了处理多种数据组合的准确性和效率,包括丢失数据的高百分比。

谷歌在多模态融合技术领域的专利布局从早期的基于统计分析的多模式数据融合系统,发展到近期的弱监督机器学习技术和自适应标记化层等创新。这些专利涵盖了提高多模态系统输入处理的可靠性与效率、优化神经网络架构、增强移动用户体验、改进对象检测和推理能力、以及更高效地处理多模态数据等多个方面。谷歌的专利布局致力于解决多模态融合中的各种挑战，如静态等

待期的低效率、多模态输入处理的复杂性、资源密集型任务处理等，推动了多模态技术在不同应用场景中的发展和应用。

3.3 主要权利人对比分析

通过对国内外重点专利权人的专利布局进行宏观分析、技术分析、代表性专利分析等，已经得到它们的专利申请量变化趋势、总体区域分布、代表性专利等具有丰富内容的信息，本节将对这些信息进行综合对比整理，以便进一步掌握目前 AI 多模态融合技术领域厂商、技术、市场等要素的配置状况。

表 3-1 AI 多模态融合技术重点权利人对比

项目	腾讯	浙江大学	百度	微软	谷歌
总部	中国	中国	中国	美国	美国
主要布局地区	CN/HK/WO	CN/WO	CN/US/EP	US/WO/CN	US/CN/WO
快速增长期	2020-2023	2020-2024	2020-2022	2014-2020	2019-2024
相关专利总量	185	153	81	64	31
近 5 年专利量	177	140	79	24	26
技术活跃度 (近 5 年/总计)	95.6%	91.5%	97.5%	37.5%	83.8%
法律状态/件	有效 71 在审 92	有效 47 在审 93	有效 25 在审 34	有效 29 在审 8	有效 5 在审 17
主要目标市场	中/美	中/美	中/美/欧	美/中/欧	美/中/欧
CN 布局量 (件)	140	146	45	10	6
主要技术布局	特征提取 特征融合	特征融合 特征提取	特征提取 特征融合	特征提取	数据融合

通过对 AI 多模态融合技术领域重点权利人的专利情况进行对比，可以发现以下特点：腾讯、浙江大学和百度作为中国主体，专利主要布局在中国，此外腾讯的布局区域还涉及香港，百度则扩展到美国和欧洲；微软和谷歌作为美国主体，主要布局以美国为基础，同时覆盖和中国，谷歌还额外关注欧洲。这样表明，上述公司的战略也将优先考虑在上述对应市场进行积极布局。

技术的快速增长方面，上述五家有共同点，也存在差异：腾讯为 2020 - 2023 年，浙江大学持续至 2024 年，百度为 2020 - 2022 年，微软为 2014 - 2020 年，谷歌从 2019 年持续至 2024 年。

而专利总量上，腾讯得申请总量最多，达 185 件；浙江大学 153 件，百度 81 件，微软 64 件，谷歌 31 件；近 5 年专利量腾讯 177 件，浙江大学 140 件，百度 79 件，微软 24 件，谷歌 26 件。结合近 5 年的技术活跃度来看，百度以 97.5% 居首，腾讯 95.6%，浙江大学 91.5%，谷歌 83.8%，微软 37.5%。专利的法律状态方面，腾讯有效 71 件、在审 92 件，浙江大学有效 47 件、在审 93 件，百度有效 25 件、在审 34 件，微软有效 29 件、在审 8 件，谷歌有效 5 件、在审 17 件。由此可见，国内主要权利人将在有效专利梳理方面会有进一步的突破，而相反的是，美国两家权利人在未来几年内能够转化授权的专利则相对较少。

国内的主要权利人腾讯和百度注重在特征提取和融合方面进行技术布局，浙江大学强调特征融合与提取；而微软侧重特征提取，谷歌关注数据融合等技术。

3.4 本章小结

本章主要聚焦于 AI 多模态融合技术领域的主要权利人及其专利布局情况，通过深入剖析国内外具有代表性的企业及高校，揭示了各主体在技术研发、专利申请及市场竞争力方面的特点与优势。

对于国内主要权利人，腾讯科技在专利申请量上占据领先地位，其专利布局广泛，主要集中在模态特征提取与融合、语义关联与内容生成、信息检索与推荐系统、知识图谱与对比学习、模型训练与优化等方向。百度则以文心大模

型为核心，积极拓展多模态技术在智能云服务、自动驾驶等领域的应用，专利申请趋势呈现稳步增长，专利主要布局主要集中在情绪识别、信息提取、特征融合、视频处理、文本识别、多模态特征编码等方向。浙江大学作为顶尖高校，在多模态融合技术领域科研成果丰硕，专利申请量位居国内高校首位，其 **OmniBind** 大模型等成果在多模态大模型开发方面具有显著优势，专利布局覆盖了复杂活动识别、肌电信号手势识别、目标检测、语音识别、视频处理和情感分析等多个方向，体现了高校在基础研究和前沿探索中的关键作用。

在分析国外主要权利人时，微软和谷歌作为全球科技巨头，其在多模态融合技术领域的专利布局和研发成果具有显著影响力。微软推出的多模态 AI 基础模型 **Magma** 以及谷歌的 **Gemini** 系列大模型等，均体现了其在该领域的技术实力和创新能力，且在全球范围内享有较高的市场认可度。

中国主体在专利申请总量及近期活跃度方面表现突出，腾讯、浙江大学和百度凭借大量的专利申请及较高的技术活跃度，在国内市场占据优势，并逐步向海外拓展，未来有望进一步巩固和扩大其在国内外市场的影响力。美国的微软和谷歌虽在专利布局上覆盖中国和欧洲市场，但近期专利申请量和活跃度相对较低，未来需加大创新投入以提升竞争力。中国主体在特征提取与融合技术上布局集中，而美国主体则侧重数据融合等差异化技术方向，预示着未来双方将在技术路径和市场应用上展开多元竞争。总体来看，中国主体凭借数量优势和较高的技术活跃度，在未来竞争中具有较强的潜力，而美国主体则需通过技术创新和专利转化来维持其市场地位。

整体来看，本章通过对各主要权利人的详细分析，展现了 AI 多模态融合技术领域不同主体的创新实力和发展动态，也反映出国内企业及高校在技术研发和专利布局方面取得的显著进步，以及在国际市场上的竞争力逐步提升。

第四章 AI 多模态融合技术专题分析之模态融合

模态融合技术是多模态应用中的关键环节，它通过整合多种模态的数据、特征或决策，打破了单一模态的局限，显著提升了系统对复杂环境的感知与决策能力。无论是智能驾驶中融合雷达与摄像头数据以精准识别路况，还是医疗诊断里结合影像与文本病历全方位判断病情，模态融合都能最大化各模态优势。本章将对模态融合技术相关专利概况进行分析，随后对模态融合相关技术进行主要技术简介，每个相关技术后将挑选的重点专利进行解读，将深入剖析模态融合的相关专利，从数据级、特征级到决策级融合的创新与应用。本章挑选重点专利时，主要依据重点专利权人近十年重点专利及关键技术代表专利、专利同族数、同族专利被引证数，兼顾专利法律状态等进行挑选和解读。

4.1 专利总量

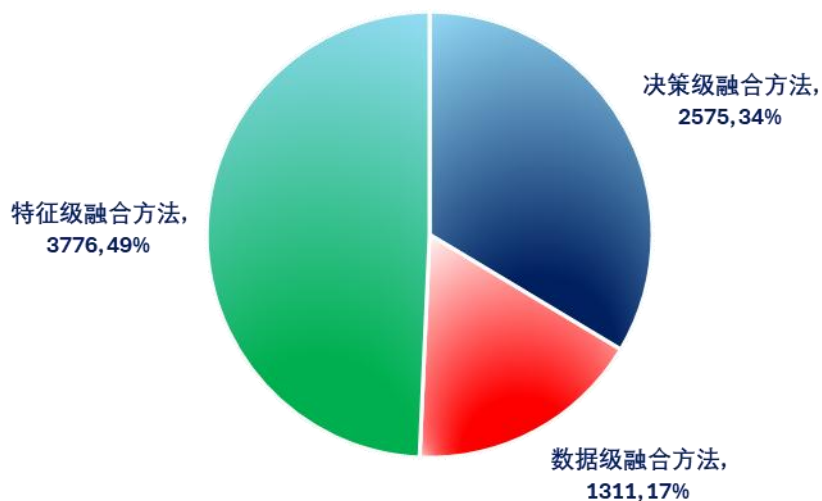


图 4-1 模态融合各分技术专利总量分布

图 4-1 展示了模态融合各分技术专利总量的分布情况。其中，特征级融合方法的专利总量最多，为 3776 件，占总量的 49%；决策级融合方法次之，专利

总量 2575 件，占比 34%；数据级融合方法的专利总量相对较少，为 1311 件，占比 17%。这表明在模态融合技术领域，特征级融合方法是当前研发和专利布局的重点方向，而决策级融合方法也占据了相当的比例，数据级融合方法虽然相对较少，但也有一定的专利积累。

