



人工智能医疗器械创新合作平台
ARTIFICIAL INTELLIGENT MEDICAL DEVICE INNOVATION AND COOPERATION PLATFORM

CAICT 中国信通院

人工智能大模型赋能 医疗健康产业白皮书 (2023 年)

人工智能医疗器械创新合作平台 智能化医疗器械产业发展研究工作组
中国信息通信研究院 云计算与大数据研究所

2023年10月

编制说明

牵头单位

人工智能医疗器械创新合作平台 智能化医疗器械产业发展研究工作组、中国信息通信研究院 云计算与大数据研究所

参编单位

北京邮电大学、北京大学、中国科学技术大学、北京大学第三医院、首都医科大学附属北京友谊医院、解放军总医院第六医学中心、温州医科大学附属眼视光医院、江苏省中医院、温州眼视光国际创新中心、国家药监局智能化医疗器械研究与评价重点实验室、网络与交换技术全国重点实验室、互联网医疗健康产业联盟、阿里云计算有限公司、腾讯医疗健康(深圳)有限公司、北京百度网讯科技有限公司、商汤科技、OPPO 广东移动通信有限公司

编写顾问

瞿佳、宋纯理、王振常、刘广东、王光宇

编写专家

高越、李曼、闵栋、程祎雯、逢淑宁、仇尚航、周少华、刘晓鸿、吕晗、李小虎、张康、徐良德、王宏、杜宁娜、李昀、周珽鲸、冯天宜、武雅文、张小洵、范佳莹、王皓、许昂兢、郑冶枫、黄超、张莹莹、吴贤、彭思翔、徐永太、贺庆、杨青、陈俊、李媛、徐捷、边成

版权声明

本白皮书版权属于人工智能医疗器械创新合作平台和中国信息通信研究院，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源：人工智能医疗器械创新合作平台和中国信息通信研究院”。违反上述声明者，编者将追究其相关法律责任。



前 言

随着全球新一轮科技革命和产业变革深入发展，以人工智能（Artificial Intelligence, AI）为代表的数字技术加速演进，成为经济增长的核心驱动力。近年来，人工智能大规模预训练模型（以下简称“大模型”）在知识、数据、算法和算力等关键要素的共同推动下，呈现爆发式增长，从自然语言处理逐步扩展、迁移到计算机视觉、多模态、科学计算等领域，增强了人工智能的泛化性、通用性，开启了人工智能发展新范式。人工智能大模型有望成为未来新型基础设施，赋能千行百业新一轮增长，落地应用和价值实现成为大模型下一步发展方向。生命科学和医疗健康是大模型等 AI 技术最重要的应用领域之一，大模型可赋能生命科学研究和新药研发，促进医疗器械创新，提升医疗智能化水平。将大模型应用于医疗健康领域，推动医疗健康行业数字化转型升级，对满足人民群众健康需求和实现经济社会绿色、智能、可持续发展具有重要意义。

本白皮书集中梳理了国内外医疗健康大模型的最新态势及发展成果，尤其是技术体系、应用场景、风险挑战、评价和监管等方面的内容。白皮书首先阐述了医疗健康大模型的概念内涵、发展优势、生态架构；对医疗健康大模型技术体系和演进趋势进行了分析；并对医疗健康大模型应用现状及优秀应用案例进行了系统性梳理。在此基础上，白皮书分析了医疗健康大模型面临的技术、应用、数据、伦理挑战，并结合医疗健康大模型标准、验证评价和监管治理情况，提出了

进一步促进医疗健康大模型发展的相关建议。

白皮书旨在为我国生命科学与医疗健康大模型技术和产业发展提供参考和引导，共同推动技术研发创新，促进行业高质量发展。由于生命科学与医疗健康大模型仍处于快速发展阶段，我们的认识还有待持续深化，白皮书仍有不足之处，欢迎大家批评指正。下一步，我们将广泛采纳各方面的建议，进一步深入相关研究，适时发布新版报告。我们诚邀各界专家学者参与我们的研究工作，积极献言献策，为促进生命科学与医疗健康大模型发展作出贡献。如您有意愿，请联系我们：gaoyue1@caict.ac.cn。

目 录

一、医疗健康大模型发展概述	9
(一) 医疗健康大模型的概念和优势	9
(二) 医疗健康大模型的生态架构	11
二、医疗健康大模型的技术体系及演进	14
(一) 大模型响应生物大数据时代的数据特征和应用需求	14
(二) 多类型基础模型为医疗健康大模型提供坚实源动力	19
(三) 四种范式助力大模型在医疗健康垂直领域的应用实践	28
(四) 模型发展呈现家族化、多模态、融合化、协同化趋势	35
三、医疗健康大模型的应用场景	39
(一) 整体发展分析	40
(二) 具体场景分析	46
四、医疗健康大模型面临的风险与挑战	68
(一) 技术风险：精度不够，尚不能完全满足医疗场景安全性可靠性需求	68
(二) 落地挑战：数据、成本、权责问题制约大模型在医疗领域落地应用	70
(三) 数据安全和隐私：个人数据滥用、隐私泄露和网络攻击风险突出	72
(四) 伦理道德问题：大模型加剧医疗偏见和有害、虚假信息传播问题	73

五、医疗健康大模型的评价验证和监管治理	74
（一）标准和指南：基础信息安全标准开始起步，领域平台规范指南需求迫切.....	74
（二）评价和验证：针对新能力、新特性、真实表现的动态评估方法有待建立.....	76
（三）政策和监管：促发展与防风险并重，奠定医疗大模型监管治理良好基础.....	78
六、医疗健康大模型发展建议	80

图 目 录

图 1 人工智能大模型+医疗健康生态架构	12
图 2 人工智能发展历程	16
图 3 医疗健康大模型的类别和实例	20
图 4 LLaVA-Med 的预训练和微调流程.....	23
图 5 Med-PaLM M 所用基准数据集的模态和任务	28
图 6 生命科学与医疗领域开发与应用大模型的四范式	29
图 7 GatorTron 模型的预训练和微调过程.....	30
图 8 BioBERT 的预训练和微调过程.....	32
图 9 ChatDoctor 模型的构建过程.....	34
图 10 Med-PaLM 2 采用多种路径提升推理能力.....	35
图 11 ClinicalGPT 的训练及对齐流程	41
图 12 商量大医模型工作原理和应用场景	43
图 13 灵医大模型输入“医-患-药”医疗健康数据和医学知识....	45
图 14 灵医大模型的商业化服务模式	46
图 15 EyeGPT 根据需求输出研究题目的文章大纲	50
图 16 未来 EyeGPT 在眼健康智能诊断一体化应用示范	50
图 17 腾讯医疗大模型多轮问询和智能问诊示意图	59
图 18 神经源性膀胱慢病管理数字疗法平台示意图	63
图 19 运动健康助手应用架构和界面	65

一、医疗健康大模型发展概述

（一）医疗健康大模型的概念和优势

人工智能大模型（AI 大模型）是“大数据+大算力+强算法”结合的深度神经网络模型，通过“预训练+微调”模式增强了人工智能的通用性、泛化性，带来人工智能研发新范式，成为迈向通用人工智能（Artificial General Intelligence, AGI）的重要技术路径。AI 大模型基于海量无标注数据进行预训练，提升模型前期学习的广度、深度和知识水平，从而能够低成本、高适应性地赋能大模型在后续下游任务中的应用。当模型参数规模足够大时，AI 大模型出现“智能涌现（Intelligence Emerging）”，例如“少样本”或“零样本”等能力¹。因此，基于 AI 大模型进行下游应用开发时，对预训练好的大模型进行领域知识微调训练（在下游特定任务上的小规模有标注数据进行二次训练）或使用提示词工程，即可高水平地完成多个应用场景的任务，实现通用的智能能力。AI 大模型是实现多种人工智能应用的通用载体，或成为未来整个人工智能生态的核心。

医疗健康大模型是面向复杂、开放医疗健康场景的基础大模型，具有大数据、大算力、大参数等关键要素，呈现涌现能力和良好的泛化性、通用性，可以根据不同的医疗健康任务，利用语言、视觉、语音乃至多模态融合的生物医学数据进行“预训练-微调”，从而为医疗健康领域提供高效、准确、个性化的服务和支持。

大模型技术的突破，为医疗人工智能的发展注入新动力。相比

传统的执行特定任务的医疗 AI 模型，大模型具有如下优势和潜力。

一是大模型可应用于多个下游任务。现有医疗人工智能模型主要采用针对特定任务的模型开发方法，在标记数据上进行训练，用于单任务学习。在美国食品和药品管理局（FDA）批准的 521 款临床医疗人工智能模型产品中，大多数仅获得了 1 或 2 个任务的批准²。大模型具有较强的泛化能力和上下文学习能力，可以灵活、直接地应用于多个医疗任务，而无需特定任务的训练和标注数据^{2, 3}。

二是大模型突破数据标注的困境。开发传统有监督的医疗人工智能模型时，需要大规模的标注数据才能形成运行良好的模型⁴。大规模、高质量的标注数据集需要医疗领域专家的参与，成本高、时间长。大模型利用自监督学习或强化学习方式，在无标注数据上进行预训练，在一定程度上减轻了标注负担和创建大规模标注数据集的工作量，也更加适配生物医学、临床和健康相关各类数据爆炸式增长的时代⁴。在一些可用数据较少的临床场景，大模型可带来显著的人工智能效果提升。

三是大模型更加用户友好和易于使用。大模型具有接受人类的自然语言（汉语、英语等）作为输入的能力，这使得外行和医疗专业人员都能够方便地访问大模型，也使得复杂的医疗信息更易于访问和理解³。大模型还提供通过自定义查询与模型进行交互的能力²。在当前实践中，AI 模型通常处理单一任务，并产生预先确定的输出（例如，当前模型可能会检测某种特定疾病，输入一种图像，并始终输出该疾

病的可能性)。相比之下，大型语言模型具有强大的生成能力，能够以自然语言的方式进行自定义查询，而不是传统检索式的查询方式，允许用户提出更加多样、更加口语化和更加自然的问题，诸如“这张头部 MRI 扫描图中的肿块更可能是肿瘤还是脓肿？”之类的问题。

四是大模型有望支持更加灵活的多种数据模态组合运用。一些严格的多模态模型通常将预定义的模态集合作为输入和输出（例如，必须始终同时输入图像、文本和实验室结果）²。相比之下，大模型具有更灵活的多模态交互方式，有望使用各种数据模态接受输入和产生输出（例如，可以接受文本、图像、实验室结果或任何组合），支持用户自由地组合各种医疗数据模态类型。

五是大模型具有更深入理解医学知识的潜力。与临床医生不同，医学 AI 模型在接受特定任务训练之前通常缺乏医学领域知识，必须仅依赖于输入数据特征与预测目标之间的统计关联，来推导数据关系。当针对特定医学任务的数据很少时，模型表现不佳。而大模型具有更复杂的网络结构和更多参数，通过结合知识图谱结构、检索式方法等技术，可以学习大量医学知识，更深入、更全面地理解医学概念和它们之间的关系，检索上下文中类似案例。这使得大模型能够推理出以前未见过的任务，并使用医学准确的语言解释相关的输出²。

（二）医疗健康大模型的生态架构

人工智能大模型赋能生命科学与医疗领域发展的生态架构主要由“上游基础层-中游模型层-下游应用层”三层构成。鉴于生物数据

的敏感性和医疗场景的严肃性，监管治理和安全能力对医疗健康大模型的负责任创新发展至关重要，贯穿于行业发展各领域和全过程。

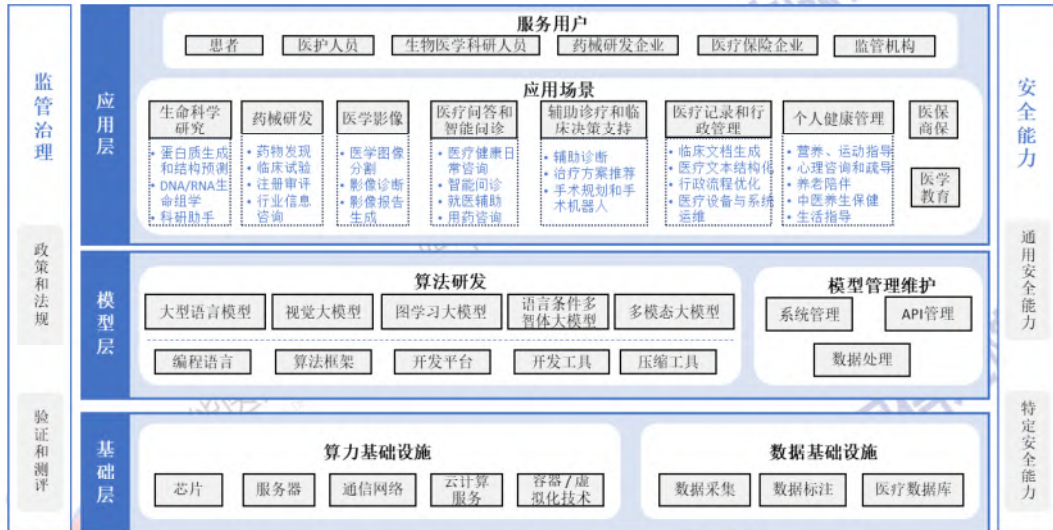


图 1 人工智能大模型+医疗健康生态架构

基础层提供算力和数据支撑资源。基础层是支撑 AI 大模型研发和应用落地的必要资源，包括算力基础设施和数据基础设施。其中，算力基础设施包括通用计算芯片、AI 计算加速芯片、计算服务器、存储服务器、通信网络、云服务、容器/虚拟化等；数据基础设施涉及数据采集和标注、生物信息学数据库、专病数据库、多模态医疗数据库资源等。

模型层形成大模型研发、管理和运维体系。模型研发是在算法研发所需的编程环境（语言）、算法框架、开发平台和工具等基础上，研发产出大型语言模型、视觉大模型、图学习大模型、语言条件多智体大模型、多模态大模型以及生物计算大模型等，完成生物医学自然语言处理、生物医学图像识别、生物医学语音语义识别、生物分子设计等任务。模型管理和运维主要包括系统管理、接口管理、数据处理等。

应用层实现药、械、医、健多场景触达用户。大模型首先赋能生命科学和药械研发，相关应用起步早、发展快、成果较为突出。同时，大模型在医学影像、医疗问答和智能问诊、辅助诊疗和临床决策支持、医学信息提取和生成、行政流程优化、个人健康管理、医保商保、医学教育等方面的应用价值日益凸显，场景探索加速。大模型技术有望为生命科学与医疗行业多个环节带来更加精确、高效、人性化的服务，提升整个生态系统的质量和效率。



人工智能医疗器械创新合作平台



人工智能医疗器械创新合作平台



人工智能医疗器械创新合作平台

CAICT 中国信通院

CAICT 中国信通院

二、医疗健康大模型的技术体系及演进

(一) 大模型响应生物大数据时代的数据特征和应用需求

1. 大模型特点突出，赋能通用医疗人工智能发展

大模型具有大参数、大数据、范式灵活、策略高效等特点。大模型是一种利用海量、多元化数据预训练的深度神经网络模型，能够在不同任务中实现信息生成和推理等能力，具有涌现能力和较强的泛化性、通用性、实用性，在医疗、交通、金融等众多垂直行业场景上，初步展现了惊艳的性能表现和巨大的发展潜力。大模型通常基于 Transformer 等主流架构，通过自注意力机制和并行计算提高性能和泛化能力。大模型的主要特点有：1) 参数规模巨大，从数亿到百万亿不等；2) 数据量庞大，无标签数据用于预训练，数据规模可达数亿至万亿 Token；3) 范式灵活，先用无监督学习预训练，再微调和适应各种下游任务；4) 学习策略高效，能够在零样本、单样本或少样本的情况下，提取、总结、翻译和生成文本信息，甚至支持推理。

大模型有望成为通用医疗人工智能的核心技术和基础设施。大模型具有高度灵活、可重复使用的技术特点，为通用医疗人工智能 (Generalist Medical Artificial Intelligence, GMAI)² 提供了新的可能性。传统医疗人工智能模型是为解决特定任务而设计、训练的，依赖大量的有特异性注释和标签的数据集进行专门的训练，这种狭隘的、针对特定任务的方法使得模型僵化，只能适用于执行训练数据集和其

标签预定义的任务。作为对比，大模型这一人工智能新范式是在大型、多样化的无标签数据集上进行自我监督学习而构建的，可应用于众多下游任务。多模态架构、自监督学习技术以及模型的上下文学习能力等新技术，使得基于大模型开发通用医疗人工智能成为可能⁵。未来医疗健康大模型有望灵活地解释不同的医学模态组合，包括来自成像、电子健康记录、实验室结果、基因组学、图表或医学文本的数据，并将产生表达能力更强的输出，例如自由文本解释、口头建议或图像注释，展示高级医学推理能力。

2. 大模型引领智能时代人工智能开发新范式新变革

人工智能技术不断演进，进入数据驱动的新发展阶段。人工智能经历了漫长的孕育期，最早可追溯到亚里士多德的演绎逻辑，并随着人类对智能的认识不断深入而不断进化。1943年，麦卡洛克和皮茨研制出世界上第一个人工神经网络模型（MP模型），开创了以仿生学观点和结构化方法模拟人类智能的途径；1948年，美国数学家威纳创立了控制论；1950年，图灵明确提出“机器能思维”的观点；1956年，麦卡锡在达特茅斯大学夏季学术研讨会上首次提出“人工智能”术语，人工智能的发展进入到了第一个小高潮时期；随后，历经罗森布拉特、瓦普尼克、沃伯斯等人的努力，多层感知机（MLP）和反向传播算法逐渐成熟并得到广泛应用；1980年，机器学习成为人工智能发展的独立分支，从数据中获取经验，打破了基于规则建模的困境；2006年，辛顿正式提出了深度学习的概念，人工智能进入了第二个小高潮时期，

