

盘点2020人工智能十大重要事件

■ 映 寒

对于人工智能产业来说,2020年是不平凡的一年,是困难与高光时刻并存的一年。

年初的疫情,是一场对人工智能发展的重要考验。AI的助力,为抗疫提供了丰富且有效的武器,业界涌现了一批优秀的人工智能优秀应用,在一定程度上对冲了疫情影响,成为疫情防控一线的使用工具。回望2020年,优异的研究成果彰显了人工智能领域的高速发展。这份2020年度AI大事盘点,让我们共同回顾那些人工智能领域的高光时刻。

AI对抗新冠病毒

AI加快了科学家对新冠病毒疫苗的搜寻,全球的机器学习研究人员争先恐后地利用AI技术来对抗新冠病毒。

例如:中国研究机构合作开发了可在CT扫描中检测Covid-19的AI模型,其准确度超过90%;美国生物技术企业莫德纳公司,其疫苗于2020年12月份获得美国食品药品监督管理局的批准,它使用机器学习来优化mRNA序列以转化为可以测试的分子;聊天机器人提供了合成的虚拟朋友,供被疫情封锁的人们聊天;视频会议公司为在线办公提供AI模型,以过滤背景噪声,并将睡衣变成虚拟的商务正装。

脸书推出最新聊天机器人

BlenderBot是脸书开源的94亿参数聊天机器人。脸书宣称,它是脸书多年研究的成果,具有包括同情心、知识和个性在内的多种会话技巧的组合。

根据评估者的看法,BlenderBot在参与度方面优于其他模型,并且感觉更人性化。这个聊天机器人包含94亿个参数,具有改进的解码技术,新颖的技能融合,参数模型数量是之前谷歌发布的聊天机器人系统的36倍。

字节跳动发布

全球最大钢琴MIDI数据集

字节跳动发布全球最大的古典钢琴数据集GiantMIDI-Piano,开发并开源了一套高精度钢琴转谱系统。

数据集中筛选并转谱了来自2786位作曲家的10854部钢琴作品,不同曲目的总时长达到了1237小时,是谷歌MAESTRO数据集的14倍,填补了计算机音乐领域一直缺少一个大规模的钢琴MIDI数据集的空白。

GiantMIDI-Piano的用途包括但不限于音乐信息检索、自动作曲等。

全国首个常态化运营 5G无人公交落地

2020年10月21日,由中国移动和轻舟智航联合部署的无人公交落地苏州高铁新城,并启动了“苏州无人公交早鸟计划”,招募一批有日常通勤需求的市民免费搭乘。

据了解,无人公交车速可达20千米/小时至50千米/小时,基于中国移动的5G网络和轻舟智航的自研技术,无人巴士在行驶过程中精准且灵活地避让、自动变道、自动转向、自动识别红绿灯,即使面对穿行人车混杂的路口、后车加塞、电动车“鬼探头”等情况,也能沉着应对。

最大语言模型GPT-3发布

2020年2月,微软发布全球最大的深度学习模型,拥有170亿参数的TuringNLP,几个月之后它就被由OpenAI打造的文本生成器GPT-3超越了。

GPT-3是一个具有1750亿参数的自然语言深度学习模型,它还收集了CommonCrawl和Wikipedia的数据集,数据集总量是之前发布的GPT-2的116倍,是迄今为止最大的训练模型。

作为GPT-2的升级版,它们功能上有什么异同呢?虽然都是基于Transformer模型,修改初始化、预规范化、可逆标记化性能也都是是一样的,但是它们的ransformer模型的类型不同,GPT-3使用了一种类似于稀疏Transformer的东西,在各层中运用了交替密集、局部带状的稀疏注意模式。

GPT-3还完美地弥补了BERT模型的两个不足之处,它既不用对领域内标记的数据过分依赖,也不会对领域数据分布过度拟合。这个强大的语言模型,不仅能够答题、翻译、算数、完成推理任务、替换同义词等。它还能够撰写新闻,写出来的新闻有理有据,难辨真假。

“阿尔法折叠”

精准预测蛋白质三维结构

人工智能辅助基础科研的表现可谓亮眼。2020年11月底,DeepMind公司传来喜讯,其人工智能系统的“阿尔法折叠”方法攻克了困扰生物学半世纪的难题,成功预测蛋白质如何从线性氨基酸链卷曲成3D形状以执行任务,这将极大加快新药研发进程。

人体拥有成千上万种不同的蛋白质,每一种蛋白质都包含几十到几百种氨基酸,这些氨基酸的顺序决定了它们之间的作用,赋予蛋白质复杂的三维形状,进而决定了蛋白质的功能。了解这些三维形状有助于研究人员设计出能在蛋白质缝隙内滞留的药物。此外,合成出拥有所需结构的蛋白质,还可以加快酶的研制进程,让生物燃料领域受益。