4.2 专利趋势

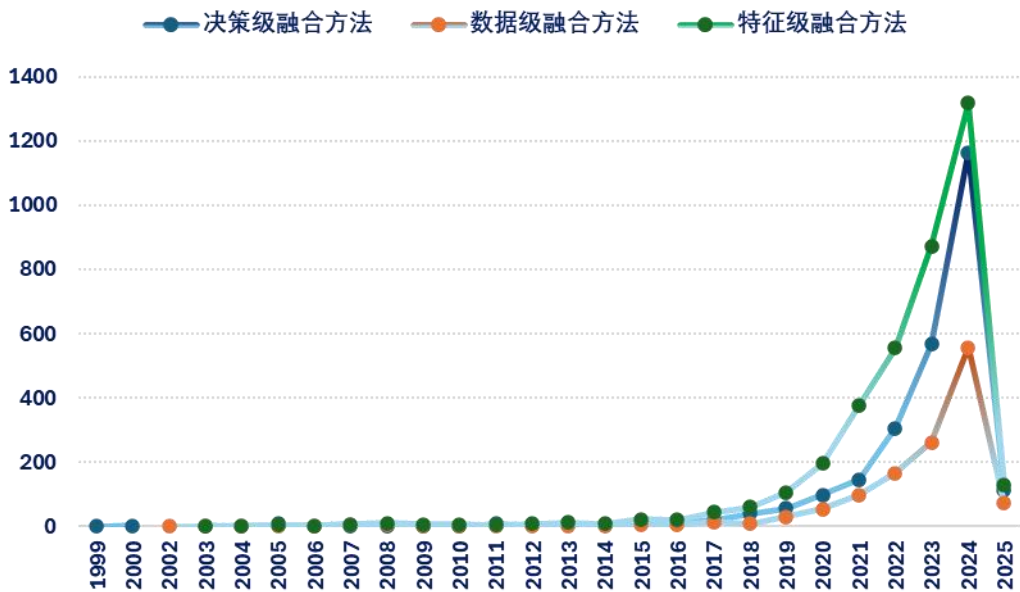


图 4-2 模态融合各分技术专利申请趋势

总体上来看，三种融合方法的专利申请量在早期（1999–2015 年）都处于较低水平，增长缓慢。从 2016 年开始，专利申请量显著增加，尤其是在 2018–2024 年期间，呈现出快速上升的趋势。2024 年后，由于公开信息延迟的原因，专利申请量迅速下降。特征级融合方法：在 2018 年后增长最为显著，2021–2023 年达到峰值，表明其在技术发展中的重要性不断提升。决策级融合方法：增长趋势与特征级融合方法相似。数据级融合方法：虽然整体申请量较少，但在 2018–2024 年期间也有明显的增长，尤其是在 2024 年达到峰值。

因此，模态融合技术在 2016–2024 年期间经历了快速的技术创新和专利申请增长，特征级融合方法和决策级融合方法是主要研发方向。

4.3 法律状态

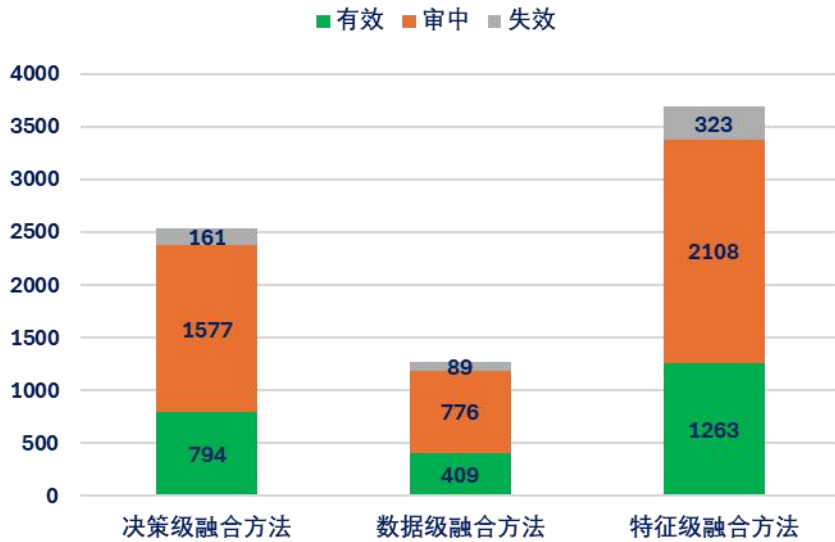


图 4-3 模态融合各分技术专利法律状态分布

图 4-4 揭示了模态融合各分技术专利的法律状态分布。特征级融合方法的有效专利最多，达 1263 件，审中专利 2108 件，失效 323 件。决策级融合方法有 794 件有效专利，1577 件审中，161 件失效。数据级融合方法有效 409 件，审中 776 件，失效 89 件。总体来看，特征级融合方法在有效专利和审中专利上均占较大比例，表明其技术成熟度较高且仍有大量专利处于审查阶段，未来可能进一步巩固其优势。决策级融合方法的审中专利数量也较多，显示出持续的研发投入。数据级融合方法的专利数量相对较少，但审中专利占比可观，表明其技术发展正在推进中。失效专利方面，特征级融合方法的失效数量最多，可能是因为其专利申请量大，部分专利因各种原因未能维持有效。

4.4 区域分布

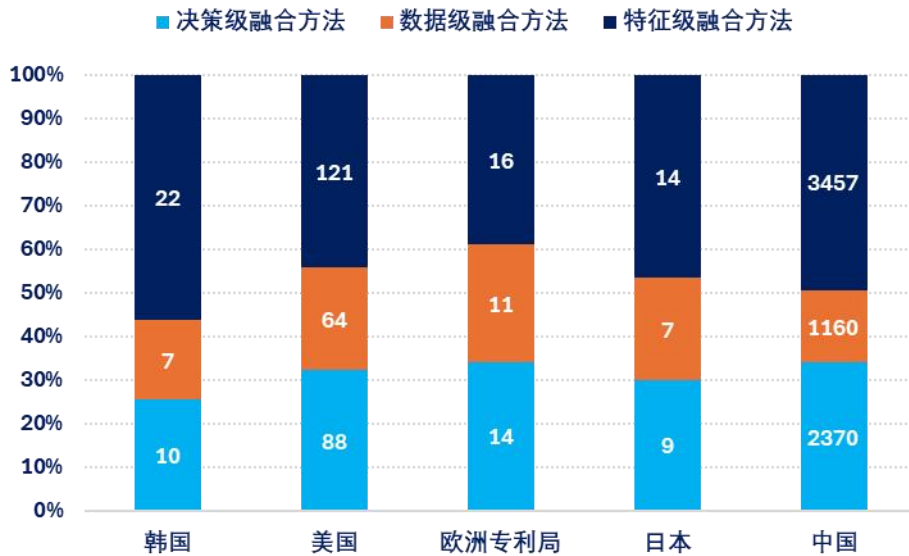


图 4-4 五大局模态融合各分技术专利占比情况

图 4-4 展示了五大局模态融合各分技术专利占比情况。在中国，特征级融合方法专利数量最多，达 3457 件，占比超 50%，决策级和数据级分别为 2370 和 1160 件。美国以特征级融合专利为主，占比超 50%，数量为 121 件，决策级和数据级为 88 和 64 件。欧洲专利局的特征级和数据级融合专利占比较高，分别为 16 和 14 件。日本的特征级融合专利占比最高，为 14 件，决策级和数据级为 9 和 7 件。韩国的特征级融合专利占比最高，为 22 件，决策级和数据级为 10 和 7 件。总体来看，特征级融合方法在各大受理局模态融合技术中占比较高，决策级融合方法次之，数据级融合方法在模态融合技术的占比则相对较低。

第五章 AI 多模态融合技术专题分析之模态对齐

模态对齐技术旨在精准匹配不同模态数据的语义，是多模态融合中至关重要的一步。它有效解决了各模态间语义差异大、对应关系难把握的问题，为多模态模型训练和应用提供基础，让模型更高效地挖掘多模态数据的深层关联。模态对齐分为显式对齐和隐式对齐。显式对齐依靠标注信息或数据特性直接建立对应关系，如监督对齐利用标注数据学习映射、无监督对齐挖掘数据内在结构；隐式对齐则借助注意力机制或语义空间，动态捕捉模态间关。本章将对模态对齐技术相关专利概况进行分析，随后对模态对齐相关技术进行主要技术简介，每个相关技术后将挑选的重点专利进行解读，将通过相关专利解读，展示这两种对齐方式在相关领域的创新与应用，助力多模态系统实现高效信息处理与语义理解。本章挑选重点专利时，主要依据重点专利权人近十年重点专利及关键技术代表专利、专利同族数、同族专利被引证数，兼顾专利法律状态等进行挑选和解读。

5.1 专利总量

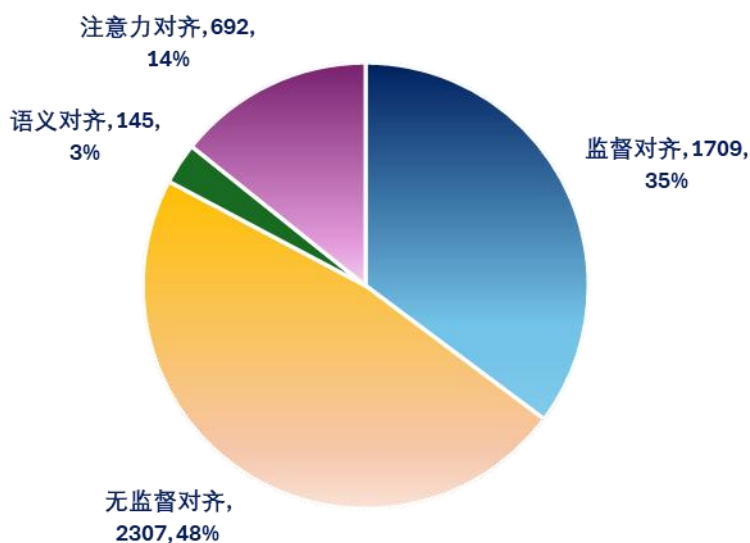


图 5-1 模态对齐各分技术专利总量分布

图 5-1 展示了模态对齐各分技术专利总量的分布情况。无监督对齐专利占比最高，达 48%（2307 件），表明其在模态对齐领域的广泛应用和研发重点。监督对齐次之，占比 35%（1709 件），显示其在特定领域的稳定需求。注意力对齐占比 14%（692 件），语义对齐仅占 3%（145 件）。这表明无监督和监督对齐是当前研发的主流方向，而语义对齐相对较少受到关注。