针对特定应用场景专门训练的神经网络，即小模型，开始不断涌现。



来源：中国信息通信研究院

图 2 人工智能发展历程

以深度学习为代表的人工智能，本质上是一种数据驱动的智能，首先进行模型训练，从大量的数据中“学习”出一些规则而生成一个模型；然后进行模型推理，运用这些规则来解决实际问题。生物进化是多层次上经历自然选择的结果，包括基因序列、蛋白质序列、蛋白质互相作用等层面，这些结果并不随机，而是蕴含一定的规律规则。通过使用包含生物信息的大规模数据进行训练，大模型能够分析生物进化的特点和规律，在一定程度上模拟生物进化的过程，进而具备预测能力，加快生命现象和生命活动规律的研究及发现，助力生物基础科研和医疗技术水平的提升。

数据的爆炸式增长对人工智能技术能力提出新需求。随着知识、数据、算法和算力等关键要素的快速发展，大数据时代出现了数据量

巨大、类型多样、价值稀疏、时效性高等特征，给人工智能的发展带来了新的挑战和需求。一方面，有标签训练集的缺乏和不平衡限制了小模型在特定任务下的性能和泛化能力；另一方面，传统深度学习方法，如循环神经网络（RNN）、卷积神经网络（CNN）等，难以存储和有效利用数据知识。“大数据+大算力+大参数”模型作为一种解决方案乘势而上，悄然酝酿着人工智能在新时代的范式转变。在生命科学与医疗领域，生物数据的复杂性、多样性和海量性，患者数据的敏感性和隐私性，以及科学研究和临床应用对准确性和可靠性的高标准，更使得传统的深度学习方法难以满足行业数字化和智能化发展的需求及挑战。

3. Transformer 架构奠定大模型主流算法架构基础

Transformer 引入注意力机制，提高了深度学习模型的性能和效率。2017 年，谷歌首次提出了 Transformer 架构，其摒弃了此前 MLP、RNN 和 CNN 的处理结构，而是利用自注意力（Self-Attention）机制来得到输入和输出之间的全局依赖关系，能够捕捉到长距离的依赖关系和上下文信息，具备并行性、灵活性和可拓展性等优点⁶。Transformer 架构引入了并行化，可以结合图形处理单元（GPU）或张量处理单元（TPU）在大规模的数据集上进行训练，从而提高了计算效率。基于 Transformer 的模型具有多用途，不局限于单一模态（文本）上，也演化出了可以用于语音、图像、视频等多模态任务的版本，灵活地进行数据融合和交互。Transformer 架构同样具备良好的可拓

展性，可以通过增加参数规模和训练数据规模，来提升模型的性能和泛化能力。

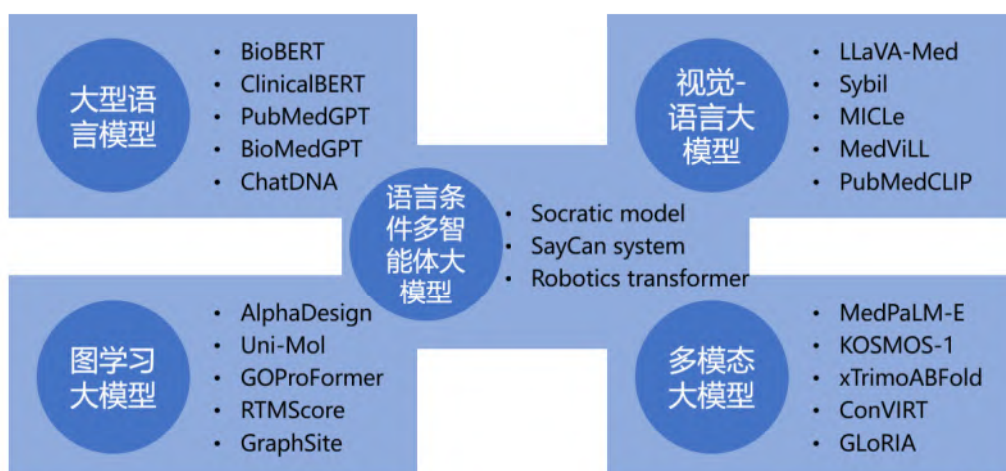
Transformer 架构为医疗健康大模型的发展提供了强大基础 and 支撑。2018 年，谷歌提出了基于 Transformer 的双向编码器表示模型 (BERT)，它是第一个基于 Transformer 架构的大模型，在 11 个自然语言理解任务上取得了当时最好的结果⁷，引发了基于 Transformer 的大模型研究和应用热潮。同样在 2018 年，OpenAI 提出了基于 Transformer 的生成式预训练模型 (GPT)⁸。2020 年，谷歌进一步提出了 Vision Transformer (ViT)，打通了自然语言处理和计算机视觉之间的壁垒⁹。这些模型不断刷新各种自然语言处理任务的性能记录，推动了人工智能领域从深度学习向大模型的范式转移。

当前，医疗健康大模型主要都是基于 Transformer 架构或者混合了 Transformer 架构的神经网络模型。基于 Transformer 架构的大模型，如 BERT 系列、GPT 系列、ViT 系列等，作为大模型的预训练基座模型，可以通过微调、领域适应或领域重训来提高在生物医学任务上的性能。例如，BioGPT、GeneGPT、scGPT、scBERT 和 MedBERT 是分别基于 PubMed 语料库、NCBI 数据库、人类细胞图谱数据、scRNA-seq 数据集和中文临床自然语言文本语料库，进行预训练而得到的生物医学大模型。此外，结合了 Transformer 架构的大模型也在不断发展。例如，麻省理工学院提出了一种灵活的基于语言模型的深度学习策略以更好地理解 and 设计蛋白质，将 Transformer 与图神经网络结合

起来构建了生成式预训练模型，可以用于预测蛋白质的二级结构含量、蛋白质溶解度和序列化任务¹⁰。

(二) 多类型基础模型为医疗健康大模型提供坚实源动力

生命科学与医疗领域涵盖医学文本、医学图像、生命组学、蛋白质工程等多种数据模态。基于不同预训练数据模态的基座模型快速发展，可完成自然语言处理、计算机视觉、图学习等多种任务，有效地应对生物医学数据的多样性、复杂性和高维性，促进领域创新和变革。例如，大型语言模型（Large-scale language models, LLMs）可以用于生成医学文本、回答医学问题、提供医学建议等；视觉大模型（Large-scale vision models, LVMs）和视觉-语言大模型（Large-scale vision language models, VLMs）可以用于识别医学图像、生成图像注释、合成图像等；图学习大模型（Large-scale graph learning models, LGMs）可以用于预测蛋白质结构、设计药物、分析基因组等；语言条件多智能体大模型（Large-scale language-conditioned multiagent models, LLMMs）可以用于实现远程会诊、智能导诊、医疗机器人等；多模态大模型（Large-scale multimodal models, LMMs）可以用于融合多种医学数据、挖掘数据价值、辅助诊断等。大模型的融合应用有望提升医疗产品创新能力和医疗健康服务水平，也将进一步推动人工智能技术的迭代进步和工程化落地。



来源：中国信息通信研究院

图 3 医疗健康大模型的类别和实例

1. 大型语言模型

大型语言模型是当前医疗健康大模型中数量最多、最为活跃的类型类型。大型语言模型在语言数据上进行预训练，并应用于语言下游任务。不同环境下的语言可以有不同的解释，如蛋白质是生命的语言，代码是计算机的语言。大型语言模型能通过自监督学习，从大量未标记数据中提取有用的语义特征和模式，随后通过微调生成符合人类期望的响应。大型语言模型通常基于 Transformer 架构打造。Transformer 可以分为编码器和解码器两部分，编码器提取输入序列的语义表示，大多用于“翻译”；解码器根据编码器的输出生成目标序列，大多用于“生成”。大型语言模型根据模型架构和执行的任务类型分为三个类别¹¹：第一类是纯编码器模型，如 BERT 及其变体。BERT 主要利用双向编码器来捕捉语言中的双向上下文信息，并利用 Transformer 编码器来提取深层的语义表示，从而将输入文本转换为一系列隐藏状态

或特征向量。这些特征向量可以用于各种下游任务，如文本分类、命名实体识别、关系抽取等。纯编码器模型通常使用掩码语言建模（MLM）或下一个句子预测（NSP）等无监督或自监督的目标进行预训练。第二类是纯解码器模型，如 GPT 系列。GPT 是一种基于 Transformer 的生成式预训练语言模型，主要利用自回归模型来捕捉语言中的上下文信息，并利用 Transformer 解码器来提取深层的语义特征，从而实现高质量的文本生成，可以用于文本生成、摘要、翻译等任务。纯解码器模型通常使用下一个标记预测（NTP）或置换语言建模（PLM）等自回归或自监督的目标进行预训练。第三类是编码器-解码器模型，如文本到文本传输转换器（T5）和双向自回归转换器（BART）。编码器-解码器模型通常使用序列到序列（Seq2Seq）或去噪自编码（DAE）等自监督或半监督的目标进行预训练。

医疗健康大型语言模型是基于通用大型语言模型的领域增强模型。为了提高大型语言模型在医疗健康领域的专业性和可靠性，通常需要在 GPT、BERT、T5 等大型语言模型基础上，利用基因或蛋白质序列信息、医学文本数据或脱敏后的患者全记录周期内的医疗代码序列，进行再训练或微调，以适应特定领域或任务的需求。例如，ChatDoctor、BioBERT¹²、MedGPT 等针对生物医学领域的预训练模型，分别基于 HealthCareMagic 网站的医患对话、PubMed 摘要和 PubMed Central（PMC）全文、MIMIC-III 临床电子健康记录进行预训练。生物医学大型语言模型已经在生物医学文本挖掘与知识发现、医学对话

系统、生物与化学序列等场景中展现出强大的应用能力。

2. 视觉大模型和视觉-语言大模型

医疗健康视觉大模型快速发展，模型架构和性能有待持续探索和优化。自然语言处理领域 Transformer 架构的突破，带动了视觉领域大模型的新发展。视觉大模型在视觉数据上进行预训练并应用于视觉下游任务。视觉大模型基于卷积神经网络(CNN)或 Vision Transformer (ViT) 架构，采用监督学习和无监督学习范式进行预训练⁴。ViT 是一种基于 Transformer 架构的纯视觉模型，借鉴了自然语言处理中的 Transformer 自注意力机制，其中将词嵌入替换成了 patch 嵌入。ViT 通过将图片信息切割成 patch 并进行线性映射，将图像转换为序列输入，从而可以使用 Transformer 进行编码和解码。ViT 在医疗健康领域应用优势突出，能够通过注意力机制有效地学习长期依赖关系，有效地整合多种医学模态，并通过多头注意力结构提供更可解释的模型，相较传统的 CNN 更加高效、更接近人类感知。

整合语言和视觉数据的医疗实际需求，催生出医疗健康视觉-语言大模型。视觉-语言医疗健康大模型可以结合视觉数据(如图像或视频)处理和生成自然语言文本，利用大量的图像-文本对进行预训练，从而学习到视觉和语言之间的对齐和关联，实现多种跨模态的任务，如图像-文本生成、检索、分类、标注等。在医疗健康场景中，视觉-语言医疗健康大模型可以自动生成医学报告、对医学图像和视频进行标注和解释、通过分析视觉信息辅助临床决策等，为智慧医疗发展提

供底层支撑，有望用于个性化预测和早期干预、患者远程监测、疾病诊断等方面。目前，典型的视觉-语言大模型 DALL-E、CLIP、ALIGN 和 Flamingo 等在医疗领域的应用探索加快，也已涌现出 MedViLL、PubMedCLIP、ConVIRT、Med-Flamingo¹³、PLIP 等针对医疗健康领域的视觉-语言大模型。

案例：LLaVA-Med¹⁴

LLaVA-Med 模型由微软推出，是一种基于 GPT-4 的视觉-语言对话助手，可以回答关于生物医学图像的开放式研究问题¹⁵。

LLaVA-Med 利用从 PubMed Central 提取的大规模、广覆盖的生物医学图题 (figure-caption) 数据集 (包括胸部 X 光、MRI、组织学、病理学和 CT 图像等) 进行预训练，使用 GPT-4 从图题中自我指导生成开放性指令遵循数据，并使用一种创新性的分阶段学习方法 (a novel curriculum learning method) 对通用领域的视觉-语言大模型进行微调。具体来说，该模型首先使用图解对齐生物医学词汇，然后使用 GPT-4 生成的指令跟随数据学习掌握开放性对话语义，广泛模拟一个外行人逐渐获取生物医学知识的过程。

LLaVA-Med 具有较强的多模态对话能力，可根据 CT、X 光图片等推测出患者病理状况，并生成有关图像的问答。LLaVA-Med 不仅可以提高疾病检测效率以及医学影像领域的智能化分析，还可以以自然语言回答用户有关生物医学图像的问题。



图 4 LLaVA-Med 的预训练和微调流程

3. 图学习大模型

图学习大模型可以有效应对生物序列数据的复杂性。生物测序数据，如蛋白质和药物分子序列，具有复杂的结构和关系，传统的机器学习方法难以有效地处理和生成此类数据。图学习大模型利用图神经网络（GNN）等图学习技术，将生物序列数据表示为图形，并通过图形算法进行分析和推理，在生物计算和药物研发领域有着重要的应用价值。Graph Transformer 是一种基于 Transformer 架构的图学习大模型，引入了注意力机制，能有效地学习长期依赖关系，并且能够与频域信息相结合，以提高表达能力。Graph Transformer 有三种架构：1）在 GNN 上加入 transformer blocks；2）交替使用 GNN blocks 和 transformer blocks；3）并行使用 GNN blocks 和 transformer blocks。通用 Graph Transformer 包含三个要素¹⁶：位置/结构编码、局部消息传递机制和全局注意力机制。近年来，将结构编码纳入模型中成为一个热点研究方向，涌现出 SAT 和 GraphiT 典型案例。

图学习模型可实现生物分子结构、药物-靶点相互作用分析等多种功能。蛋白质序列与自然语言存在相似之处，也存在一定差异，将图学习模型引入蛋白质、药物分子相关任务，可以预测蛋白质功能、评估蛋白质质量、预测蛋白质-配体结合位点和蛋白质-DNA 结合位点、预测药物-靶点相互作用等。例如，ADesign 针对从 3D 结构预测蛋白质序列问题，使用简化的图形 Transformer 编码器（SGT）和置信感知蛋白质解码器（CPD），提高蛋白质设计效率¹⁷。MHTAN-DTI 是一种基于 Transformer 和层次图注意力网络的药物-靶点相互作用预测模型

¹⁸，可用于发现作用于特定蛋白质的潜在药物，助力药物重定位、药物副作用预测、多重药理学和耐药性的研究。

4. 语言条件多智能体大模型

语言条件多智能体模型有望成为未来医疗机器人的心脏和基石。

语言条件多智能体模型是一种利用语言作为多个大模型之间的中介接口的新技术，可以将不同类型和领域的大模型串联起来，形成一个强大的智能体，能够完成单个模型难以完成的任务，代表性案例有 Socratic model、SayCan system、Robotics transformer。语言条件多智能体模型可能使用包括大型语言模型、视觉-语言模型、音频-语言模型、视觉-导航模型等不同的基础大模型，来执行更复杂和多模态的任务，如人机对话或多方协作等。例如，基于语言条件多智能体模型打造的虚拟医疗助理和手术机器人对真实世界理解、交互能力将显著提升，前者可以用来与患者进行远程交流、诊断、预约、检查、治疗等，后者可以用来接收和解释医生的指令、处理和生成手术图像和声音、控制机械臂进行手术操作等。未来，医疗健康语言条件多智能体大模型可以更自然、更智能、更灵活地与人类医生和患者沟通协作，提高医疗质量和效率，降低医疗成本和风险；也可以利用医疗大数据，从海量的医学文献、图像、视频等数据中学习新知识新技能，为医学创新和发现提供支持；还可以根据不同的场景和需求，动态地调整自己的行为 and 策略，以适应复杂和不断变化的医疗环境。

5. 多模态大模型

多模态大模型综合分析各种类型的医学数据，实现架构、模态、任务统一。多模态大模型在多种模态数据上进行预训练，并应用于各种单模态或多模态下游任务。随着医疗信息化的快速发展以及医疗设备的更新迭代，海量且类型多样的医学数据应运而生，涵盖临床文本、影像和波形、音频、视频、生物组学数据等多种类型。医疗多模态大模型能够融合和分析各种类型的医学数据，将多模态任务表达为序列到序列生成的形式，结合任务特定的指令在经典的 Transformer 架构中实现**架构统一、模态统一和任务统一**。架构统一是指预训练和微调使用同一 Transformer 编码-解码器；模态统一是指将自然语言处理、计算机视觉和多模态任务统一到同一框架和训练范式中；任务统一是指将任务统一表达成序列到序列的形式，预训练和微调均使用生成范式进行训练，模型可以同时学习多任务，让一个模型通过一次预训练即可获得多种能力，包括文本生成、图像生成、跨模态理解等。

多模态大模型提升医疗诊断准确性，是实现临床落地和智能化价值的关键。获取患者相关数据的每类方式均为一种数据模态，不同模态的医学数据都从特定的角度提供了患者的诊疗信息，信息间既有重叠又有互补。以往单模态的模型只能分析疾病某一层面的信息，具有较大局限性，极大限制了人工智能的医疗应用，而多模态大模型结合多种模态的医学信息，进一步提高了诊断治疗的准确性，是人工智能诊疗产品临床落地的关键。目前国内外涌现多个医疗多模态大模型案例，服务于患者诊断、手术导航、康复训练、影像报告生成等场景。

案例：Med-PaLM M¹⁹

Med-PaLM-M 由谷歌 Research 和 DeepMind 团队共同研发，是一个多任务多模态的通用生物学大模型，可以处理包括临床文本、医学图像和基因组学数据在内的多种医疗健康数据。Med-PaLM M 基于 PaLM-E 多模态模型构建，通过指令微调和生物学领域对齐，在 MultiMedBench 上进行了训练；还在通用领域语料库（英语 Wikipedia 和 BooksCorpus）上进行了预训练，学习了跨模态数据之间的关联和知识。

