



AI考生和北大学子同考一张试卷 试题由北大多个专业团队联合设计 出题过程就像在游戏中打怪升级 174名北大学生能否考过AI? 结果很意外



▶ [相关新闻]

AI医生“出诊”记

“孩子这半个月天天晚上睡不好，夜里总醒。”近日，一位家长带着一岁多的孩子，满心焦虑地走进北京市房山区燕山医院求助。为精准高效找到病因，接诊的基层医生引导家长借助“AI儿科医生”辅助诊疗。

“孩子每日作息是否规律，夜间入睡、晨起及白天午睡时间是否固定？”“孩子白天活动量与户外活动时长是多久？”……“AI儿科医生”在协助排除病理性问题后，围绕核心症结展开追问，诊疗逻辑清晰。

根据“AI儿科医生”的梳理引导，医生最终确诊孩子为儿童行为性睡眠障碍，并参考AI给出的建议，为家长制定了调整作息、优化喂养习惯等切实可行的干预方案。

早在今年6月，北京北儿窠店儿童医院的一名“AI儿科医生”正式“上岗出诊”，是这一前沿人工智能医疗应用在北京基层医院落地的关键一步。

落地北京北儿窠店儿童医院的“AI儿科医生”基层版，整合了北京儿童医院300余位知名专家的临床智慧与脱敏后的高质量病历数据，构建了覆盖儿童常见病及疑难病的立体化诊疗知识库。

依托自然语言处理技术，AI医生能模拟专业儿科医生的循证思维，进行多轮病情询问，为患儿提供个性化诊疗建议。北京北儿窠店儿童医院专家认为，“AI儿科医生”判断详细全面，给出的建议具体且贴合实际，为患儿后续诊疗提供了有效参考。

北京儿童医院副院长赵成松表示，北京儿童医院后续将在12家北京市儿科医联体成员单位启动“AI儿科医生”推广工作，以国家区域医疗中心、儿科医联体、县级三甲医院、社区医院等多种场景为试点，通过云端部署或本地化部署，以“AI儿科医生+真人医生”的“双医模式”，赋能全国儿科诊疗能力提升。

本组稿件据新华社

在北京大学化学与分子工程学院，期中考前突如其来的一条通知，却让这场考试的气氛变得不同寻常：考场里迎来的一批“特殊考生”。它们不需要落座，也无需纸笔。它们是GPT、Gemini、DeepSeek……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，北大化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。

说考生

一场特殊的期中考试

AI考生和北大学子同考一张试卷

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上随手可得的主题库，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 pre-

diction，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——其中不乏奥运金牌得主——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。

说试题

一场北大人的学术共创 出题过程就像在游戏中打怪升级

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。

说结果

当最强大脑遇上“北大难度” AI成绩仅与低年级本科生相当

考试成绩揭晓。

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进阶任务。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。

▶ [科普一下]

SUPERChem项目

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：从聊天机器人，到专业科学助手，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。

9岁男孩因作业“气晕”妈妈 之后报警求助堪称“模范操作”

医生提醒：呼吸性碱中毒严重时会发生致命的心律失常

近日，云南省昆明市嵩明县公安局接到一名9岁男孩的报警。孩子在电话中焦急地说妈妈呼吸很困难，但家里只有他和几个月大的妹妹，急需警察叔叔的帮助。

报警男孩称：“有个坡下来，下来有两个路口，往最下面那个路口，然后左转就是我家。妈妈的呼吸太急促了。”

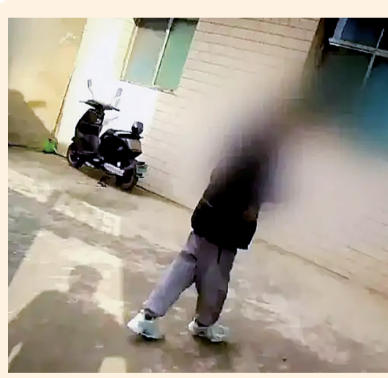
电话中，男孩精准地为民警导航自己家的位置。为了能第一时间找到男孩家，民警和他约定在巷子口的商店碰面，并拨打了120急救电话。

民警到约定地点后，在男孩的带领下，民警来到他家中。民警询问男孩妈妈：“你现在哪里不舒服？不怕，医生马上就到了。小朋友把火炉关了，把电源断了。”

嵩明县公安局杨林派出所民警董开剑回忆：“孩子妈妈倒在沙发边上，呼吸比较困难，不能讲话，手呈鸡爪状，就像抽筋一样。”

因为男孩家中还有一个襁褓中的婴儿，父亲又在外地上班，民警电话联系上男孩的奶奶，请她尽快回家。

民警问男孩：“是不是和爸爸打电话吵架了？”男孩则说出母亲呼吸困难的原因：“不是，她说我作业做得不好。”



男孩给民警带路



男孩母亲被抬上救护车

什么是呼吸性碱中毒？

人们出现生气、焦虑和紧张的情绪时，心率和血压会上升，呼吸也会不由自主地加快，所以当呼吸过快、换气过度的时候，二氧化碳的生成速度低于排出速度，二氧化碳排出较多，就会导致体内的二氧化碳的含量下降，血液中的二氧化碳浓度随之降低。一旦缺少足够的二氧化碳，人体的酸碱平衡就会被打破，就会出现呼吸性碱中毒，也称过度通气综合征、呼吸神经综合征、高通气综合征。

呼吸性碱中毒有哪些症状？

在体内呈碱中毒的情况下，血钙和血钾会下降，脑血管收缩。呼吸性碱中毒的主要表现为手、足、面部特别是口周麻木并有针刺样感觉，胸闷、胸痛、头晕、恐惧，甚至四肢抽搐，包括意识障碍（昏迷）。

呼吸性碱中毒该如何急救？

一旦出现呼吸性碱中毒的症状，首先要帮助病人平静一下心情，然后使用一个纸袋或塑料袋，套住病人的嘴巴，让呼出的二氧化碳重新吸回到体内去增加二氧化碳的吸入。如果没有纸袋，可以用一张报纸，卷成一个喇叭状，大头套着嘴巴，小头向外进行回吸，就是把自己呼出的气体尽量回吸到体内。

非病理性呼吸性碱中毒一般不会引起死亡，可自我缓解；但因为碱中毒可导致患者体内血钾下降，使心脏电生理异常，从而发生致命的心律失常。

医生提醒，对于不能缓解的急性症状，建议马上到医院就诊。

据新民晚报