5.2 专利趋势

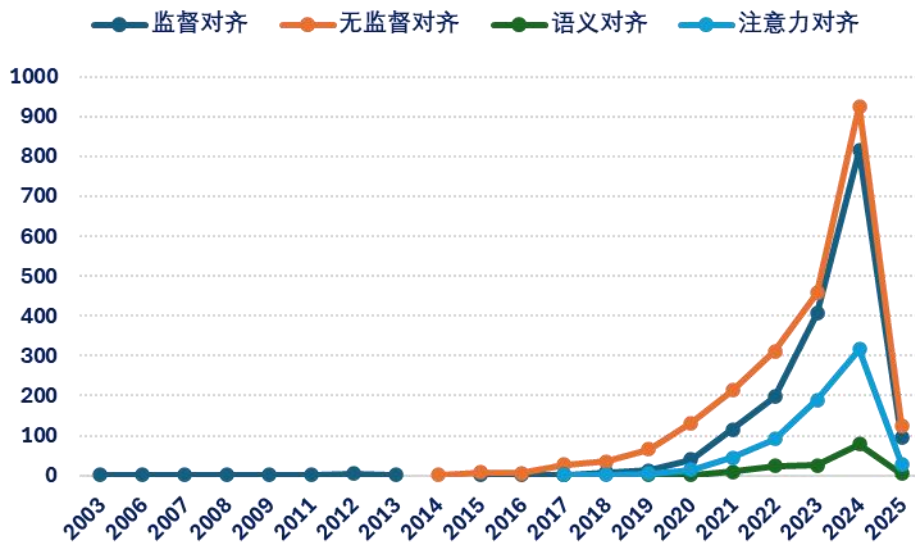


图 5-2 模态对齐各分技术专利申请趋势

总体趋势来看，2003–2015 年：所有技术的专利申请量都处于较低水平，增长缓慢，表明模态对齐技术在这一时期处于初步探索和发展的阶段。2016–2024 年：专利申请量显著增加，尤其是无监督对齐和监督对齐技术，显示出这一时期模态对齐技术的快速发展和广泛应用。

各技术具体分析，无监督对齐方面：2003–2015 年：专利申请量接近于零，表明无监督对齐技术在早期几乎没有得到关注。2016–2024 年：专利申请量迅速上升，2024 年达到峰值，超过 900 件。这一趋势表明无监督对齐技术在这一时期得到了广泛的研究和应用。监督对齐方面：2003–2015 年：专利申请量较

低且波动不大，表明监督对齐技术在早期阶段的研发活动有限。2016–2024 年：专利申请量稳步增长，2024 年达到峰值，接近 800 件。虽然申请量少于无监督对齐，但显示出持续的研发投入。注意力对齐方面：2003–2015 年：专利申请量接近于零，表明在早期阶段关注度低。2016–2024 年：专利申请量逐渐增加，2024 年达到较高水平，约为 700 件。表明在这一时期逐渐受到关注。语义对齐方面，2003–2015 年：专利申请量几乎为零，表明早期几乎未受到关注。2016–2024 年：专利申请量缓慢增长，2024 年达到约 100 件。表明在这一时期开始受到关注，但总体申请量仍然较少。

模态对齐技术在 2016–2024 年经历了快速的发展和专利申请增长，无监督对齐技术在这一时期占据主导地位，显示出其广泛的应用潜力和研发重点。监督对齐和注意力对齐技术也表现出显著的增长趋势，表明它们在特定领域的应用受到关注。语义对齐技术虽然整体申请量较少，但在 2016–2024 年期间也有一定的发展。2024 年后，专利申请量的下降可能受到专利信息延迟公开影响。

5.3 法律状态

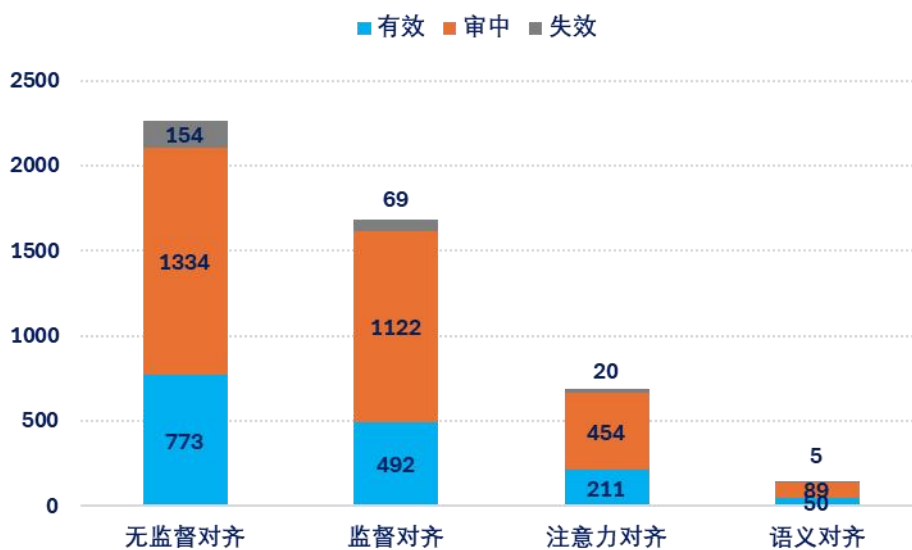


图 5-3 模态对齐各分技术专利法律状态分布

图 5-3 展示了模态对齐各分技术专利的法律状态分布情况。无监督对齐的审中专利最多，达到 1334 件，有效专利 773 件，失效专利 154 件。监督对齐的审中专利为 1122 件，有效专利 492 件，失效专利 69 件。注意力对齐的审中专利 454 件，有效 211 件，失效 20 件。语义对齐的审中专利 89 件，有效 50 件，失效 5 件。总体来看，无监督对齐和监督对齐在审中专利数量上占据较大比例，表明这些技术仍在快速发展和扩展。有效专利方面，无监督对齐和监督对齐也占据优势，显示出其技术成熟度较高。失效专利数量相对较少，表明这些技术的专利维护情况较好。

5.4 区域分布

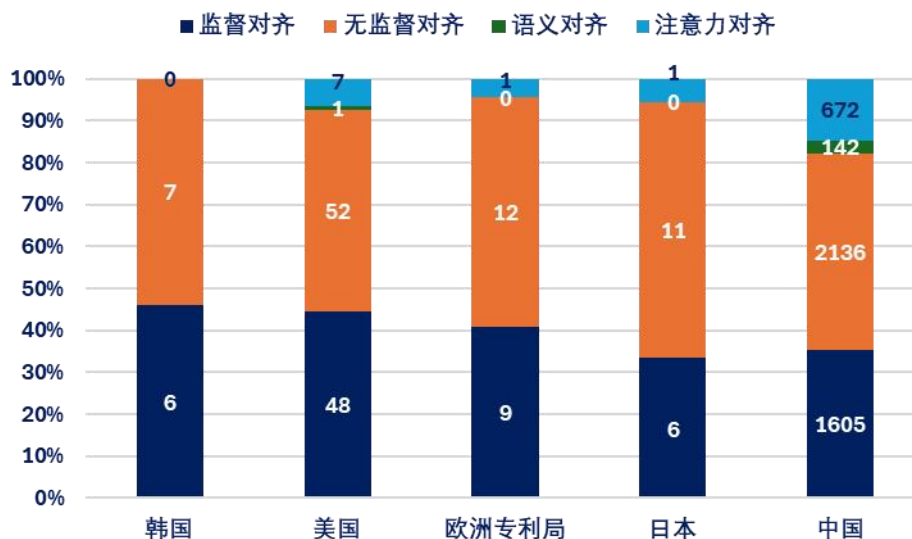


图 5-4 五大局模态融合各分技术专利占比情况

图 5-4 展示了模态对齐各分技术专利在不同地区的分布情况。无监督对齐在中国的专利数量最多，达到 2136 件，占比超过 50%。此外，无监督对齐在美国和欧洲专利局的专利数量占比均较大，专利数量分别为 52 和 12 件。监督对齐方面，是五大局地区模态对齐技术第二关注的重点技术方向。注意力对齐在中国有 672 件专利，美国 7 件专利，韩国和日本各有 1 件。语义对齐在中

国有 142 件专利，美国有 1 件。总体来看，显性对齐中的无监督对齐技术和监督对齐技术在五大局地区占据主导地位；而隐性对齐中的注意力对齐和语义对齐在中国也有一定的专利积累，而在其他地区的在此方面的专利布局则相对较少。

5.5 本章小结

专利总量分布，无监督对齐：在专利总量中占主导地位，占比 48%，显示出其在模态对齐领域的广泛应用和研发重点。其中，中国的专利数量最多，达到 2136 件，占该类专利总量的绝大部分。监督对齐：专利总量占 35%，位居第二。美国在该类方法上的专利数量相对较多，显示出其在监督对齐技术方面的研发实力。注意力对齐：专利总量占比 14%，表明其在模态对齐技术中也有一定的应用和研发关注。语义对齐：专利总量最少，仅占 3%，显示出其在模态对齐领域相对较少受到关注。

专利申请趋势方面，早期阶段（2003-2015 年）：专利申请量整体较低，技术发展处于初步探索阶段。快速增长期（2016-2024 年）：专利申请量显著增加，尤其是无监督对齐和监督对齐技术，在 2018-2024 年期间达到峰值。无监督对齐在 2024 年达到最高申请量，随后急剧下降。

法律状态方面，无监督对齐的有效专利数量最多，达到 773 件，显示出其技术成熟度较高且在市场上具有较强的法律保护。监督对齐的有效专利为 492 件，注意力对齐为 211 件，语义对齐为 50 件。审中专利：无监督对齐的审中专利数量最多，为 1334 件，显示出其技术仍在快速发展和扩展。监督对齐的审中专利为 1122 件，注意力对齐为 454 件，语义对齐为 89 件。失效专利：

无监督对齐的失效专利数量最多，为 154 件，可能由于部分早期专利因技术更新或维护成本等原因失效。监督对齐的失效专利为 69 件，注意力对齐为 20 件，语义对齐为 5 件。

区域分布分析方面，中国：在无监督对齐和注意力对齐上占据显著优势，无监督对齐专利数量达到 2136 件，注意力对齐为 672 件，表明中国在这些技术方向上的强大研发能力和市场应用。美国：在监督对齐上具有明显优势，专利数量为 48 件，显示出其在监督对齐技术方面的研发实力。欧洲专利局：在监督对齐上有一定的专利积累，显示出其在该技术方向上的研发重点。韩国和日本：在注意力对齐上各有少量专利，分别为 1 件，显示出其在该技术方向上的持续关注。