PaLM-E 是谷歌于 2023 年发布的迄今最大的多模态具身视觉语言大模型²⁰，拥有 5620 亿个参数，在需要多模态理解和推理的零样本任务上表现良好。PaLM-E 使用不同的编码器将来自不同模态的信息映射到语言嵌入空间，然后将这些模态、状态向量整合到一个大型语言模型中。主要的模态、状态向量包括 2D 图像（使用 ViT 进行编码）以及 3D 感知信息（使用对象场景表示变换器进行编码）。除了传统的语言生成任务外，PaLM-E 还可以用于连续的机器人控制规划、视觉问答、图像描述、知识获取等多模态理解和推理任务上都取得了最先进的结果。

MultiMedBench 是谷歌自建的多模态医学测试基准。该基准由 12 个开源数据集和 14 个单独的任务组成，用于测量通用生物学 AI 执行各种临床任务的能力。其中 12 个数据集共包含了六种生物学数据模式（文本、放射学（CT、MRI 和 X 光）、病理学、皮肤病学、乳房 X 光检查和基因组学），14 个任务则涵盖五种类型（问题回答、报告生成和摘要、视觉问题回答、医学图像分类和基因组变异识别）。

Med-PaLM-M 在 14 个不同的医疗任务上接近或超过了现有的最先进模型，包括医疗问答、影像分类、基因预测等。Med-PaLM-M 为医疗领域提供了强大、灵活的生成模型，可处理多种数据类型和任务，促进医疗知识的发现和应用。Med-PaLM-M 能灵活地编码和解释多模态的生物学数据，并且具备仅使用基于语言的指令和提示就能准确识别和描述未见过的图像中的医疗状况的能力。

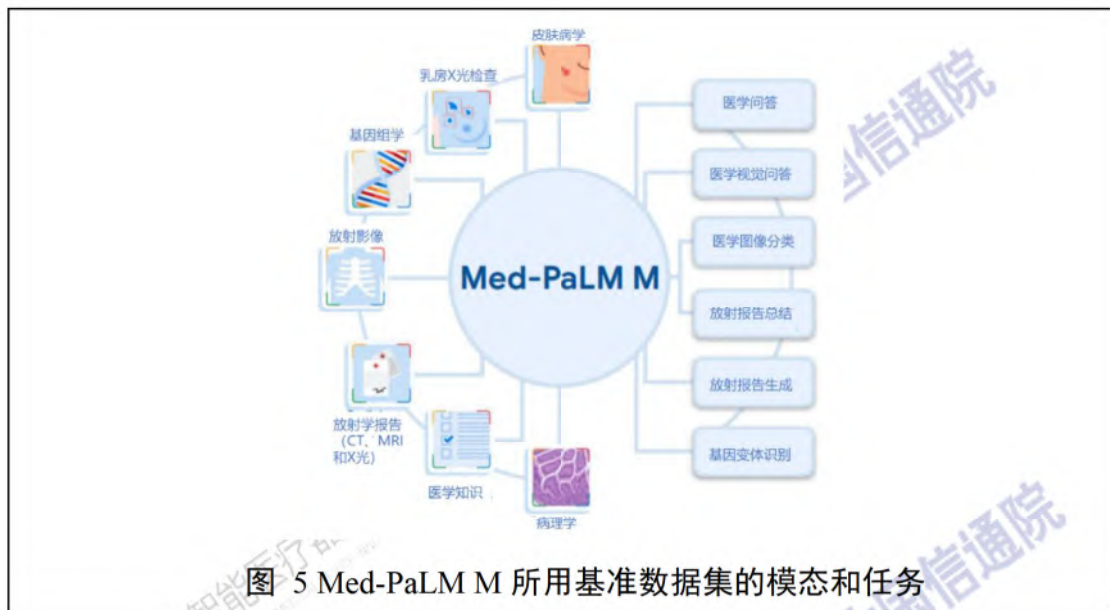
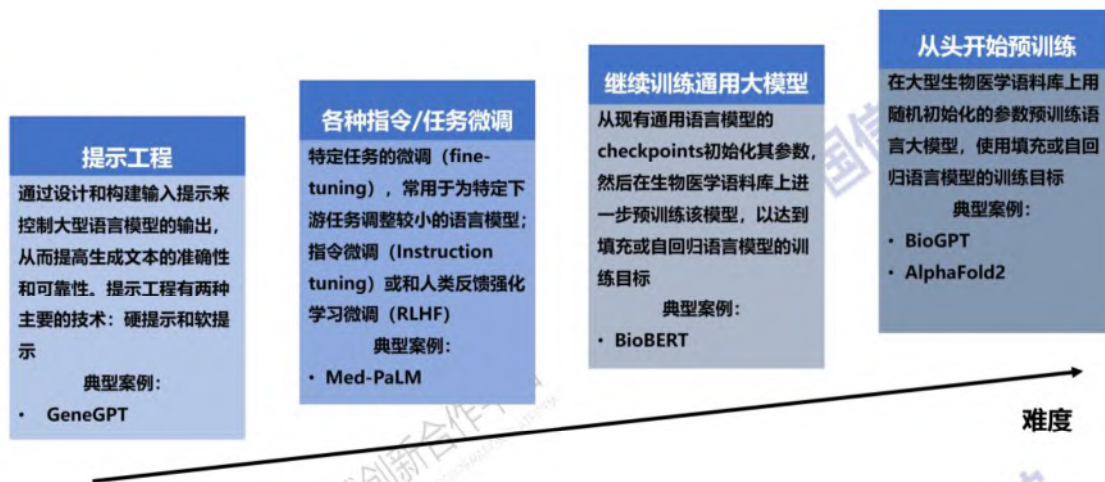


图 5 Med-PaLM M 所用基准数据集的模态和任务

(三) 四种范式助力大模型在医疗健康垂直领域的 应用实践

目前，有多种策略可使通用基础模型适应生命科学与医疗健康领域的特定任务。总体而言，在生命科学与医疗领域应用大模型主要有四种技术范式²¹，由难到易分别从头开始预训练、继续训练通用大模型、各种指令/任务微调和提示工程。当有大量数据、计算资源和专业知识时，可以通过从头开始或继续训练通用大模型，来开发生命科学与医疗健康领域特定模型，但成本相对昂贵。各种指令/任务微调和提示工程更具成本效益。不同范式可以单独或组合使用，以满足生命科学与医疗健康场景需求。



来源：中国信息通信研究院

图 6 生命科学与医疗领域开发与应用大模型的四种范式

1. 从头开始的预训练

从头开始进行预训练是创建医疗健康大模型的最具挑战性的范式。这种范式不依赖于通用领域的预训练模型，而是直接在大型生物医学数据库上用随机初始化的参数预训练大模型，这样可以避免领域不匹配的问题，提高模型在下游任务上的泛化能力。同时它可以根据不同的任务需求，选择不同的预训练目标和策略，例如使用掩码语言模型（MLM）或自回归语言模型（ARLM）作为训练目标，使用课程学习或对比学习等方法进行优化。典型的大模型案例有生物医学文本生成和挖掘类模型 BioMedLM（以前被称为 PubMed GPT）以及临床语言模型 GatorTron²²。

案例：GatorTron²³

GatorTron 由佛罗里达大学（UF）与英伟达的研究人员合作开发，是一种经过电子健康记录中超过 900 亿字文本训练的大型语言模型，可以快速准确地从大量临床数据中提取见解来加速医学研究和支持医疗决策。

临床语言大模型 GatorTron 在 UF health 临床笔记（超过 820 亿字文本）、PubMed 文章（60 亿字文本）、维基百科（25 亿字文本）和 MIMIC-III（5 亿字文本）的大型数据集上进行了训练。UF Health 临床数据集收集了 2011-2021 年超过 126 个科室的临床笔记，大约 200 万患者的 5000 万次住院、门诊和急诊环境下的就诊记录。MIMIC 数据库（Medical Information Mart for Intensive Care）是麻省理工学院（MIT）下属管理的一个大型、单中心、可供自由使用的公共临床数据，MIMIC-III v1.4 版由 2001 年~2012 年在 Beth Israel Deaconess 医疗中心重症监护室中住院的 4 万多病人的 200 万条临床记录数据组成。

研究人员在五个临床自然语言处理任务（临床概念提取（或临床命名实体识别）、医疗关系提取、语义文本相似性、自然语言推理和医学问答）上评估了 GatorTron 模型，并发现其性能优于现有的生物医学和临床笔记训练的 Transformer 模型。具有更多参数的较大模型会比相对较小的模型表现更好，并且通过增加用于训练模型的数据量能够提高其性能。

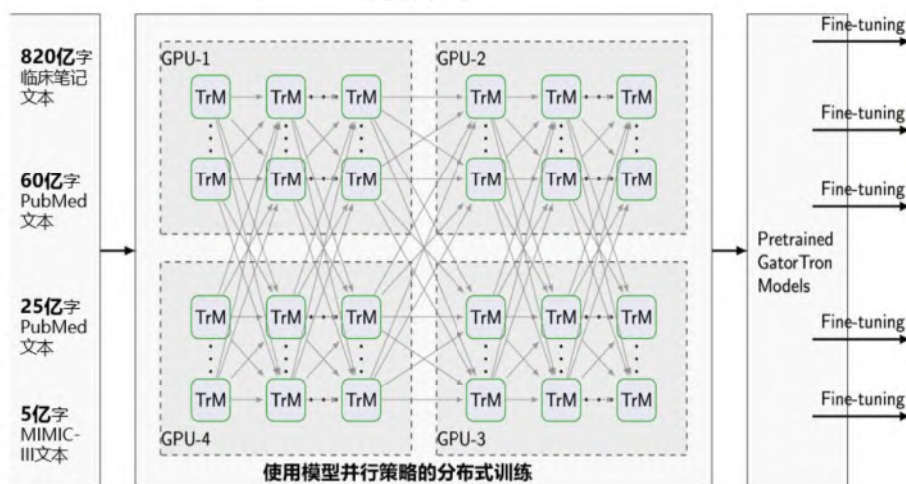


图 7 GatorTron 模型的预训练和微调过程

2. 继续训练通用大模型

继续训练通用大模型是一种兼顾了时间、成本和多功能性的技术范式。该范式充分利用现有通用基础模型的知识 and 能力，从现有通用

语言模型的 checkpoints 初始化模型参数，然后在生物学数据库上进一步预训练，以达到掩码语言模型或自回归语言模型的训练目标。这种策略可以利用通用基础模型已经学习到的通用知识和能力，同时在特定领域的的数据上进行适应、优化和增强，提高模型的性能和泛化能力。继续训练通用大模型可以节省预训练的时间和成本，同时保留通用基础模型的多功能性，可以实现多种功能，例如文本生成、文本理解、文本分类、文本检索、文本摘要等，为生命科学与医疗领域提供多样化的服务和应用。典型的大模型案例有生物学文本挖掘大型语言模型 BioBERT 和医疗问答大型语言模型 PMC-LLaMA，前者基于 BERT 在 PubMed 等生物学语料库上进行进一步预训练而得到，后者通过在 PubMed Central 文章上进一步预训练 LLaMA-7B 模型而得到。

案例：BioBERT¹²

BioBERT 是由韩国 NAVER 公司和韩国科学技术研究院 (KAIST) 的研究人员合作研发，是一种用于生物学文本挖掘的生物学大型语言模型。

BioBERT 在大型生物学语料库 (PubMed 摘要和 PMC 全文文章) 上对 BERT 进行预训练, 适应生物学领域的术语和表达方式。BioBERT 利用 BERT 的强大的语言表示能力，提高了下游任务的性能和泛化能力。

BioBERT 在生物学命名实体识别、生物学关系抽取和生物学问答这些生物学文本挖掘任务上取得了最先进的性能。BioBERT 为生物学领域提供了一个通用和高效的语言模型，可以应用于各种场景和需求。

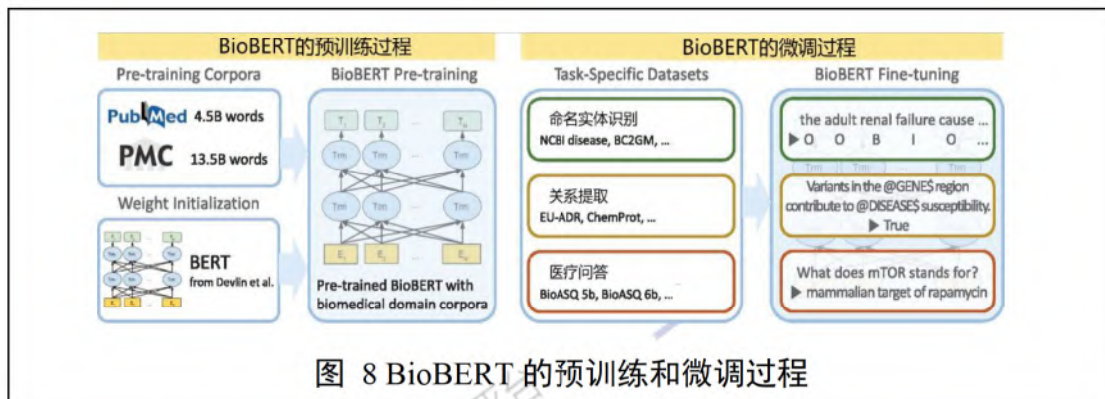


图 8 BioBERT 的预训练和微调过程

3. 各种指令/任务微调

各种指令/任务微调是医疗健康大模型中应用最为广泛的技术范式。该范式可以在已经预训练好的大模型的基础上，对其进行不同程度和方式的优化，以适应不同的下游任务和场景。各种指令/任务微调有两种方式，一种是特定任务的微调（fine-tuning），常用于为特定下游任务调整较小的模型。这种策略是先在大规模未标注数据集上训练一个大模型，然后在较小的任务特定数据集上对其进行性能提升。但微调需要很多任务特定的样本，只能在训练过的任务上运行，并且可能会导致灾难性遗忘。例如，BioGPT 就是在从头开始预训练后，又在特定任务数据上进行了微调。

另一种则是指令微调（instruction tuning）或人类反馈强化学习微调（Reinforcement Learning from Human Feedback, RLHF）。该范式通过有监督学习和/或强化学习，在指令-响应对的数据上对模型进行微调，从而使大模型与更好的指令响应特性或领域知识保持一致。指令微调使用一组多样化的指令对模型进行微调，强调知识理解。指令微调可以使模型同时学习多个任务，并有效地处理未见过的任务。人类

反馈强化学习微调则使用人类偏好数据集来训练一个奖励模型，可以预测奖励函数，并通过强化学习算法进行优化。

案例：ChatDoctor²⁴

ChatDoctor 是一个主要在 LLaMA 上微调的医学领域的大型语言模型。ChatDoctor 在大量医学文献上训练，可理解医学术语、程序和诊断情况。患者可以通过聊天接口与 ChatDoctor 模型互动，询问有关健康、症状或医疗状况的问题。

ChatDoctor 在 LLaMA 模型的基础上进行微调训练，样本数据包含真实医患对话，同时加入自主知识检索能力，比如 Wikipedia 或者疾病数据库。通过微调训练，模型在理解患者需求、提供合理建议并在各种医疗相关领域提供帮助方面能力显著提升。自主知识检索能力能够实时访问权威信息，提示模型回答的准确性。

具体来说，ChatDoctor 首先基于 Alpaca 的数据集微调了 LLaMA，产生一个通用对话模型，然后在从在线问答医疗咨询网站“HealthCareMagic”获得的 100k 条真实的医患对话数据集上进行了微调。该项目还从在线医疗咨询网站“iCliniq2”收集了大约 10k 条医患对话用于评估 ChatDoctor 模型的性能。此外，为了提高模型的可信度，该项目还设计了一个基于 Wikipedia 和医疗领域数据库的知识大脑，它可以实时访问权威信息，并根据这些可靠信息回答患者的问，这对容错率较低的医疗领域至关重要。

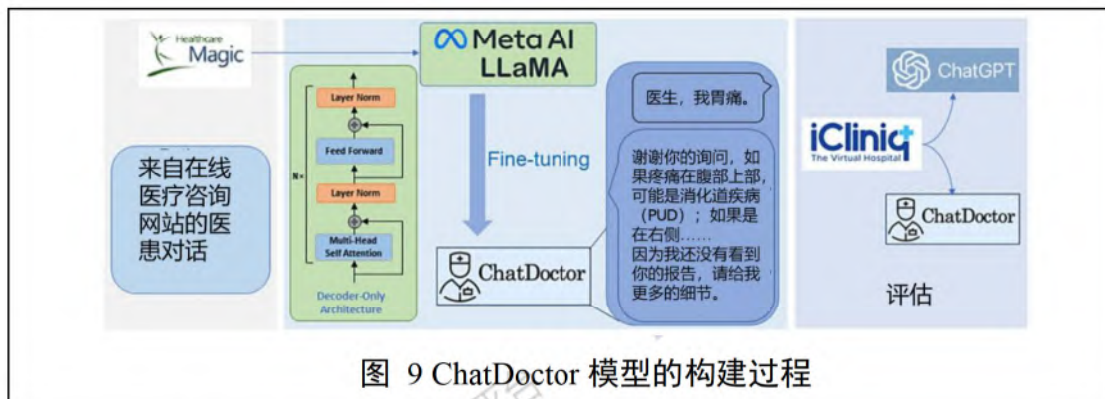


图 9 ChatDoctor 模型的构建过程

4. 提示工程

提示工程通常与其他范式结合使用，增强大模型在生命科学与医疗特定领域和任务上的表现。提示工程是一种利用大模型的能力和潜力来完成特定的下游任务的方法，通过设计和构建输入提示来控制大模型的输出，从而提高生成内容的准确性和可靠性。提示工程有两种主要的技术：硬提示和软提示。硬提示是由人工设计的由具体词汇组成的提示，是人类可读的提示；而软提示是由机器自动优化生成的由特征向量组成的提示，不是人类可读的提示。硬提示需要人工根据不同的任务和模型进行尝试和调整，而软提示可以通过梯度搜索等方法自动学习最佳的向量表示。软提示通常比硬提示更有效地利用了预训练语言模型的潜力。提示工程可以减少所需的任务特定样本的数量，提高少样本性能，但是需要进行提示设计，并且可能会受到上下文长度的限制。提示工程已经在多个领域和场景中得到了应用，例如 Med-PaLM 模型是通过软提示调整将 Flan-PaLM 应用于生物医学领域的典型案例。在 Flan-PaLM 的基础上，Med-PaLM 使用了一个小规模生物医学语料库进行微调，并使用了一个基于梯度下降法优化的软提示

