

一场特殊的“期中考”

174名北大学生能否考过AI

在北京大学化学与分子工程学院,有机化学考试是许多同学痛并快乐着的挑战。然而,期中考前突如其来的一条通知,让这场考试的气氛变得不同寻常:“请注意,本次考试范围不仅限于有机化学。”

但比起考试范围的变化,考场里迎来的一批“特殊考生”,更让人意想不到。

它们不需要落座,也无需纸笔。它们是GPT、Gemini、Deep-Seek……这些当下世界上最聪明的AI,正在云端与174位北大化学与分子工程学院的大二学生同场竞技。

这是一场精心设计的“图灵测试”,也是北大科研团队为大语言模型投下的一块“试金石”。

近日,北京大学化学与分子工程学院联合北大计算中心、计算机学院、元培学院团队,发布了最新成果SUPERChem。近期,他们以一套“北大试卷”为标尺,冷静丈量着AI在科学推理上的真实边界。



考试现场

一场特殊的“期中考”

打开SUPERChem的题库,一种“压迫感”扑面而来。晶体结构的精细解析、反应机理的深度推演、物化性质的定量计算……这500道题目并非来自网络随手可得公开题库,而是源于对高难度试题和前沿专业文献的深度改编。

为什么要费尽周折重新出题?

“因为大模型太会‘背书’了。”团队成员解释道。互联网可及的测试题大多已被博闻强识的AI在训练阶段熟读。而化学,恰恰是一门不能只靠死记硬背的学科。它既有严密的逻辑推演,又充满了对微观世界的空间想象。“我们非常好奇,大语言模型的一维next token prediction,能否解决二维甚至三维空间中的复杂推理问题。”

要设计一套让AI“没见过”、必须靠硬实力推理的题目,难度极高。然而,这正是北大化院的独特优势所在。近百名师生——其中不乏奥林匹克金

牌得主——集结起来,决定给AI出一套高门槛、重推理、防作弊的试卷。他们要考的,是AI是否真的“懂”化学。

一场游戏中的学术共创

设计题目往往是枯燥的,但这群年轻的北大人却把它变成了一场“游戏”。

为了构建这套高质量评估集,团队搭建了一个专属协作平台。在这里,出题、审题、修题从单调的任务变成了一套循序渐进的“通关”流程。成员们在平台上协作,互相审阅、彼此“找茬”,让严谨的科学探讨与活跃的思维碰撞交织共融。

团队还引入了积分激励系统,让出题过程就像在游戏中打怪升级。一道题目需历经编写初稿、撰写解析,再通过初审与终审的严格审核,每个环节均由不同的同学把关,并发放相应的积分。终审通过的题目,甚至最多迭代过15个版本。

在化学学院的一间间会议室里,灵感的火花不断闪现、碰撞、淬炼,最终沉淀为一套含金量极高的基准测试。

当最强大脑遇上“北大难度”

在这场精心设计的考试中,人类展现出了复杂的科学直觉。作为基线,参与测试的北大化院本科生取得了40.3%的平均准确率。这个数字本身,就足以说明这套题目的“硬核”程度。

而AI的表现如何?即便是接受测试的顶尖模型,其成绩也仅与低年级本科生的平均水平相当。

让团队感到意外的是视觉信息带来的困惑。化学的语言是图形,分子结构、反应机理图蕴含着关键信息。然而对于部分模型而言,当引入图像信息时,其准确率不升反降。这说明,当前的AI在将视觉信息转化为化学语义时,仍存在明显的感知瓶颈。

然而,即使选对了答案,解题步骤也可能经不起推敲。因此,团队为每一道题目都标注了详细的评分规则。在SUPERChem这台“显微镜”下,AI是真懂还是装懂,一目了然。

团队发现,AI的推理链条往往断裂于产物结构预测、反应机理识别以

及构效关系分析等高阶任务。当前的顶尖模型虽然拥有海量的知识储备,但在处理需要严密逻辑和深刻理解的“硬核”化学问题时,仍显得力不从心。

通向理解世界的一小步

SUPERChem的诞生,填补了化学领域多模态深度推理评测的空白。

团队发布这项成果,并非为了证明AI的短板,而是为了推动它走得更远。SUPERChem就像一个路标,它提醒我们:从通用的聊天机器人,到能够理解构效关系、推演反应机理的专业科学助手,中间还有很长的一段路要走。那是从“记住知识”到“理解物理世界”的跨越。

目前,SUPERChem项目已全面开源。团队希望这套源自北大的“试卷”,能成为全球科学与人工智能领域的公共财富,去催化下一次技术的爆发。或许在不久的将来,当我们再次打开这张试卷时,AI能交出一份满分的答卷。那将是化学与人工智能共同的惊喜。(人民日报官微)

六虎同框 野生东北虎母虎带五崽罕见亮相

12月26日,据世界自然基金会(瑞士)北京代表处(WWF)消息,11月,野生东北虎母虎带五只幼崽出现在东北虎豹国家公园珲春保护地,这是全球首次野生东北虎“六虎同框”的影像记录。

根据对个体体态、行走步态及体尺特征的分析判断,成年雌虎年龄约9岁,五只幼崽年龄在6—8个月。野生东北虎通常一胎产1到4只幼崽,即使在猎物丰富的地区,出现五只幼崽的情况也极为罕见。在猎物数量仍在恢复的区域内,五只幼崽能存活到这个年龄亦极其罕见。这段影像为研究老虎繁殖提供了宝贵资料,也标志着中国野生东北虎种群的恢复取得了积极进展。

同样在11月的珲春保护地,巡护员们还记录到另一个母虎带四崽的老虎家庭,成年雌虎年龄在5—8岁,幼崽年龄在6—8个月。“五虎同框”和“六虎同框”属于两个独立的老虎家庭。东北虎作为顶级掠食者,对栖息地的完整性、食物链结构均极为敏感。珲春保护



六虎同框(图片来源:东北虎豹国家公园)

地持续记录到多个老虎家庭以及老虎的健康繁殖,突显了该地区长期保护工作的成功,说明老虎栖息地质量良好,栖息地保护、巡护能力建设、反盗猎管理和人兽冲突管理等保护措施长期有效,生物多样性恢复趋势明显。

WWF全球老虎恢复计划(TAI)负责人表示:“这段非凡的影像说明通过持续的保护努力可以取得良好的保护效果。中国致力于恢复老虎栖息地、减少盗猎、管理人兽冲突、与当地社区合作提升保护成效,保护工

作已取得了实实在在的进展。这些图像也提醒着我们,持久和长期的保护努力对恢复老虎种群和保护栖息地至关重要。”

“六虎同框”的珍贵影像是东北虎豹国家公园范围内记录到的,也进一步说明建设东北虎豹国家公园提升保护能力对于促进东北虎、东北豹野生种群数量稳步增长,维护生态系统的原真性、完整性与连通性更加具有非凡意义。东北虎豹国家公园内现有野生老虎约70只,相比2010年的约20只有了显著增长。

WWF北京代表处副总干事周非表示:“中国记录到全球第一个‘六虎同框’的画面反映了中国的保护行动的有效性,也生动地展示了中国对全球老虎保护和恢复的贡献。多年的保护工作表明,在当前全球自然损失加速的情况下,‘虎啸归山林’为其他濒危物种及其栖息地的恢复和自然保护带来了希望和力量。”(澎湃)