

人工智能大模型 产业创新价值研究报告

速途网 x 大模型之家 联合发布

目录

第一节 大模型发展概况

- 大模型的缘起
- 大模型发展要素
- 一场关于“工程”的革命

第二节 大模型发展趋势分析

- 大模型市场规模预测
- 大模型产业发展趋势
- 大模型发展格局分析

第三节 国内大模型产业分布

- 大模型产业发展架构体系
- 大模型产业链全景图
- 大模型商业产业价值分析
- 大模型落地场景案例

第四节 大模型融资版图

- 2023大模型融资状况
- 资本机构观点汇总

第五章节 大模型AQUA评价体系

前言

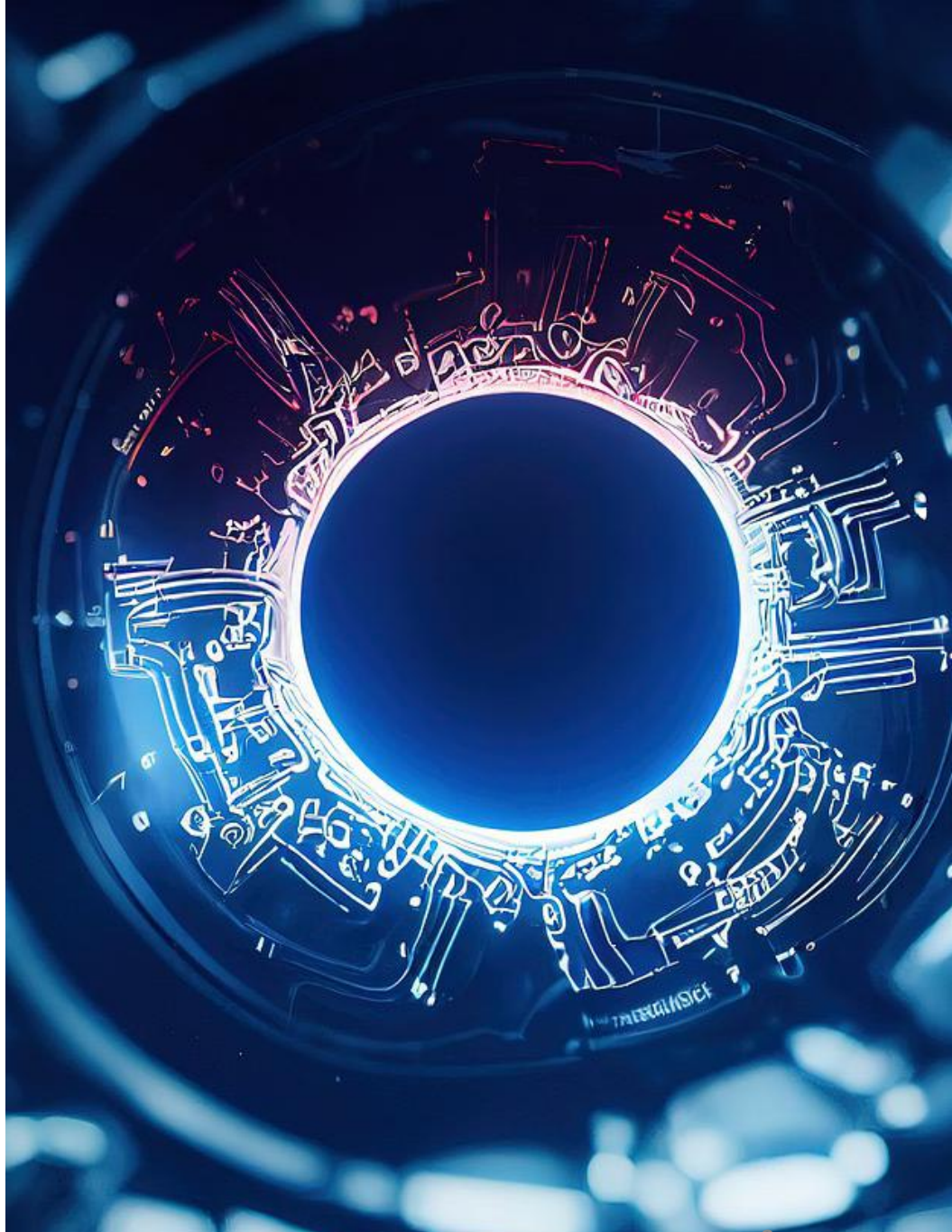
这是一个属于AI的时代。在2022年底，诸如ChatGPT、Midjourney、Stable Diffusion等大型模型的相继亮相，掀起了人工智能大模型的发展热潮。

国内的企业也积极投入到大模型的开发与落地中来。百度、商汤、360、云知声、科大讯飞等知名企业纷纷加入到这一领域，并取得了重要的成果，推动了人工智能技术在各行各业的落地应用。

从语言理解到图像识别再到自然语言处理，AI大模型在千行百业中都展现出了其巨大的潜力和迅猛的发展趋势。大模型开始对千行百业带来了潜移默化的影响，也承担起产业结构升级的重任，进一步促进数实深度融合。

人工智能大模型必将引发新一轮的效率革命。在这个充满机遇和变革的时代，我们需要深入理解人工智能大模型的发展趋势和应用场景，以应对未来的挑战并抓住机遇。

为此，速途网联合AI大模型前沿媒体大模型之家，以及速途元宇宙研究院专业分析师团队策划编写《AI大模型产业创新价值研究报告》，全方位展现AI时代大模型前沿趋势，解构大模型在产业侧的创新价值。



核心观点

概论

人工智能大模型是指在机器学习和深度学习领域中，利用大规模数据和复杂网络结构构建的庞大神经网络模型。

人工智能大模型的发展，在很大程度上是一场工程革命，而非单纯的科学革命。

全球市场

根据大模型之家测算，预计2023年，全球人工智能大模型市场规模将达到**210亿美元**，并在2028年使大模型市场规模达到**1095亿美元**。

国内市场

根据大模型之家测算，预计2023年，中国大模型产业市场规模将达到**147亿元人民币**，并在2028年达到**1179亿元人民币**。

规模

过去十余年中，大型语言模型的数据量增长了**数百万倍**，这种指数级增长的趋势在其他领域的大模型中也同样存在。当大模型数据量的量变积累到一定程度，便会引发通用性的质变。

落地

在实际应用中，参数规模并非越大越好，需要综合考虑多个因素来确定最适合的模型规模。针对不同行业的场景特点，进行有针对性的知识增强在解决现阶段问题中将发挥重要作用。

展望

随着技术的演进，大模型产业将朝着推进构建和部署模型的自动化进程，降低行业用户获得人工智能能力的门槛的方向发展。

核心观点

算力 激增

据大模型之家预测，全球AI芯片市场规模有望在2025年突破千亿大关。同时，算力芯片领域的大国角力，预示推动中国人工智能算力的高质量发展成为当务之急。

渗透 率

根据大模型之家的调研结果显示，各行业对于大模型的接受程度和渗透速度存在差异，但整体上与互联网时代的行业渗透率相似。目前，虽然仅有6.4%的企业表示无意涉及人工智能领域，但仍有46.3%的企业表示仍处在观望阶段，且仅有12.4%的企业表示全面拥抱人工智能。中国人工智能大模型仍然拥有极大的发展潜力。

投融资

据大模型之家统计，截至2023年6月30日，我国人工智能行业公开投融资规模近300亿元，人工智能领域共有286件融资交易事件，多集中于天使轮、A轮等早期融资，大模型创业热潮正处在进行时。

发展 现状

截至2023年6月30日，国内数据量超10亿的大模型已经超80个，涉及产学研多个领域，主要集中在通用领域，涵盖自然语言处理（NLP）、计算机视觉（CV）和语音识别等技术领域，聚焦通过创新来提高效率、降低成本。

核心观点

2大路径推动大模型场景落地

链接知识库，打造行业大模型

知识库是一种静态、结构化的知识表示方式，可以存储和组织各种领域的专业知识。通过将大模型与知识库相链接，可以使大模型具备更强的专业知识和推理能力，从而提高大模型在特定领域的表现和适应性。

例如，在医疗领域，可以将大模型与医学知识库相链接，使大模型能够理解和使用医学术语、概念、规则等，从而提高大模型在医疗场景中的准确性和可信度。

增强联网能力，增强时效与灵活性

联网是一种动态的知识获取方式，可以使大模型不断地从互联网上获取和更新最新的知识和信息。通过增强联网的能力，可以使大模型具备更强的时效性和灵活性，从而提高大模型在不同场景中的应变和创新能力。

例如，在搜索领域，可以使大模型能够实时地从互联网上获取和分析最新的新闻事件、舆情、趋势等，从而提高大模型在新闻场景中的敏锐性和多样性。

核心观点

大模型解决人工智能产业的3大痛点

1

用泛化能力解决数据不足

2

优化开发环境降低开发成本

3

多模态能力满足多元场景需求

核心观点

大模型将引领产业的**5大革命**

- 1 大模型将引发数据的革命
- 2 大模型将引发算力的革命
- 3 大模型将引发人工智能多模态、多场景的革命
- 4 大模型将引发产业效率的革命
- 5 大模型将引发用户习惯的革命

核心观点

大模型产业升级的6大预判

- 1 从专家到通才：多MoE模型构成超级大模型的实现路径
- 2 “具身AI”或成为大模型发展的新技术方向
- 3 全球算力市场争夺“一触即发”
- 4 云计算、分布式计算将灵活解决算力需求
- 5 MaaS将成为新的商业模式
- 6 大模型产业价值在于与场景结合，降低AI使用门槛，提升行业效率

第一章 从量变到质变的奇迹

大模型发展概况

AI大模型概述

人工智能大模型是指在机器学习和深度学习领域中，利用大规模数据和复杂网络结构构建的庞大神经网络模型。大模型是在智算算力驱动下最为典型的重大创新。得益于模型泛化能力强、长尾数据的低依赖性以及下游模型使用效率的提升，大模型被认为具备了“通用智能”的雏形，并成为业内探索实现普惠人工智能的重要途径之一。

相较于传统人工智能，大模型拥有如下特点：

参数规模巨大

人工智能大模型的参数规模远超过传统的人工智能模型。传统机器学习算法，如逻辑回归、支持向量机等，参数数量相对较小。大模型基于深度学习技术，通过增加网络层数和参数数量来提高模型的表达能力。

表达能力更强

基于海量的参数规模，大模型具有更强大的表达能力，能够更好地捕捉数据中的复杂模式和特征，从而提高模型的准确性和性能。

预训练与微调

大模型在预训练阶段，使用大规模无标签数据对模型进行初始化，使其具备一定的语言和知识理解能力。再通过微调阶段使用特定领域的标签数据来进一步优化模型，在具体任务上能够展现出更好的性能。

上下文理解与生成

大模型在自然语言处理任务中拥有更强大的能力，相比传统的基于规则或统计的方法，大模型能够更好地处理语言的上下文信息，生成更自然、连贯、富有逻辑的文本结果。

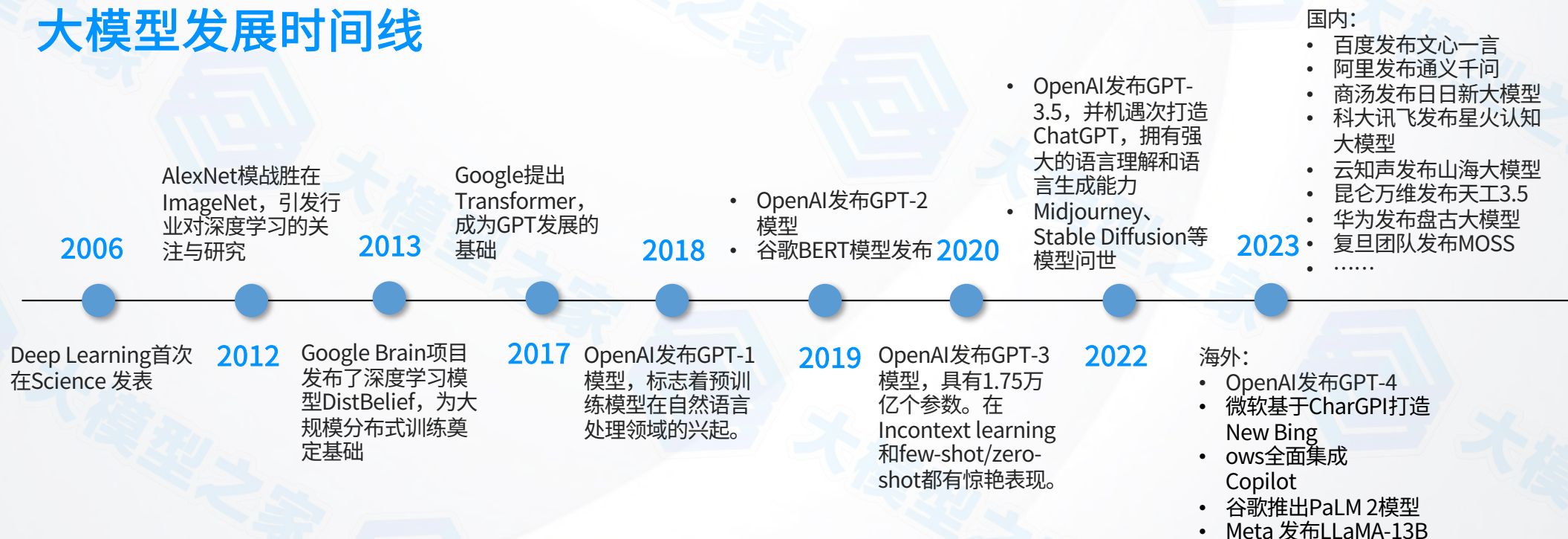
资源和计算开销更多

大模型的巨量参数规模，使其计算资源和存储空间的要求也相应增加。训练和部署大模型需要强大的计算设备和高速存储系统。这在一定程度上增加了开发和应用人工智能大模型的成本和技术挑战。

AI大模型缘起

从深度学习的突破开始，大规模的神经网络模型在图像识别、语音识别和自然语言处理等领域取得了显著进展。随后，预训练模型的兴起引领了自然语言处理的革命，如GPT和BERT等模型的问世改变了文本生成和理解的方式。近年来，大模型的规模和性能不断提升，推动着人工智能大模型的快速发展，为实现更强大的智能应用和人机交互打下了基础。