向量来调节 Flan-PaLM 的参数。通过这种方式，Med-PaLM 在多个生物学自然语言处理任务上都取得了最先进或接近最先进的结果。这表明，提示工程可以有效地利用现有通用语言模型的知识能力，并在特定领域和任务上进行适应和优化。

案例：Med-PaLM 2²⁵

Med-PaLM 2 由谷歌研究和 DeepMind 团队共同研发，是一种用于回答各种医学领域问题的医学领域微调大型语言模型。

Med-PaLM 2 基于谷歌的大型语言模型 PaLM 2，进行针对医学领域的微调训练。Med-PaLM 2 利用了一种新的提示策略——集成精炼（ensemble refinement），它可以通过多个推理路径来提高大型语言模型的推理能力。

Med-PaLM 2 在医学考试问题上持续表现出“专家”医生水平。Med-PaLM 2 在多个医疗问答数据集上实现了最先进的结果，并且在人工评估中得到了医生和消费者的高度认可。例如，Med-PaLM2 在 MedQA 数据集上得分高达 86.5%，比 Med-PaLM 提高了超过 19%；对消费者医学问题的长篇回答进行人类评估显示，Med-PaLM2 的回答在与临床效用相关的九个轴线中有八个轴线优于临床医生和 Med-PaLM 的回答，例如事实性、医学推理能力和低危害可能性等。Med-PaLM 2 为医学考试、消费者健康和医学研究等场景提供了一个更强大和更灵活的问答系统。

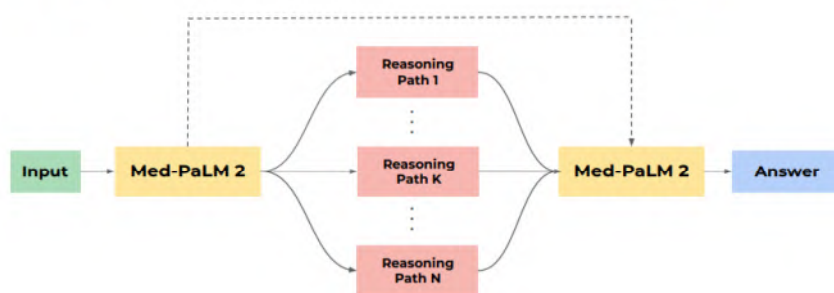


图 10 Med-PaLM 2 采用多种路径提升推理能力

（四）模型发展呈现家族化、多模态、融合化、协同

化趋势

1. 家族化、系列化大模型将持续涌现，迭代升级不断加快

医疗健康大模型的横向和纵向拓展，提供医疗人工智能多样化发展路径。横向来看，业界出现以 BERT 系列、GPT 系列、ViT 系列为代表的基础模型，以 BERT 系列、GPT 系列、PaLM 系列、ViT 系列为代表的基础模型引领大模型在生命科学与医疗领域的应用，同时该领域也出现了 Megatron²⁶、Vicuna²⁷等其他的模型系列。这些模型在自然语言处理、自然语言生成和视觉识别乃至多模态任务等领域具有强大的能力。通过微调、提示工程等技术，这些基础模型可以适应不同的医疗任务，涌现出了大量以 GPT、BERT 为后缀的家族化医疗健康大模型，例如 scGPT、MolReGPT、CancerGPT、MedGPT、MedBERT 等。纵向来看，业界领先的医疗健康大模型的系列化升级和迭代也是大势所趋，例如 Med-PaLM 系列逐代升级，不断刷新医疗健康细分领域的最先进水平。

基础模型的升级显著地提升医疗健康大模型的性能和功能。例如 GPT 系列从 GPT-2 到 GPT-4，参数规模从 1 亿数量级增加到 10,000 亿数量级，训练数据规模更是大幅度提升。参数规模和训练数据规模的增加，使得 GPT-4 相比于 GPT-2 和 GPT-3，在语言能力、泛化能力和多模态能力上有了显著的提升，阐述更准确，且可实现多模态输入，推动了其在医疗领域更广泛的应用。基于 GPT-4 打造的医疗健康大模

型因此可以更好地理解和生成复杂、多样、专业的文本、图像、语音、视频等医疗健康数据，实现了更好的数据融合、信息挖掘、人机交互等高级功能²⁸。

2. 生物医学数据复杂多样，催化大模型多模态化发展提速

医疗健康大模型的多模态发展，迈向通往通用医疗人工智能的关键一步。由于医疗健康数据和应用场景的复杂性，多模态、跨尺度的数据格式和处理需求推动了医疗健康大模型多模态融合发展。一是多模态医疗健康模型的规模将更大，能处理的医学模态将更丰富。目前的多模态医疗健康大模型主要涉及视觉和语言两种模态，未来可以融合更多模态进行大规模预训练，结合各种数据类型（文本、图像、视频、音频、数据库）和尺度（分子、基因、细胞、组织、患者、群体），进一步释放大模型在科学发现和临床诊疗方面的潜力。二是多模态医疗健康大模型的训练将加速。多模态大模型需要统一架构、模态和任务，复杂多样的生物信息和医学健康数据处理对算法和算力的要求很高，因此需要采用更高效、更经济的训练方法和技术。类似 FastMoE 的优化算法将不断涌现，大模型的计算效率和训练速度将进一步提高。三是多模态医疗健康大模型将走向“真正统一”，将能够适应多种不同类型、模态和层次的生物医学数据，实现有效且鲁棒的信息编码、解释、生成等操作的能力，还将适应多重医疗场景和诊疗需求，服务大健康产业。当前，微软 KOSMOS-1 和谷歌 PaLM-E 都是适应多模

态数据的通用人工智能模型，经过医学领域微调后可以打造生物医学领域的通用大模型，这些技术的进步与迭代使得人们对于未来完全通才的医疗健康人工智能有了更清晰的具象。

3. 知识融合趋势日渐显现，推动多维生物信息全面理解分析

医疗健康大模型将进一步融合不同领域、任务和模态的知识，并将其创新地应用于各种生命科学研究和医疗健康场景中。**跨领域知识融合方面**，医疗健康大模型将整合多个领域和学科的知识，如基因组学、生物信息学、药物化学、临床医学、公共卫生、物理化学等，以实现多模态、多层次、多维度的生物医学数据的全面理解和综合分析。**跨任务知识融合方面**，医疗健康大模型将具备跨任务的泛化能力，能够在不同的医疗健康任务之间进行知识迁移和共享，如临床问题问答、消费者医学问题回答、X 光报告生成等。**跨模态知识融合方面**，医疗健康大模型将具备跨模态的表达能力，即能够处理和生成包括文本、图像、声音、视频等在内的多种类型的医疗健康数据，实现不同模态数据之间的语义对齐和信息互补，从而更好地服务科学家、临床医生和病患。

4. 大小模型协同进化，性能与成本平衡是未来重点研究方向

目前大多数医疗 AI 企业都是基于小模型的模式，未来大模型和

小模型将长期共存发展。人工智能模型不一定参数越大越好，也不是所有的医疗健康场景都需要部署使用大模型，满足业务场景需求、平衡成本和模型性能、便捷化部署、可靠足量数据获取等在实践应用中也十分重要。小模型更易在窄域场景实现较高精度，比如医疗影像分析等对精度要求高、对泛化能力要求低的场景更合适小模型，但大模型更适合泛化和通用能力要求高的广域场景，但往往对精度要求没那么高²⁹。小模型更适合要求低成本、低时延的场景，而大模型目前还是适合对成本和时延要求没那么敏感的领域。未来，大模型和小模型有望协同推动人工智能的发展，大模型向小模型输出模型能力，小模型更精确地执行任务，再将相关数据与结果反哺给大模型，促进大模型迭代更新，从而达到降低能耗、提高整体模型精度的效果³⁰，高效率、低成本地解决业务问题。

三、医疗健康大模型的应用场景

中国信通院通过桌面调研、案例征集、专家访谈等方式，收集并整理了一个医疗健康大模型应用案例库（截至 2023 年 9 月底），囊括国内外超过 260 个典型案例（已发布或在研），涵盖生命科学研究、医药研发、医疗器械、医疗服务、健康管理、行业支撑等多个细分领域，涉及高校、新型科研机构、科技企业、智能医疗企业等多个创新主体。基于此案例库，结合大模型发展潜力，中国信通院对大模型赋能生命科学与医疗健康应用发展态势进行了整体分析，并由此梳理出发展较快、潜力较大的应用场景（一级场景和二级场景）。

（一）整体发展分析

科研院校大多基于开源模型进行微调，以 LLaMA 为底座的模型呈引领之势。开源大模型成本更低、迭代速度快、定制化上限高，目前学术界已经出现了以不同开源基础模型底座发展起来的一众医疗大模型。其中，以 Meta 的 LLaMA 为底座的模型最多，约占 50%，如 MedAlpaca、NHS-LLM、ChatDoctor、DoctorGPT，这些模型可完成医疗问答任务。其次是以 ChatGLM 为底座的模型，如 DoctorGLM、SoulChat 等，这些模型可完成医疗问答、慢性病和心理咨询任务。第三是以 bloomz 为底座的模型，如 ClinicalGPT、MING 等，这些模型可完成医疗问答任务。此外，少数医疗健康大模型使用 Baichuan、元语智能（ChatYuan）作为底座。

案例：ClinicalGPT

ClinicalGPT 由北京邮电大学信息智能与数字健康研究所研究团队研发，是面向医学场景的基础大模型。

ClinicalGPT 使用超过数千万条医学记录及数百万例患者的大规模电子病历进行了增量预训练和指令微调，并通过人类反馈强化学习进行了人类医生行为和医疗安全性的对齐。目前，ClinicalGPT 包括 7B、13B、176B 的版本，其中 ClinicalBERT、ClinicalGPT 7B 已经开源，上线 Hugging Face 以来，平均单月下载超过 10,000+。

ClinicalGPT 在多种临床任务，包括多轮问诊互动、医学检查、疾病及鉴别诊断、医学考试、以及医学知识问答上均体现出了优秀的能力。此外，通过结合图像语义理解，ClinicalGPT 也具备了多模态的理解能力，能够基于人脸进行多种疾病风险预测，通过上传人脸的照片即，可给出包括心血管疾病、代谢类疾病、免疫系统疾病等个体疾病风险的评估与预测结果。基于 ClinicalGPT

的智能问诊系统与中医智能导诊系统已在多家知名医院进行了临床及医疗服务场景的应用落地。

在通用医学方面，ClinicalGPT 具备结合病史进行多轮问诊能力，并且在智能导诊、症状分析、疾病类型诊断、疾病鉴别诊断等方面具备了出色的能力，其辅助临床决策能力覆盖了呼吸、消化、泌尿、精神病学、神经病学、妇科和血液学等科室。

在中医临床方面，基于中医理论、辨证施治、经络穴位、中药配方、术语准确性、剂量准确性、区分中西医概念等维度，ClinicalGPT 覆盖了包括心脏、肿瘤、针灸等 16 大科室，能够根据用户提供的主诉、疾病、症状、体征信息，给出辨证施治结果和中医方剂方案，同时能够根据患者的主诉及病历信息提供健康、食疗、推拿、针灸等多维度调理方案。

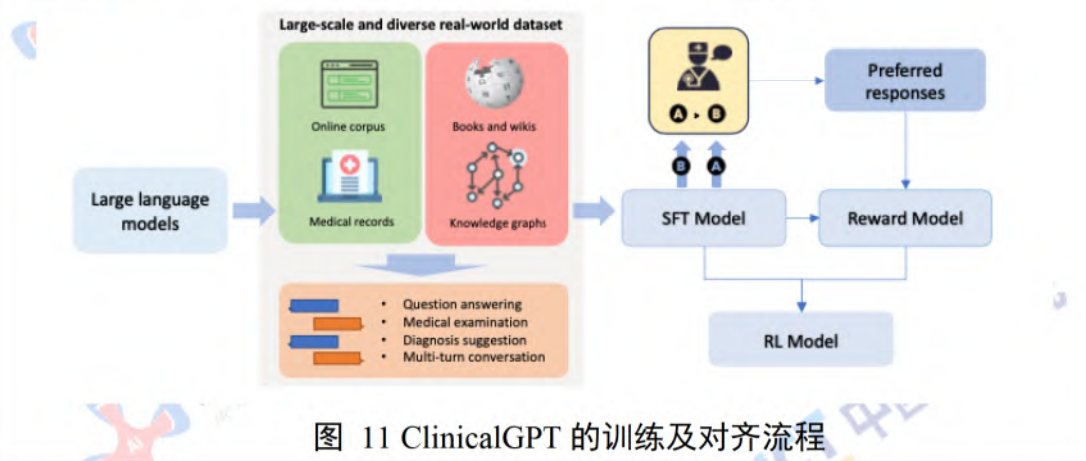


图 11 ClinicalGPT 的训练及对齐流程

大型科技公司纷纷自研通用大模型，平台化赋能生命科学与医疗健康行业发展。科技巨头依托算力、资金优势，在自研通用大模型之上，加快推出生命科学与医疗垂直行业通用基础模型及平台，赋能下游细分应用发展。英伟达推出生物医药基础模型云服务平台 BioNeMo，提供生成式化学、蛋白质语言和结构预测相关开源模型，支持蛋白质、DNA 和生物化学数据处理，加速药物研发。英特尔研发万亿参数的科学生成式模型 Aurora genAI，根据生物学、化学、材料学、物理学、

医学和其他来源的通用文本、代码、科学文本和结构化科学数据进行训练，可加速科学研究和新药研发。腾讯基于“混元”大模型底座的“腾讯医疗大模型”赋能文案生成、智能问答、病历结构化和检索、影像报告和辅助诊断等场景。百度基于国产“文心”大模型底座的“灵医大模型”已全面向公立医院、药械企业、互联网医院、连锁药房等机构开放体验，并支持其他生态合作方进行 API 调用、插件集成。阿里基于“通义”大模型底座的“通义医疗行业大模型”助力医疗、医药业务场景模型构建和全生命周期二次训练、推理、评测和模型加速应用服务。

案例：商量大医

商量大医是由商汤科技研发的医疗健康大型语言模型，以商汤自研的千亿参数规模的大型语言模型“商量”为基座（拥有万亿 token 预训练语料），利用超 200 亿 tokens 的高质量医学知识数据训练而成，数据范围涵盖医学教材、医学指南、临床路径、药品库、疾病库、体检报告等资料，以及 4000 万真实病历、医患问答和对话等。

在增量预训练、指令调优、奖励模型构建以及基于执业医师反馈的强化学习训练基础上，商汤自主研发了长程记忆存取、医学知识库查询、医学计算器等实用性插件功能，使得“大医”能够精确回答医疗健康领域专业问题。

“大医”聚焦智慧大健康、智慧患者服务、智慧临床以及数智建设四大应用领域，已覆盖智能自诊、体检咨询、健康问答、导诊、预问诊、用药咨询、诊后随访管理、智慧病历、诊室听译机器人、智慧医助、智慧随访、影像报告结构化及病历结构化共 13 个细分医疗健康场景，实现模型功能与具体场景的精准匹配，推动医疗健康全产业链数智化转型。



图 12 商量大医模型工作原理和应用场景

在行业赋能方面，“大医”支持私有化部署、API 接口调用两种合作模式，针对医疗健康大模型落地时面临的个性化需求，“大医”提供“医疗模型 DIY”功能及定制化服务，包括一键调整提示工程，以及构建专属知识库插件等模式，为机构打造紧密贴合其需求的专属医疗健康大模型，为行业提供了更多元的解决方案，推动了医疗健康大模型与实际需求场景的紧密融合。

为了满足差异化部署需求，商汤推出多个“大医”版本，参数量从千亿到百亿级别不等，既可以公有云服务方式为合作机构提供服务，也可帮助合作机构高效完成模型私有化部署。借助创新的模型量化技术，可降低“大医”落地部署的硬件需求，降低了医疗健康大模型的部署门槛。

目前“大医”已面向医疗健康产业链上下游机构客户开放服务，并将进一步探索与营养保健、健康管理等领域企业、机构的合作，为产业链高质量发展赋能。此外，商汤科技还与行业伙伴合作，推出了医疗影像大模型、生信大模型等多种垂类基础模型群，覆盖 CT、MRI、超声、内镜、病理、医学文本、生信数据等不同医疗数据模态。

药械医健类企业具有行业数据优势，以调用接口或基于开源模型自研方式切入。制药企业、医学影像设备企业、CXO 企业（生物医药外包服务企业）、互联网医疗企业、医疗 AI 创业企业等通常积累了丰富的生物医学行业数据、用户资源和细分领域专业能力，一是可以通过直接调用先进大模型 API 接口或基于相关大模型微调改进自身产

品,快速打造出具有领域特色和竞争力的智能化产品及服务,如 Ferma AI、Nuance、Doximity、Bionic Health、Wondercise 集成 ChatGPT/GPT-4 大模型,提供更加便捷、智能的生物技术及制药咨询、自动化病历生成、医学文档处理、个人健康管理功能;二是可以基于开源大模型和自有行业数据开发研制医疗大模型产品,利用开源模型技术能力和研发资源,充分发挥自有数据价值,不断迭代提升模型效率和水平,但该模式对软/硬件算力基础设施要求较高。目前在虚拟健康助手、注册审评咨询、互联网问诊、医保商保、医药信息情报领域都涌现一批典型产品。