模态对齐技术领域，无监督对齐在专利总量和有效专利上占据主导地位，显示出其在技术成熟度和市场应用中的优势。监督对齐紧随其后，表明其在特定领域的广泛应用。注意力对齐和语义对齐虽然总量相对较少，但也有一定的发展潜力。专利申请趋势表明，2016-2024 年是模态对齐技术的快速发展期，2024 年后申请量的下降可能预示着技术进入成熟期或市场饱和状态。在法律状态方面，无监督对齐和监督对齐的审中专利数量较多，显示出其技术仍在快速发展中。不同地区的专利分布反映了各自的技术研发重点和市场策略，显性对齐中的无监督对齐技术和监督对齐技术在五大局地区占据主导地位；而隐性对齐中的注意力对齐和语义对齐在中国也有一定的专利积累，而在其他地区的在此方面的专利布局则相对较少。

第六章 AI 多模态融合技术应用分析

在人工智能技术的浪潮中，AI 多模态融合技术快速，巧妙地整合多种模态的数据，极大地拓展了人工智能的应用版图。鉴于其巨大的应用潜力，对 AI 多模态融合技术在不同领域的应用进行深入探究具有极为重要的意义。本章将聚焦于该技术在医疗健康、智慧交通、工业制造、消费电子以及金融投资等关键领域的实际应用场景，剖析其如何优化传统流程、创造全新价值。同时，深入解读相关专利，展现技术创新的法律保护及行业引领作用，呈现 AI 多模态融合技术在多元领域的落地成果与知识产权布局，把握其在各领域的实践进展。

6.1 医疗健康

AI 多模态融合技术在医疗健康领域的应用显著提升了诊断精度并重构了临床决策流程。在医学影像诊断方面，通过整合多模态影像数据和非结构化数据，可构建三维动态诊疗模型，实现精准诊断，如瑞金医院的 RuiPath 病理大模型提高了癌症诊断准确率并缩短人工复核时间。在解决医疗数据孤岛问题上，联邦学习与隐私计算技术成为关键突破，如昆山市第一人民医院的跨区域多模态医疗平台提升了阿尔茨海默病的诊断敏感度和特异性。在药物研发领域，AI 多模态融合技术颠覆传统模式，如华为云的盘古大模型加速了新冠口服药研发进程。

1) 医疗健康应用专利解读之 CN110580695B

申请号	CN201910727072.9	申请日	2019-08-07
公开号	CN110580695B	法律状态/事件	授权
标题	一种多模态三维医学影像融合方法、系统及电子设备		
申请(专利权)人	深圳先进技术研究院		
发明目的	通过多任务生成对抗网络的设计，结合对抗损失函数和交叉熵损失，实现了 MRI 影像与 PET 影像之间的关联特征学习和跨模态影像合成，解决了现有技术中多任务问题的收敛点冲突和任务关联性问题，提升		

	了影像生成和分类诊断的精度。
主要附图	
第一权利要求	1.一种多模态三维医学影像融合方法，其特征在于，包括以下步骤： 步骤 a：构建多任务生成对抗网络，所述多任务生成对抗网络包括生成器、判别器和分类器； 步骤 b：根据受试者的 MRI 影像、PET 影像和诊断标签信息对所述多任务生成对抗网络进行训练，使所述多任务生成对抗网络自动学习 MRI 影像和 PET 影像之间的关联特征； 步骤 c：将待检测者的 MRI 影像输入训练好的多任务生成对抗网络，所述生成器根据 MRI 影像合成对应的 PET 影像，并将待检测者的 MRI 影像与合成的 PET 影像输入分类器，所述分类器将待检测者的 MRI 影像与合成的 PET 影像进入融合后输出待检测者的疾病分类预测标签； 在所述步骤 b 中，所述多任务生成对抗网络进行训练具体包括： 步骤 b1：构建所述多任务生成对抗网络的对抗损失函数；训练过程的对抗损失用改进的极大极小代价函数进行表示： 上述公式中，(C,G,D)分别表示分类器、判别器和生成器，(x,y,z)分别表示 MRI 影像、PET 影像和诊断标签信息；$\alpha \in (0,1)$ 是一个常量，用于控制分类器和生成器损失在训练过程中所占比重，$E(x,y,z) \sim p(x,y,z)[\log D(x,y,z)]$ 表示判别器将来自于真实数据分布中的样本判定为真实样本；表示判别器识别出有分类器输出数据空间中的伪样本对；表示判别器将自生成器的伪样本标签对识别出来，其中 x 表示受试样本的 MRI 模态影像，z 表示样本标签，$G(x, z)$ 表示条件生成网络合成的 PET 模态影像； 步骤 b2：对分类器引入监督学习下的交叉熵损失 $K_c = E(x, y, z) \sim p(x, y, z)[-\log p_c(x, y, z)]$，将分类器的样本分布 $p_c(x, y, z)$ 的收敛点限定在 $p(x, y)$ 附近，使模型的全局最优满足生成器 G 和分类器 C 产生的样本分布与真实数据分布相同； 步骤 b3：引入生成器监督损失，利用目标图像与生成图像之间的梯度互信息作为相似性度量： $K_g = NI(A, B) = G(A, B) \cdot I(A, B)$ 上述公式中，$I(A, B)$ 和 $G(A, B)$ 分别表示生成图像与目标图像之间的梯度信息和梯度差值。

发明专利CN110580695B 主要针对现有的生成对抗模型在处理多任务问题时存在样本生成和模式分类收敛点的冲突，且跨模态影像合成方法未考虑样本标签信息，独立设计的模型无法优化任务之间关联性的问题进行探究。

深圳先进技术研究院运用设计了一种多任务生成对抗网络,包括生成器、判别器和分类器，通过对抗损失函数、交叉熵损失和梯度互信息的结合，训练网络以学习 MRI 影像和 PET 影像之间的关联特征，并在数据驱动模式下实现跨模态影像合成和多模态融合分类的协同训练，解决了传统模型在生成器和分类器性能兼顾时的收敛点冲突，实现了跨模态影像合成和多模态融合分类的协同训练，提升了图像生成误差和分类诊断性能，避免了 PET 采集的高昂成本和辐射暴露风险，提高了疾病分类的精度和模型的准确率。

2) 医疗健康应用专利解读之 US10842379B2

申请号	US15/416102	申请日	2017-01-26
公开号	US10842379B2	法律状态/事件	授权 权利转移
标题	Multi-modality image fusion for 3D printing of organ morphology and physiology		
申请(专利权)人	西门子医疗股份公司		
发明目的	3D 打印中多模态图像融合和先进材料的集成通过创建整体的、解剖学上准确的模型、增强临床应用和教育工具来解决单模态成像的局限性。		
主要附图	<pre>graph TD; 202[RECEIVE MEDICAL IMAGES FROM MULTIPLE MEDICAL IMAGING MODALITIES] --> 204[FUSE MEDICAL IMAGES FROM MULTIPLE MEDICAL IMAGING MODALITIES INTO SINGLE COORDINATE SYSTEM]; 204 --> 206[SEGMENT ORGAN COMPONENTS FROM FUSED MEDICAL IMAGES]; 206 --> 208[ESTIMATE PHYSIOLOGICAL PARAMETERS FROM FUSED MEDICAL IMAGES]; 208 --> 210[3D PRINT HOLISTIC MODEL];</pre>		
第一权利要求	1. A method for generating a patient-specific 3D printed model of a target organ from multiple medical imaging modalities, comprising: fusing a plurality of medical images of a target organ of a		

	patient from different medical imaging modalities; generating a holistic mesh model of the target organ by segmenting the target organ in the fused medical images from the different medical imaging modalities, wherein generating a holistic mesh model of the target organ by segmenting the target organ in the fused medical images from the different medical imaging modalities comprises: segmenting a mesh model of the target organ from one or more of the plurality of medical images from the different medical imaging modalities, segmenting tissue substrate information for the target organ in a medical image in the plurality of medical images from the different medical imaging modalities other than the one or more of the plurality of medical images from which the mesh model of the target organ is segmented, and mapping the segmented tissue substrate information to the mesh model of the target organ, resulting in the holistic mesh model of the target organ; estimating one or more spatially varying physiological parameters from the fused medical images and mapping the estimated one or more spatially varying physiological parameters to the holistic mesh model of the target organ; and 3D printing the holistic mesh model of the target organ including a representation of the estimated one or more spatially varying physiological parameters mapped to the holistic mesh model.
--	--

发明专利 US10842379B2 主要针对目前 3D 打印在医学应用中仅限于单模态成像,无法整合多种医学成像模态的形态、结构、动态和功能信息,限制了其在临床决策、治疗计划和患者教育中应用的问题进行探究。

西门子医疗股份公司运用了一种多模态图像融合的方法和系统,使用图像分析和先进材料结合来自各种医学成像模态的信息,创建器官的整体 3D 打印模型,结合形态、基质和生理学,从而能够表示空间变化的生理学 通过语义图像配准和动态材料 3D 打印来调整参数,能够创建全面的 3D 打印模型,通过集成不同的医学成像数据来促进临床决策、治疗计划、患者教育和设备开发,从而

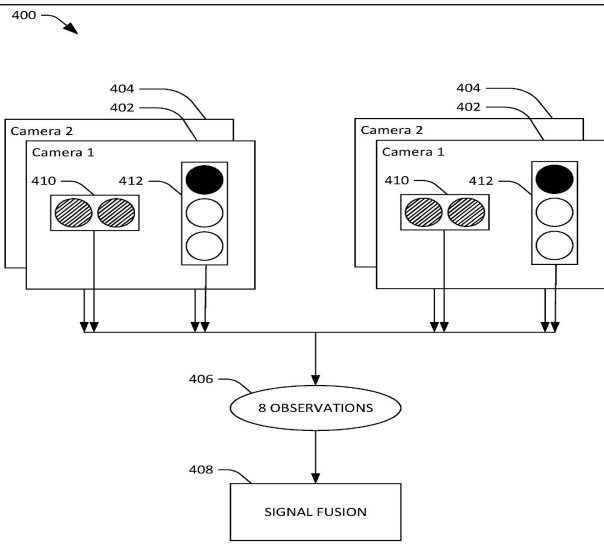
提高解剖表示的准确性和稳健性。

6.2 智慧交通

AI 多模态融合技术正推动智能交通系统向全局协同演进。在自动驾驶与车路协同领域，整合多源数据构建高精度感知体系，如特斯拉的 HydraNet 架构通过跨模态特征对齐生成鸟瞰图，实现全方位感知与快速决策；百度 Apollo 系统融合 5G-V2X 路侧单元数据，提升定位精度，扩大无人驾驶运营范围。在交通流优化方面，香港科技大学的 LightGPT 模型整合多模态信息预测流量变化，动态调整信号灯配时，提升通行效率；青岛“数见”智能交通体通过数据融合建模，预测区域交通压力，助力路网规划。智能座舱方面，华为 DriveONE 平台携手语音、手势与表情分析，构建全感官交互座舱；蔚来 NOMI 助手结合手势与高精地图信息，提供实时语音导览。