大模型发展时间线



大模型发展阶段

人工智能大模型的发展经历了多个关键阶段。最初的人工智能主要依赖于基于规则和统计的算法迭代。然而，自2012年以来，深度学习的突破开启了人工智能的新篇章，进入了快速发展期。2017年，Transformer模型的问世引领了预训练模型的兴起，推动了自然语言处理领域的革新，如Seq2Seq、GPT和BERT等模型的问世改变了文本生成和理解的方式。随后，2020年GPT-3大模型的出现带来了突破性的理解和生成能力，还让大模型首次拥有了颠覆性的涌现能力。如今，随着国内外越来越多的玩家入局大模型，为大模型的发展按下了“加速键”，在推动技术进步的同时，还推进大模型面向千行百业的应用拓展，开启人工智能大模型发展的新篇章。随着技术的演进，大模型产业将朝着推进构建和部署模型的自动化进程，降低行业用户获得人工智能能力门槛的方向发展。



大模型是一场工程革命

人工智能大模型的发展，在很大程度上是一场工程革命，而非单纯的科学革命。

大模型的成功并非只依赖于科学理论的突破，工程因素在其中起到了至关重要的作用。通过工程实践的不断探索和优化，这包括数据集的构建和清洗、计算资源的扩展和优化、训练策略的改进等方面大模型基于现有的深度学习框架和技术，通过模型设计、训练策略、硬件优化等方面的工程手段，不断提升模型的规模、性能和应用领域。

同时，人工智能大模型的应用场景需要考虑实际需求的需求，进行模型的定制和调整，使其能够在现实场景中发挥最佳效果。



第二章节

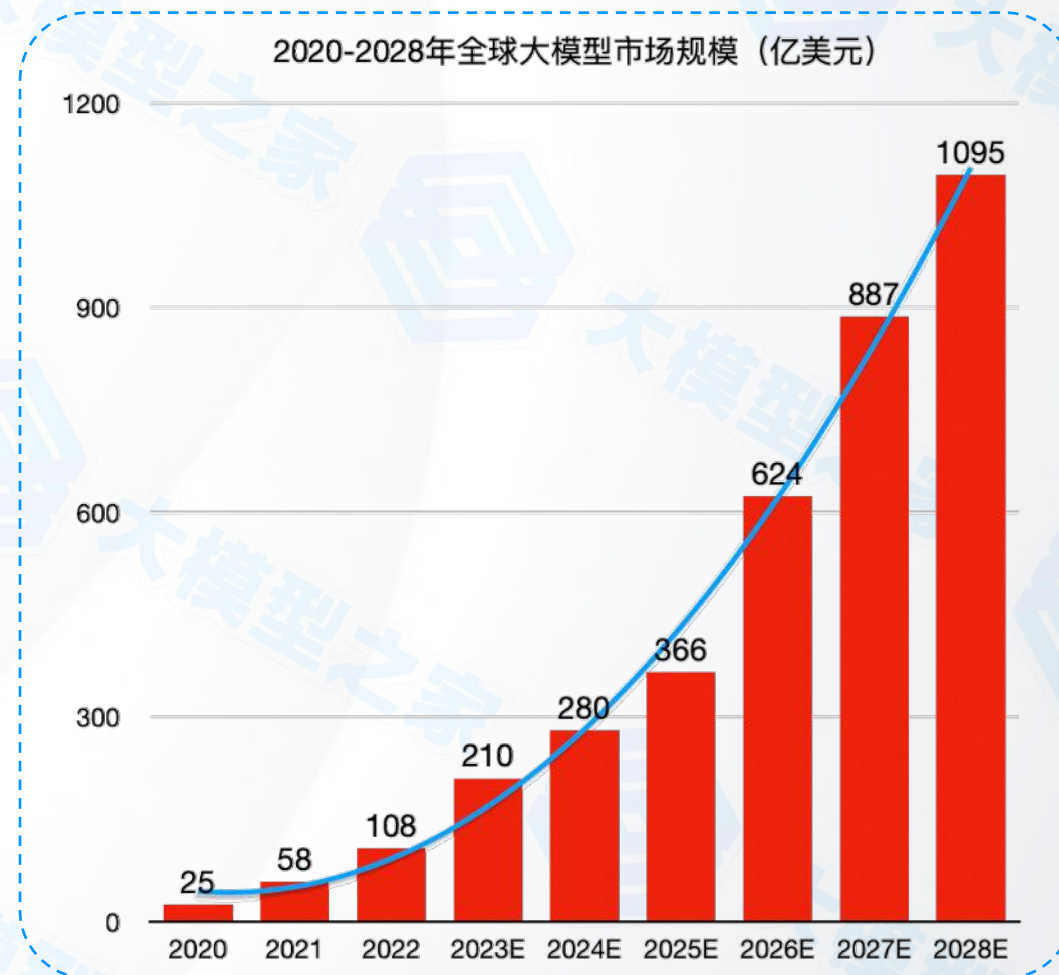
百川朝海，流行不止

大模型发展趋势分析

全球大模型市场规模2028年将突破千亿美元

在过去几年，人工智能大模型市场经历了快速增长。从2020年起，大型预训练模型在涵盖了自然语言处理、计算机视觉、语音识别、推荐系统等多个领域展现出的卓越性能，引发行业广泛关注，推动市场迅速扩大。

根据大模型之家测算，预计到2023年，全球人工智能大模型市场规模将达到210亿美元，并且预计随着大模型的进一步发展和技术的不断创新，为企业和组织提供了更强大的数据分析、预测能力和智能化解决方案，为人工智能领域带来更多商机和发展空间，持续为市场贡献增长动力。这一趋势将推动人工智能领域的快速增长，并在**2028年使大模型市场规模达到1095亿美元**。

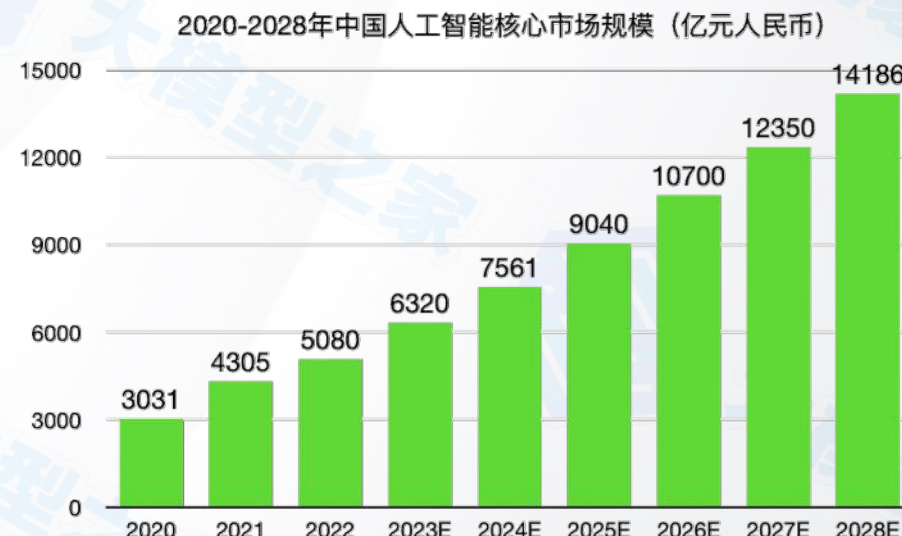
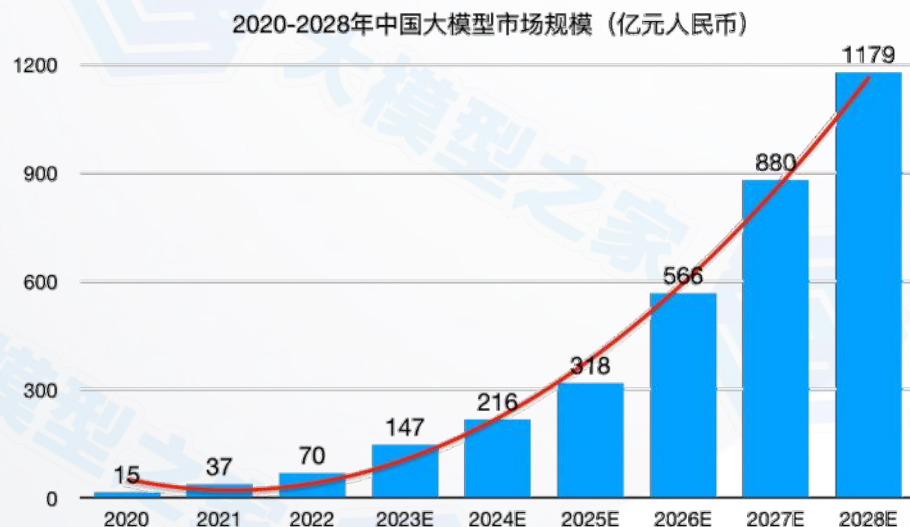


中国市场规模2028年将达1179亿元人民币

世界人工智能发展看中国，中国庞大的市场需求和丰富的人才储备为发展大模型提供了有利的客观条件。
根据大模型之家测算，中国大模型产业市场规模将达到147亿元人民币，并在2028年达到1179亿元人民币，平均增速高于世界平均增长速率。

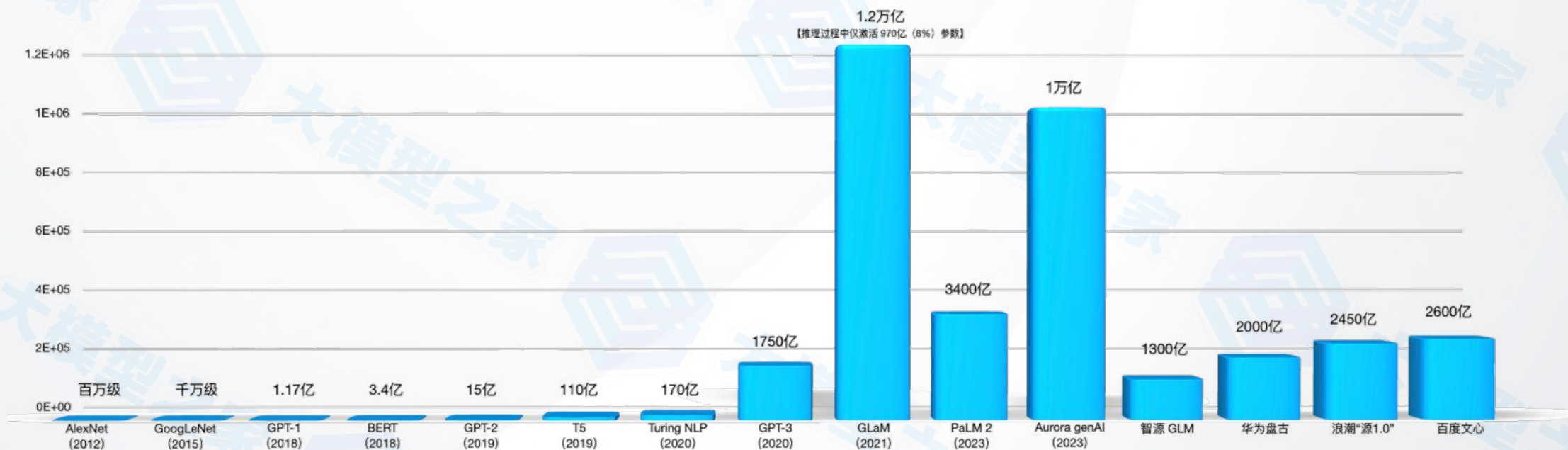
目前，中国和美国在大模型领域的研发数量占据全球总数的80%以上，其中中国的大模型数量仅次于美国。

同时，政府在人工智能领域的支持和投资，以及科技企业的积极推动，加强了人才培养和引进，促进产学研用结合，为中国大模型产业的发展提供了强有力的支持。随着技术的进一步突破和创新，中国将有望在大模型领域取得更多的成果，并与全球领先国家共同推动人工智能大模型的发展和应用。



大模型整体参数规模持续增长，十年增长百万倍

随着数据的积累和采集技术的进步，大模型所使用的数据集规模也不断扩大。以自然语言处理领域为例，**过去十余年中，大型语言模型的数据量增长了数百万倍**，这种指数级增长的趋势在其他领域的大模型中也同样存在。从2012年的AlexNet的百万级参数，到如今模型参数规模正在逐步进入万亿级时代。当大模型数据量的量变积累到一定程度，便会引发通用性的质变，使得大模型能够通过增量式学习和持续训练的方式不断吸收新的数据并更好地学习和理解不同领域的知识和规律，及时适应新的数据和场景变化，保持其性能的稳定和持续提升。



数据的质量与处理能力将成为大模型决胜因素

大模型将引发数据的革命。

数据的质量以及数据清洗的工程化能力会显著拉开大模型预训练阶段的效果差距将成为决胜关键因素。

大模型的发展将为数据相关产业和行业带来多重机遇：

- 1.数据分析和洞察力：**大模型可以处理更大规模、更复杂的数据集，并从海量的原始数据中学习提取出有用的特征和知识，实现更精准和全面的模型表达。
- 2.数据清洗和预处理：**大模型对数据质量的要求更高。因此，数据清洗和预处理的工程化能力将变得更为重要。
- 3.数据驱动的创新：**大模型可以通过对海量数据的学习和分析，为企业和行业带来全新的创新机会。
- 4.数据安全和隐私保护：**随着大模型使用的普及，数据安全和隐私保护将成为一个重要的议题，保护个人隐私和敏感信息的技术和法律措施将得到进一步加强，以确保数据的合规性和安全性。
- 5.数据驱动的行业转型：**大模型的引入将促使许多行业向数据驱动的模式转型。从金融、零售和制造业到医疗保健、能源和交通等领域，企业将更加重视数据的价值，并通过大模型的应用来实现业务的创新和改进。



Do more with less

虽然参数量的增多可以显著提升大模型的通用性，但行业普遍认识到参数规模的增长也带来了一些挑战和限制。首先，随着参数规模的增加，对计算资源的需求也随之增加，这可能导致训练和推理过程变得更加耗时和昂贵。其次，模型部署的复杂性也会增加，需要更多的工程和技术支持来实现高效的部署和集成。