医疗大模型商业化落地尚处早期探索阶段,部分科技巨头初试商业变现路径。当前,国内医疗大模型产品大多处于研发内测或定向体验阶段,国外部分医疗健康 AI 应用通过集成 ChatGPT/GPT-4 大模型来增强对外服务能力。整体来看,由于技术成熟度和医疗健康领域特殊性,大模型产品实际部署使用范围较小,未来有较大拓展空间。商业模式上,医疗基础模型接口调用收费、大模型定制开发收费、数据标注和处理服务收费、模型训练加速和优化服务收费等模式或将兴起。国外 GPT-4 大模型通常按调用接口的服务量收费;国内百度灵医大模型已面向医疗机构、药械企业、药房等用户群体提供不同服务,尝试多元变现路径。除模型调用和模型开发外,目前已出现一些工具型、平台型方案,通过提供医学数据库、医疗专用智能计算平台、系列基础模型组合平台、模型训练降本提效技术服务而盈利。未来,随着医

疗大模型技术的创新发展、应用的加速落地，相关商业模式将更加清晰和多元化。

案例：百度灵医大模型

灵医大模型是百度集团研发推出的“产业级”医疗大模型，是基于百度文心大模型底座和百度智能云千帆算力资源，经过算法训练和精调，结合灵医智慧临床脱敏数据和医学知识、百度健康线上优质就医问诊数据和 GBI 医药数据等，打造的知识增强大型语言模型。

灵医大模型是以自研国产、知识增强的文心大模型为底座，通过海量医疗健康优质语料进行预训练和指令精调，以及循证医学对齐后获得的医疗行业大模型。**算法方面**，灵医大模型基于文心大模型，打造千亿参数的大模型，并内置了知识增强、检索增强和上下文增强等多项增强技术，提升了大模型生成的准确性和多样性。**算力方面**，灵医大模型基于万卡算力集群和全生命周期的模型开发工具链的保障，进行无标注的预训练，不断优化底层参数。**数据方面**，灵医大模型使用了千亿 Token 的训练语料数据，包括海量临床脱敏数据、海量医学知识图谱、300 万+例多模态影像数据，6 亿+条健康科普内容，70 万+临床试验研究信息，丰富的语料数据提高了模型的全产业链服务能力。



图 13 灵医大模型输入“医-患-药”医疗健康数据和医学知识

在服务模式上，灵医大模型可划分为四层架构，按需为不同用户提供服务。

应用层：主要为患者、医院、企业等用户提供 AI 原生应用，目前以灵医 BOT 为助手，聚焦在智能健康管家、智能医生助手、智能企业服务三大方向，满足“医-患-药”各自的特定需要。智能健康管家提供智能导诊、预问诊、健康咨询等能力，为患者提供就医咨询引导。智能医生助手从辅助诊断、病历生

成、文献速览等方面为医生提供服务，节省医生时间，提高工作效率。智能企业服务从运营助手、职业培训和知识服务等方面为企业客户提供能力，创新营销方式，帮助企业快速完成新药上市后的推广。

能力层：主要以 API 或 AI 插件的方式，为生态合作伙伴提供高质量的 AI 服务，帮助合作伙伴进行二次开发，打造 AI 原生应用。

模型层：主要根据不同的应用需要和部署资源，提供旗舰版、Lite 版和定制版服务。旗舰版提供公有云服务，用户无需担心部署成本。Lite 版面向医院客户或对私有数据较为重视的客户提供模型服务，以私有化方式部署，分档设置十亿和百亿参数量级的模型。定制版针对自有高质量数据且具有一定研发能力的客户，需针对具体场景，如专科专病，提供定制化模型训练或调优服务。

算力层：重点面向有私有化部署需求的用户，提供三个等级的软硬一体的算力支持能力。一体机版，将模型和算力进行了一体化封装，主要针对对算力要求高且预算充足的客户，可在内部私有化环境中直接部署使用，也可基于内部数据进行训练和微调，更好地满足业务场景需要。信创版，针对国产化需要的客户，基于百度自研的昆仑芯芯片，提供全栈国产的算力支持。CPU 版，针对算力和预算紧张的客户，提供开箱即用的能力，无需训练和微调模型，有限算力下，直接使用大模型推理能力，满足特定任务的应用需要。



图 14 灵医大模型的商业化服务模式

(二) 具体场景分析

1. AI 大模型在生命科学研究领域的应用

生命科学领域的 AI 大模型发展较为成熟，起步早、模型数量较多，迭代发展快。生命科学研究生物大分子（如 DNA、RNA、蛋白质）在生命过程中的作用，描述分子的结构、功能和合成等各种生命活动和现象，研究成本高、实验周期长、生产难度高。传统的信息处理和分析方法往往受限于庞大的数据量和高度的复杂性，生物数据如基因组序列、蛋白质结构、临床数据和生物图像等需要强大的计算能力和数据分析工具。由于参数量大且学习能力强，大模型在处理多维复杂生物数据上具有独特优势，极大助力蛋白质语言表征、生命组学计算，并可通过科学文本检索和提取来辅助科研工作。

AI 大模型可完成蛋白质语言理解和生成任务，助力蛋白质结构预测和从头设计合成。大模型可以预测蛋白质的结构、功能以及最优催化温度、催化效率、稳定性等属性，完成蛋白质理解任务；也可根据不同条件设计对应的蛋白质，完成蛋白质生成任务，如根据给定的功能标签生成能实现该功能的蛋白质，或根据给定的蛋白质结构，设计一段可折叠成该结构的氨基酸序列。基于 Transformer 架构的 AlphaFold 2、ESMFold 及其衍生模型，可以快速而准确地预测蛋白质结构。蛋白质语言模型 ProGen 能够生成跨大型蛋白质家族、具有可预测功能的蛋白质序列³¹，其在包含 19000 个家族的 2.8 亿个蛋白质序列的公开数据集上进行训练，可用于从头设计蛋白质。

AI 大模型赋能 DNA/RNA 等生命组学计算，为病因推断、疾病预测、精准医疗提供新思路。生命多组学研究通常包括在脱氧核糖核

酸 (DNA) 复制、转录、翻译、翻译后修饰的过程中, 产生的全部基因(基因组学)、基因表达的广泛变化(表观遗传组学)、核糖核酸(RNA, 转录组学)和蛋白质(蛋白质组学), 以及下游的小分子代谢产物(代谢组学)。大模型以生物医学大数据作为输入, 不断进行算法训练、迭代, 输出模拟真实生命系统的结构与功能特征, 有助于从分子、细胞微观层面捕捉、理解疾病发生机制, 推动精准医疗发展。生物计算工具 Geneformer 在约 3000 万个单细胞转录组的大规模语料库上进行预训练, 可通过迁移学习应用于多种下游任务, 以加速发现生物网络关键调节因子和候选治疗靶点³²。单细胞生物学大型语言模型 scGPT 在单细胞生物学下游任务上表现优异, 包括 multi-batch 整合、多组学整合、细胞类型注释、遗传扰动预测和基因网络推断等, 可以有效地捕获有关基因和细胞的关键生物学见解³³。

AI 大模型作为科研助手辅助生物医学研究开发工作, 优化科研流程。大模型具有生物医学信息检索、医学文本摘要、信息提取、科学知识推理等能力, 可以帮助生物科学家、临床科研人员、医药产品研发人员等快速读取生物医学论文、生成研究综述、总结临床研究报告 (CSR)、整理药物数据、生成代码, 提升科研工作效率。Meta AI 发布的面向科学领域的大型语言模型 Galactica, 在人类科学知识的大型语料库上进行训练, 可以总结学术文献, 解决数学问题、生成百科文章、编写科学代码、注释分子和蛋白质³⁴。基于大模型的 AI 临床科研大数据平台可以实现患者临床数据、心电图、心电时序、发病时水

文气象环境等多模态数据融合处理、解析与理解，完成高维度深度学习建模，助力临床科研数据处理，减小人工筛选数据的工作难度与耗时。

案例：EyeGPT

EyeGPT 由温州眼视光国际创新中心（中国眼谷）开发，是眼科专用大型语言模型，目前主要应用于科研场景以及临床医疗辅助等医学相关内容，以交互式问答形式提供服务，未来将进一步向眼健康早期检测和智能诊断场景拓展。

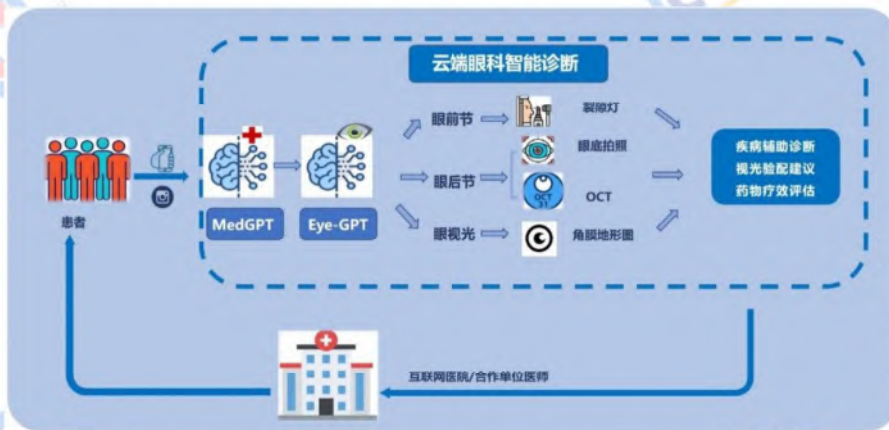
EyeGPT 基于大量自然语料以及医学专业语料进行预训练，并针对医学场景使用专业数据进行微调而得到。训练内容不仅涵盖了通用医学知识，还针对眼科专科电子病历信息，如视网膜疾病、近视、青光眼等眼科疾病相关信息进行了训练。该眼科专用大型语言模型将为眼科知识图谱建立、分诊问诊系统开发、用药治疗建议提供重要工具。

EyeGPT 已在研究性论文、病历报告生成、医学学术文章润色、封面信件及审稿意见辅助撰写、医疗文档写作、医学相关知识问答等方面有较为成熟的落地应用。EyeGPT 可以根据某研究题目生成一篇关于该研究方法的论文大纲，可以根据需要信息创建一篇病例报告，可以根据输入的需要润色的文章部分进行进一步阐述，可以辅助进行封面信件及审稿意见的撰写，可以辅助进行 SOAP 病历、交班记录、实习生急诊交班记录、临床诊疗计划等医疗文件写作以及制作 PPT 文件，也可以辅助医学生进行医学相关考试的复习，可以进行医学相关知识的讲解。





EyeGPT 通过整合自然语言处理和计算机视觉，为医学专业人士和患者提供了智能辅助工具，用于更好地理解、分析和应用医学及眼科专业知识，在眼科医学领域的诊断、治疗、研究和教育方面具有巨大的潜力。



2.AI 大模型在药械研发领域的应用

AI 大模型技术可服务于药品和器械从研发到上市的各个环节，包括药物发现、临床前研究、临床试验、注册申请、上市后再评价等，实现提速降本增效。同时，针对药械企业的医药信息情报和行业知识问答大模型也不断出现，进一步赋能药械产品研发创新。尤其在制药领域，新药研发面临成本高、周期长、成功率低的痛点难题，而大模

型具备更高的预测能力，实现药物设计、筛选、优化、验证关键环节效率和效果的双重提升，节省药物研发时间，降低研发费用。

从分子性质预测到靶点发现, AI大模型助力药物研发早期阶段。

在分子性质预测方面，通过使用 SMILES 字符串或分子图表征分子结构来预测分子性质，如 GROVER 通过运用分子图来表示分子形式，在 100 多万未标记的分子数据上训练了超 1 亿个参数，能够获取丰富的分子结构和语义信息。在分子生成方面，通过药物分子生成模型，可以在短时间内生成大量的分子供药物学家进行筛选，加快筛选到有效的分子。如 MolGPT 模型基于 Transformer 解码器模块构建，能够生成具有特定分子性质的分子，或者含有用户指定骨架的分子。在靶点发现方面，大模型通过来自细胞、动物模型、病人身体组织的数据进行 AI 建模，助力生物标记物发现和靶点发现。此外，大模型在药物-靶标相互作用 (DTI) 预测、药物筛选、先导化合物优化等方面也具有较大潜力。目前，业内开始开发具有药物发现全过程优化能力的大模型及相关平台。

AI 大模型可提高临床试验执行效率，节省时间和成本。在临床试验方案设计上，大模型可以检索分析大量文献，快速总结近年来药物临床试验设计和结果，辅助自动撰写临床研究文档，帮助研究者节省时间，还可以根据数据及时预测试验风险，生成优化建议。在临床试验患者招募上，大模型可分析临床数据，识别潜在的临床试验候选者，提高患者-试验匹配效率，提升临床试验入组率和成功率。如大型语言

模型 CTmatch、TrialGPT 通过分析患者的病史和治疗计划，可匹配与患者病情相关的临床试验。在**临床试验质量控制**上，大模型可对大量病例进行快速分析，发现并校验不良事件，根据监测数据及时预测试验风险，生成优化建议，辅助临床试验质量控制，提高临床试验的效率和准确性。例如，跨国生物制药公司阿斯利康在临床试验的设计、患者招募、跟踪进展和分析数据方面开始使用 GPT-3 大模型。

案例：通义医疗行业大模型

通义医疗行业大模型由阿里云研发，是以通义生成式语言大模型为基底，在通义 Qwen 整体训练数据超过 3 万亿 token 基础上，融入海量医学知识文献与医疗数据所训练出来的行业大模型；并针对医药行业知识密集以及严肃医学的特性，使模型在医疗领域具备更强大的行业知识的推理、认知等能力，包括知识问答准确性、单轮问诊权威性、多轮问诊场景化、领域制式文本生成以及领域文献理解等。除了在 MMLU、C-Eval、GSM8K、MATH、GaoKao-Bench 等 12 个权威评测中取得优秀的的成绩外，在医学数据集评测中也具有优异表现。

通义行业大模型通过 API 与交互式问答形式提供服务，并提供用于模型二次训练与评测的完整操控平台，已经开始在行业应用层面探索落地。

在**医学智能问询**方面，针对药物说明书、文献等严肃医学场景进行智能问询，常见问题生成。在**科研文献理解**方面，针对医学文献进行特定格式的摘要总结，例如通过临床不良反应的论文与学术文章解读，生成用于审批与上报的内容，提升药械企业办公效率。在**多指标诊断**方面，通义医疗行业大模型可以提升体检行业主检医生的总结效率，实现检测报告的自动识别与解读、多指标组合等能力，从而在专业预防、精准筛查、慢病管理等领域为客户提供更高品质的健康管理服务。

同时，通义医疗行业大模型服务平台还提供用于**安全防范**的安全卫士能力，根据生成式人工智能技术的健康发展规范，对生成内容进行了安全管理，回答内容积极向上，并提供内容安全应急管控机制。

AI 大模型助力药械注册和审评自动化，提升药械企业和监管机构办公效率。药品和医疗器械领域的标准规范和监管体系文件通常十分繁多、复杂，且需要高度准确性，涉及大量的专业术语、法规要求以及详细的操作指南。**在申报端**，以大模型为驱动力的类 ChatGPT 应用，可以提供药品和医疗器械法律法规解读、预审咨询，产品分类与同类产品对照，注册申请材料自动生成和完整性评估等功能，为药械研发企业提供合规支持，赋能研发、临床试验、注册、准入等环节，加速产品上市和产业创新发展。例如，中国信通院联合相关单位研发 BME Docs AI 医疗器械标准与体系文件智能服务平台，提供医疗器械标准解释、文件生成、风险评估等服务。**在审评端**，大模型能够提供审评问答咨询，辅助药品和医疗器械安全性和有效性审查及决策，以数字化手段提升监管效能。

AI 大模型与生物医药行业数据相结合，驱动行业信息咨询服务新发展。大模型具有强大的自然语言理解与生成能力，交互方式更加灵活，在生命科学与医疗垂直领域数据上训练后，可深入挖掘行业信息，提供更准确、更及时、更有价值的行业知识和情报咨询服务。例如，一些医药垂类对话机器人可依托药品数据库、医疗器械数据库、研究人员数据库等专业数据，结合大模型技术提高数据获取、治理、精准搜索与匹配能力，为企业提供定制化分析服务；一些应用基于综合临床试验数据库、生命科学会议摘要、生物医学期刊出版物、生命科学新闻等数据，为制药或生物技术公司的生命科学专业人士提供生

命科学问题咨询服务；还有一些应用基于临床研究数据、生物专利数据进行训练，在应用工具、临床知识理解、行业信息收集、专利情报信息以及数据决策分析等领域提供帮助。

3.AI 大模型在医学影像和图像领域的应用

医学图像具有独特成像方式和特点，基于 SAM 的医学图像分割模型成热点方向。医学中的成像方式与自然图像有很大不同，主要基于物理特性和能量来源（如光、电子、激光、X 射线、超声波、核物理和磁共振）。医学图像分割使图像中解剖或病理结构的变化更加清晰，对提高诊断的效率和准确性至关重要。由于多样的成像模式、精细的解剖结构、不明确且复杂的边界以及广泛的物体尺度等，医学图像分割面临较大的挑战。图像分割通用基础模型 SAM（Segment Anything Model）由 Meta AI 发布，具有用于各种图像分割任务的潜力，但在医学影像分割任务中的表现目前还不稳定。基于 SAM 的改进模型不断出现，如 SAM-LST、MedSAM、SAM-Med2D、SAMUS 医学图像分割大模型将 SAM 从自然领域适应到医学领域，更适合临床应用。

案例：面向多模态、多器官的影像通用分割大模型

面向多模态、多器官的影像通用分割大模型由北京邮电大学网络与交换技术国家重点实验室研发，是一种基于通用 SAM 的医疗大模型，同时支持多模态、多任务的医学影像分割，解决“单一模型只能为单一任务、单一模态服务”的问题。

研究团队为 SAM 构建了一个面向医疗场景的大模型适配器。适配器包含

先验知识池与多适配器专家两个模块，使模型在训练过程中显式地得到影像的任务和模态信息并且适配于任一模态、任一任务的医学分割场景。在训练时，只需要微调大模型适配器的参数，就可以将医学领域知识整合到分割模型中，使 SAM 适配于医学的多任务和多模态场景。

研究团队构建了一个包含 40 万张医学影像和相应的分割标签大型医学数据集 SAM-MEDData，对模型进行微调和验证，统计各个模态、任务上的实验性能，并计算平均指标。结果显示通用医疗大模型适配器在多模态、多任务的医学分割应用上取得了较好的成绩，从而解决了“单一模型只能为单一任务、单一模态医学影像服务”的问题，这也为多模态、多任务医学分割场景提供了一种新的学习范式。