几十年来,研究人员利用X射线晶体学或低温电子显微镜等实验技术来破译蛋白质的三维结构,但这种方法可能需要数月甚至数年,且未必见效。目前,在生命体内发现的2亿多个蛋白质中,只有17万个蛋白质的结构被破解。

美国马里兰大学结构生物学家约翰·穆尔特表示,这是一个有50年历史的问题,“阿尔法折叠”改变了游戏规则,实验学家将能使用精确的结构预测来理解不透明的X射线和低温电磁数据;药物设计者也可借此迅速厘清新冠病毒等新出现的危险病原体中每种蛋白质的结构,从而更快研制出相关药物。



人工智能生成作品 受著作权法保护

AI作者的才华已经辐射到越来越多的创作领域,与此同时,AI的作品属于谁还是个有待讨论的问题,这对包括著作权法在内的知识产权制度带来不小的挑战。2020年2月,人工智能写作领域第一案的判决结果为相关问题提供了一个判例。按照腾讯公司状告“网贷之家”的判决结果,AI生成作品属于著作权法保护范围。判决书显示:涉案文章由原告主创团队人员运用Dreamwriter软件生成,其外在表现符合文字作品的形式要求,其表现的内容体现出对当日上午相关股市信息、数据的选择、分析、判断,文章结构合理、表达逻辑清晰,具有一定的独创性。

独创性,正是案子的核心争议点之一。在中国社会科学院科学技术和社会研究中心主任段伟文看来,这是个复杂的问题。中国法律体系的主要渊源是制定法,但对于一些全新的法律冲突也需要通过具体判例来推行。“我国相关法律里目前尚无关于人工智能写作或智能写作是否具有知识产权的说明,这个判例具有一定的开创性。”

清华大学发布 首个自动图机器学习工具包

2020年12月21日,清华大学朱文武教授带领的网络与媒体实验室发布了全球首个开源自动图学习工具包——AutoGL。该工具支持在图数据集上全自动进行机器学习,并且支持图机器学习中最常见的两个任务:节点分类任务与图分类任务。

AutoGL工具包首先使用AutoGL Dataset维护图机器学习任务所需数据集。AutoGL Dataset导入了大规模图表示学习工具包和图神经网络库中的数据集模块,并添加对OGB数据集的支持,同时还添加了一些支持以便集成auto solver框架。不同的图机器学习任务可以通过不同的AutoGL Solver得到解决。AutoGL Solver使用四个主要模块自动化解决给定任务,分别是特征工程、图学习模型、超参数优化,以及模型自动集成。每个部分在设计时都引入了对图数据特殊性的考虑。

《国家新一代人工智能标准 体系建设指南》印发

2020年8月5日,为加强人工智能领域标准化顶层设计,推动人工智能产业技术研发和标准制定,促进产业健康可持续发展,国家标准化管理委员会、中央网信办、国家发展改革委、科技部、工业和信息化部联合印发了《国家新一代人工智能标准体系建设指南》。

据《国家新一代人工智能标准体系建设指南》,到2021年,明确人工智能标准化顶层设计,研究标准体系建设和标准研制的总体规则,明确标准之间的关系,指导人工智能标准化工作的有序开展,完成关键通用技术、关键领域技术、伦理等20项以上重点标准的预研工作。

到2023年,初步建立人工智能标准体系,重点研制数据、算法、系统、服务等重点急需标准,并率先在制造、交通、金融、安防、家居、养老、环保、教育、医疗健康、司法等重点行业和领域进行推进。建设人工智能标准试验验证平台,提供公共服务能力。

四城市建设国家新一代 人工智能创新发展试验区

科技部2020年3月9日发函,支持济南、西安、成都、重庆建设国家新一代人工智能创新发展试验区。其中,要求成都在智能空管、普惠金融、智慧医疗等场景加强应用示范,培育以行业融合应用为引领的人工智能新业态新模式,推动构建开放型产业体系;要求重庆开展智慧旅游、智慧物流、智慧交通、智慧生态保护应用示范,打造具有山城特色场景的智慧城市。要求西安发挥在智能感知处理、智能交互等方面的研发基础和人才优势,在先进制造、文旅旅游、商贸物流等方面形成一批有效的行业解决方案,打造创新驱动发展的新引擎。要求济南发挥人工智能应用场景丰富、算力基础和数据资源雄厚等优势,在制造、农业、交通等重大场景中的创新应用。

科技部方面表示,将积极配合试验区建设,协调研究解决相关政策问题,加强工作指导和资源对接,及时总结典型经验和政策措施并予以推广。早在2019年9月,《国家新一代人工智能创新发展试验区建设工作指引》就提出推进人工智能基础设施建设,到2023年建设20个左右试验区。