因此，在实际应用中，并不是参数规模越大，模型就一定更好用，需要综合考虑多个因素来确定最适合的模型规模。

Do More with Less

节约成本

对于数据量较小的任务或场景，使用较小的模型可以更快地进行训练和推理，同时也更节省计算资源。

部署的灵活性

一些需要将模型部署在本地设备或边缘计算设备上的场景，要求模型具有较小的参数规模，以适应有限的计算能力和存储容量。

使用需求的限制

并非所有应用都需要极高的通用大模型能力。有些任务可能只需要满足特定的需求，而不需要涵盖全部的知识 and 功能。

行业专用大模型

为了适应特定的专业场景，可以针对性地开发行业专用的大模型，为了适应特定的专业场景，可以针对性地开发行业专用的大模型。

算法的迭代

更为先进的工程算法和训练方法能够大大减少对数据量的依赖。通过优化算法和改进训练方法，可以在较小的数据集上训练出性能优秀的大模型，从而降低了对海量数据的需求。这种技术的进步为更广泛的应用提供了可能性，并且使得大模型的发展不再受限于数据量的稀缺性。

更高效的分布式训练和推理技术以及更低成本的计算资源成为刚需

大模型将引发人工智能算力的革命。 大模型参数量的增加导致训练过程的计算需求呈现指数级增长。为了快速训练大规模模型，需要强大的计算能力来支持高效的分布式训练和并行计算。随着硬件设备的进步，如高性能计算机和分布式计算平台的普及，将成为支持更大规模的模型训练和迭代的重要方式。

更高的运算需求

更高效的分布式训练技术

更先进的硬件加速技术

更高性能的存储网络

更海量合规数据共享机制

更多模态的输入与输出

更低的成本消耗

弹性计算资源满足不同需求

云服务降低计算成本

边缘计算减少延迟与带宽损耗

高效率高密度低功耗的服务器

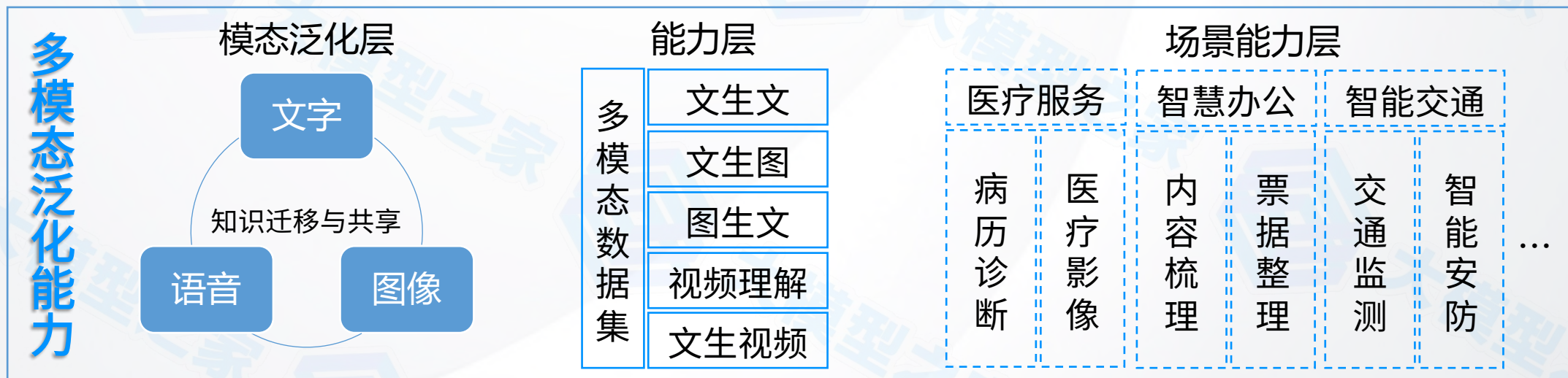
模型剪枝和量化

大模型将涉及更多的模态和场景

大模型将引发人工智能多模态、多场景的革命。

随着传感器技术的发展和物联网的兴起，大量的多模态数据（如图像、视频、声音等）被广泛采集和应用。大模型可以利用这些多模态数据进行跨模态学习，从而提升其在多个感知任务上的性能和表现。通过充分利用大模型的泛化能力、构建多模态数据集、解决融合和对齐问题，以及提供强大的计算资源支持。

大模型具有强大的泛化能力，可以在不同模态和场景之间实现知识的迁移和共享，将大模型的应用扩展到不同的领域和场景。



大模型将引发产业效率的革命

大模型将引发产业效率的革命。

通过结合多模态数据和智能算法，大模型能够赋能多个行业，为行业提质增效提供助力，推动数据与实体的融合，改变行业发展格局。

- **大模型将成为智能语言处理的核心引擎**，为各种语言相关的应用提供强大的支持。例如，在法律领域，大模型可以作为智能合同生成器，根据用户的需求和规范，自动生成合法和合理的合同文本。
- **大模型将成为AIGC的主要生产力**，为各种内容相关的应用提供丰富和多样的输出。例如，在娱乐领域，大模型可以作为智能剧本编剧，根据用户的喜好和风格，自动生成有趣和吸引人的剧本故事。
- **大模型将成为智能数据处理的专家**，为各种数据相关的应用提供准确和高效的解决方案。例如，在工业领域，大模型可以作为智能质量控制器，根据生产数据和标准，自动检测和纠正产品质量问题。
- **大模型将成为智能知识管理的领航者**，为各种知识相关的应用提供完善和系统的服务。例如，在教育领域，大模型可以作为智能学习平台，根据知识图谱和学习路径，自动推荐和组织学习资源。

一线城市用户对大模型的欲望更加强烈

大模型将引发用户习惯的革命。

根据大模型之家的统计数据，目前已有很高比例的用户了解大模型，并且有相当比例的用户已经使用过大模型。特别是在一线城市，用户对大模型的了解和使用的欲望更加普遍和强烈。



更多的智能化服务和个性化体验
吸引更多用户的关注和使用。

根据用户的喜好、兴趣和需求进行精准推荐
改变用户对技术和产品的期望和使用习惯。



推动用户习惯从传统媒体向数字化媒体的转变
用户越来越倾向于通过大模型获得定制化的新闻、信息和娱乐内容，而非传统的媒体渠道。

一静一动，推动大模型场景落地

目前，大模型产业仍然处在发展的早期阶段，但为千行百业打开了效率提升的新路径。针对不同行业的场景特点，进行有针对性的知识增强在解决现阶段问题中发挥着重要作用。通过链接知识库进行专业知识增强，打造行业大模型，或是通过增强联网的能力实时扩充大模型知识储备，大模型能够深入了解各个行业的专业知识，并将其融合到模型的学习和推理中。

链接知识库，打造行业大模型

知识库是一种静态、结构化的知识表示方式，可以存储和组织各种领域的专业知识。通过将大模型与知识库相链接，可以使大模型具备更强的专业知识和推理能力，从而提高大模型在特定领域的表现和适应性。

例如，在医疗领域，可以将大模型与医学知识库相链接，使大模型能够理解和使用医学术语、概念、规则等，从而提高大模型在医疗场景中的准确性和可信度。

增强联网能力，增强时效与灵活性

联网是一种动态的知识获取方式，可以使大模型不断地从互联网上获取和更新最新的知识和信息。通过增强联网的能力，可以使大模型具备更强的时效性和灵活性，从而提高大模型在不同场景中的应变和创新能力。

例如，在搜索领域，可以使大模型能够实时地从互联网上获取和分析最新的新闻事件、舆情、趋势等，从而提高大模型在新闻场景中的敏锐性和多样性。

从专家到通才：多MoE模型构成超级大模型的实现路径

随着模型越来越大，训练使用的样本数据越来越多，训练的开销越来越接近成本与算力的极限，而MoE（Mixture of Experts,混合专家系统）可以动态激活部分神经网络，从而实现在不增加计算量的前提下大幅度增加模型参数量。MoE 技术目前是训练万亿参数量级模型的关键技术。

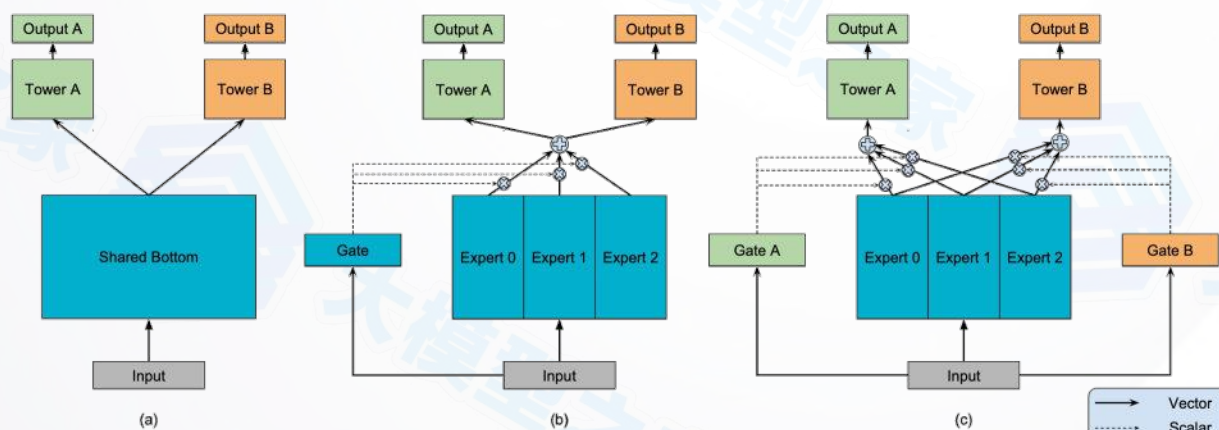


Figure 1: (a) Shared-Bottom model. (b) One-gate MoE model. (c) Multi-gate MoE model.

MoE 是一种人工智能大模型中常用的架构和方法。MoE模型的基本思想是将大模型分解为多个专家子模型，每个子模型专注于处理特定的任务或领域，专家子模型可以并行工作，或通过技术集成共同为整个大模型提供预测和决策能力。

多个MoE模型可以组合成一个极为通用的大模型。通过将多个MoE模型连接在一起，形成一个层级结构，可以构建一个更加强大和全面的大模型。每个MoE模型可以专注于不同的任务或领域，而整个大模型则可以在多个任务和领域上展现出优异的性能。

在现阶段，企业可以采取基于稍小样本量的通用大模型，并通过知识强化和整合MoE模型的方法，可以帮助企业实现对多样化场景的全面覆盖，提升解决方案的适应性和效果。这种渐进式的发展策略可以在初期快速建立起基础能力，并逐步提升和拓展模型的应用范围。

“具身AI” 或成为大模型发展的新技术方向

具身AI是指人工智能系统具备感知和互动，并能够在现实世界中与环境 and 用户进行实时交互和合作的能力。这一技术方向的发展将进一步提升大模型的智能水平，拓展其应用场景，以及深化与人类的互动。

具身智能对于AI领域的重要性在于，它代表了从简单的图像识别等机器学习任务向学习如何在多维度的现实环境中执行任务的转变。这种能力将使得AI智能体能够更自主地适应和解决不断变化的环境中的问题。

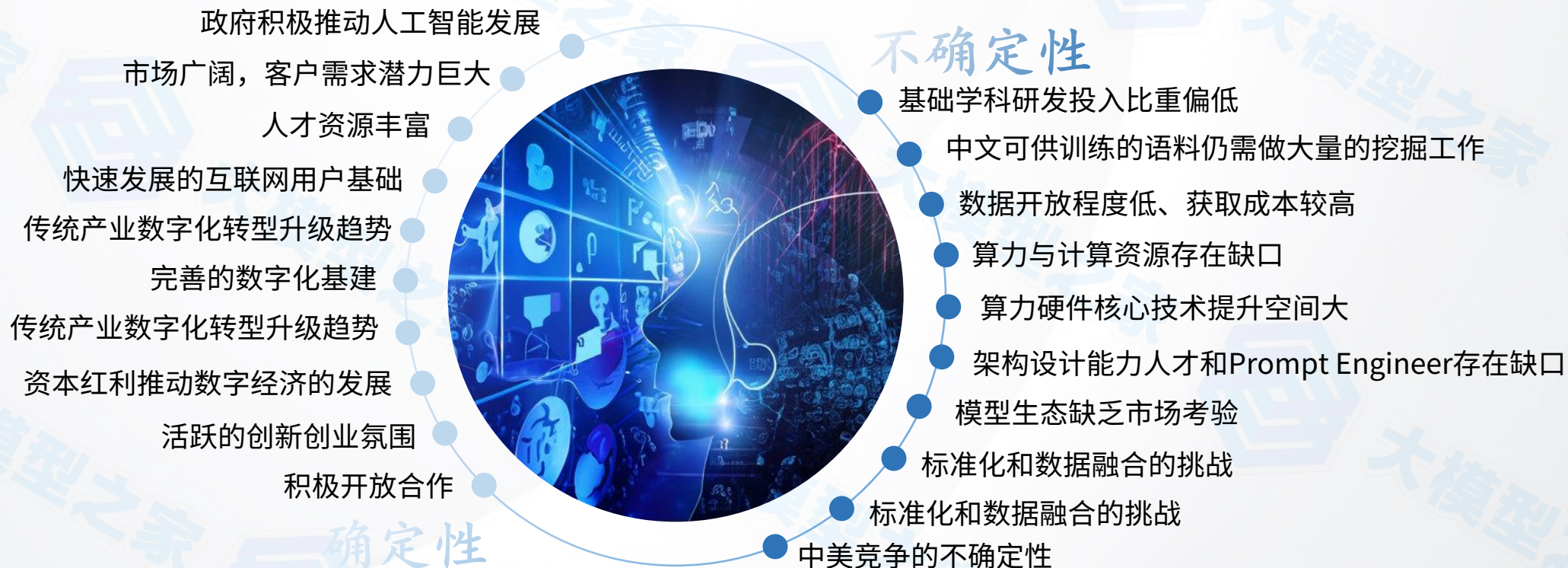
以自然语言交互的大模型为例，它们不仅具有对话的能力，还能够感知和理解环境中的物体和事件，以及根据人类的意图和需求做出合适的反应。通过人机对话，大模型可以给机器人发命令、理解机器人的反馈、分解任务变成动作、帮助机器人处理多模态的数据（比如图像、声音等）、指导机器人改正错误、提高机器人的学习能力。