1) 智慧交通应用专利解读之 US10852743B2

申请号	US16/124239	申请日	2018-09-07
公开号	US10852743B2	法律状态/事件	授权 期限延长
标题	Multimodal multi-technique signal fusion system for autonomous vehicle		
申请(专利权)人	通用汽车环球科技运作有限责任公司		
发明目的	自动驾驶车辆中的多模态多技术信号融合系统通过融合传感器信号和物体检测器输出来解决单摄像头系统的可靠性问题,通过不同的传感器集成和指令的概率合并来提高交通灯检测精度和车辆控制。		

主要附图	
第一权利要求	1. An autonomous vehicle, comprising: - a mechanical system; - at least one sensor system, comprising: - a fixed exposure camera system configured to output a first sensor signal; - and an auto exposure camera system configured to output a second sensor signal; - a computing system in communication with the mechanical system and the at least one sensor system, wherein the computing system comprises: - a processor; - and memory that stores instructions that, when executed by the processor, cause the processor to perform acts comprising: - generating a first independent directive and a second independent directive that each define a vehicle maneuver based upon an illuminated configuration of a traffic light captured in the first sensor signal and the second sensor signal; - fusing the first independent directive and the second independent directive to output a fused directive; - and - controlling the mechanical system based upon the fused directive.

发明专利 US10852743B2 主要针对仅依靠摄像头传感器系统进行交通灯检测的自动驾驶车辆很容易因障碍物或不正确的物体检测而出现单点故障,从而导致车辆控制指令可能出现错误的问题进行探究。

通用汽车环球科技运作有限责任公司运用了多模态多技术信号融合系统,结合多个传感器系统和物体检测器模块(包括固定曝光和自动曝光相机)的输出,

生成用于控制车辆的融合指令,使用概率技术合并独立指令并提高检测精度,达到了通过降低单点故障的风险来提高交通灯检测的准确性和可靠性,通过集成不同的传感器系统和检测技术来确保精确的车辆控制。

2) 智慧交通应用专利解读之 US11983625B2

申请号	US16/911100	申请日	2020-06-24
公开号	US11983625B2	法律状态/事件	授权 期限延长
标题	Robust multimodal sensor fusion for autonomous driving vehicles		
申请(专利权)人	英特尔公司		
发明目的	使用 B-DNN 的不确定性感知多模态贝叶斯融合框架通过估计和管理预测不确定性来解决自动驾驶汽车中传感器噪声和不确定性的挑战,从而提高自动驾驶汽车传感器融合的可靠性和安全性。		
主要附图			
第一权利要求	1. An electronic control unit (ECU), comprising: a plurality of a neural networks, each one of the plurality of neural networks having an input coupled to a respective autonomous vehicle (AV) sensor and configured to output distribution data representing a set of different classes used in accordance with an AV application environmental model; a plurality of uncertainty estimation units, each one of the plurality of uncertainty estimation units being configured to calculate uncertainty estimate values comprising an aleatoric uncertainty value and an epistemic uncertainty value using the distribution data output from a respectively coupled one of the plurality of neural networks, the		

	aleatoric uncertainty value being with respect to data provided by the AV sensor coupled to the respectively coupled one of the plurality of neural networks; and a control unit configured to: gate a contribution of data provided by each AV sensor to the distribution data output by each one of the plurality of neural networks in accordance with a gating function that utilizes respective uncertainty estimate values calculated for each one of the plurality of neural networks; and generate the AV application environmental model using the gated distribution data.
--	--

发明专利 US11983625B2 主要针对由于噪声和不确定性,自动驾驶车辆在可靠地融合来自多个传感器的数据方面面临挑战,导致潜在的错误和故障,特别是在遇到新数据或错误数据时,因传统方法无法有效捕获传感器漂移或提供可靠的置信度分数的问题进行探究。

英特尔公司运用了使用贝叶斯深度神经网络 (B-DNN) 实现不确定性感知多模态贝叶斯融合框架来估计预测不确定性,计算任意和认知不确定性,并根据这些估计动态门控传感器贡献,确保稳健的传感器融合,实现更安全的驾驶控制和路径规划,这种方法通过识别故障传感器和新数据来实现透明、可靠的决策,将精确召回 AUC 提高 10.2%,并提供强大的机制来处理噪声和不确定的传感器输入,从而提高自动驾驶车辆操作的安全性和效率。

6.3 工业制造

AI 多模态融合技术正在重塑工业制造的生产全流程。在设备预测性维护中,通过融合振动传感器、红外热成像和声学信号,实现故障早期预警。例如,华为盘古大模型在汽车冲压车间的应用显著减少了设备停机时间和维护成本。西门子利用 AR 眼镜叠加设备三维模型和实时振动频谱,大幅缩短了检修作业时间。在生产质量管控方面,天准科技和三星显示通过整合多模态数据,提升

了缺陷检测的准确性，减少了质量损失。此外，在流程优化领域，杉数科技的智能排产系统通过多模态强化学习，为钢铁企业节省了大量成本。

1) 工业制造应用专利解读之 CN114707551B

申请号	CN202210354875.6	申请日	2022-04-06
公开号	CN114707551B	法律状态/事件	授权
标题	面向工业互联网设备故障诊断的多模态数据融合方法		
申请(专利权)人	西北大学		
发明目的	通过多模态数据融合方法，利用 MvDA 和 CNN 等技术，解决了单一传感器在复杂环境下故障诊断准确率低的问题，实现了更高的故障诊断准确率和鲁棒性，适用于工业互联网设备的故障诊断。		
主要附图			
第一权利要求	1.一种面向工业互联网设备故障诊断的多模态数据融合方法，其特征在于，包括以下步骤： 步骤 1，收集大型设备的故障数据，进行数据预处理，将其作为训练样本； 步骤 2，对步骤 1 中整理分类后得到的大型设备故障训练样本，针对数据类型或特性不同的传感器采集到的数据，从不同角度进行多模态数据融合，提取不同传感器数据的特征矩阵和主成分矩阵； 步骤 3，对于所有传感器，通过多模态数据融合的方法建立多模态数据融合模型，首先定义求解多模态数据融合模型的映射矩阵，然后使用一种基于 Ratio Trace 方法的迭代算法求得多模态数据融合模型的全局最优解；具体包括如下子步骤： 步骤 3.1，对于每个样本，定义： 式中，为设备故障的数据集，d_{ijk} 为第 i 种故障中第 j 个传感器的第 k 个样本，I 为故障的种类，J 为传感器的总数，q_o 为传感器 o 的特征维数，k_{ij} 为第 j 个传感器的第 i 种故障的样本个数； 步骤 3.2，对于所有传感器，可以通过多模态数据融合 MvDA 的方式建立多模态数据融合模型，定义求解多模态数据融合模型的映射矩阵为：		

	式中，分别是不同传感器的最优映射矩阵，Lin 为公共空间中的投影的类内散布矩阵，Lout 类间散布矩阵； 两个散布矩阵由下式 5 和式 6 表示： 式中，为所有传感器中第 i 类故障样本的平均矩阵； 为第 i 种故障所有传感器的样本数； 为所有样本在公共空间的平均矩阵； K 为所有故障样本的数目，有 F_{ijk} 为 d_{ijk} 样本映射到公共空间上所对应的样本值； 为了方便计算，类间散射矩阵 Lout 和类内散射矩阵 Lin 被写为： $\text{Lout} = \text{STWS}$(式 7) $\text{Lin} = \text{STVS}$(式 8) 式中：ST 为所有传感器映射矩阵转置的集合，S 为所有传感器映射矩阵的集合，有 $S = \{S_1, S_2, \dots, S_J\}$； W 和 V 都为关于的块矩阵，其定义如下： 式中，为第 j 个传感器中第 i 种故障所有样本的平均矩阵，为的转置，k_{ij} 为第 j 个传感器的第 i 种故障的样本个数；为第 i 种故障所有传感器的样本数；为第 m 个传感器中第 i 种故障所有样本的平均矩阵；为的转置；k_{im} 为第 m 个传感器的第 i 种故障的样本个数；为第 i 种故障所有传感器的样本数； 步骤 3.3，步骤 3.2 中式 4 需要求解 J 个传感器的线性变换视图，通过将目标函数转换为迹比问题，可以简化问题的求解，式 4 目标函数的迹比解又可以被写为： 以上目标函数可以被看作一个迹比问题，奇异值分解很难获得多模态数据融合模型的全局最优解，为了解决这个问题，使用一种基于 Ratio Trace 方法的迭代算法来解决迹比问题； 步骤 3.4，首先为了保持全局结构的可解性，加入正交约束 $\text{STS} = E$；其中，E 为单位矩阵，并且让则式 11 又可写为： 已知式 12 满足由于模型使用了 PCA 主成分分析法已经对传感器数据进行处理，所以不存在为 0 的空间，则可以用以下公式转换 式中：$\Theta = [\theta_1, \theta_2, \theta_3, \dots, \theta_C]$，$\theta_c > 0$，$c = 1, 2, \dots, C$；$C$ 为中正奇异值的数量； 令 $S = QB$，d 为 S 的秩，目标函数可被表示为： 其中，$W(u) = BTWB$， 通过奇异值分解考虑到，式 14 中存在：即使有闭式解但仍可能与最优解存在较大误差； 步骤 3.5，将式 14 转化为迹差问题： 其中 λ_n 是从上一步的投影矩阵 B_{n-1} 计算出的轨迹比值，具体算法步骤为： a) 将 B_0 初始化为任意列正交矩阵； b) 从 $n = 1, 2, \dots, N$ 开始，通过计算矩阵 B_{n-1} 计算 θ_n； c) 将迹差问题构造为： d) 使用特征值分解法求解迹差问题： 式中，为中第 r 大的特征值，为对应的特征向量； 为了保持正交，重新构造投影矩阵，设 d 为低维特征的秩，让 $V(b)$
--	--

