

AI 成为数学家得力助手还要多久

从计算机科学到医学再到国家安全,数学是众多关键应用的基石,但数学领域的进步可能需要数年时间才能实现。

几百年来,数学家使用的工具依然简朴:一张纸、一支笔。他们凭借逻辑与灵感,在符号的迷宫中寻找通往真理的路径。

为了打破这一局面,美国国防高级研究计划局今年4月启动了“指数性数学”计划,旨在开发一种能极大提升数学研究效率的人工智能(AI)“合著者”系统。

几十年来,数学家借助计算机进行辅助计算或验证命题,如今的AI或许能更上层楼,挑战那些人类长年未解的难题。不过,从能解高中题的AI,到能协助攻克前沿数学难题的AI,中间仍隔着一道鸿沟。

解决重大难题仍力不从心

大型语言模型(LLM)并不擅长数学。它们常常出现“幻觉”,甚至可能被误导相信 $2+2=5$ 。但新一代大型推理模型,如OpenAI的o3、Anthropic的Claude 4 Thinking等,展现出的进步令数学家眼前一亮。

今年,这些模型在美国数学邀请赛中的表现接近优秀高中生水平。不同于过去“一锤定音”式的输出,这些模型开始尝试模拟数学家逐步推理的思考过程。

同时,一些将LLM与某种事实核查系统相结合的新型混合模型也取得了突破。例如,谷歌“深度思维”的AlphaProof系统将语言模型与棋类AI——AlphaZero结合,成为首个取得与国际数学奥林匹克竞赛银牌得主成绩相当的系统。今年5月,谷歌的AlphaEvolve模型更进一步,在

多个长期未解的数学与计算难题上找到优于人类现有方案的解法。

美国《麻省理工科技评论》指出,尽管这些AI成绩亮眼,但专家们普遍认为,它们仍不具备真正的协助科研的能力。竞赛题虽难,却更像是智力游戏,有一定“套路”。真正的数学研究则更开放、更复杂。面对“P vs NP”“黎曼猜想”等重大难题时,AI仍力不从心。

为了更准确地评估AI的能力,初创公司Epoch AI去年推出了FrontierMath测试,联合60多位数学家设计出全新高难度题目,避开模型已见过的训练数据,结果LLM几乎集体“交白卷”。

这些测试表明,AI在数学道路上虽已迈步,但离“合著者”角色仍有很长一段路要走。

需攻克“超长推理链”

仔细观察数学问题会发现,它们在某些方面类似:解决问题需完成一系列连续步骤,关键在于找到这些步骤。

美国加州理工学院谢尔盖·古科夫指出,困难的差异往往体现在路径的长度上。高中数学可能只需10到40步,而像黎曼猜想这样的难题,路径可能长达百万步。

这类“超长路径”极难处理。就像下围棋时寻找一条制胜序列,AI必须在指数级增长的可能路径中找到正确解法。而在数学中,这个复杂度要远超棋类游戏。

据物理学家组织网今年2月报道,为应对这一挑战,古科夫团队开发了一种方法,将多个步骤打包成“超级步骤”,相当

于穿上“巨人靴”跨越大段路程。他们设计了一个系统,其中强化学习模型负责提出超级步骤,另一个模型负责验证其合理性。

该策略在经典未解难题——安德鲁斯—柯蒂斯猜想上取得了突破。虽然尚未证明或推翻该猜想,但借助AI,科学家推翻了一个40年来被广泛引用的“反例”。这曾被视为证明该猜想错误的依据。

牛津大学数学家马丁·布里森对此表示肯定:“排除错误路径,是科研中非常有价值的一步。”

古科夫相信,这种“压缩路径”的思路适用于所有需要推理链条的领域。他希望,这种方法不仅能推动AI跳出固有模式,也为数学研究带来新突破。



(图片为AI生成)

真正的创新和突破仍属人类

跳出思维定式,正是数学家攻克难题的关键。数学常被看作机械推理,而高等数学则更像是一场实验,充满一波三折的试错与灵光乍现的顿悟。

这正是AlphaEvolve等AI工具的优势所在。它通过LLM不断生成并改进解题代码,配合第二个模型评估每一轮结果,最终提出比人类更优的解法。这种方法不仅能独立探索,也支持人类随时介入,提供灵感和指令。

澳大利亚悉尼大学数学家乔迪·威廉姆森强调,探索性思维是数学的核心。据英国《新科学家》网站报道,他与Meta合作开发了PatternBoost AI系统,可根据一个数学想法生成相似概念,帮助激发灵感。他说:“这就像是这里有一堆有趣的东

西,我不知道是怎么回事,但你能再生成一些类似的东西吗?”