大模型发展的确定性与不确定性

中国发展人工智能大模型的优势在于政策的支持、丰富的人才资源、资本红利和活跃的创新创业氛围。千行百业数字化转型需求创造了丰富的场景。此外，中国的数字化基建完善，积极开放合作，为研发与落地创造有利条件，推动大模型技术创新和经济发展。

中国大模型发展也面临着挑战。基础学科研发投入比例偏低，中文语料数据、以及算力资源的缺口，仍然有很大的提升空间。与此同时，还要面对标准化和数据融合挑战，解决安全与伦理问题，以及中美竞争不确定等共性问题。



第三章 数实融合新机遇

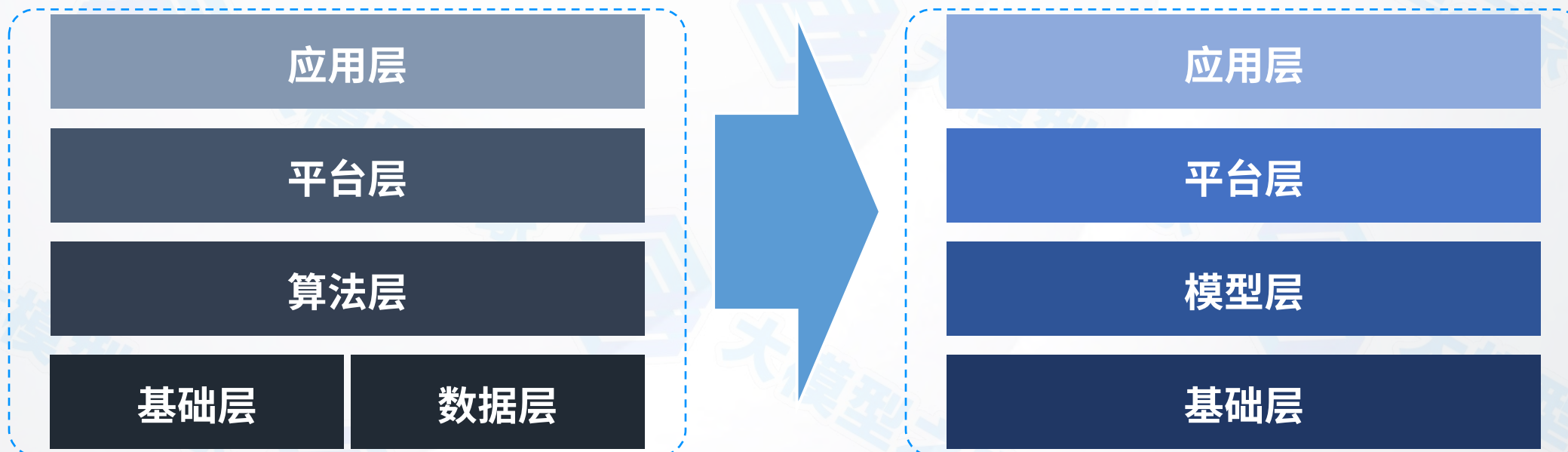
国内大模型产业分布

大模型引发的产业结构升级

大模型的出现也带来了革命性的变化。在数据层，不再需要进行繁琐的标注和加工，而是直接将海量的原始数据输入到大模型中。通过自动学习数据中的规律和知识，大模型能够获得深刻的理解和洞察。

在算法层，不再需要为不同的任务和领域设计专门的小模型，而是利用大模型的强大泛化能力。通过少量的微调或适配，大模型可以完成多种复杂的功能，极大地提高了效率和灵活性。

大模型成为了一个通用的、高效的、智能的平台，为各种应用提供强有力的支持。它的出现改变了传统算法开发和应用的方式，加速了人工智能技术的推广和落地。大模型的发展为各行各业带来了无限的可能性，为企业和组织带来了更多创新和增长的机会。

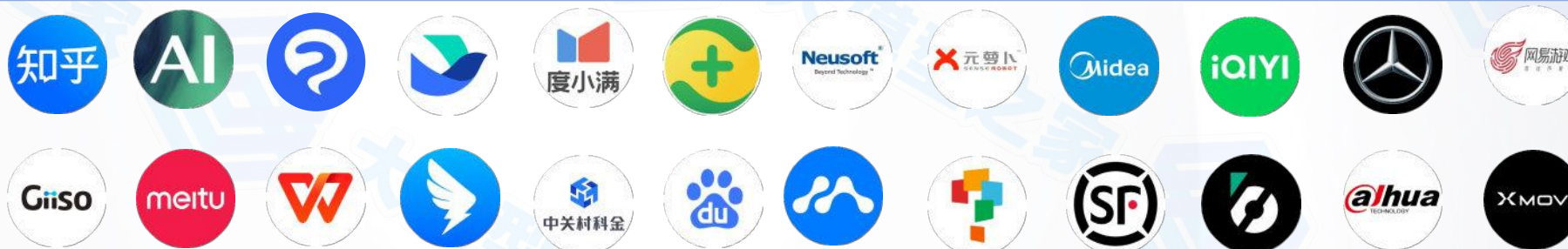


大模型产业架构简图



国内大模型产业架构图

应用层



平台层



模型层



基础层



注：以上排名不分先后

大模型产业价值在于降低AI使用门槛，提升行业效率

大模型产业价值的关键，在于降低人工智能的使用门槛，将其特征与能力与各种场景结合，以实现场景效率的提升。

大模型拥有数据的强大处理和分析能力，能够处理和分析大规模、复杂的数据，从中发现隐含的模式和规律。这使得企业能够更准确地了解市场需求、优化业务流程，做出更明智的决策，提高效率和竞争力。

大模型的多模态能力为行业提供了更多的可能性，能够处理语音、图像、文本等不同类型的数​​据，并将它们综合起来进行分析和理解。这使得在跨行业合作和创新方面有了更广阔的空间，例如在医疗行业，大模型可以结合多种医疗数据，提供更精确的诊断和治疗方案。

大模型还具备持续学习和自适应能力，随着不断的训练和迭代，大模型可以不断提升自身的性能和准确度，适应不断变化的业务需求和环境。这种灵活性和适应性使得大模型成为行业创新和优化。

大模型

直接基于大模型形成独立的产业体系，形成智能能力的生产和消费模式。

作为智能系统中的一个组成部分，集成于已有的产品或服务中，实现某一个环节的效率提升和成本降低。

大模型可以根据不同行业的特点和需求，进行定制化开发和优化，形成行业大模型。

大模型可以作为一种公共基础资源，像电力或自来水一样随取随用。

大模型解决人工智能产业三大痛点

基于海量的数据与训练，人工智能大模型拥有更强的泛化能力，能够有效解决人工智能产业面临的三大痛点：数据不足、开发成本高和跨场景应用能力弱。

用泛化能力解决数据不足

许多产业领域的数据量有限，或者数据质量不高，导致传统的人工智能模型难以训练出好的效果。人工智能大模型可以通过预训练和迁移学习的方式，利用海量的通用数据和知识，提升模型的泛化能力和通用性，降低对标注数据的依赖，从而解决数据不足的问题。

优化开发环境降低开发成本

传统的人工智能模型开发需要大量的人力、物力和时间投入，而且每个任务或领域都需要定制化的模型，导致开发成本高昂。人工智能大模型可以通过提供通用或专用的平台和工具，为开发者提供便捷高效的开发环境和资源，如数据预处理、模型训练、模型部署、模型优化等，从而降低开发门槛和成本，提升开发效率和质量。

多模态能力满足多元场景需求

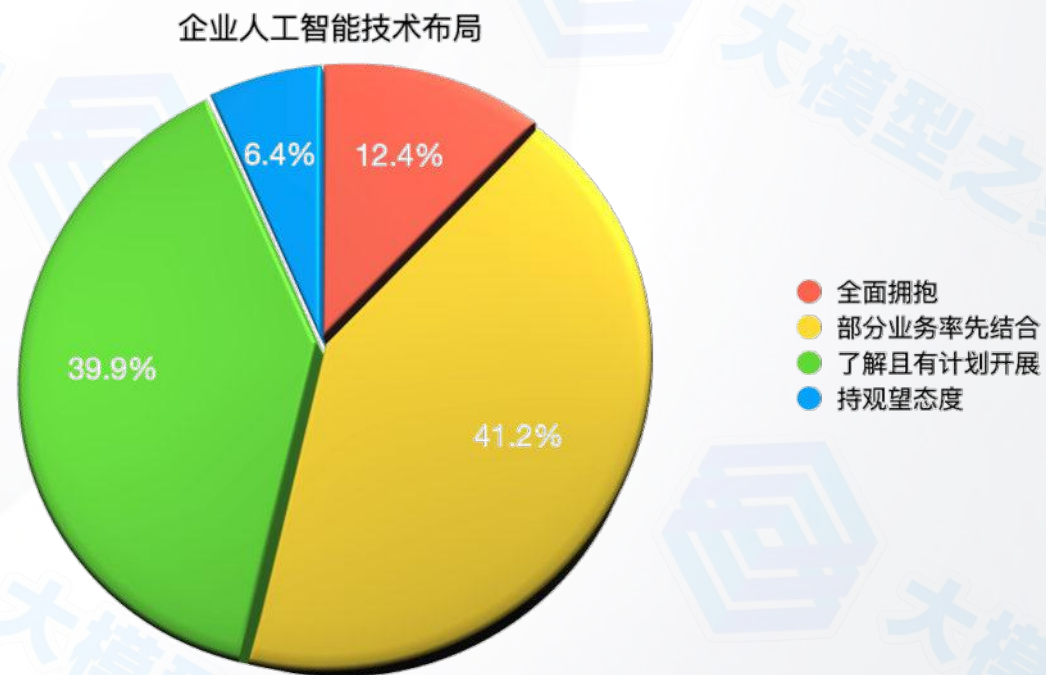
传统的人工智能模型往往只能在特定的任务或领域中表现良好，而在复杂多变的实际应用场景中，难以满足用户的需求和期望。人工智能大模型可以通过融合多种模态（如文本、图像、语音等）和多种任务（如分类、生成、检索等）的信息，实现跨模态理解和生成能力，从而提升应用效果，满足更多元场景的需求。

行业对大模型的接受程度与互联网时代渗透率相似

在互联网时代，互联网技术的普及和应用对各行业产生了深远的影响。类似地，大模型作为人工智能的前沿技术，也具有广泛的应用前景和潜力，正在改变各行业的发展方式和商业模式。中国虽然对于大模型研究起步较晚，但市场规模广阔、场景丰富，为大模型发展提供了肥沃的土壤。

根据大模型之家的调研结果显示，各行业对于大模型的接受程度和渗透速度存在差异，但整体上与互联网时代的行业渗透率相似。

大模型之家还观察到，一些先进的企业已经开始重视人工智能对于行业的重塑，并率先探索大模型技术与产业的融合。这些企业从全面或部分业务入手，利用人工智能的辅助，实现数据管理和决策的优化，从而提升企业的整体运营效率。

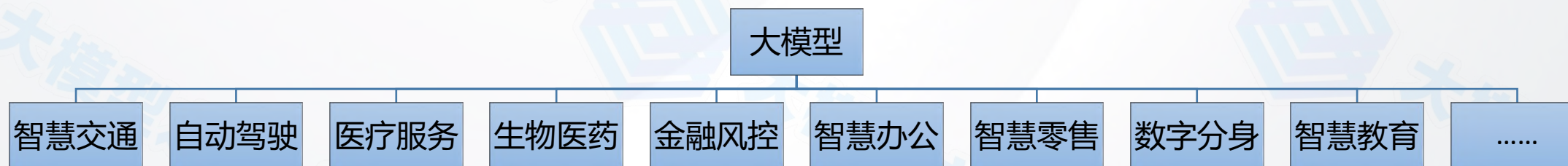


从赋能行业，到赋能百业——大模型的产业触点

随着深度学习和大规模数据的训练的加持，大模型能够实现多模态、生成式、可解释及对话式的智能交互。

而在面向更加细分的领域与场景，同时也需要通过知识图谱、迁移学习和联合学习等技术，高效融合不同垂直领域的专业知识，**构建具有领域专长和业务逻辑的专业大模型**，能够针对各行各业的特定场景和问题提供智能解决方案，从根本上降低大模型的下游应用成本和门槛，让更多的企业和机构能够便捷地接入大模型的强大能力，提升自身的效率和创新力。

以大模型为关键驱动的数字经济，与实体经济深度融合，将做强做优做大实体经济。汽车制造、能源、交通等行业，大模型可以深入核心业务场景，在智能客服、供应链、系统调度等版块创新，促进行业的数字化转型和智能化提升。



百度文心产业级知识增强大模型

2021年，百度发布“文心”系列产业级知识增强大模型，推动技术自主创新。文心从大规模知识和海量数据中融合学习，效率更高，效果更好，并具有更好的可解释性。同时，文心还通过跨语言学习，具备同时理解多种语言的能力；通过跨模态学习，实现复杂场景的多层次、高精度语义理解。

2023年5月，全球首个一站式企业级大模型生产平台“文心千帆大模型平台”发布，以文心一言为核心，提供大模型服务，帮助客户改造产品和生产流程。同时作为一个大模型生产平台，还提供大模型开发和应用的整套工具链，企业可以在文心千帆上基于任何开源或闭源的大模型，开发自己的专属大模型。

文心产业级知识增强大模型

工具 与平台	EasyDL-大模型		BML-大模型		大模型 API		
	大模型套件						
	数据标注与处理	大模型精调	大模型压缩	高性能部署	场景化工具		
文心 大模型	行业大模型						
	国网-百度·文心	浦发-百度·文心	航天-百度·文心	人民网-百度·文心	冰城-百度·文心	电影频道-百度·文心	
	深燃-百度·文心	吉利-百度·文心	泰康-百度·文心	TCL-百度·文心	辞海-百度·文心		
	NLP 大模型		CV 大模型		跨模态大模型		生物计算大模型
	文心一言 ERNIE Bot		OCR图像表征学习 VIMER-StrucText		文图生成 ERNIE-ViLG	文档智能 ERNIE-Layout	化合物表征学习 HelixGEM
	对话 PLATO	搜索 ERNIE-Search	多任务视觉表征学习 VIMER-UFO				蛋白质结构分析 HelixFold
	跨语言 ERNIE-M	代码 ERNIE-Code	视觉处理多任务学习 VIMER-TCIR	自监督视觉表征学习 VIMER-CAE	视觉-语言 ERNIE-ViL	语言-语言 ERNIE-SAT	
	语言理解与生成						
	ERNIE 3.0 Tiny ERNIE 3.0 鹏城百度·文心 ERNIE 3.0 Zeus						

商汤“大模型+大装置”多模态赋能产业升级

商汤大模型体系在其“大模型+大装置”的AGI战略布局之下，持续推动AI基础设施能力的跃进提升，不仅打造通用能力更加强大的基础模型。能够大幅降低人工智能生产要素的成本，从而实现高效率、低成本、规模化的AI创新和赋能。

“商汤日日新SenseNova”大模型体系全面升级：

	商汤商量 SenseChat 2.0	突破模型输入长度的限制，并推出不同参数量级模型，按需部署降低成本
	商汤秒画 SenseMirage 3.0	模型参数从10亿提升至70亿量级，能够实现专业摄影级的图片细节刻画
	商汤如影 SenseAvatar 2.0	语音和口型流畅度提升30%以上，4K高清视频效果，AIGC形象及数字人歌唱功能
	商汤琼宇 SenseSpace 2.0	空间重建效率提升20%，渲染性能提升50%
	商汤格物 SenseThings 2.0	对小物体的纹理及材质还原达到毫米级精细度，突破对高反光和镜面物体采集难题

.....