	$= B_n(B_n)TV(u)B_n(B_n)T$ ，进行奇异值分解，有： $V(b) = B_n\Theta b(B_n)T$ (式 19) e)如果有则退出循环； 输出 $B = B_n$ ； 步骤 4，利用步骤 3 中获得的多模态数据融合模型，将传感器特征投影到一个新的空间，使用分类器对数据进行分类和预测。
--	--

发明专利CN114707551B 主要针对现有工业物联网设备在复杂工业环境中使用单一传感器进行故障诊断时，容易受到外界干扰导致失灵，导致诊断准确率低和鲁棒性弱的问题进行探究。

西北大学运用了采用多模态数据融合方法，通过多视图判别分析（MvDA）建立多模态数据融合模型，使用 Ratio Trace 迭代算法求解，结合小波图像特征和卷积神经网络（CNN）提取传感器数据特征，并通过 PCA 降维，投影到公共空间进行分类和预测，提高了故障诊断的准确率和鲁棒性，增强了对复杂工业环境下的设备故障诊断能力，降低了对单一传感器的依赖，提高了诊断系统的稳定性和可靠性。

2) 工业制造应用专利解读之 CN118733660B

申请号	CN202411235791.6	申请日	2024-09-04
公开号	CN118733660B	法律状态/事件	授权
标题	基于迁移学习的跨领域多模态工业数据抽取与融合方法		
申请(专利权)人	广东恒电信息科技股份有限公司		
发明目的	通过基于迁移学习的方法进行多模态工业数据抽取与融合，解决了跨领域数据应用的难题，实现了高效、准确的数据处理和融合，提升了模型的适应性和性能。		
主要附图	<pre>graph TD; S1[收集多领域的工业数据，生成多个数据集，构建数据抽取模型，数据抽取模型获取数据集的特征集，基于特征集在数据集抽取关键信息，基于关键信息生成标准数据] --> S2[将任一数据集对应的标准数据设定为训练集，基于训练集构建数据迁移模型，将目标数据库对应的数据集的关键信息设定为目标标签，数据迁移模型基于目标标签将源数据库对应的标准数据迁移至目标数据库，生成迁移数据]；S2 --> S3[比较源数据库与目标数据库中间一标准数据之间的数据相似度，生成数据迁移模型的第一准确率，基于第一准确率校正剩余数据集中标准数据对应的关键信息，生成数据标签]；S3 --> S4[构建多模态融合模型，基于多模态融合模型将目标数据库对应的标准数据与迁移数据进行融合生成融合数据，重复执行此步骤，至所有数据集对应的标准数据均融合至目标数据库中，生成多领域的融合数据集]；</pre>		
第一权利要求	1.基于迁移学习的跨领域多模态工业数据抽取与融合方法，其特征在		

	于, 所述方法包括如下步骤: S1: 收集多领域的工业数据, 生成多个数据集, 构建数据抽取模型, 所述数据抽取模型获取所述数据集中的特征集, 基于所述特征集在所述数据集中抽取关键信息, 基于所述关键信息生成标准数据; S2: 将任一数据集对应的所述标准数据设定为训练集, 基于所述训练集构建数据迁移模型, 将目标数据库中对应的数据集的所述关键信息设定为目标标签, 所述数据迁移模型基于所述目标标签将源数据库对应的所述标准数据迁移至所述目标数据库, 生成迁移数据, 其中, 基于预训练的 Transformer 模型作为迁移学习的起点, 将所述训练集输入所述 Transformer 模型中, 训练生成适配的神经网络模型, 将所述神经网络模型设定为所述数据迁移模型; S3: 比较所述源数据库与所述目标数据库中同一所述标准数据之间的数据相似度, 生成所述数据迁移模型的第一准确率, 基于所述第一准确率校正剩余数据集中所述标准数据对应的所述关键信息, 生成数据标签; S4: 构建多模态融合模型, 基于所述多模态融合模型将所述目标数据库中对应的所述标准数据与所述迁移数据进行融合生成融合数据, 重复执行此步骤, 至所有数据集对应的所述标准数据均融合至所述目标数据库中, 生成多领域的融合数据集; 其中, 所述构建多模态融合模型包括以下步骤: 设置一个多任务学习框架的技术模型, 基于数据生成类型同时处理多种模态的数据, 并优化资源分配, 通过所述技术模型的解构分散策略和神经架构搜索技术, 自动设定最优的技术模型配置, 引入注意力机制和图卷积网络, 并设置元学习技术, 获取所述技术模型的收敛指数, 基于所述收敛指数模拟和合成数据生成技术, 以训练所述技术模型, 将训练后的所述技术模型设定为所述多模态融合模型; 生成所述融合数据包括以下步骤: 将所述目标数据库中包含的所述标准数据和所述迁移数据均设定为待融合数据, 基于所述多模态融合模型提取所述待融合数据的所述数据生成类型和时间信息, 将所述待融合数据基于所述时间信息分类为第一类数据和第二类数据, 若所述融合数据存在所述时间信息, 则分类为所述第一类数据, 否则, 分类为所述第二类数据; 在所述第一类数据中基于变分近似方法分别生成与相同数据生成类型包含的所述时间信息对应的潜在变量分布, 标记所述多模态融合模型的隐藏状态, 基于所述潜在变量分布更新所述隐藏状态, 将更新后的所述隐藏状态和所述潜在变量分布输入所述多模态融合模型中, 基于所述神经架构搜索技术输出所述潜在变量分布对应的结果数据, 将所有所述结果数据设定为所述第一类数据对应的第一融合数据; 将所述第二类数据基于所述数据生成类型和所述关键信息依次汇总生成第二融合数据, 将所述第一融合数据与所述第二融合数据组合设定为所述融合数据。
--	---

发明专利CN118733660B 主要针对现有技术难以实现跨领域的多模态工业数据抽取与融合, 传统机器学习方法在不同领域间应用困难, 且数据隔离导

致难以共享的问题进行探究。

广东恒电信息科技股份有限公司采用基于迁移学习的方法，通过数据迁移模型将多领域数据集的标准数据迁移至目标数据库，构建多模态融合模型进行数据融合，利用特征相似度和差异特征生成标准数据，优化资源分配和神经架构搜索技术以提高模型性能，显著提高了工业数据抽取与融合的效率 and 准确性，减少了数据标注成本，提升了模型的泛化能力，能够在实时环境中高性能运行，并提供模型可解释性。

6.4 消费电子

AI 多模态融合技术正在重塑消费电子产品，通过跨模态特征对齐与动态权重分配，实现自然交互与场景化服务升级。在智能终端交互中，雷鸟科技的 V3 AI 眼镜和华为 DriveONE 平台分别通过整合多种交互数据，提升了跨境电商导购和座舱交互的体验。在智能家居领域，海尔 HomeGPT 系统和 EMQX 平台优化了设备联动和本地化响应效率，实现场景自适应调节和快速语音指令解析。AI 硬件创新方面，宇树科技的 G1 人形机器人、Rokid Glasses 和小米 AI 眼镜等产品通过多模态数据融合，显著提升了仓储效率、直播场景交互和意外事故发现速度。

1) 消费电子应用专利解读之 CN116185182B

申请号	CN202211731584.0	申请日	2022-12-30
公开号	CN116185182B	法律状态/事件	授权
标题	一种融合眼动注意力的可控图像描述生成系统及方法		
申请(专利权)人	天津大学		
发明目的	通过融合眼动信息和 Transformer 模型，结合 AR 眼镜，解决了现有技术中图像描述生成中交互效率低的问题，实现了可控的图像描述生成，适用于无法发声的特殊人群，提高了交互效率和系统的实践性。		

主要附图	<pre> graph TD User[用户] <--> ImageDisplay[图像显示模块] User <--> EyeAcq[眼图采集模块] ImageDisplay <--> ImageAcq[图像采集模块] ImageAcq --> ImageAcq2[图像采集模块] EyeAcq --> EyeCoord[眼动坐标识别模块] ImageAcq2 --> FeatureFusion[特征融合模块] EyeCoord --> FeatureFusion FeatureFusion --> LangGen[语言描述生成模块] LangGen --> Result[可控的图像描述结果] subgraph System ImageAcq2 ImageDisplay EyeAcq EyeCoord FeatureFusion LangGen end </pre>
第一权利要求	1.一种融合眼动注意力的可控图像描述生成系统,其特征在于,包括:依次相连的图像采集模块、图像预处理模块及图像显示模块;依次相连的眼图采集模块及眼动坐标识别模块;分别与图像预处理模块和眼动坐标识别模块相连的特征融合模块,与特征融合模块相连的语言描述生成模块; 图像采集模块用于实时采集环境图像; 图像预处理模块用于对采集的环境图像进行预处理; 图像显示模块用于显示经过预处理后的环境图像; 眼图采集模块用于通过 AR 眼镜采集一段时间内用户的眼灰度图像序列并保存; 眼动坐标识别模块用于对采集的眼灰度图像序列进行注视角度识别,并将注视角度转化为与环境图像相对应的以像素为单位的二维坐标值序列; 特征融合模块用于提取文本特征、图像特征及眼动坐标识别模块输出数据的眼动特征,并将文本特征、图像特征及眼动特征进行融合生成多模态特征; 其中,所述特征融合模块包括:特征拼接模块及分别与其连接的文本特征提取模块、图像特征提取模块、眼动序列处理模块; 文本特征提取模块用于对上一时刻语言描述生成模块生成的文本信息进行特征提取; 图像特征提取模块用于对输入图像进行特征提取,并将三维的图像特征展开为二维形式; 眼动序列处理模块用于将二维坐标值序列转化为二维高斯分布图;其中高斯图的均值为某一时刻的眼动坐标点(a,b),a表示眼动坐标点的横坐标值,b表示眼动坐标点的纵坐标值;方差设定为1,高斯图中的像素点符合二维正态分布,即: (x,y)表示高斯图中的像素坐标,f(x,y)为像素坐标(x,y)对应的坐标值;x的取值在0至w之间;y的取值在0至h之间;w为输入图像的宽,h为输入图像的高;每个像素坐标(x,y)都有一个对应的坐标值f(x,y);