AI大模型辅助医学影像诊断分析，并可自动生成影像诊断报告。

大模型基于医学影像的图文对数据集进行训练，可以将视觉理解与文本知识相结合，以对话的方式解释胸部 X 射线等医学图像，回答相关问题，并生成影像诊断报告。典型的面向医学影像问答的多模态医疗微调模型有 Visual Med-Alpaca、XrayGLM。胸部 X-ray 基础模型 KAD 在大规模医学影像和放射报告数据上进行预训练，不依赖于人工注释，在下游胸部 X-ray 诊断任务上，无需任何监督微调，即展现出较高的精度。病理语言-图像预训练模型 PLIP 学习了社交媒体上超过 20 万张病理图像，可以根据图像或文本检索相似的内容，使用大模型理解视觉图像（病理图像）和文本描述（自然语言），帮助人类病理学家寻找类似的病例作为参考。

案例：面向多场景的智能医疗多模态大模型

面向多场景的智能医疗多模态大模型由北京大学多媒体信息处理全国重点实验室研发，是一种新颖的医疗影像报告生成多模态大模型，关注视觉问答

和报告生成。

该模型同时支持多种医学影像模态和器官作为输入，生成临床应用级别的高质量影像报告。该模型也具有医疗视觉问答和对话功能，能够为患者提供可靠的治疗方案建议和日常健康贴士。除此之外，研究团队还设计了一种基于 GPT4 的影像报告评估指标，使该模型能够从多个维度全面比较生成报告和参考报告，以填补报告质量评估领域空白。

该模型在视觉语言多模态大模型 LLaMA-Adapter-V2 的基础上，设计了以下几种新颖模块：

1. 基于 SAM 增强的医学视觉编码器预训练：利用 SAM 分割病灶作为补充输入，提升视觉编码器的病理信息提取能力。

2. 基于混合模态专家适配器的大型语言模型（LLM）微调策略：针对不同医学模态（X-ray、CT 和 MRI）设计模态特异和模态共享的适配器，解决不同模态诊断疾病不同，涉及知识库差异较大的问题；

3. 基于相似报告检索增强的训练策略：模仿医生写报告会参考相似案例的特点，从图像和文本两方面进行相似度检索，找到类似报告作为模型输入的上下文信息，减少 LLM 幻觉。

模型在 MIMIC-CXR 公开数据集上进行了医学影像报告生成任务的测试，相较于同时期医疗多模态大模型有较为明显的优势，报告生成的相关性和流畅性更高。未来，该模型可面向医生落地临床影像报告生成、病灶辅助定位、治疗方案参考等场景应用，面向患者落地 AI 私人医生、病例管理、用药建议、健康预警、医疗知识问答等场景应用。

4. AI 大模型在医疗问答和智能问诊领域的应用

AI 大模型以对话方式回答用户的医疗健康问题，支持连续自由对话和多场景功能。针对用户日常生活中的医疗健康问题，AI 大模型经过丰富的医学数据和知识的训练，具备通过多个轮次对话了解用户疑问和想法，并给出专业、详实回答的能力，支持与用户连续多轮对

话、病情初步预测、建议挂号科室、推荐医院和医生、提供健康科普和给出用药建议等多场景功能。例如，医疗 AI 搜索引擎 MediSearch 采用“大模型+医疗文献”方案，回答医疗健康或生物科学问题，其生成的每条答案都会给出信息来源，不仅兼顾了医疗领域问答的科学性、准确性，同时还能用对话方式提供超越普通搜索引擎的用户体验。还有一些大模型产品在更多样化的医疗数据上进行训练，能够掌握医疗专业术语知识，结合用户的自我症状描述、健康记录、过往检查和诊疗数据等，灵活地对病症进行解释和建议，提供轻问诊服务。

AI 大模型提升问诊类产品的准确性和智能化程度，对话式“数字医生”发展提速。传统在线或远程问诊类产品积累了大量在线问诊数据和经验，包括真实医患互动对话、检验检测和病例信息等，引入大模型进行数据深度挖掘和分析，再结合真人医生反馈监督微调训练等方式，能有效提升模型的疾病特征判断与模式识别能力，提高问诊准确性。一些企业将 ChatGPT 等大模型技术整合到其问诊类产品中，允许用户用自己的话描述自身症状，包括非医学术语和口语表达，在与用户完成互动问答后，还会生成病情摘要和可能的疾病诊断建议，供医生后续诊疗参考，节省了医生的沟通成本，将医生从之前的重复性流程中解放出来，使其能更专注于解决患者健康问题。目前已出现多款基于大模型的对话式数字医生、智能全科医生、智能问诊类产品，但此类产品生成的内容需人工监督或审核，目前所有产品均不能完全代替人类医生的判断和建议。

案例：腾讯医疗大模型

腾讯医疗大模型由腾讯医疗健康（深圳）有限公司以全链路自研混元大模型为基座继续训练而来。

混元大模型拥有超千亿参数规模和超 2 万亿 tokens 预训练语料，具备强大的中文创作能力、复杂语境下的逻辑推理能力以及可靠的任务执行能力。在此基础上，持续加入涵盖 285 万医学实体、1250 万医学关系，覆盖 98% 医学知识的医学知识图谱和医学文献，使大模型进一步掌握专业的医学知识。经过 3000 万覆盖患者、医生、药厂等场景及医疗流程的问答对话进行多任务微调，以及 36 万组专家医生标注数据的强化学习，使得腾讯医疗大模型在处理医疗需求时更专业、精准，同时也兼顾了患者关怀，更接近人类医生。

目前，腾讯医疗大模型包括文案生成、智能问答、病历结构化和检索、影像报告和辅助诊断等场景大模型，可嵌入医疗环节全流程，在科室导诊、医生推荐、预问诊、医患对话、病历自动生成和智能院务客服等应用场景中实现医疗服务水平和质量的全面提升。

依托腾讯医疗大模型，包括医学智能问答、数字人就医助手、智能自由问诊、病历自动生成、AI 合理用药、智能化随访管理、患者全流程管理在内的一系列 AI 应用进行了全面升级，为医疗决策全流程提供更精准的辅助，助力患者就医体验以及临床医生、药剂师服务效率和质量的双向提升。诊前阶段的应用除了医学智能问答，还升级了数字人就医助手，能提供 7*24h 的智能客服及专业科普服务；诊中环节，基于大模型的能力学习百万级医患对话及 3000 多疾病推演解析，升级了模拟自由对话真实度更强的辅助问诊、更符合书写规范的病历自动生成以及能给出可解释性依据的辅助诊疗等应用；诊后阶段采用人工智能+互联网医疗的新型模式升级了智能化随访管理和患者全流程管理等应用。此外，基于腾讯医疗大模型的家庭医生助手提升基层服务能力，家庭医生助手利用微信、企业微信打通医患的沟通渠道，将为医生提供更加智能的签约、咨询、随访、宣教能力，也为居民提供精细化的健康管理。

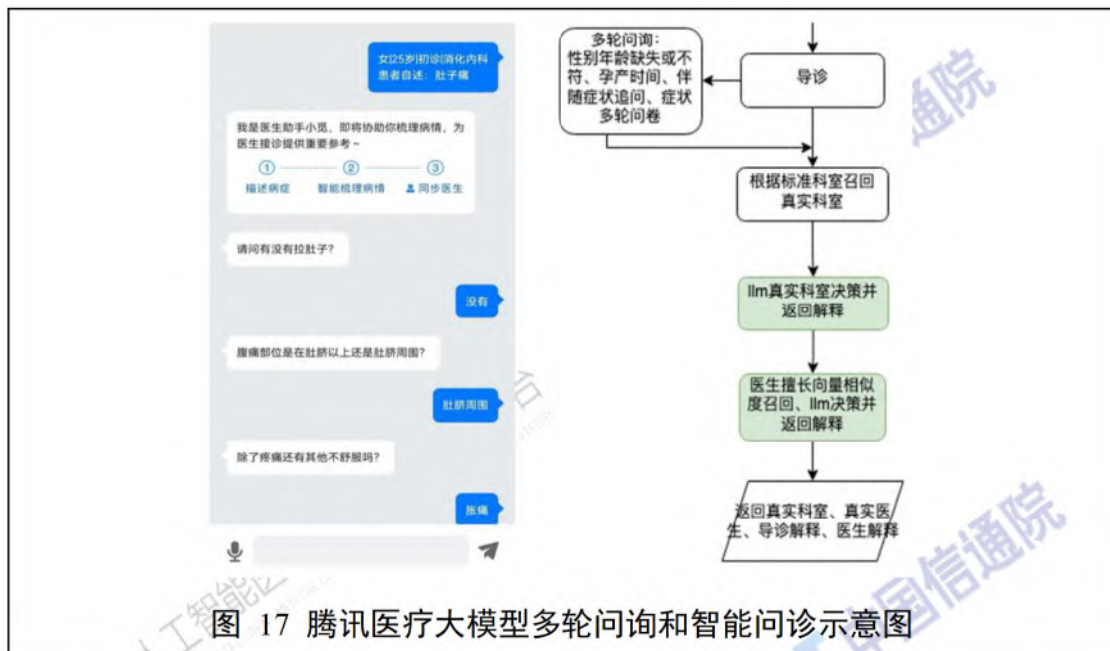


图 17 腾讯医疗大模型多轮问询和智能问诊示意图

5.AI 大模型在辅助诊疗和临床决策领域的应用

AI 大模型可预测疾病风险、生成诊断和治疗建议，为临床决策提供支持。AI 大模型可以处理和分析大量的病理学、临床记录和基因组学数据，从海量的医学文献和临床病例数据库中学习知识，并结合患者的个人信息和检测检验结果进行综合分析，提供解释结果、生成诊断和治疗建议、评估治疗选择、预测疾病风险和紧急预警等方面的帮助，为医生和研究人员提供决策支持。医疗大模型具有强大的推理能力，可以对患者未来的疾病风险作出合理预测，例如 MedGPT 可以预测患者可能的疾病，Med-BERT 能够预测糖尿病患者未来出现心力衰竭的可能性，NYUTron 可以预测患者的再入院风险。大模型也可以根据病历信息和临床数据，自动化生成医学解释、临床诊断建议和治疗方案，从而辅助医生进行临床决策，提升现有的临床决策支持系统（CDSS）的智能化水平。

案例：固生堂中医大模型和 AI 智能装备

固生堂中医大模型由固生堂中医连锁管理集团联合百度灵医智惠共同研发，提供中医智能诊疗、中医方剂推荐等智能化中医服务，改善患者的就医体验。

固生堂引入百度灵医大模型能力，开发 AI 智能装备，通过将舌诊仪、面诊仪转化为二类医疗器械，在诊疗过程当中进行数据采集，从而为中医 AI 智能化体系建设奠定数据基础。在患者端，固生堂中医大模型将通过 AI 大数据搭建生成式对话大模型，并以固生堂中医 App 为载体，为患者提供病情咨询、快速找医生、智能导科室、用药指导、智能客服等多种服务。在医生端，固生堂中医大模型集辅助问诊、辅助诊断、智能开方、文献搜索、中医病案信息化管理等功能于一体，为医生提供临床诊疗和学术研究数字化赋能。临床诊疗上，固生堂计划推出辅助诊疗设备，将传统中医诊疗服务中的“望闻问切”与人工智能、大数据等技术相结合，推动中医诊断标准化，提升看诊效率。

AI 大模型有望改善机器人视觉和交互能力，赋能术中导航和病灶识别判断。虽然大模型在医疗机器人中的应用刚刚起步，但它将对改善医疗机器人的视觉、交互和自主性产生重大影响。通过在手术机器人中结合 AI 大模型，可以增强机器人的病灶分割能力和 3D 视野中的导航能力，融合视觉、触觉多模态信息，使得机器人在脑手术中对病灶有更精准的认识和判断能力。AI 大模型还能赋能康复和陪伴类机器人，加强机器人对人类意图、手势、语音和情绪的理解，提升康复群体和老年人护理服务质量。大模型的适应性和泛化能力已经用于提高通用机器人的自主性，未来可能会推动医疗机器人技术向更高自主性的方向发展。此外，大模型还可改善机器人的手术流程分析能力，更准确、客观地预测并发症或成功的可能性，从而帮助外科医生更好地规划手术。大模型还可生成和模拟外科手术，使外科医生提前练习和完善技能。

6.AI 大模型在医疗记录和行政管理领域的应用

AI 大模型助力临床文档生成和医疗文本结构化，提升医学信息处理智能化水平。大模型能够分析医患实时对话、获取对话关键信息，自动生成入院记录、病历摘要和报告、手术记录、出院小结和病后随访计划等，减轻医生的文书工作负担，优化患者就诊体验。目前大模型已实现结合语音、图像、视频、文字等多种模态形式信息，自动转录、分析并生成就诊摘要、病情描述、鉴别诊断、治疗方案等高质量内容，供医护人员审核和编辑。此外，大模型能够助力医疗文本结构化，提高医疗数据的可读性和可理解性。例如，Truveta 大型语言模型 (TLM) 将电子健康记录中含有拼写错误和不同的术语或缩略语的非结构化数据转换为标准化数据点，进一步用于药物、医疗设备和疾病研究。Augmedix Prep 采用非结构化输入来生成结构化输出，根据就诊类型和患者以前的医疗记录，在就诊前为医生准备好规范化、清晰明了的患者健康记录内容，减轻人工负担，方便医生查看患者信息，直接进入看诊环节。

AI 大模型优化医疗设备和系统运维管理，提升组织机构内部行政事务处理效率。嵌入 ChatGPT 等大模型的生命科学电子质量管理体系，能够优化药械企业的质量管理流程、自动化相关任务，提高合规性、文档和报告方面的效率，并实现更高质量的结果，增强机构合规性。大模型在医疗设备管理环节可以为运维人员提供准确的故障诊断、智能化运维方案和实时数据监测，强大的运算能力能够精准预测设备故障，提前进行维护和修复，减少意外停机时间，提高设备运行

效率。用大模型来构建患者院内患者管理系统和财务系统等，能够快速处理各种入院手续等流程，有效地为更多的患者群体提供服务。大模型还可以帮助医生处理行政流程文档，减轻医生行政负担，提高行政流程效率。

7.AI 大模型在个人健康管理领域的应用

个人健康管理领域的大模型分布广泛，模型种类繁多。此类模型多针对于个人用户，旨在帮助个人在非医院场景中解决健康问题。可穿戴设备、营养与运动、心理健康、中医保健、生活指导等相关模型和应用不断涌现，并渗透到预防、诊后、居家、室外、养老院等多场景中，实现预防、咨询、预约、康复的全周期智能化健康管理，是未来推动个人健康发展的重要力量。

AI 大模型推动个人健康管理迈向主动化、个性化、智能化。基于大模型的应用产品通过分析个人病史、基因构成、生活方式和其他因素，可为患者生成个性化的健康管理计划，助力居家场景下的个人健康管理更加精准、高效、智能。大模型可以将专业的诊后管理和康复指导延伸到院外，为出院后的患者提供病情随访、用药指导、慢病管理、健康知识科普以及患者咨询等个性化、智能化服务。Google Health 使用大型语言模型来帮助患者进行健康管理，通过跟踪用户的病史、药物服用情况和症状，生成报告以帮助患者更好地了解自己的健康状况。个人基因数据分析应用 Livewello 接入 ChatGPT 大模型，能够让用户使用自然语言查询基因数据，询问有关自身健康状况、药物作用、

疾病遗传风险等的问题。

案例：神经源性膀胱慢病管理数字疗法平台

神经源性膀胱慢病管理数字疗法平台由江苏省中医院（南京中医药大学附属医院）联合曼博迪健康(苏州)有限公司研发，通过和病人的互动，动态生成对应的智能体，实现从监测、管理到干预的数字化智能膀胱慢性病管理。

该平台引入了类似于 ChatGPT 的 Transformer 模型，采用人类反馈强化学习微调（RLHF）方式，根据医生、病人以及其他使用了大模型的用户们的反馈，进行进一步优化。该平台能够针对膀胱问题进行智能解答，实现基于膀胱慢性病管理医学指南的人机对话以及个性化的指导。该平台通过将一系列参数输入应用程序和智能设备，实现对神经源性膀胱患者数据的连续监测、异常值预警、膀胱容量和残余尿量的监测，对下尿路功能障碍患者提供无创、客观的下尿路症状测量方法。该平台能够实现基于循证的线上人机互动交流，为临床慢病管理提供长期随访评估数据，构建从监测、管理到干预的数字化智能膀胱慢性病管理。

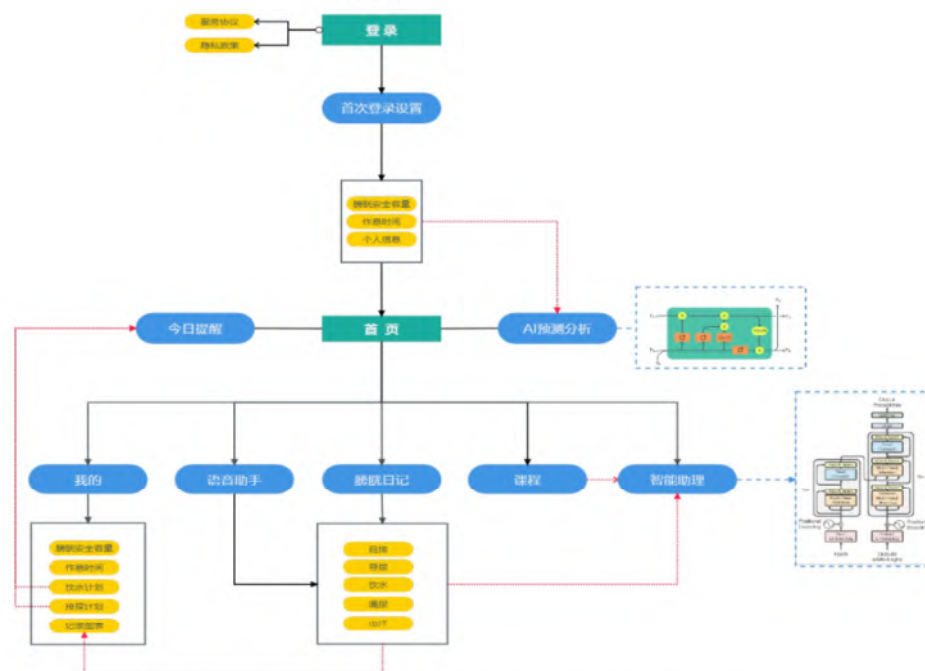


图 18 神经源性膀胱慢病管理数字疗法平台示意图

AI 大模型提供营养、运动辅导服务，多方面支持个人智能健康管理

理。大模型与可穿戴设备融合应用，可深入挖掘健康监测数据的价值，提供个性化的运动指导方案，实现有效的自我健康管理。一些整合了 ChatGPT 的智能手表手环、减重 APP，可以对运动数据、身体指数进行分析处理，从饮食、作息以及运动方式等多方面生成专业建议，逐步改善用户的健康状况，帮助用户养成健康习惯，提升用户参与度与满意度。一些营养健康 AI 大模型在多个营养健康权威机构的高质量素材进行训练，已通过国内外注册营养师、公共营养师、运动营养师等多个营养健康专业认证考试，可提供健康评估、营养计划制定、个性化营养建议、营养配餐、智能提醒、过程辅导和激励、目标和计划动态调整等一系列营养保健服务。