“商汤日日新SenseNova”模型体系赋能产业：

领域	应用
金融	接入大语言模型能力，提供投研分析、研报撰写等功能
医疗	“大医”大模型提供导诊、问诊、健康咨询、辅助决策等能力
移动终端	针对信息获取的问答交互，针对生活场景的知识交互，针对语言和图像生成的内容交互等
线下场景	为巡检带来长尾故障识别、复杂缺陷判断等智能解决方案。基于琼宇2.0的空间重建，打造实景空间的数字孪生。
短视频	共建“云+AIGC+短视频直播”生态，打造更高效、低成本、便捷易用的AI视频和营销工具
智能汽车	通过大模型打造“车舱大脑”，支持快速定制的数字人进行拟人化交互

大模型与千行百业推进程度（部分）

对于大模型接受程度更高的行业通常有以下特点：

- 数据量大、质量高、多样性强，可以为大模型提供充足的训练和微调数据。
- 技术需求高、创新能力强、竞争激烈，可以为大模型提供强烈的需求和动力。
- 规范完善、监管严格、社会责任高，可以为大模型提供良好的应用和推广环境。
- 客观、偏理性，需要更准确、更有逻辑的分析和建议的场合。

搜索引擎 ★★★★★

数据训练 ★★★★★

云服务 ★★★★★

金融保险 ★★★★★

数字员工 ★★★★★

医疗服务 ★★★★★☆

艺术设计 ★★★★★☆

数字电商 ★★★★★☆

在线教育 ★★★★★

智慧办公 ★★★★★

网络媒体 ★★★★★

人力资源 ★★★★★

生物医药 ★★★★★

智慧交通 ★★★★★

自动驾驶 ★★★★★☆

先进制造 ★★★★★☆

社交软件 ★★★★★☆

影视传媒 ★★★★★☆

直播演艺 ★★★★★

营销服务 ★★★★★

航空航天 ★★★★★

房地产 ★★★★★☆

关怀陪伴 ★★★★★☆

农业生产 ★★★★★☆

大模型与智慧交通行业融合

大模型利用强大的计算能力和深度学习算法，能够从海量的交通数据中提取有效的特征，并实现智能交通管理的精确化和自动化。通过对交通流量、道路状况、车辆行为等数据进行分析 and 建模，大模型能够准确预测交通拥堵情况、实现智慧的交通规划。

大模型在车辆安全与预警领域也能够起到重要作用。通过对交通数据和车辆传感器数据的实时监测和分析，大模型能够识别危险驾驶行为、预测交通事故概率，并通过智能预警系统及时发出警告，有效降低交通事故的发生率，提高整个智慧交通系统的可靠性和安全性。

此外，大模型在交通预测与调度方面也有广泛的应用前景。通过对历史交通数据和实时信息的分析，大模型能够准确预测交通拥堵情况、人流量等，为交通管理者提供合理的路网调度和交通资源分配方案。

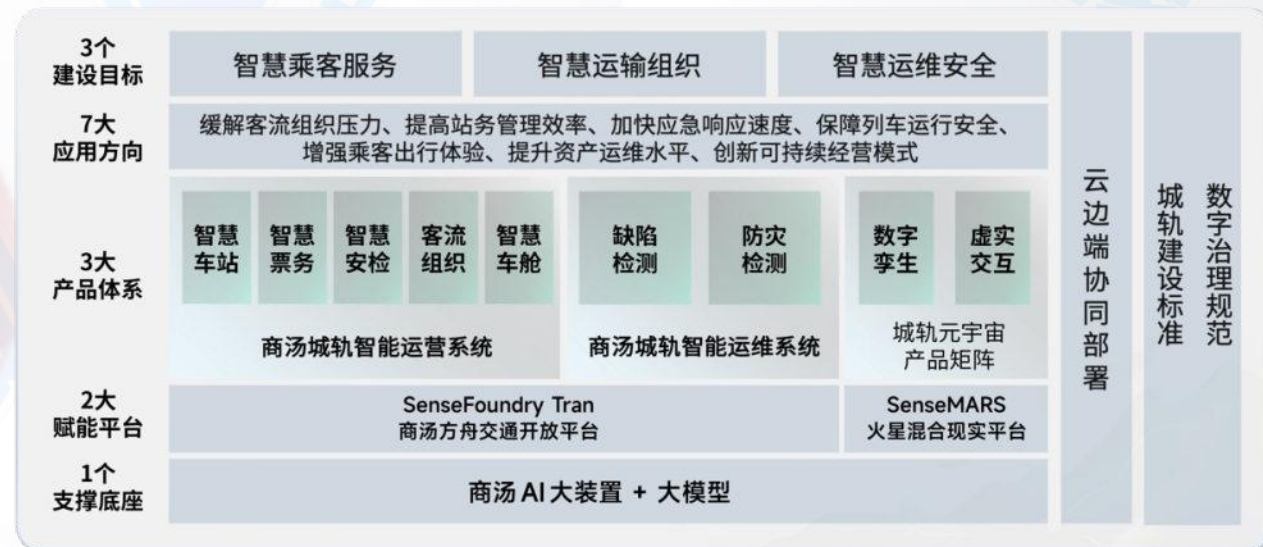


商汤日日新大模型体系赋能智慧城轨高质量发展

商汤科技依托于大装置SenseCore和大模型日日新SenseNova大模型体系，构建了开放式平台架构模式，实现了智慧城轨的视觉引擎。利用大模型，可以实现对车厢内的打架斗殴、人员摔倒、物品和人员滞留等异常/违规行为的主动提醒。

基于大装置构建的SenseFoundry Tran方舟交通开放平台和SenseMARS火星混合现实平台，可分别实现高效、低成本的多场景算法生产满足城轨可持续经营目标。商汤利用视觉AI解决方案，能够解决城轨的视觉感知问题，包括客运组织压力、站务管理效率、乘客出行体验、列车运行安全、应急响应速度、资产运维水平和创新经营模式，降低了应用建设门槛，满足城轨交通场景的多样化需求。

目前，商汤科技城市轨道交通智能化建设，已为郑州、哈尔滨、西安、成都、太原等城市的超过30条线路、超过640个车站实现了多种场景和方向的智能化升级。



大模型与自动驾驶行业融合

在自动驾驶领域，大模型可以实现端到端的系统，即直接从传感器数据作为输入，输出期望的驾驶行为，如转向、加速和刹车。这样可以避免传统的分层设计，简化系统复杂度，提高系统鲁棒性。

提高感知和决策的准确性：人工智能大模型可以利用自监督学习、强化学习等技术，从大量的标注和未标注数据中学习更丰富和泛化的特征表示，从而提高对场景、物体、行为等的感知和决策的准确性。

降低数据标注和开发的成本：人工智能大模型可以利用预训练和微调的方式，实现对不同场景和任务的快速适应，从而降低数据标注和开发的成本。

创新应用和商业模式：人工智能大模型可以利用生成式、对抗式等技术，实现对真实场景的仿真和增强，从而创造新的应用和商业模式。



百度文心大模型扩充自动驾驶感知能力

百度推出的文心大模型-图文弱监督预训练模型，背靠文心图文大模型数千种物体识别能力，大幅扩充自动驾驶语义识别数据，如：特殊车辆（消防车、救护车）识别、塑料袋等，自动驾驶长尾问题解决效率实现了指数级提升；此外，得益于文心大模型-自动驾驶感知模型10亿以上参数规模，通过大模型训练小模型，自动驾驶感知泛化能力也显著增强。

该方案的数据提纯路径利用车端小模型和云端大模型，实现高效率数据挖掘和自动化标注；数据消化架构实现自动化训练，具备联合优化和数据分布理解的能力，有效地利用高纯度数据进一步提升自动驾驶系统的整体智能水平。

百度Apollo还基于文心大模型能力，开发出场景对话式交互、策略规划、知识问答和新车导购等多项能力。在大模型融入智舱场景后，百度Apollo智能座舱将演变为“第三生活空间”的核心载体。

高提纯、高消化是数据闭环的核心驱动力

基于小模型和大模型的车-云协同挖掘策略和自动化标注，提高数据纯度

实现数据、模型和指标的集中式、端到端整合是数据消化的重要途径

通过训练、推理和数据分布，构建基于数据-模型-指标的消化反馈机制



毫末智行DriveGPT雪湖·海若

毫末智行Haomo.ai发布的自动驾驶生成式大模型——DriveGPT雪湖·海若，基于MANA数据智能体系，可以通过从人类反馈中进行强化学习，提升自动驾驶的认知决策模型。可有效节省90%的标注成本，将算法对于模糊车道线的识别能力从40%左右提升到90%以上，将驾驶轨迹的预测误差降低近3倍。

在预训练阶段，通过引入量产驾驶数据训练初始模型，再通过引入驾驶接管 Clips 数据完成反馈模型（Reward Model）的训练，然后再通过强化学习的方式，使用反馈模型去不断优化迭代初始模型，形成对自动驾驶认知决策模型的持续优化。

同时，DriveGPT 雪湖·海若还会根据输入端的提示语以及毫末 CSS 自动驾驶场景库的决策样本去训练模型，让模型学习推理关系，从而将完整驾驶策略拆分为自动驾驶场景的动态识别过程，完成可理解、可解释的推理逻辑链生成。

截至2023年4月，毫末 DriveGPT 参数规模达到1200亿，预训练阶段引入4000万公里量产车驾驶数据，RLHF 阶段引入5万段人工精选的困难场景接管 Clips。后续还将持续整合毫末多个大模型的能力。



大模型与自动智慧零售融合

在智慧零售领域，人工智能大模型可以实现对消费者行为、市场趋势、竞争动态等的深度分析和预测，从而提供个性化、高效、安全的零售体验。

提高客户满意度和忠诚度：人工智能大模型可以利用行为分析和客户智能等技术，实现对客户需求、偏好、反馈等的全面理解，从而提供定制化的商品推荐、优惠券、促销活动等。

优化库存管理和供应链：人工智能大模型可以利用预测分析和机器学习等技术，根据市场、消费者和竞争者的数据，预测行业变化，并对营销、商品和业务策略进行主动调整。

提升安全性和合规性：人工智能大模型可以利用图像识别和自然语言处理等技术，实现对店内外的视频监控、商品标签、发票等的智能分析和识别，从而防止盗窃、欺诈、假冒等风险，并符合相关法规。



京东深植数智供应链打造产业大模型

京东云旗下言犀人工智能应用平台将整合产业实践和技术积累，推出千亿规模言犀大规模预训练语言模型，并基于“125”路线图，以产业AI力量，加速中国人工智能技术开发与落地，推动实体经济的发展。

零售场景：实现了千万级自营SKU在供需不确定场景下的精准预测、智能决策以及高效协同，采购自动化率达到85%，助力京东库存周转降至31.2天。

物流场景：通过供应链中台、数据算法中台和云仓、物流等供应链生态，有效提升拣货效率达3倍以上。

营销服务场景：智能客服系统基于言犀人工智能平台的AI能力，通过“咨询+服务”的交互模式，可自动化应答90%的服务咨询。



京东探索研究院院长、京东科技智能服务与产品部总裁何晓冬：大模型将重塑所有的商业模式，未来通往AGI的终局，多模态智能是必经之路。

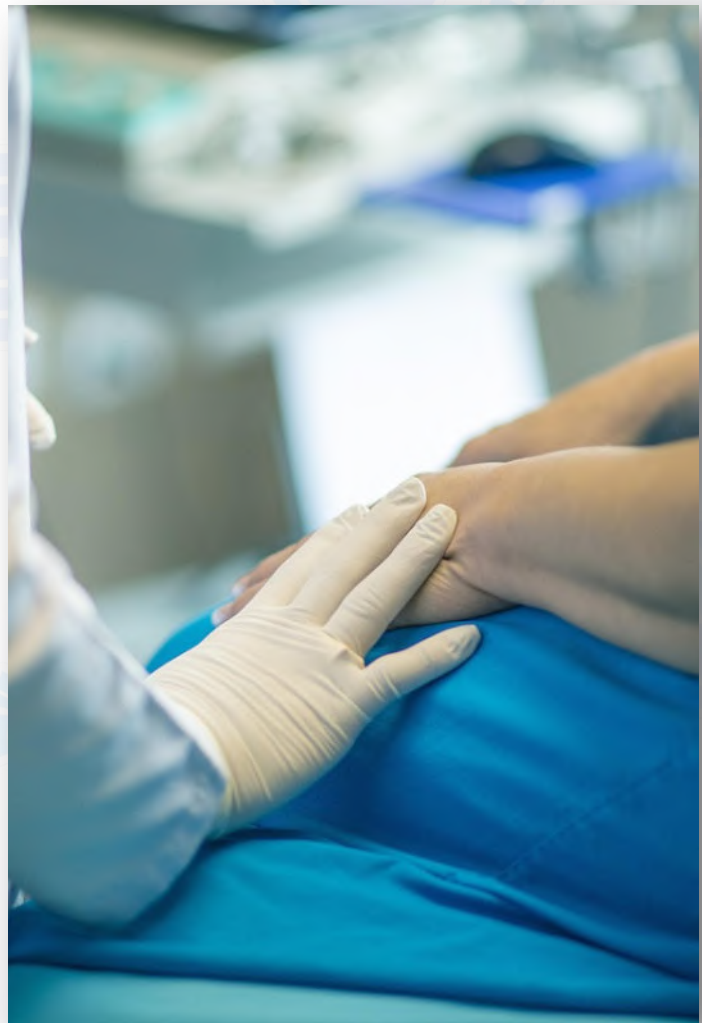
大模型与医疗服务行业融合

医疗产业面临着数据量大、数据质量差、数据隐私保护、医生资源不足、诊断效率低等痛点。大模型在医疗健康领域能够提高医疗信息化效率、改善在线问诊体验、实现实时监测预警以及助力药物研发，为医疗行业带来了变革性的推动，为患者提供更好的诊疗体验与个性化管理。