	特征拼接模块用于将文本特征提取模块、图像特征提取模块及眼动序列处理模块三者的输出进行拼接，生成多模态特征； 语言描述生成模块用于由多模态特征生成文本信息。
--	--

发明专利CN116185182B 主要针对现有的人机交互系统在图像描述生成中难以有效利用眼动信息，尤其是在交互效率低、无法发声的特殊人群和现实交互场景中，且缺乏整体交互系统的设计和应用实践。的问题进行探究。

天津大学运用设计了一种融合眼动注意力的可控图像描述生成系统，包括图像采集、预处理、显示模块、眼动采集和坐标识别模块，通过特征融合生成多模态特征，并利用 Transformer 模型生成文本信息，结合 AR 眼镜实现交互。达到了实现了基于 AR 眼镜的可控图像描述生成，提高了交互效率，提供了真实可行的人机交互系统，避免了模型欠拟合问题，并适用于无法发声的特殊人群。

2) 消费电子应用专利解读之 CN116880701B

申请号	CN202311148693.4	申请日	2023-09-07
公开号	CN116880701B	法律状态/事件	授权
标题	基于全息设备的多模态交互方法及系统		
申请(专利权)人	深圳优立全息科技有限公司		
发明目的	通过在全息设备中采集和处理红外反射信号、视觉动线和脑电波数据，实现了多模态数据的时间序列对齐和特征提取，解决了虚拟现实空间中交互体验不够真实的问题，提供了更高效和个性化的虚拟现实体验。		

主要附图	<pre>graph TD 100 --> 101[从金融内容平台中采集待风控检测的目标金融文档] 101 --> 102[基于预训练的语义大模型对目标金融文档进行多模态语义特征提取， 得到文本模态的文本语义空间和图像模态的图像语义空间] 102 --> 103[根据所述文本语义空间中各个文本语义特征在上下文的关系语义构建文 本要素的语义关联树，进而由所述语义关联树和所述文本语义空间中 文本语义特征的句法结构关系确定文本模态的语义关联图] 103 --> 104[通过目标金融文档中图像的整体语义特征和图像内视觉对象的句法特征 将所述图像语义空间中各个图像语义特征之间的语义结构关系转换为 图像模态的语义关联图] 104 --> 105[对所述文本模态的语义关联图和所述图像模态的语义关联图进行模态间 细粒度结构融合，得到多模态融合图] 105 --> 106[从所述多模态融合图中提取多模态特征，对提取的多模态特征进行异常 分类，若分类结果为异常，则确定目标金融文档存在风险]</pre>
第一权利要求	1.一种基于全息设备的多模态交互方法，其特征在于，所述基于全息设备的多模态交互方法包括： 对目标用户的手持手柄设备进行红外反射信号采集，得到红外反射信号； 通过预置的虚拟现实眼镜采集所述目标用户的视觉动线，得到目标视觉动线，并通过所述虚拟现实眼镜对所述目标用户进行脑电波数据采集，得到脑电波数据集合； 对所述红外反射信号、所述目标视觉动线以及所述脑电波数据集合进行时间序列对齐处理，得到目标融合数据； 对所述目标融合数据进行多类型特征数据提取，得到位置特征数据、关注位置特征数据以及情绪特征数据；具体包括：对所述目标融合数据进行数据滤波处理，生成滤波融合数据；对所述滤波融合数据进行数据分层，确定多层待处理滤波融合数据；分别对每层所述待处理滤波融合数据进行多类型特征数据提取，得到所述位置特征数据、所述关注位置特征数据以及所述情绪特征数据；其中，对每层所述待处理滤波融合数据进行分层标准匹配，确定每层所述待处理滤波融合数据对应的数据类型；基于每层所述待处理滤波融合数据对应的数据类型进行特征提取算法匹配，确定每层所述待处理滤波融合数据对应的特征提取算法；通过每层所述待处理滤波融合数据对应的特征提取算法分别对每层所述待处理滤波融合数据进行多类型特征数据提取，得到所述位置特征数据、所述关注位置特征数据以及所述情绪特征数据；其中，对于位置特征数据，服务器提取用户在虚拟场景中的运动轨迹、注视点集合以及交互位置集合，位置特征数据可以帮助服务器了解用户在虚拟场景中的活动范围、感兴趣的区域以及与虚拟对象的交互情况； 通过所述位置特征数据以及所述关注位置特征数据对预置的三维显示空间进行空间分割，得到对应的目标空间区域；具体包括：对所述位置特征数据进行三维虚拟点云构建，得到第一候选三维虚拟点云；对所述关注位置特征数据进行三维虚拟点云构建，得到第二候选三维

	虚拟点云；对所述第一候选三维虚拟点云进行虚拟三维空间映射，得到第一虚拟三维空间；对所述第二候选三维虚拟点云进行虚拟三维空间映射，得到第二虚拟三维空间；对所述第一虚拟三维空间以及所述第二虚拟三维空间进行空间分割点匹配，确定空间分割点集合；基于所述空间分割点集合对所述三维显示空间进行空间分割，得到所述目标空间区域； 采集所述目标用户的语音控制指令，并通过所述情绪特征数据对所述语音控制指令进行交互指令生成，生成目标交互指令； 通过所述目标交互指令构建全息显示空间坐标，生成全息显示空间坐标集，并通过所述目标交互指令进行全息设备匹配，确定目标全息设备集合；具体包括：对所述目标交互指令进行交互位置映射，确定交互位置数据；基于预置的空间坐标系对所述交互位置数据进行坐标映射，生成所述全息显示空间坐标集；对所述目标交互指令进行设备标识分析，确定目标设备标识集；通过所述目标设备标识集进行全息设备匹配，确定所述目标全息设备集合；其中，目标用户通过交互方式表达了自己的意图，用户通过手势指向特定位置，或者使用语音指令描述所需操作；将对目标交互指令进行处理，并从中提取交互位置数据，通过传感器数据的捕捉和解析，以获得用户指定的交互位置坐标；使用预置的空间坐标系来对交互位置数据进行坐标映射；预置的空间坐标系是全息显示空间的坐标系，用于描述全息投影的布局 and 位置；通过将交互位置数据映射到这个空间坐标系中，可以得到全息显示空间的坐标集，其中包含用户交互的具体位置信息；同时，还会对目标交互指令中的设备标识信息进行分析，以确定用户希望在哪些设备上交互，设备标识信息可以告诉用户的设备偏好，将根据设备标识集进行全息设备匹配，确定适合用户交互意图的全息设备集合；通过匹配目标设备，可以确保内容在正确的设备上显示或传输，为用户提供定制化的全息交互体验； 基于所述全息显示空间坐标集，通过所述目标全息设备集合构建三维虚拟现实空间，并将所述三维虚拟现实空间传输至所述目标用户的虚拟现实眼镜进行全息影像显示；具体包括：通过所述目标全息设备集合对所述全息显示空间坐标集进行几何元素提取，确定几何元素集合；对所述几何元素集合进行虚拟空间生成，得到所述三维虚拟现实空间；将所述三维虚拟现实空间传输至所述目标用户的虚拟现实眼镜进行全息影像显示；其中，目标全息设备集合包含虚拟现实眼镜；根据目标全息设备集合对全息显示空间坐标集进行几何元素提取，全息显示空间坐标集包含用于构建全息影像的所有空间位置信息，几何元素提取是将空间坐标集中的位置信息转换为具有几何形状的元素，几何元素集合包括点、线、面，这些几何元素集合将作为构建三维虚拟现实空间的基础，随后，利用提取的几何元素集合进行虚拟空间生成，在虚拟空间生成过程中，会根据几何元素的位置和形状来构建三维虚拟现实空间，这个虚拟现实空间是由几何元素组成的，可以模拟现实世界中的物体、场景和环境；将生成的三维虚拟现实空间传输至目标用户的虚拟现实眼镜进行全息影像显示，虚拟现实眼镜是用户与虚拟现实环境进行交互的设备，将三维虚拟现实空间投影到用户的视觉感
--	--

	知中，通过虚拟现实眼镜的显示，用户可以体验全息影像，并与虚拟现实空间进行互动和沉浸式体验。
--	---

发明专利 **CN116880701B** 主要针对现有的多模态数据处理技术在时间序列对齐和特征提取方面存在准确性不足，导致虚拟现实空间中的交互体验不够真实和自然的问题进行探究。

深圳优立全息科技有限公司运用了通过在全息设备中采集红外反射信号、视觉动线和脑电波数据，进行时间序列对齐处理和多类型特征数据提取，生成位置特征、关注位置特征和情绪特征数据，用于构建三维虚拟现实空间并进行全息影像显示，提高了多模态交互的显示时效性和准确度，提供了更加沉浸式、自然和个性化的交互体验，增强了用户与虚拟环境的互动效果。

6.5 金融投资

AI 多模态融合技术正重构金融行业的底层逻辑，通过跨模态数据整合与动态决策优化，推动业务全链条智能化升级。在智能风控与反欺诈领域，中金支付的风控平台融合多模态数据，利用 **Transformer** 架构实现特征对齐，风险预警准确率突破 **89%**；某商业银行通过联邦学习技术解决数据孤岛问题，降低跨境交易欺诈误报率 **45%**；**DeepSeek-R1** 金融大模型解析海量政策文件与财报数据，实现毫秒级异常交易模式识别。合规监管体系也在演进，某银行的多模态合规审查系统整合 **OCR**、**NLP** 与声纹验证技术，大幅缩短金融宣传材料审核时间。这些实践表明，AI 多模态技术已深度融入金融价值链重构，覆盖风险识别、运营效率、资产增值等多重维度。