这种头脑风暴在数学中至关重要,它是新想法产生的源泉。以二十面体为例——古希腊人通过纯粹推理发现了它,其形状并不存在于自然界中,却深刻影响了数学的发展。威廉姆森希望,AI未来也能协助发现类似的“新数学对象”。

不过,目前AI仍缺乏真正的创造力。让AI赢棋是一回事,让它发明围棋游戏则是另一回事。像AlphaEvolve和PatternBoost这样的工具或许能作为人类直觉的“侦察兵”,帮助人们发现路径、避开死路,但专家普遍认为,真正的创新与突破,仍然属于人类。

(科技日报)

当心“三无防晒衣” 选购认准国标标识



(图片为AI生成)

据日前央视财经频道《财经调查》报道,市场上标有“防晒”字样的产品品类繁多,但部分防晒衣实际无法有效阻挡紫外线,不符合国家标准。消费者可能花了钱,却买到了没有防晒效果的衣服。针对热门事件,记者走访本地市场调查时发现,尽管许多防晒服饰及其外包装上,均标有明显的防晒指数标识,但由于导购人员以及消费者对这类标识的认知程度比较参差,“是否真有防晒效果”的认知仍然存在盲区。

UPF>40的产品才具备防晒效果

防晒衣作为夏日防晒的“第一道防线”,消费者首先要看懂符合国家标准的防晒指数标识。防晒衣,也即防紫外线服装,其防晒主要通过面料吸收或反射紫外线来减少紫外线的透射率。国家标准GB/T18830-2009《纺织品防紫外线性能的评定》规定,当产品的紫外线防护系数(UPF)>40,且长波紫外线(UVA)透射率<5%时,方可称之为“防紫外线产品”。

近日市场监管总局也发布了相关消费提示,指导消费者科学挑选防晒用品,其中包括防晒衣与防晒帽两部分的贴士。在选择防晒衣时,首先需要留意标签信息,查看吊牌是否有标注防紫外线功能及有无UPF值,UPF值40+为合格,UPF50+的产品防紫外线性能更好。其次看面料,织物密度

越高,防紫外线效果相对越好,尽量挑选厚度适中、面料致密的防晒衣。再次看颜色,浅色面料一般能反射更多可见光,减少热量吸收,体感温度比深色面料低。

在选购防晒帽时,可关注三个要点。其一,明示具有防晒性能的帽子,一般会在标签上标注标准编号和UPF值。UPF>40的产品才具备防晒效果。其二,结合实际使用场景选购合适的款式。其三,选择适配个人的尺寸。

警惕商家直播间的“测试把戏”

“我看到许多商家在直播间中,用紫外线感应卡测试防晒衣效果,真的可以吗?”市民刘女士向记者提出疑问。业内人士向记者表示,这种方法并不能测试出防晒衣的品质,实际上是商家玩的一种“测试把戏”,选购时需要警惕。

广东省服装服饰行业协会副会长陈韶通告诉记者,广州作为一线城市和重要的服装集散地,其主流商场、品牌专卖店销售的防晒服饰产品整体较为规范。这些渠道对入驻品牌和商品的资质审核相对严格,产品标签标识、吊牌信息(尤其是防晒标准标识UPF值)相对齐全,基本能保障产品来源可追溯、功能有依据。陈韶通建议,消费者在选购防晒用品时,务必通过正规渠道购买,仔细查看产品吊牌信息是否齐全,产品工艺是否精良。优先选择口碑良好或经过专业认证的品牌产品,注重产品的质量和功能性,切勿盲目跟风或轻信虚假宣传。

(广州日报)