例如，可以辅助医生进行疾病诊断与预测，通过学习医学数据进行影像分析和风险评估。在药物研发中，大模型能加速筛选、预测药效，帮助发现新治疗方法。个性化医疗方面，大模型根据基因组和临床数据提供个性化诊疗方案与健康管理建议。

然而，模型解释性、数据隐私与安全等问题仍需关注。首先，大模型仍然存在“幻觉”，即虽然能够生成令人信服的结果，但其决策过程仍然是不透明的。这种不确定性可能会对辅助治疗产生影响，甚至可能增加人身生命安全的隐患。

其次，医疗行业的数据涉及患者隐私和国家安全等敏感信息，因此安全风险较大。大模型需要大量的数据进行训练，但医疗数据的开放性相对较低，数据获取和共享面临着许多障碍。

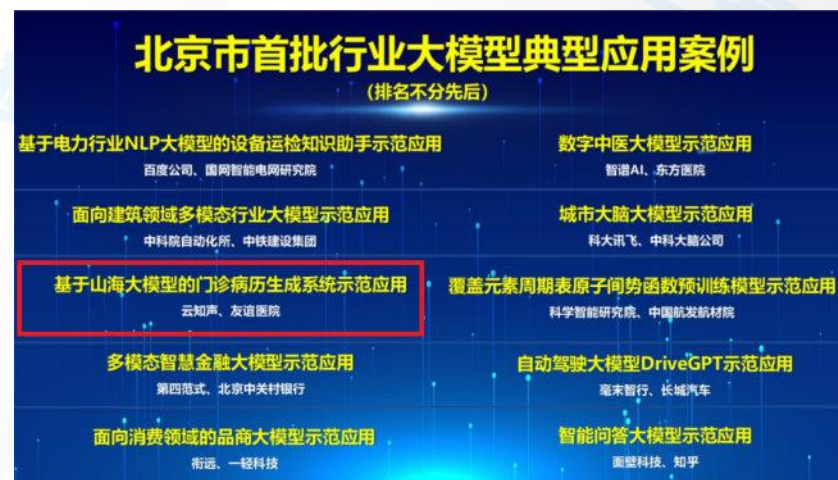
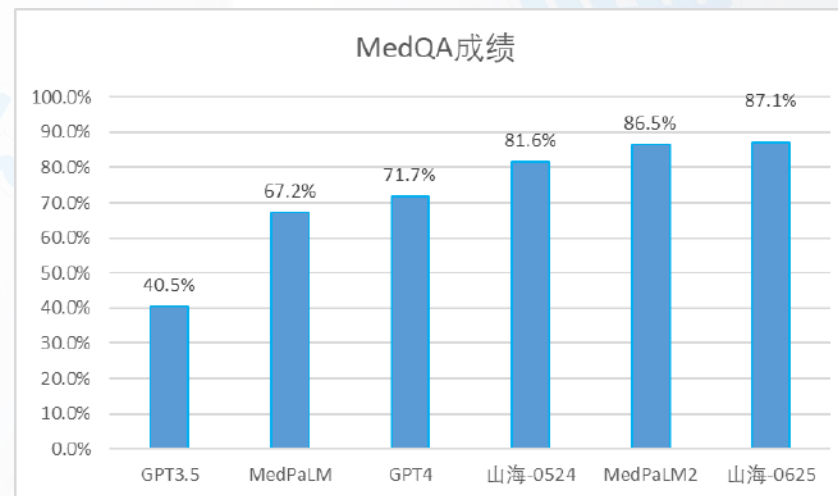


云知声基于山海大模型的门诊病历生成系统

云知声山海大模型技术在智慧医疗领域有三大核心应用，分别是手术病历撰写助手、门诊病历生成系统和商保智能理赔系统。这些应用都能够通过语音识别和自然语言处理技术，实现医疗信息的自动化采集、分析和生成，提高医疗效率和质量，减轻医生负担，优化患者体验。

医疗领域为例，山海大模型学习了大量教材、百科等高质量医学文献，使其能够提供更加全面、专业的医疗信息支持。在扩展了医疗大模型的领域知识的同时，山海大模型仍旧保持原有的高性能、高准确率等特点。目前该模型在MedQA任务上提升到了87.1%，超越Med-PaLM 2，临床执业医师资格考试提升到了523（总分600分），超过了99%的考生水平。

6月27日，北京市科委、中关村管委会举办北京市人工智能策源地引领推介活动的首期活动，并发布北京市首批10个人工智能行业大模型应用案例，由云知声和北京友谊医院共同开发的基于山海大模型的门诊病历生成系统示范应用成功入选。

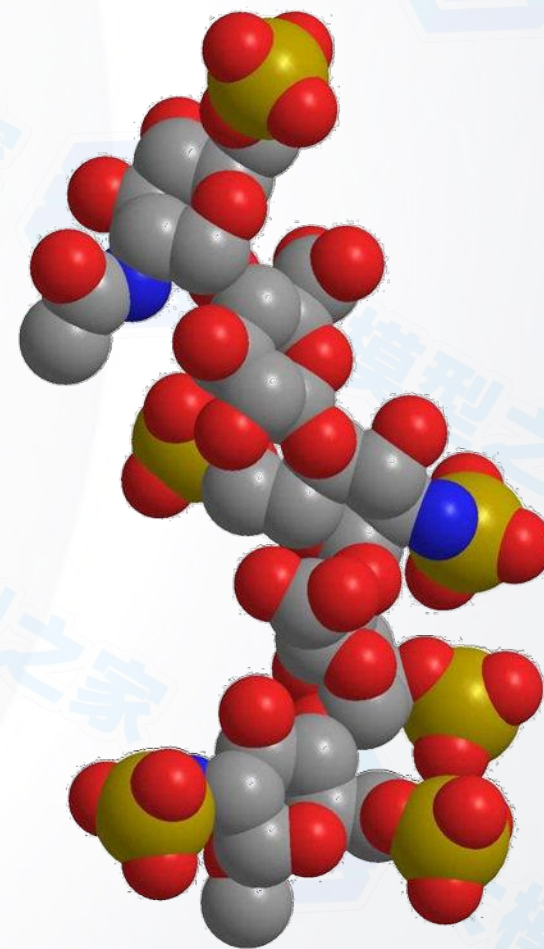


大模型推动蛋白质生成，改进生物医药格局

人工智能大模型可以利用生成式、对抗式等技术，实现对药物分子的设计、筛选、优化和评估。例如，加速蛋白质生成技术的发展进程，并推动其在生物医药、合成生物学等领域应用，进而改变生物经济的格局。

NewOrigin大模型通过学习千亿级多模态大数据，可实现多模态定向生成，单模型就能满足序列生成、结构预测、功能预测、从头设计等蛋白质生成全流程需求，解决产业应用所需的特定功能蛋白质生成难题，并在真实的产业环境中评估效果与价值。

AI蛋白质生成大模型瞄准创新药设计、合成生物学等真实产业应用需求，将用一个模型满足蛋白质生成全流程需求，未来大分子药、新生物材料等蛋白质设计可实现“一键定制”。



大模型与金融保险行业融合

金融产业的核心竞争力，在于不断降低金融风险、与不断提高行业的服务水平。

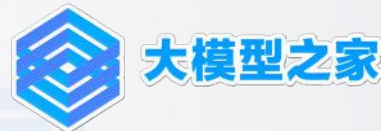
在金融行业，大模型可以用于风险管理和欺诈检测，通过分析大量的交易数据和用户行为，提供更准确的风险评估和欺诈检测能力，帮助金融机构减少损失。

同时，大模型在投资决策方面发挥着重要作用，能够分析市场趋势、预测股票价格和优化投资组合，为投资者提供更可靠的决策依据。

此外，大模型还可以应用于信贷评估、客户服务和金融市场监管等领域，提高金融机构的运营效率和服务质量。然而，金融行业对于数据安全和隐私保护的要求极高，因此在应用大模型时需要严格遵守相关法规和保护措施。总体而言，大模型在金融行业的应用将进一步提升行业效益和客户体验，推动金融科技的发展。



众安AIGC中台灵犀助力大模型内嵌行业专业知识库



在保险领域，大模型可以利用自然语言处理、语音识别等技术，实现对客户需求、偏好、反馈等的全面理解，从而提供定制化的金融保险产品、服务、咨询等。

此外，大模型还可以利用预测分析、异常检测等技术，实现对金融市场、客户交易、保险理赔等数据的智能分析和识别，从而防止欺诈、洗钱、假冒等风险，并符合相关法规。

众安保险自主研发了AIGC中台灵犀。通过适配国内外主流大模型，灵犀可以实现“1个MaaS（Model as a Service）平台，2种应用场景策略”，可以允许机构用户在大模型内嵌入行业专业领域知识库，实现AIGC在保险垂直领域的应用快速适配。灵犀也支持把企业内部应用工具包装成大模型插件，让大模型更加贴近业务应用场景，通过AIGC助推企业业务能力提升。



大模型与智慧办公行业融合

在智慧办公领域，大模型能够通过自动化和智能化技术改进办公流程。

首先，**大模型可以用于自然语言处理和智能对话系统**，例如，通过大模型进行文档摘要和自动化翻译，提高文档处理的效率；通过语音识别和自然语言处理技术实现语音助手，帮助用户处理日常任务和安排会议；通过自动化的数据分析和处理，提供决策支持等，促进团队间的沟通和合作。

其次，**大模型在信息管理和知识管理方面也发挥重要作用**，能够实现智能搜索、文档自动分类和知识图谱构建，帮助员工更轻松地获取和利用内部和外部的知识资源。

此外，**人工智能大模型可以应用于智慧办公的单据管理和报销流程**。通过图像识别和文本分析，大模型可以自动识别和分类各种单据，如发票、收据、报销申请等。



WPS AI为用户提供智能文档能力

WPS AI是由金山办公发布的具备大语言模型能力的人工智能应用，为用户提供智能文档写作、阅读理解和问答、智能人机交互的能力。

WPS AI能够通过语音或文字与AI进行自然沟通，能够帮助实现文档的创建、编辑、转换等功能，自动生成工作总结、广告文案、社媒推文、文章大纲、招聘文案、待办事项、创意故事、旅行游记等结构化文本。用户还可以插入一篇或多篇已有的文档，作为AI内容生成的参考素材，使生成的内容更具相关性。

同时，WPS AI可以对长文档进行归纳总结，提取关键信息，帮助用户快速掌握文档要点。WPS AI还可以对用户提出的问题进行智能回答，支持多轮对话，帮助用户解决创作问题。

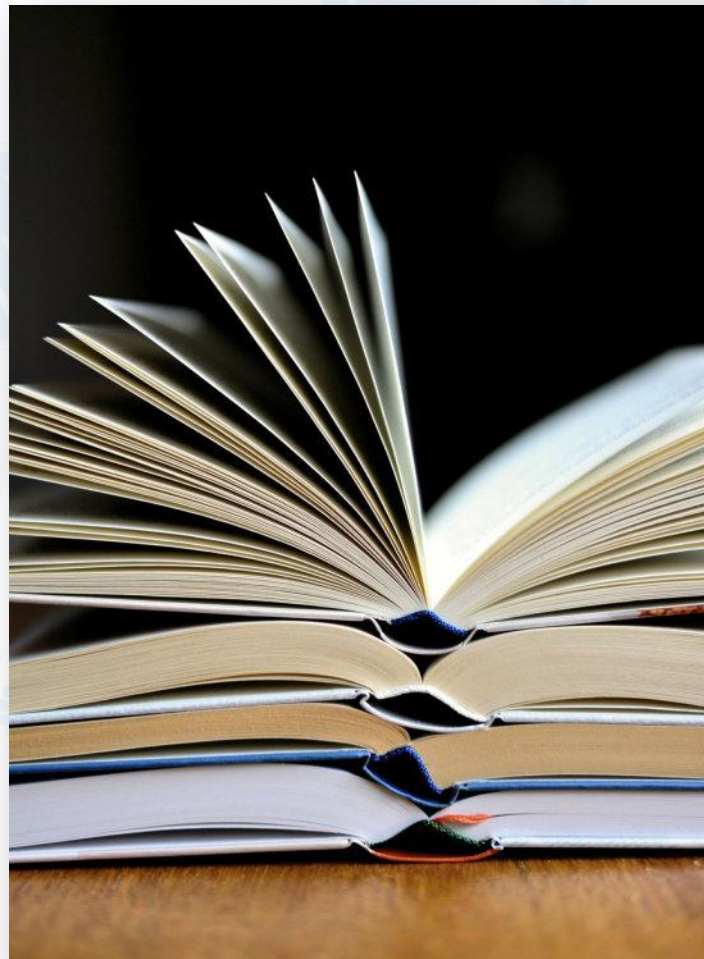
此外，WPS AI还可以将文字指令转化成表格公式、演示文稿，并能够识别和纠正表格中的文本、公式和语法错误。