案例：运动健康助手

运动健康助手由 OPPO 广东移动通信有限公司研发，旨在为用户提供健身指导和健康支持。该健身助手基于大型语言模型，能够提供个性化健身计划、健康数据分析和健康社交互动等功能，帮助用户实现健康生活方式的转变。

研究团队依托健康知识图谱训练大型语言模型 andesGPT，再结合健康大数据感知与分析技术，进一步增强通用大型语言模型的数据分析能力，打造运动健康领域数据专属解决方案。

该运动健康助手基于大模型卓越的语言运算处理能力与强大的信息生成能力，可为用户的健康运动问题提供多样化且高度准确的回答；可调用智能健康终端（手机、手环、手表等）接口，结合用户个人健康数据和同步接受系统推送的天气信息、用户偏好场所，为用户提供定制化的运动方案，并及时进行反馈和调整；可生成健康报告，以图表或图形的形式呈现用户的健康数据分析结果，如身体指标、运动成效等；可以利用大模型强大的推理能力与总结能力，检索知识条目，搭建知识网络，帮助用户理解健康报告出现的专业术语；还具

备快照功能，利用大型语言模型的摘要能力，结合用户的多模态数据信息（运动时长、项目、消耗能力），就用户当日的运动表现提炼专属的关键词，更形象地展示用户的运动表现和运动趋势；还具备社交功能，提供用户之间的互动和交流，打造以健康运动为核心的社群。



图 19 运动健康助手应用架构和界面

AI 大模型提供更拟人化的情感疏导和支持，同时带来心理健康评估、疗愈新工具。大模型可以识别用户在对话中流露出的情绪和心理状态，诸如焦虑、抑郁或其他不良情绪，能够以共情方式倾听用户烦恼、担忧和困惑，用理解和温暖的话语回应，缓解用户情绪压力，同时还能提供工作压力、人际关系、个人成长等生活建议和指导，帮助用户更好地应对压力和挑战。大模型可以根据患者的言语行为，辅助进行心理健康评估，识别患者是否患有焦虑、抑郁、创伤后应激障碍等心理健康问题，并提供治疗支持和指导性建议。例如，心理大模型 MindChat 在多轮心理对话数据进行训练，涵盖工作、家庭、学习、

生活、社交、安全等多个方面，帮助人们纾解心理压力与解决心理困惑，提高心理健康水平。中文儿童情感陪伴大模型 QiaoBan 面向中小學生及家長群体，具有陪伴、益智和教育功能。

AI 大模型赋能传统医学，生成中医药处方或多维度中医养生方案。大模型在大规模中医知识图谱数据、中医古籍和文献数据、中医专家医案数据，中医临床诊疗数据和脉象、舌象、经络、穴位数据上进行训练，可根据用户症状、体征信息，给出辨证结果、中药处方或个性化中医健康状态辨识结果，以及食疗、茶饮、推拿、艾灸等多维度养生调理方案。目前已有数字中医大模型产品面向中医领域名医经验挖掘整理需求，开发中医问答功能，根据症状生成中医处方，并提供处方主治症候医学解释等辅助诊疗功能，探索中医临床经验的智慧化复制新模式。还有一些中医药大模型具有一定的中医药方面的知识及回答医学咨询的能力，可以根据症状推荐方剂/中草药，赋能中医药传承。

基于 AI 大模型的智慧康养应用前景广阔，满足老年人陪伴、看护、社交娱乐等多样化需求。老年人由于身体机能下降、社交活动减少等原因，对陪伴的需求更加强烈。传统养老陪伴、聊天机器人仅能实现特定语句的反馈，在情感交互、内容生产、机器自学习能力上还有所欠缺。大模型强大的语言理解和文本对话生成能力，使得养老机器人以自然语言进行多轮对话、拟人化的深度聊天，并整合个人喜好和历史聊天信息，更好地理解老年人内心需求，提供精神慰藉和陪伴。

居家看护机器人通过加载多模态大模型，融合智能感知、大数据、云计算等技术，可用于解决独居老年人生活照料、看护难题。基于大模型的智能创作和社交娱乐产品能让老年人拥有更佳的视效体验，且易于理解和操作，如将原有图文健康报告一键转换为具有解说词、配音、画面的视频报告，或根据语音/文字输入自动生成创意内容，满足老年人的文娱需求。此外，大模型技术有望破解适老化改造中 App 智能客服体验感不佳、家庭机器人“并不智能”等难点问题。

8.AI 大模型在医疗保险领域的应用

AI 大模型助力医疗保险数据处理自动化和信息咨询，落地场景向智能核保核赔延伸。大模型在医保政策法规、医保基础知识、商业健康险产品和保险条款、医保违法违规案例等数据上进行训练，可向保险从业人员、参保客户等提供保险信息问答、产品细节解读和决策参考服务，赋能保险产品智能营销。大模型还可助力保险公司索赔处理、文档处理和客户服务解决方案，增强其端到端的业务流程自动化能力。目前已有医疗保险大模型产品实现智能医疗审核和小额快赔，在票据识别过程中提高了药品和费用归集处理的准确率，同时还具有信息修整、辅助药品和费用归集处理等多个功能。随着数据积累和技术发展，大模型赋能的保险业务有望从前端销售、业务问答、文档处理等场景，向个性化产品定价、核保核赔、风控减损等场景拓展延伸。

9.AI 大模型在医学教育领域的应用

AI 大模型可以模拟不同类型的病人与医生进行对话，带来提高学生知识、技能和能力的新机会。大模型可充当虚拟患者或虚拟测试对象，为医学生模拟临床环境，提出问题，解释响应并提供反馈，使医学生能够在安全和受控的环境中练习临床推理、决策和沟通技巧，帮助医学生为未来的医疗实践做更好的准备。大模型也可作为虚拟导师，通过提供即时反馈和个性化指导来帮助学生将医学理论知识应用于现实世界的情境。大模型还可以用于课程开发、个性化教学计划和材料、学生测试和评估、医学写作协助等方面，辅助临床医生培养工作。例如，Hippocratic AI 产品利用大模型的能力模拟患者，完成和人类医生的对话，其模拟的患者不仅具有不同疾病、性格、情绪和疾病史，还能为医学生的临床诊断技能提供反馈评价。

四、医疗健康大模型面临的风险与挑战

（一）技术风险：精度不够，尚不能完全满足医疗场景安全性可靠性需求

大模型“幻觉”问题成核心关切，模型准确性、可靠性有待提升。大模型能够产生听起来合理、措辞连贯但不准确的内容，出现“幻觉”问题。当此类内容用于提供医疗建议或临床决策时，或将带来严重后果。在生命科学研究中，尽管 AI 大模型能针对选题生成综述和参考文献，在研究方案设计时提供新思路，但 AI 大模型可能伪造参考文献，由此生成的内容难以辨别真伪。一方面，训练数据的准确性、完整性有待确认和验证，大模型训练需要输入大量的生物医学数据，人

工检查数据质量难以实现，且可能出现训练数据集和测试数据集重叠而导致模型准确性的过度预测现象；另一方面，大模型没有经过“理解”训练，仅限于学习单词间的概率性关联，而不是如人类一样理解了输入或输出的信息，大模型尚无法掌握医学知识和临床决策的全部复杂性，也无法完全复制临床医生的经验和细致的判断。

大模型缺乏可解释性和透明度，难以取得用户的信任。一方面，医疗健康大模型问答过程或决策逻辑的可解释性不足，使其较难获取用户、专业人员和监管机构的信任。大模型通常采用深度神经网络，拥有多个隐藏层，参数超过亿级别，使用多种策略实现并行加速，因而难以追踪单一数据在模型中的处理过程，也很难获得模型产生对应推理结果的有效解释³⁵，也不清楚训练数据集的哪部分被用于生成结果。另一方面，医疗大模型透明度和开源发展不足。在一般领域，许多用于预训练 AI 大模型的大型数据集仍然闭源，大模型代码通常不会公开发布，甚至有些大模型本身只限于研究人员接触，业内很难独立验证和建立以前的结果。由于生物医学数据通常价格昂贵并且受到更严格的审查，上述情况预计将在生物医学、临床和健康领域更加常见。透明度和可验证性的不足会进一步引发用户的疑虑和不信任心理。

大模型的输出具有不可预测性，较难满足一致性、可控性合规需求。大模型理论（如上下文学习）处于黑箱阶段，人类暂时无法通过大模型机理实现对输出的完全可控。指令微调虽能引导大模型输出人类倾向的结果，但无法精准或完全控制其输出。大模型能够接受多种

多样、灵活的输入，但大模型的推理有一定的不可预测性，输出并不是唯一确定的，输入措辞的微小变化有可能导致输出结果的巨大变化。输入的灵活多变和输出的不确定性使得在医疗保健环境中应用大模型时难以确保一致性。而医疗领域对于系统的输出结果有严格的控制要求，如果模型不能满足可控和合规的需求，其应用价值就难以体现。

（二）落地挑战：数据、成本、权责问题制约大模型在医疗领域落地应用

大模型预训练缺乏高质量医疗健康数据集，训练可能出现偏差和过度代表性。数据质量是影响大模型性能的重要因素之一，如果训练数据的准确性、完整性、一致性、可靠性不高，那么大模型预训练可能会出现错误或者偏差。尤其是医疗健康大数据，更是有类别多样性、模态跨尺度、未标记数据的海量性和标记数据的稀缺性等特点。大模型预训练需要考虑三方面因素，包括通过数据过滤选择高质量的数据、删除重复的数据以避免记忆和过拟合、确保数据的多样性以促进模型的泛化。在斯坦福大学统计的 80 多种临床语言基础模型和电子病历基础模型中³⁶，几乎所有的临床语言基础模型都是在 MIMIC-III 公共数据集上进行训练，造成一定的过度代表性。因为 MIMIC-III 是研究人员少数可用的临床记录公共数据集之一。大多数电子病历基础模型在小型公共电子病历（EHR）数据集或单个私立医院的 EHR 数据库上进行训练。

大模型训练、推理成本高昂，在一定程度上制约了医疗应用落地。

大模型在金钱、时间、算力、硬件设施和环境方面具有高昂的开发和运营成本。即使 GPU（图形处理器）的计算能力不断增强，这些硬件也可能无法跟上大模型在网络参数、网络深度和数据量的快速增长。例如 GPT-4 的训练使用了约 $2.15e25$ 的 FLOPS 算力，涉及 2.5 万个 A100 GPU，耗时 90~100 天，训练成本高达数千万美元。即使使用新一代 H100 GPU 进行预训练，时间也要 55 天左右³⁷。日常运营和模型迭代中也会消耗大量算力，同时带来较大的电力、碳排放等环境成本，给医院私有化部署带来挑战。未来仍需进一步探索更高的硬件算力、更快的软件算法、更优的分布式计算策略、更强的软硬件协同、更好的技术模式，降低医疗健康大模型研发和运营成本。实际医疗健康场景部署中，需进一步做好模型大小选择、成本和收益权衡。

医疗健康大模型引发新的权责问题，问责制仍待探讨完善。尽管大模型具有较强的生成能力和支持人类医疗决策的潜力，但它们也存在限制个人自主权和产生新义务、新纠纷的风险。个别患者可能会过度依赖大模型等人工智能系统，失去对自身健康的自主理解或控制，因而也可能需要承担新的责任；AI 大模型参与生成的医学论文中，AI 能否署名作者以及生成内容的原创性、准确性由谁负责的问题仍需探讨，AI 生成内容可能存在知识产权侵权风险，其来源和生成内容也存在产权权属不清、取证和损害认定困难等问题；允许大模型协助或参与哪些医学、医疗任务，在各任务中赋予 AI 大模型何种自主权（作为自主、半自主或完全从属工具）等问题也有待探讨和达成共识。此外，

权责界定模糊会降低潜在用户的使用意愿，大模型的开发、实施和使用过程涉及生态链中多方角色，如基座模型提供者、应用程序开发者、销售商、医护服务提供者、患者等，一旦出现问题，难以确定与AI相关医疗错误的确切原因，也很难认定多个参与者的责任。

（三）数据安全和隐私：个人数据滥用、隐私泄露和网络攻击风险突出

大模型训练数据具有多种不同来源，存在隐私泄露风险。通用大模型的训练语料库通常包含不同来源的各种数据，其中可能包括私人信息，而且大模型可能会尝试根据用户输入来预测患者的性别、种族、收入或宗教信仰，最终侵犯个人隐私。用于医疗健康大模型训练的生物学和临床文本数据包含患者特征信息，如健康状况、疾病诊疗情况、临床监测数据、生物基因信息等，不仅涉及患者隐私，还具有特殊的敏感性和重要价值，一旦泄露，可能给患者带来身心困扰和财产损失，甚至对社会稳定和国家安全造成负面影响。未经知情同意，患者数据也可能被用于其他用途。在训练模型之后，人工智能系统仍然存在面临隐私攻击的风险，有时只需检查生成的模型就可以重建训练中使用的原始数据点，从模型层参数中还原出原始的输入信息。如果在训练之前对输入进行加密，则可以更好地保护患者数据免受此类攻击，但这种方法通常以模型可解释性为代价³⁸。

大模型面临多种网络数据安全风险，易遭受投毒攻击和对抗性攻击。在数据收集阶段，可能会面临数据被篡改、数据偏差、虚假数据、

数据泄露等安全风险；在数据预处理阶段，攻击者篡改图像并滥用人
与机器之间的（视觉）认知差异，以实现欺骗和逃避攻击；在模型训
练阶段，AI 大模型最易受到投毒攻击，攻击者可以通过注入大量“错
误”的数据来搅乱数据的分布，调整大模型向其所期望的道路偏移，
甚至是利用后门，在 AI 系统内深度隐藏，以待最佳的攻击时机；在
推理阶段常常面临对抗攻击，即对输入样本故意添加一些人为无法察
觉的细微干扰，导致模型以高置信度给出一个错误的输出。对抗攻击
与投毒攻击不同的是，对抗攻击并不会改变模型。在使用阶段，要避
免由大模型引起的数据泄露、模型接口等安全风险³⁹。

（四）伦理道德问题：大模型加剧医疗偏见和有害、 虚假信息传播问题

大模型会强化医疗领域偏见现象，加剧歧视和社会不公平性。由
于性别、年龄、种族、收入、教育和地理差异，世界上大多数国家在
医疗保健方面仍存在不平等、不公平现象。大模型会延续和放大导致
医疗保健不公平的系统性差异和人类偏见。一是随着大型语言模型的
规模和能力不断增长，偏见的可能性和严重性也在不断增加⁴⁰。例如，
Gopher（2800 亿参数的语言模型）比相对较小的模型表现出更高的偏
见和毒性水平⁴¹。二是预训练数据规模庞大，难以实现人工收集、标
注和检查，训练数据集中可能存在样本不平衡（如某些人群的代表性
不足、过分强调特定治疗或过时的医疗实践）、歧视、偏见、歪曲表述
等内容，由此产生的模型可能会无意中学习和传播这些偏差，输出对

某些群体的刻板印象或负面关联信息。三是算法设计中，模型开发者如何优化模型性能，在模型设计和测试中在多大程度上考虑了特定的敏感属性，以及如何定义和衡量期望的结果等⁴²，都可能会进一步加剧不公平性。

大模型有可能生成有害内容，且可能传播虚假、错误的信息。在生成有害内容方面，随着医学领域大型语言模型的兴起和发展，患者可能会无意中接触到可能导致严重情感伤害的话题。虽然患者有其他方式获取信息（谷歌等搜索引擎），此类问题不是大型语言模型独有的，但大型语言模型通常提供类似聊天对话的交互界面，更容易获得患者的信任，产生更大的风险。这些模型拟人化程度高，但往往缺乏额外的个性化情感支持的能力²²。在虚假信息传播方面，大模型生成的文本与人类书写的文本越来越难以区分，虽然使用大模型帮助临床医生完成行政任务，可以帮助减轻文档负担，但这也可能导致其他人恶意使用大模型来生成虚假、错误、误导性的文档。

五、医疗健康大模型的评价验证和监管治理

（一）标准和指南：基础信息安全标准开始起步，领域平台规范指南需求迫切

大模型基础通用的信息安全标准开始起步制定。在通用 AI 大模型标准方面，全国信息安全标准化委员会（TC260）从信息安全的角度，针对生成式人工智能服务相关的训练数据安全、人工标注安全、防范虚假信息等方面已经立项了《信息安全技术 生成式人工智能人

工标注安全规范》、《信息安全技术 生成式人工智能预训练和优化训练数据安全规范》等系列国家标准，并发布了《网络安全标准实践指南-生成式人工智能服务内容标识方法》，用于指导生成式人工智能服务提供者提高安全管理水平。