1) 金融投资应用专利解读之 **CN118965279B**

申请号	CN202411435306.X	申请日	2024-10-15
公开号	CN118965279B	法律状态/事件	授权
标题	一种基于大模型的金融内容风控方法及系统		

申请(专利权)人	北京网智天元大数据科技有限公司 网智天元科技集团股份有限公司
发明目的	通过基于大模型的多模态语义特征提取和融合技术，解决了现有技术中金融文档虚假信息检测准确性不足的问题，实现了对金融文档的细粒度多模态特征融合，提高了风控检测的准确性。
主要附图	<pre>graph TD 100 --> 101[从金融内容平台中采集待风控检测的目标金融文档] 101 --> 102[基于预训练的语义大模型对目标金融文档进行多模态语义特征提取， 得到文本模态的文本语义空间和图像模态的图像语义空间] 102 --> 103[根据所述文本语义空间中各个文本语义特征在上下文的关系语义构建文 本要素的语义关联树，进而由所述语义关联树和所述文本语义空间中文 本语义特征的句法结构关系确定文本模态的语义关联图] 103 --> 104[通过目标金融文档中图像的整体语义特征和图像内视觉对象的句法特征 将所述图像语义空间中各个图像语义特征之间的语义结构关系转换为 图像模态的语义关联图] 104 --> 105[对所述文本模态的语义关联图和所述图像模态的语义关联图进行模态间 细粒度结构融合，得到多模态融合图] 105 --> 106[从所述多模态融合图中提取多模态特征，对提取的多模态特征进行异常 分类，若分类结果为异常，则确定目标金融文档存在风险]</pre>
第一权利要求	1.一种基于大模型的金融内容风控方法，其特征在于，包括下述步骤： 从金融内容平台中采集待风控检测的目标金融文档； 基于预训练的语义大模型对目标金融文档进行多模态语义特征提取， 得到文本模态的文本语义空间和图像模态的图像语义空间； 根据所述文本语义空间中各个文本语义特征在上下文的关系语义构建 文本要素的语义关联树，进而由所述语义关联树和所述文本语义空间 中文本语义特征的句法结构关系确定文本模态的语义关联图； 通过目标金融文档中图像的整体语义特征和图像内视觉对象的句法 特征将所述图像语义空间中各个图像语义特征之间的语义结构关系 转换为图像模态的语义关联图； 对所述文本模态的语义关联图和所述图像模态的语义关联图进行模 态间细粒度结构融合，得到多模态融合图； 从所述多模态融合图中提取多模态特征，对提取的多模态特征进行异 常分类，若分类结果为异常，则确定目标金融文档存在风险。

发明专利CN118965279B主要针对现有金融内容风控技术难以在金融文档中充分利用文本与图像的协同信息，导致虚假信息检测的准确性不足，尤其在复杂金融场景中难以识别跨模态信息和深层次语义关系的问题进行探究。

北京网智天元大数据科技有限公司联合网智天元科技集团股份有限公司运用了采用基于大模型的方法，通过预训练的语义大模型提取金融文档的多模态语义特征，构建文本和图像的语义关联树和关联图，进行模态间细粒度结构融

合，生成多模态融合图，并对其进行异常分类以判断虚假信息，提高了金融文档的虚假信息检测准确性，能够更准确地捕捉潜在的虚假信息和其传播路径，增强了对金融文档的整体理解能力，提升了风控检测的准确性。

2) 金融投资应用专利解读之 CN110598699B

申请号	CN201910869274.7	申请日	2019-09-16
公开号	CN110598699B	法律状态/事件	授权
标题	一种基于多光谱图像的防伪票据鉴伪系统和方法		
申请(专利权)人	华中科技大学		
发明目的	通过多光谱图像采集和预处理技术, 结合 SURF 特征点匹配和多特征融合, 解决了现有技术中防伪票据鉴伪准确性和实时性问题, 实现了高效的票据鉴伪和信息识别, 确保了数据管理的统一性。		
主要附图	<pre>graph TD A[防伪票据图像采集] --> B[图像预处理 (色彩补偿、几何校正、滤波降噪)] B --> C[防伪检测] C --> D[红外图像防伪油墨检测] C --> E[紫外图像防伪油墨检测] C --> F[透射图像红蓝光 纤维检测] D --> G{上述检测是否全部通过} E --> G F --> G G -- 否 --> H[提示报错信息] H --> I[上传到票据数据库进行备份] G -- 是 --> J[信息识别] J --> K[票据类型识别] J --> L[所属结构识别] J --> M[票号识别] J --> N[账号识别] J --> O[大小写金额识别] K --> P[显示结果] L --> P M --> P N --> P O --> P P --> Q[上传到票据数据库进行备份]</pre>		
第一权利要求	1.一种基于多光谱图像的防伪票据鉴伪方法，其特征在于，所述基于多光谱图像的防伪票据鉴伪方法包括： 步骤一，对防伪票据进行图像采集，获得不同光谱下的防伪票据图像； 步骤二，对获取的多幅票据图像进行预处理，包括色彩补偿、几何校正、滤波降噪； 步骤三，在红外和紫外图像上提取防伪特征，并融合多特征判断防伪票据真伪； 步骤四，在白光图像上检测文本区域，并识别票面文本信息，所述票面文本信息包括票据种类、所属机构、票号、账号、金额； 步骤二对获取的多幅票据图像进行预处理的方法具体包括： 第一步，对防伪票据图像进行色彩补偿：利用黑白校正纸，确定校正系数，对白光正面反射图像逐点进行色彩补偿，对色彩补偿后的防伪票据图像进行几何校正；		

	第二步，防伪票据图像几何校正：基于目标区域的四个角点与校正后的四个角点获得透视变换矩阵，并进行透视变换，得到校正后的票据图像； 第三步，对校正后的防伪票据图像进行滤波降噪处理； 第一步具体包括： 1)利用黑白校正纸，确定校正系数，CIS 传感器感光元器件有 N 个，利用线性模型 $y = kx + b$ 对图像的每个像素点进行色彩补偿，其中 k，b 均为校正系数，x 为采集到的像素值，y 表示实际的像素值，对每一个感光元器件，计算校正系数 k_i 和 b_i，产生 $2N$ 个校正系数； 首先采集校正纸黑色部分图像，获得对应图像，即获得当 $x = 0$ 时，y 的像素值，然后采集校正纸白色部分图像，获得对应图像，获得当 $x = 255$ 时，y 的像素值，再计算获得对应像素的校正系数； 2)对白光正面反射图像逐点进行色彩补偿， 根据校正系数和线性模型 $y = kx + b$，对白光正面反射图像逐点进行色彩补偿，计算得出实际的像素值，其中 x 为采集到的像素值，y 为实际的像素值； 第二步具体包括： a)获取色彩补偿后的防伪票据图像，并对其进行灰度化处理； b)对原始灰度图像，沿着四条边的垂直方向以一定间隔进行采样搜索，确定目标区域即票据图像区域的边界采样点； c)对每一个方向的边界采样点进行最小二乘拟合，确定目标区域的边界直线； d)求取四条边界直线的交点，即为目标区域的四个角点； e)基于目标区域的四个角点与校正后的四个角点获得透视变换矩阵，并进行透视变换，得到校正后的票据图像； 第三步，对校正后的防伪票据图像进行滤波降噪处理中，利用自适应中值滤波对校正后的防伪票据图像进行滤波处理； 步骤三具体包括： A)获取预处理后的防伪票据图像，包括红外光正面反射图像、红外光透射图像、紫外光正面反射图像、紫外光透射图像、白光正面反射图像； B)在红外图像上提取待检测区域的特征，并与模板图像进行 SURF 特征点匹配，若匹配点数超过给定的第一阈值，则判断匹配成功，该处存在真实的防伪特征；否则防伪检测不通过，返回上层应用并提示报错信息； C)在紫外图像上提取待检测区域的特征，并与模板图像进行 SURF 特征点匹配，若匹配点数超过给定的第二阈值，则判断匹配成功，该处存在真实的防伪特征；否则防伪检测不通过，返回上层应用并提示报错信息； D)将紫外图像转换到 HSV 通道，提取红蓝颜色分量并进行阈值分割，然后在二值图像上使用形态学开运算滤除细小物体，并采用连通域分析法搜索定位目标，计算每个可能目标的尺寸参数，并与给定阈值比较，判断是否为荧光纤维丝； E)若上述所有防伪特征均检测通过，则进行下一步票面信息识别，否
--	--

	则提示报错信息，并将错误信息和对应防伪票据图像传输至票据数据库进行备份。
--	--------------------------------------

发明专利 **CN110598699B** 主要针对现有技术防伪票据鉴伪中利用防伪特征种类单一，导致鉴伪准确率和可靠性差，匹配速度慢，无法实现实时检测，且缺乏全面的票面信息识别和统一数据管理的问题进行探究。

华中科技大学运用了采用多光谱图像采集和预处理技术，提取红外和紫外光谱下的防伪特征，并结合白光图像识别票面文本信息，通过 **SURF** 特征点匹配和多特征融合实现快速鉴伪，建立票据数据库进行数据管理，提高了防伪票据的鉴伪准确率和可靠性，实现了实时检测，能够有效识别多种票据类型和票面信息，确保了数据的统一管理。

6.6 本章小结

从多模态融合技术应用场景来看，医疗健康领域利用 **AI** 多模态融合技术提升诊断准确性和效率，相关专利多涉及医学影像分析和数据融合技术。智慧交通领域，该技术助力自动驾驶和车路协同，优化交通管理，主要专利集中在自动驾驶感知和决策系统。工业制造中，它实现生产过程的智能化监控和优化，相关专利多与工业机器人和生产流程管理相关。消费电子领域，技术提升产品交互体验，专利多涉及智能语音助手和图像识别。金融投资则借助该技术进行风险评估和市场预测，相关专利集中在金融数据分析和智能投顾系统。整体而言，**AI** 多模态融合技术在各领域展现出强大的应用潜力，推动多个行业向智能化发展。