大模型与智慧教育融合

人工智能大模型与智慧教育行业融合，创造了多个创新的应用场景，使得智慧教育变得更加个性化、高效和智能化，提升学习体验和教学水平。

- **个性化学习**：根据学生的兴趣和知识水平提供定制化的学习计划；
- **智能助教**：回答学生问题并解释复杂概念；
- **智能测验和批改**：对客观题和主观题一并进行评分和评估；
- **生成教育内容**：自动撰写课程讲义、设计习题和实验方案等；
- **分析学习情感和心理状态**；
- **辅助招生决策**：预测潜在学生的表现和适应性；
- **为教师提供培训建议和评估**；
- **挖掘教育数据**，改善教学和预测学生表现。



科大讯飞打造学科专业化大模型及产品

科大讯飞星火认知大模型具备跨领域多任务上类人的理解和生成能力，可以模拟人类大脑对自然语言的理解和生成过程，实现对语音、图像、文本等多种信息形态的理解和处理，为用户提供更加智能、个性、高效的服务和体验。

在智慧教育领域，基于丰富的行业数据和经验，科大讯飞打造了针对不同学科和年级的专业化大模型，例如在在线教育场景中，在语文、数学、英语等主要学科上进行了深度优化和应用，与网易云课堂合作，提供了智能课程开发、智能课堂辅助、智能课程评估等产品。

在产品侧，科大讯飞推出了全球首款认知大模型AI学习机——T20系列，将通用认知智能的大模型接入智能硬件实体机，实现作文批改、口语陪练、数学辅学等方面的类人互动辅学功能在AI产品的落地。

科大讯飞在教育领域拥有覆盖教、学、考、管四大主场景的智慧教育全栈产品体系，包括智慧课堂、智能阅卷、智能评测、智能辅导、AI学习机等多种产品和解决方案。



科大讯飞AI学习机T20 Pro

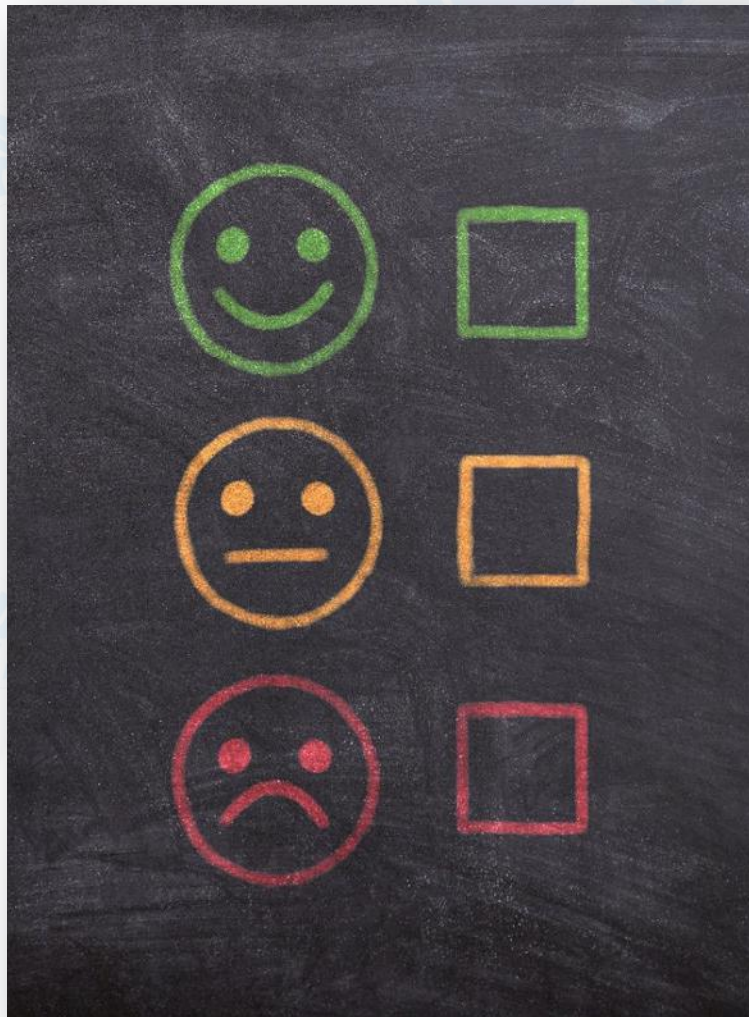
大模型与媒体营销服务融合

大模型还可以在媒体与营销领域相融合，产生1+1大于2的合力。

首先，大模型可以用于目标受众的精准定位和个性化推荐。通过分析海量的用户数据和行为模式，大模型可以帮助媒体企业更好地了解用户的兴趣、偏好和需求，从而提供个性化的内容推荐和广告定向投放，提升广告的点击率和转化率。

其次，大模型可以应用于舆情监测和情感分析。通过对社交媒体、新闻网站和论坛等平台的数据进行实时监测和分析，有助于媒体企业及时了解市场反馈、调整营销策略，并做出针对性的回应。

此外，大模型还可以用于内容生成和创意辅助。媒体营销行业需要大量的创意内容来吸引用户和传递信息，而大模型可以通过学习和模仿人类的写作风格和创作思维，生成高质量的广告文案、新闻报道、社交媒体内容等。这不仅可以提高创作效率，还可以帮助媒体企业实现内容的个性化和多样化。



智能问答与内容创作应用——AI问客

智能问答与内容创作应用——“AI问客”，依托最新的AI大模型技术，提供了近百个垂直化应用场景。

用户可以在职场、生活、教育等细分领域中更高效、便捷地获取信息。在内容创作方面，AI问客具备强大的功能。根据不同的文章类型和风格需求，它提供定制化的创作帮助，让您轻松撰写新闻稿、微信稿、小红书文案，甚至润色、续写和校对文章，激发用户的创作灵感。

与传统的AI辅助工具相比，“AI问客”还具备联网功能，获取更多实时性的网络信息。



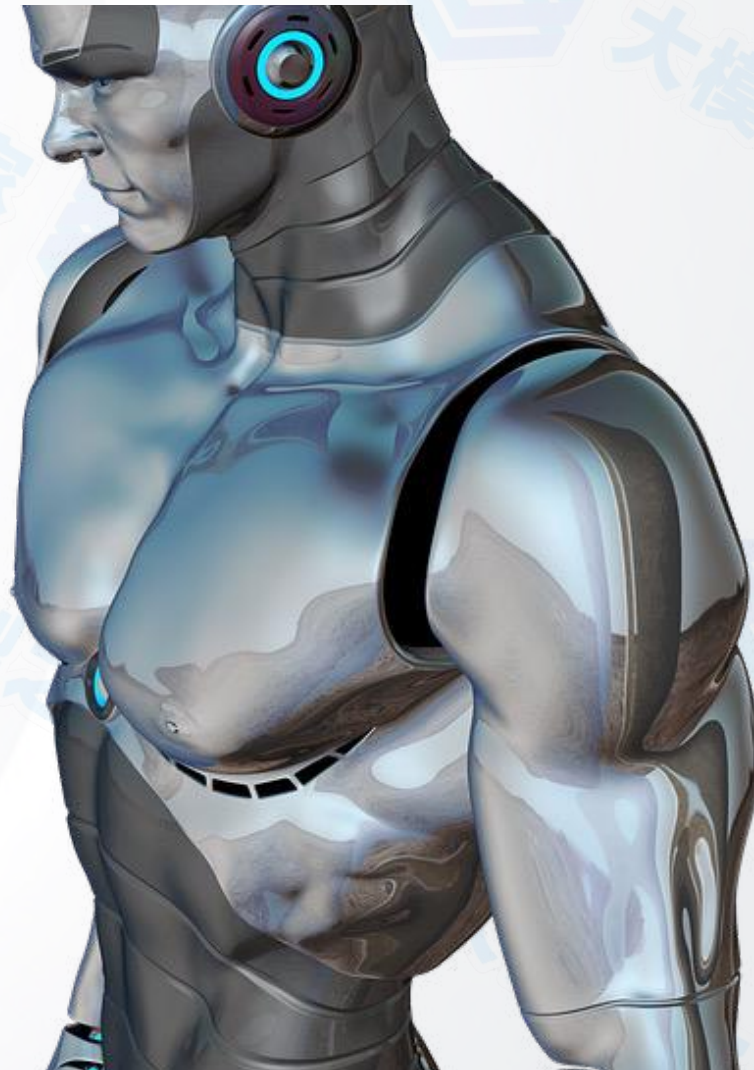
大模型与数字分身融合

在数字人/数字分身领域，人工智能大模型可以实现对真实人类的高度仿真和模拟，从而提供更自然、更真实、更多样的人机交互体验，甚至实现数字时代的永生。

数字人仿真增强：人工智能大模型可以利用图像识别、语音识别、自然语言处理等技术，实现对用户的面部表情、声音、语言等的智能分析和识别，最终实现仿真和增强，从而生成更加真实的数字人/数字分身。

降低成本和时间投入：人工智能大模型可以利用预训练和微调的方式，实现对不同场景和任务的快速适应，从而降低了创建和维护数字人/数字分身所需的成本和时间。

多维度的仿真能力：人工智能大模型可以利用多模态数据（如文本、图像、视频、音频等）进行联合学习和生成，实现对语言、举止等多维度的仿真。



360智脑实现AI数字人“私人订制”

AI数字人作为360在大模型上的核心创新之一，不同于传统数字人，360AI数字人具有记忆、人设和性格，能够复刻真实人类的思维方式和人生经历，是未来人工智能大模型的最重要的应用入口。

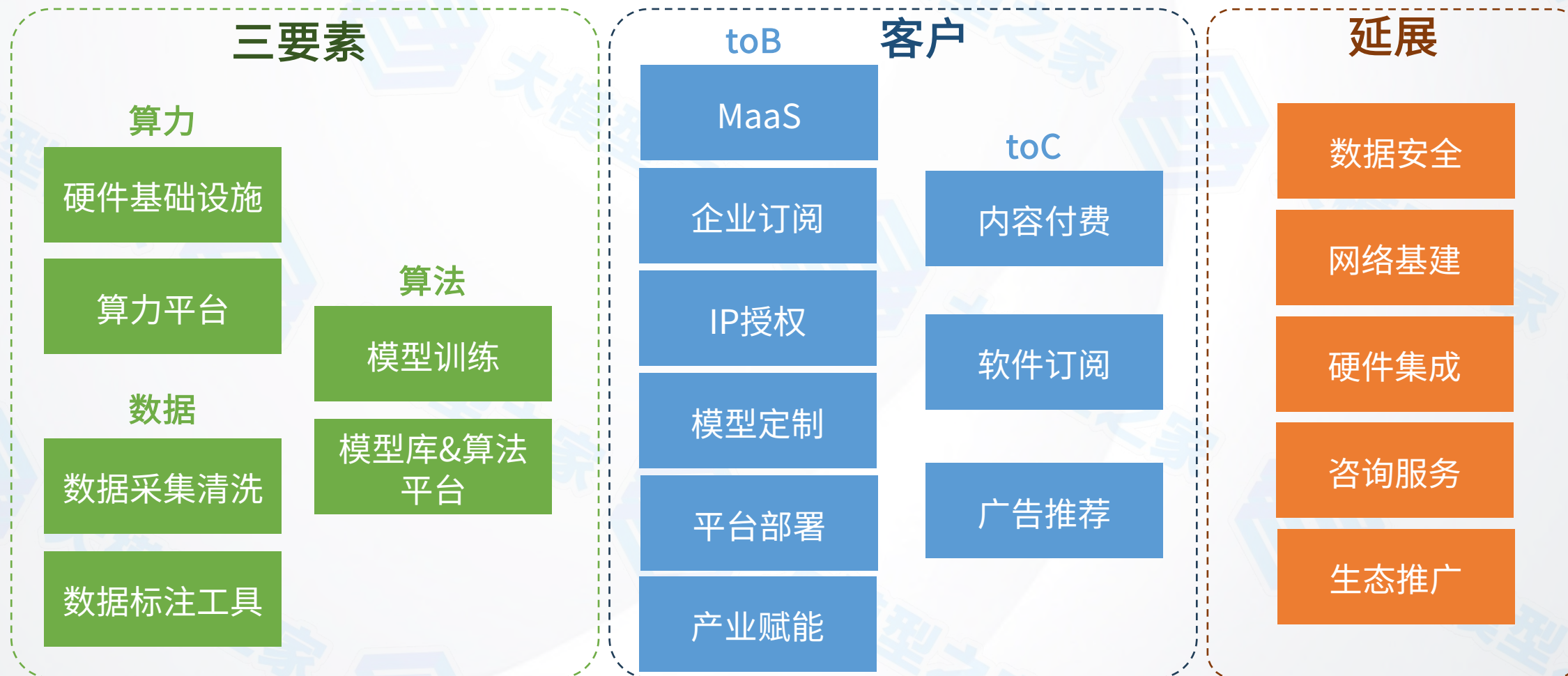
360AI数字人有两类角色：数字名人和数字员工。数字名人可以让用户与历史人物、偶像明星、文学形象等进行对话和交流；数字员工可以为企业提供专业的知识和办公服务。

360AI数字人未来将拥有更强的多模态生成能力，能够生成声音、视频，并具备长期记忆、目标规划和执行力。依托于“360智脑4.0”的强大支撑，每个人都能通过上传私人知识库，低成本生成自己的专属数字人。



大模型打开人工智能产业变现路径

大模型能够为各行各业降低成本、提高效率，从而产生实际的商业价值。同时，大模型不仅仅是一种技术工具，它还催生了新的产品和服务模式，也为人工智能行业自身开启了新的盈利模式。



全球算力市场争夺 “一触即发”

人工智能大模型的训练需要消耗大量的算力资源，如GPU、TPU等。随着模型的参数量和数据量的不断增长，训练成本也随之上升。

大模型的火爆也带动了算力市场的竞争。其中，GPU作为重要的计算资源在人工智能领域中发挥着关键作用。特别是GPGPU技术的应用，使得GPU在云端进行大模型训练算法时能够显著缩短训练时长，减少能源消耗，降低人工智能的应用成本。

在全球算力市场中，英伟达作为高算力芯片的龙头企业，占据着领先地位。其A100、H100显卡适用于支持各种机器学习模型，具备强大的数据处理能力，使其成为目前主流的AI芯片。目前，全球GPU市场中英伟达占据近90%份额。**据大模型之家预测，全球AI芯片市场规模有望在2025年突破千亿大关。**