大模型在药械领域的合规应用指南需求迫切。在生物制药与医疗器械领域，IT 技术应用于行业内的数据处理需要满足严格的 GxP 法规体系，用于规范药物/医疗器械的研发、临床、制造和流通等。因此出于对患者安全的重视，为了更好地支持和保障业务，生命科学行业的 IT 基础设施通常要有更高的安全性、可用性及保密性需求。针对 AI 大模型在药械领域的应用，业界迫切需求面向 GxP 合规的 AI 大模型系统评价标准、指南作为指导，包括 AI 大模型云基础设施、AI 大模型算法、AI 大模型应用等，需要提出适用于不同细分领域、规模的企业最佳实践，并结合最佳实践经验形成标准、指南和规范，为政策制定者、监管方、企业、研究机构提供有益参考。

大模型在健康领域的平台建设规范开始有所探索。在医疗健康 AI 大模型标准方面，中国信通院联合高校、医院、科技巨头企业、医疗人工智能企业成立生命科学与大模型工作组，推动国内标准研制和国际标准转化、评估框架和验证平台搭建、医学数据集平台建设、医疗大模型伦理安全发展工作。中国通信标准化协会（CCSA）移动互联网+健康标准工作组从信息技术支撑医疗健康系统平台建设的角度，积极推动医疗健康大模型行业系列标准研制，用于指导医疗健康领域

AI 大模型的系统平台实践，共同促进产业发展。

（二）评价和验证：针对新能力、新特性、真实表现的动态评估方法有待建立

医疗大模型评估以学术数据集为主，无法衡量大模型在现实世界中的表现。现有的医疗大模型主要是在学术数据集（如 MedQA、PubMedQA 等）和为人类设计的医学考试（如 USMLE）上进行评估，但此类数据集具有局限性，并不能确切反映大模型在真实临床任务中的表现。临床实践并不等同于正确回答各类主观客观测试题和考试问题，找到适当的基准来衡量大模型的临床潜力是一项巨大的挑战。未来有必要进一步探讨评估医疗大模型的方法，建立反映真实需求的多样化、复杂和真实的基准数据集，引入新的测评方式，更好地理解 and 验证医疗大模型在真实世界应用中的有效性和实用性。

人类与 AI 或将建立新型合作关系，大模型应用于医疗的价值主张有待探讨。人工智能的出现使得治疗过程更加便捷、医患对话进一步加深，有可能将传统上家长式的临床-患者关系转变为决策过程中的联合伙伴关系。在医疗健康领域引入大模型的目标是增强人类的决策力、判断力，而非取代人类。需要进一步思考一个人加上这些大模型可以一起做什么，从而产生更好的医疗保健结果。目前对医疗大模型的评估没有考量人类和 AI 之间新型合作的好处和价值。未来需明确定义大模型应用于医疗领域的价值，并围绕相关价值主张设立对应的测量指标，以验证大模型在医疗领域应用的好处和价值。

医学问答为医疗大型语言模型评测重点，但对交互特性关注不足。

医疗健康领域对准确可靠答案的需求较高，大型语言模型具有强大的自然语言处理和推理能力，已被广泛用于医疗问答系统，以提供准确及时的行业信息。目前大多数医疗大型语言模型的评估都是在医学问答数据集展开，如谷歌构建的最新基准 MultiMedQA，是由七个医学问答数据集组成的基准，包括 6 个现有数据集 MedQA、MedMCQA、PubMedQA、LiveQA、MedicationQA、MMLU 临床主题和 1 个全新数据集 HealthSearchQA。在测试方法和题型设置上，已有评测体系大多选用单轮对话、选择题等形式，对大模型的整体理解能力及生成能力方面考量不足。现有评测体系主要关注大模型生成内容与预期答案的匹配程度，忽略了对话类医疗大模型的交互特性，其中封闭式问题居多，输出简洁（如多项选择），无法反映基于大模型的聊天助手的典型使用情况。

动态的评价体系有待开发建立，以满足医疗大模型不断发展的能力评估需求。现有医疗人工智能任务的评估大多依赖于静态和公共基准（评测数据集通常是公开的）。一方面，鉴于大模型技术的快速发展，大模型的能力可能随着时间的推移而增强，现有的静态基准无法准确、公平地评估大模型不断发展的能力。另一方面，随着大模型的规模和训练数据集的增大，评估数据集有可能被包含在训练数据集中，导致潜在的训练数据集污染问题。因此，需要开发建立动态的、不断发展的医疗大模型评价体系，实现对大模型公平公正的评价。

（三）政策和监管：促发展与防风险并重，奠定医疗大模型监管治理良好基础

算力、数据和应用政策布局加速，推动医疗大模型技术和产业发展。韩国科学与信息通信部发布《超大型 AI 竞争力提升方案》，通过部署技术产业基础设施、创新生态系统、创新制度和文化加速 AI 大模型发展，提出医疗等五大领域超大型 AI 旗舰项目，关注健康检查和医疗意见撰写、既往病历总结、治疗方法和处方推荐等诊疗业务，以及个人健康管理等医疗咨询两大方向。国内来看，地方政府加快布局大模型在智能诊疗、生命科学研究、药物研发领域的应用。北京率先发力，发布促进通用人工智能创新发展的若干措施，布局算力、数据、技术、场景和监管，提出挖掘多模态医疗数据、蛋白质和分子药物实验数据，探索通用 AI 在医疗和科学研究的应用。深圳发布第一批“城市+AI”应用场景清单，涵盖智慧医疗等 26 个场景。上海重点发展生命科学、药物设计大模型技术和平台。成都着力推动医疗行业大模型技术创新，布局 AI+医疗、AI+科研（如药物研发、基因研究）。

国际社会加强前瞻性研讨，重视医疗大模型的风险防范和有效监管。联合国支持成立国际性 AI 监管机构，计划设立高级 AI 咨询机构，支持各国最大限度地发挥人工智能的作用，减轻现有和潜在风险，并建立全球协调一致的监测和治理机制。世界卫生组织（WHO）呼吁在医疗领域谨慎使用大型语言模型工具，这些技术“迅速”崛起，其中许多仍处于实验阶段，需要仔细研究它们对医疗保健和科学研究关

键价值观(包括透明度、包容性、公众参与、专家监督和严格评估)构成的风险。美国商务部下属的国家标准与技术研究院(NIST)计划成立一个新的人工智能公共工作组,研究生成式 AI 的机遇和挑战,指导应对生成式 AI 相关的风险。美国食品药品监督管理局(FDA)局长指出大型语言模型有着重要的医疗应用,可以改善医疗保健,但需要“灵活”的监管方法⁴³。英国药品和健康产品监管局(MHRA)软件和人工智能主管表示,为特定医疗目的(如提供诊断、治疗或临床决策支持)而开发或调适的大型语言模型可能要纳入医疗器械范畴,受到审批监管。

医疗大模型监管行动任重道远,面临诸多挑战。一是质量控制与标准化面临挑战,大模型具有较高的复杂性,评估验证训练数据来源及质量、算法科学性、结论可解释性的标准、方法缺失,难以有效支撑监督审查,确保 AI 大模型生成的医疗建议的可靠性和一致性;**二是**大模型比其他形式的人工智能更难监管,其输入和输出范围更加广泛,内容更不可控;**三是**大模型的来源无法控制,当其用作构建医疗设备的基座模型或医疗设备组件时,无法监测和控制其来源,一些开源大模型可能会被视为来源不明的软件(SOUP),面临合规挑战;**四是**大模型具有实时适应特性,可以根据用户输入和不断变化的环境实时调整其响应,这种动态行为要求持续的监督监测和评估机制,以确保负责任的使用和对伦理准则的遵守。

六、医疗健康大模型发展建议

加快医疗大模型基础研究和技术研发，实现技术自主创新突破。加大基础研究投入，加快医疗大模型算法、框架等基础性原创性技术突破，提升医疗大模型在泛化性、准确性、可解释性、公平性等方面的能力，突破技术局限；加快研发提升模型计算效率的技术，加大研究剪枝、量化、知识蒸馏等模型压缩技术，降低大模型开发和使用成本；鼓励产学研医联合开发相关技术，整合 AI 科技企业、制药企业、CRO 企业、医疗机构、科研机构、智算超算资源，通过传统医学人才和 AI 大模型技术人才的联合攻关，突破新理论、新技术和新算法。

推进医疗大模型数据库和算力设施建设，夯实产业持续发展基础。加快构建兼具质量和数量的医疗数据集，支撑 AI 大模型训练、测评；建设电子病历、基因、蛋白质、试验数据、专科病种多元化数据集，保障数据质量、数据共享和信息标准化；建设医学知识信息公共平台，对不受版权限制的医学文献、报告进行开放获取；面向医疗健康领域需求，开发 AI 大模型算力基础设施相关技术和核心器件、设备，加快构建高性能、低能耗、多层次计算基础设施；向高校、研究机构、中小企业和初创企业提供算力产品和计算资源支持。

强化医疗大模型场景应用牵引，推动“技术-应用-产业”良性循环。鼓励有条件的机构探索大模型技术在生物医药、医疗健康实际场景中的应用，充分发挥应用场景的纽带作用，推动医疗大模型技术向产业转化，促进医疗大模型成果落地和产业链协同发展；加快大模型

在医疗健康领域的融合发展，实施产业跨界融合示范工程，在生命科学研究、新药研发、疾病预防、诊断治疗、康养护理、健康管理等领域建设一批创新场景，积极培育新模式新业态；依托“揭榜挂帅”“试点示范”等机制，鼓励协同攻关，拓展应用场景，加速关键核心技术应用迭代与产业化。

研制医疗大模型标准规范和测评基准，构建动态、实用的评价体系。探索建立医疗大模型安全性、有效性和可靠性评估标准，研制具体技术标准和评估准则，确保大模型在实际应用场景中稳定可靠运行；面向医疗健康领域，建设动态更新、规模更大及模态更多的基准测试数据集，促进大模型技术问题发现和改进，提升技术指标，确保大模型具有符合应用需求的性能；建设标准验证、测试和仿真等服务平台，对医疗大模型性能以及伦理、网络安全等潜在风险进行评估，推动相关产品及服务的评价和优化，满足产业发展需要；加强国际合作，推进全球性的医疗大模型技术标准的制定和合作交流。

建立健全医疗大模型监管机制，保障行业规范有序、高质量发展。鉴于医疗大模型的交叉融合特性，要通过跨部门、跨领域协同，建立全方位、多层次、立体化的监管体系，提高监管效能；采取分类管理、风险分级策略，根据医疗大模型的预期用途和风险等级进行监管，确立不同类型医疗大模型的使用范围和限制；加强训练数据的来源及质量审核，加快数据确权、共享和安全防护机制研究，明确数据收集和使用的合规要求，防止数据滥用和隐私泄露；探讨建立医疗大模型问

责制度，明确研发、部署、使用方的权利和义务；明确医疗大模型代码、微调技术及其生成内容的版权归属。

完善医疗大模型公共服务体系，营造更具活力的创新生态环境。鼓励医疗大模型开源社区、公共算力平台建设，降低大模型的获取成本；建设医疗大模型数据集公共服务平台，整合多源医疗健康数据，提供数据收集、处理、标注等的工具，促进数据要素价值释放；推广模型训练、微调和优化工具，建立指令微调数据集，降低大模型开发和部署门槛；探索建立公共服务平台多方参与、合作共赢的商业模式，推动公共服务平台市场化、专业化运营，实现平台自我造血；鼓励医疗大模型领域的龙头企业强化产业生态布局，提供第三方开发能力和解决方案，带动中小企业协同发展。



人工智能医疗器械创新合作平台
ARTIFICIAL INTELLIGENCE MEDICAL DEVICE INNOVATION COOPERATION PLATFORM

CAICT 中国信通院

参考文献

- 1 Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).
- 2 Moor, Michael, et al. "Foundation models for generalist medical artificial intelligence." *Nature* 616.7956 (2023): 259-265.
- 3 Xie, Qianqian, et al. "Faithful AI in Healthcare and Medicine." *medRxiv* (2023): 2023-04.
- 4 Qiu, Jianing, et al. "Large ai models in health informatics: Applications, challenges, and the future." arXiv preprint arXiv:2303.11568 (2023).
- 5 Moor, Michael, et al. "Foundation models for generalist medical artificial intelligence." *Nature* 616.7956 (2023): 259-265.
- 6 Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- 7 Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- 8 Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).
- 9 Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
- 10 Buehler, Markus J. "Generative Pretrained Autoregressive Transformer Graph Neural Network applied to the Analysis and Discovery of Novel Proteins." arXiv preprint arXiv:2305.04934 (2023).

- 11 Yang, Jingfeng, et al. "Harnessing the power of llms in practice: A survey on chatgpt and beyond." arXiv preprint arXiv:2304.13712 (2023).
- 12 Lee, Jinhyuk, et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." *Bioinformatics* 36.4 (2020): 1234-1240.
- 13 Moor, Michael, et al. "Med-Flamingo: a Multimodal Medical Few-shot Learner." arXiv preprint arXiv:2307.15189 (2023).
- 14 Li, C., et al. "LLaVA-Med: training a large language-and-vision assistant for biomedicine in one day. ar**v [Preprint]. 2023 [cited August 21, 2023]."
- 15 Li, Chunyuan, et al. "Llava-med: Training a large language-and-vision assistant for biomedicine in one day." arXiv preprint arXiv:2306.00890 (2023).
- 16 Rampášek, Ladislav, et al. "Recipe for a general, powerful, scalable graph transformer." *Advances in Neural Information Processing Systems* 35 (2022): 14501-14515.
- 17 Gao, Zhangyang, Cheng Tan, and Stan Z. Li. "Alphadesign: A graph protein design method and benchmark on alphafolddb." arXiv preprint arXiv:2202.01079 (2022).
- 18 Zhang, Ran, et al. "MHTAN-DTI: Metapath-based hierarchical transformer and attention network for drug–target interaction prediction." *Briefings in Bioinformatics* 24.2 (2023): bbad079.
- 19 Tu, Tao, et al. "Towards generalist biomedical ai." arXiv preprint arXiv:2307.14334 (2023).

- 20 Driess, Danny, et al. "Palm-e: An embodied multimodal language model." arXiv preprint arXiv:2303.03378 (2023).
- 21 Meyer, Jesse G., et al. "ChatGPT and large language models in academia: opportunities and challenges." *BioData Mining* 16.1 (2023): 20.
- 22 Omiye, Jesutofunmi A., et al. "Large language models in medicine: the potentials and pitfalls." arXiv preprint arXiv:2309.00087 (2023).
- 23 Yang, **, et al. "A large language model for electronic health records." *NPJ Digital Medicine* 5.1 (2022): 194.
- 24 Li, Yunxiang, et al. "ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge." *Cureus* 15.6 (2023).
- 25 Singhal, Karan, et al. "Towards expert-level medical question answering with large language models." arXiv preprint arXiv:2305.09617 (2023).
- 26 Shoeybi, Mohammad, et al. "Megatron-lm: Training multi-billion parameter language models using model parallelism." arXiv preprint arXiv:1909.08053 (2019).
- 27 Chiang, W. L., et al. "Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, Mar. 2023."
- 28 Nori, Harsha, et al. "Capabilities of gpt-4 on medical challenge problems." arXiv preprint arXiv:2303.13375 (2023).
- 29 智源研究院林咏华: 未来十年的产业落地, 大模型和小模型

将共存[EB/OL]. (2023-05-29) [2023-09-24].

https://www.sohu.com/a/679964625_350699 .

30 IDC. 2022 中国大模型发展白皮书[EB/OL]. (2023-02-16) [2023-09-24].

<http://www.xinhuanet.com/tech/20230216/c3a0ae25dd134a9aae2e13f26901d8ca/c.html> .

31 Madani, Ali, et al. "Large language models generate functional protein sequences across diverse families." *Nature Biotechnology* (2023): 1-8.

32 Theodoris, Christina V., et al. "Transfer learning enables predictions in network biology." *Nature* (2023): 1-9.

33 Cui, Haotian, et al. "scgpt: Towards building a foundation model for single-cell multi-omics using generative ai." *bioRxiv* (2023): 2023-04.

34 Taylor, Ross, et al. "Galactica: A large language model for science." *arXiv preprint arXiv:2211.09085* (2022).

35 智源研究院. 超大规模智能模型产业发展报告[EB/OL]. (2021-09-26) [2023-09-24].

https://www.sohu.com/a/492116495_121015326 .

36 Wornow, Michael, et al. "The shaky foundations of large language models and foundation models for electronic health records." *npj Digital Medicine* 6.1 (2023): 135.

37 终极“揭秘”：GPT-4 模型架构、训练成本、数据集信息 [EB/OL]. (2023-08-16) [2023-09-24].

<https://www.toutiao.com/article/7267753977444794914/> .

38 Kaissis, Georgios A., et al. "Secure, privacy-preserving and

federated machine learning in medical imaging." Nature Machine Intelligence 2.6 (2020): 305-311.

39 拥抱 AI，网络安全的机会在哪里[EB/OL]. (2023-05-27) [2023-09-24]. <https://www.secrss.com/articles/55068> .

40 Zhang, Daniel, et al. "The AI index 2021 annual report." arXiv preprint arXiv: 2205.03468 (2022).

41 Rae, Jack W., et al. "Scaling language models: Methods, analysis & insights from training gopher." arXiv preprint arXiv:2112.11446 (2021).

42 Chen, Irene Y., et al. "Ethical machine learning in healthcare." Annual review of biomedical data science 4 (2021): 123-144.

43 Speech by Robert M. Califf, M.D. to the National Health Council's 2023 Science for Patient Engagement Symposium -- Patient Empowerment in the Digital Health Era.[EB/OL]. (2023-05-08) [2023-09-24]. <https://www.fda.gov/news-events/speeches-fda-officials/speech-robert-m-califf-md-national-health-councils-2023-science-patient-engagement-symposium-patient> .

人工智能医疗器械创新合作平台

地址：北京市海淀区气象路 50 号院 1 号楼

网址：www.aimd.org.cn



中国信息通信研究院

地址：北京市海淀区花园北路 52 号

网址：www.caict.ac.cn