目前，我国数据中心、智能计算中心加快部署，2021年基础设施算力规模达到140 EFLOPS，位居全球第二，算力总规模达到202 EFlops，全球占比33%，增速达到50%，其中智能算力成为增长驱动力，增速达到85%。不过，在高端人工智能算力芯片领域，我国发展受到了海外的封锁和制约，美国限制英伟达和AMD向中国企业出售两款顶级GPU计算芯片A100和H100，因此，推动中国人工智能算力的高质量发展成为当务之急。

云计算、分布式计算将灵活解决算力需求



人工智能大模型的发展也催生了新的算力需求和供给模式。例如，一些创业公司和研究机构可以通过租用或共享算力资源来进行模型的开发和测试，而不需要自己购买昂贵的硬件设备。另一方面，一些算力平台和中介也可以为算力需求方和供给方提供更灵活和高效的匹配和交易服务。

人工智能大模型的应用也需要相应的算力支持，各个企业和组织纷纷增加对于高性能计算设备的投入。云计算服务商也积极提供更强大的计算资源，同时灵活调配，按需购买的方式，也满足企业对于大模型应用的需求。这导致算力市场规模不断扩大，各种计算设备和解决方案不断涌现。

尤其是在云端部署和提供服务的场景下，例如ChatGPT、Midjourney等基于LLM的应用需要在云端处理用户的请求和响应，以及提供持续的更新和优化。因此，云计算厂商如亚马逊、微软、阿里云、百度等可以获得更多的客户和收入。

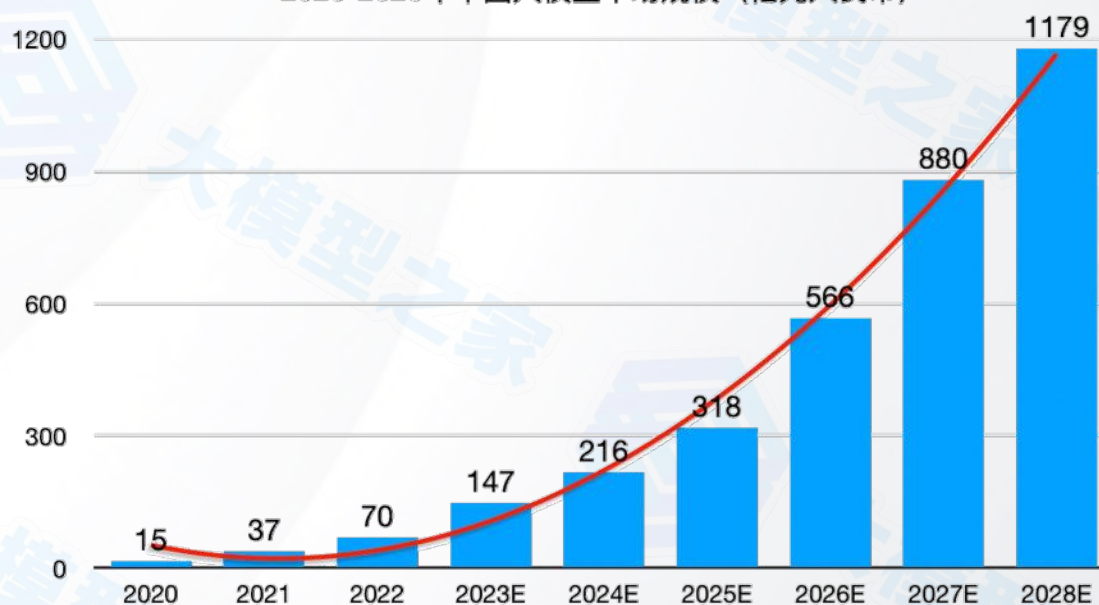
同时，在人工智能大模型之下，云服务商可以提供更丰富和智能的云计算服务，例如对话生成、编程、事实问答、文生图等功能，将吸引更多的开发者和企业使用云计算平台来开发和部署AI应用。

MaaS将成为新的商业模式

随着大模型技术的发展，MaaS（模型即服务）将成为一种全新云计算模式，它提供了一种将机器学习模型作为服务的方式，允许用户在不需要拥有自己的硬件设备或专业技能的情况下，使用高质量的机器学习算法和模型。MaaS有望成为云服务厂商继IaaS和PaaS之后的新的商业模式，为客户提供更便捷、高效和全面的人工智能模型服务。

- MaaS按需订阅的模式，可以灵活降低AI应用的门槛和成本，让更多的开发者和企业能够快速利用大模型的能力，实现智能化的创新和转型。
- MaaS可以提高AI应用的效率和质量，让开发者和企业能够根据不同的场景和需求，灵活地选择和调用适合的大模型，实现多模态和跨领域的智能交互。
- MaaS可以促进AI应用的生态和创新，让开发者和企业能够基于大模型进行二次开发和微调，实现个性化和定制化的智能解决方案。

2020-2028年中国大模型市场规模（亿元人民币）



第四章

涓涓之流，积成江河

大模型领域融资版图

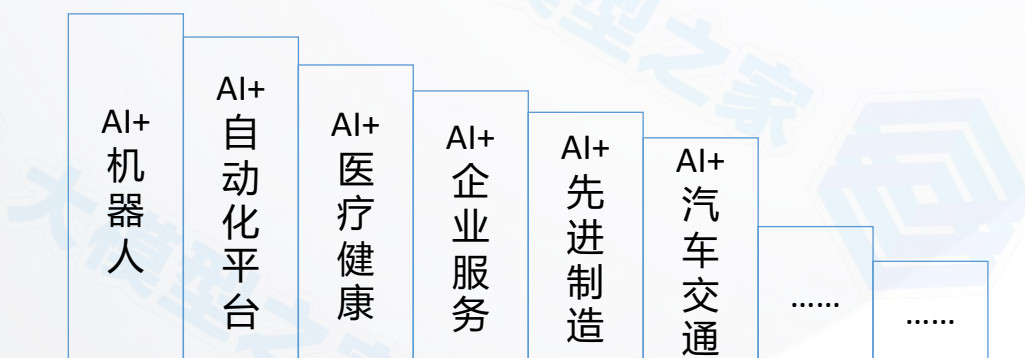
大模型融资版图

大规模人工智能模型的出现引发了对人工智能领域的极大兴趣和投资热潮。据大模型之家统计，截至2023年6月30日，我国人工智能行业公开投融资规模近300亿元，人工智能领域共有286件融资交易事件，多集中于天使轮、A轮等早期融资，大模型创业热潮正处在进行时。

风险投资公司、私募股权基金和科技巨头纷纷投资专注于开发和应用大规模人工智能模型的创业公司 and 研究项目。资金涌入人工智能创业公司 and 研究项目，促进了大模型领域快速发展，并推动了大规模人工智能模型的采用。

综合融资金额与融资次数，排名前列的行业：

其中，规模较大的融资事件有：



2.5亿美元A轮

名之梦（上海）科技有限公司

7亿元D3轮

云知声

6亿元A+轮

飞诺门阵

5亿元C+轮

杉数科技

资本观点

华创资本
熊伟铭

国内整体的策略还是追赶策略，从全球来看，OpenAI一季度在这个领域里面处于领先地位，Open AI在这个领域的投资也是全球最高的。我觉得从基础设施投资角度来讲，真的就是一分钱一分货，不投到那个钱，就出不来这个结果。在GPU的采购以及一些开源的大模型没有受到更多的限制之前，应该说大家差距可能现在还没有那么大。

生成式AI落地到应用场景中，将是80%存量应用升级+20%新应用探索的“二八定律”局面。也将给正在演进中的技术，如元宇宙、自动驾驶等，带来渐进式能力推动。

云启资本
合伙人
陈昱

华映资本
管理合伙人
章高男

大模型解决问题的效果首先取决于训练数据的优质性，因为所谓通用大模型进入很多细分行业领域里数据是很稀疏的。所以通用大模型在细分领域中效果未必是最好的，有很多垂直行业的细分大模型的机会。

大模型产业六大趋势洞察

大模型产业六大趋势

- 1、资本热度：**大模型的兴起为创业者带来了利好，吸引了资本的高度关注和投资。但资本热潮推动不可避免会引发行业泡沫，需要仔细甄别和评估投资机会。
- 2、发展方向：**技术推动领先企业与模型形成数据飞轮，推动大模型能力快速提升、成本降低、门槛减低。
- 3、竞争策略：**大模型领域竞争者争夺技术、市场和人才先发优势，直接在大模型上开发应用很难形成有效的竞争壁垒，需要寻找其他差异化策略。
- 4、马太效应：**由于模型、算力和数据三大支柱向大型企业集中，未来可能出现马太效应，加剧市场竞争不平衡。
- 5、机会分配：**未来的巨头机会在于底层架构和通用大模型领域，而小企业和创业者则可基于大模型开发各种行业应用，实现大模型的民用化。
- 6、数据治理与安全性：**企业掌握海量数据，对数据治理和安全性将提出更高要求。

第五章节

权，然后知轻重

大模型评价体系

国内数据量超10亿的大模型已超80个

随着人工智能大模型的快速发展和广泛应用，通过建立和使用合适的评价体系，我们可以更好地评估和比较不同大模型的性能和效果。一个有效的大模型评价体系应该具备任务导向性、客观性和可重复性、多维度评估、合适的数据集选择以及持续迭代和改进的特点。

目前，国内数据量超10亿的大模型已经超80个，涉及产学研多个领域，主要集中在通用领域，涵盖自然语言处理（NLP）、计算机视觉（CV）和语音识别等技术领域，聚焦通过创新来提高效率、降低成本。



大模型之家专业大模型“AQUA”评价体系

大模型之家也注意到，随着国内专业大模型的推出，通过一套“通用”的模型试题很难准确评价针对不同行业进行知识强化的诸多大模型，因此行业需要一个较为泛化的评估体系。

模型能力
(Model Capability)
20%

模型规模
(Model Size)

训练数据量
(Training Data Size)

训练算力
(Training Compute Power)

任务处理能力
(Task Processing Capability)
25%

感知能力
(Perception Capability)

认知能力
(Cognitive Capability)

生成能力
(Generation Capability)

泛化能力
(Generalization Ability)
25%

多语言支持
(Multi-language Support)

多模态支持
(Multi-modal Support)

多场景支持
(Multi-scenario Support)

应用扩展能力
(Extension Capability)
15%

应用场景扩展
(Application Scenario Extension)

垂直行业应用
(Vertical Industry Applications)

适应性和定制性
(Adaptability and Customization)

开放性指标
(Openness Metrics)
10%

开发环境
(Open Source Contribution)

部署方式
(Training Data Size)

标准化支持
(Standardization Support)

应用生态
(Application Ecosystem)
5%

社区支持
(Community Support)

合作伙伴
(Partnerships)

训练算力
(Training Compute Power)

国内主流大模型



文心
百度



通义
阿里巴巴



浑元
腾讯



日日新SenseNova
商汤科技



讯飞星火
科大讯飞



360智脑
三六零



盘古
华为



山海
云知声



从容
云从科技



言犀
京东



天工3.5
昆仑万维



源1.0
浪潮信息



孟子
澜舟科技



炎帝
硅基智能



ChatGLM
智谱AI



玉言
网易伏羲



知海图
知乎



紫东.太初
中科院自动化所



悟道3.0
智源研究院



MOSS
复旦大学



式说
第四范式



元乘象
智子引擎



天燕AiLME
麒麟合盛



序列猴子
出门问问



好未来
mathGPT



Baichuan-7B
百川智能

注：以上排名不分先后

拥抱AGI是一代代人工智能从业者所追求的星辰大海。

人工智能大模型技术发展迅猛，它正逐渐落地于各个产业，并创造出丰富的价值。然而，我们也必须清醒地认识到，人工智能的发展之路依然漫长。尽管我们已经取得了巨大的进步，但从弱人工智能走向强人工智能的道路上，我们仍然只是处于起点。我们需要继续不断努力，推动大模型技术的进一步创新和应用，不断探索创新的应用场景和商业模式，解决现实世界中的复杂问题，实现人工智能在更广泛领域的全面应用。

同时，我们也要认识到人工智能的发展过程中可能面临的挑战和风险，需要积极探索和制定相应的规范和道德准则，确保人工智能的发展始终符合人类的价值观和伦理标准。只有在这样的前提下，人工智能才能真正成为人类社会进步和发展的有力工具。

在未来的旅程中，我们必须保持开放和合作的精神，促进学术界、产业界和政府之间的紧密合作，共同推动人工智能的发展和应用。通过共同努力，我们可以开创出一个智能化、创新化的未来，让人工智能为人类带来更大的福祉和发展。大模型技术的崛起为我们打开了一扇门，迈出坚实的一步，但这只是我们前进的开始。让我们继续探索、努力前行，迎接人工智能未来的挑战和机遇。

关于大模型之家

大模型之家是国内专业的人工智能垂直媒体，聚焦于最新、最全面的人工智能产业信息和专业技术干货，紧密关注人工智能领域的动态，深度解读最新技术成果和产品应用，分享业内专家的观点和经验。

作为国内AI大模型前沿媒体，大模型之家秉持着开放、创新、分享和共赢的价值观，为读者与企业提供前沿、实用的内容服务，持续为产业创造价值。

版权声明

《人工智能大模型产业创新价值研究报告》的知识产权归大模型之家及所属公司所有。可供个人学习及研究使用，但未经授权不得复制传播。个人或单位如需商用，应联系大模型之家获得授权后，可按照授权的要求使用。违反上述声明者，本公司将追究其相关法律责任。

免责声明

《人工智能大模型产业创新价值研究报告》中涉及到的内容与行业数据主要为大模型之家分析师整理相关公开数据分析撰写，本报告相关结论与预测基于分析师个人观点，但受研究方法、数据来源资料等多方面限制，大模型之家不保证其准确性或完整性。大模型之家并不对使用本报告或其所包含内容产生的任何直接或间接损失或与此有关的其他损失承担任何责任。

编撰

鸣谢



官网网址: www.sootoo.com

微信公众号: 大模型之家

微博: @大模型之家

官方邮箱: largemodels@sootoo.com

感谢观看