

“院士班主任”的毕业班会

——中国海洋大学探索培养服务国家重大战略需求的基础学科创新人才

■本报记者 廖洋 通讯员 冯文波

6月7日清晨,中国工程院院士、中国海洋大学教授包振民开车早早来到学校,随车还带了一箱水果。穿着崭新的衬衣,头发梳理得整齐,他高兴地说:“今天是个大日子,衣着容貌一定要得体大方、干净整洁……”

包振民口中的大日子,是他担任班主任的海洋生命学院2020级生物科学(强基计划)班的学生即将本科毕业,今天他要给学生开一场毕业班会。

4年时光,匆匆而过,包振民回忆着与25名同学初次见面、开班会、授课、作讲座以及指导科研的场景,一幕幕仿佛就在昨天。

院士当班主任,既好奇又惊喜

2020年,教育部在部分高校开展基础学科招生改革试点(也称强基计划),主要选拔培养有志于服务国家重大战略需求且综合素质优秀或基础学科拔尖的学生。中国海洋大学作为首批36所试点高校之一,于2020年9月在生物科学(强基计划)专业招收了首批25名同学。

如何办好这个班?课程体系怎么建设?培养方案如何制定?面对“强基班”这一全新的班集体,当时担任院院长的包振民心头始终萦绕着这些问题。

“没有可借鉴的经验,其他学校也在摸索。学院党委研究决定,由我来做第一个班的班主任,体现学院对办好这个班的高度重视。这既有利于快速建立相关教学培养体系,又有利于和学校相关部门协调,推进相关工作落实。”谈起4年前的决策,包振民说,不存在大材小用,只是出于快速建立教学培养体系的目的作出的决策。

“院士做班主任的初衷是,通过院士全程参与本科生培养,与学生共同研究、交流沟通,指导他们做好学业计划和人生规划,在学生科学精神的培养、家国情怀的塑造等方面产生激励作用,树立大格局、大情怀、大志向。”中国海洋大学海洋生命学院党委副书记、副院长张晓燕表示。

“4年前,初人大学,当我得知自己的班主任是贝类遗传学和育种学专家包振民院士时,既好奇又惊喜,也觉得很幸运。”在毕业班会上,2020级生物科学(强基计划)班学生王浩然回忆起初次听说院士当班主任时如是说。

海洋生命学科是中国海洋大学历史最悠久的学科,名师、学院或系领导担任班主任的传统由来已久。作为1978级海洋生物学专业的学生,包振民依然记得自己大学时的班主任分别由系支部副书记刘文浩和系团总支书记庄肃敬担任。“两位班主任都非常认真负责,管理服务到位,工作热心周到,关心学生的方方面面。这种良好的教风、作风我们应该传承下去。”包振民说。

“此次这个‘强基班’在我国高等教育改革历程中和中国海洋大学发展史上都是非常特殊的,算是‘黄埔一期’,他们承载着服务国家重大战略使命,必须保证培养条件的稳定。知道我很忙,学院又请黄晓婷老师协助我,由她来担任班主任。”包振民表示。

“当接到学院领导让我当‘强基班’副班主任的电话时,内心忐忑,很有压力,普通班的班主任我都担心干不好,更别说‘强基班’了。包院士就鼓励我,不用怕,有他在前面挡着。”中国海洋大学海洋生命学院教授、2020级生物科学(强基计划)班副班主任黄晓婷说。

做好班主任是个功夫活,需要时间慢慢磨

“做好班主任是个功夫活,需要时间慢慢磨。”包振民坦言,在40余年的从教生涯中,这是第三次当班主任。面对这群“00后”的孩子,如何与他们相处,他也一度发愁。

为了拉近与学生之间的距离,在繁忙的工作中,包振民会关注“2020生命强基班”微信群里的动态,了解同学们感兴趣的话题,适时参与互动并给予指导;暑假来临,他会叮嘱大家“把安全放在第一位,君子不路险地”;学生获得奖励或荣誉,他会表示“祝贺”,并勉励他们“继续努力”。黄晓婷的电脑里保存着10个版本的强基



2024年6月7日,参加毕业班会的师生合影留念。



►2023年4月26日,包振民与同学们交流。



◀2024年6月19日,包振民与同学们一起拍摄毕业照。中国海洋大学供图

计划人才培养方案,还有关于动态调整、导师制、本研转段等历次研讨的会议记录。“纯粹是摸着石头过河,大到培养方案制定,小到哪一位老师授课,包院士都亲自参与设计或过问。”

黄晓婷说,强基计划采用“本硕博”衔接式培养,本科阶段结束后,如何转入研究生阶段,“保研”“本研连读”“本研贯通培养”都是他们探索过程中用过的词,直到教育部下发通知才确定为“本研转段”。

包振民觉得数学作为基础学科对强基人才

培养十分重要,于是联系数学科学学院选派优秀教师为学生单独授课,并要求把难度提升。英语也是如此,让学生大直接学《大学英语三、四》的课程。专业课大部分采用双语教学。

无论是“开学第一课”“生命科学导航”课,还是“赓续红色基因,点亮蓝色生命”的思政大课,只要有机会,包振民都会亲自登台授课,为同学们答疑解惑,启迪洞见。

“包老师是讲故事的高手,看问题的格局很大、视野宽广,善于把道理讲深、讲透、讲活,潜移默化

默化中给我们人生的启迪和帮助。”2020级生物科学(强基计划)班团支部书记袁兰蕊说。

“三文鱼养殖还缺人吗?咱们能去吗?养了能吃吗?”毕业班会上,有同学围绕包振民团队刚刚突破的三文鱼繁育养殖技术难题发问。

“不仅缺养鱼的人,更缺吃鱼的人。只有吃鱼的人多了,我们的养殖才有希望。消费是拉动产业发展的最重要动力,产业需求又牵动了科技进步。”包振民循循善诱,引领同学们把视野打开,不要局限于找一份谋生的工作,强基班学生肩负着建设蓝色粮仓、破解困扰国家深远海养殖事业发展瓶颈问题的重大使命。

“包老师每年都会给我们开班会,针对同学们在专业学习、培养政策、深造发展等方面的困惑进行耐心解答,既有长辈对孩子般的温情关怀,也有老师为人治学的精神引领。”2020级生物科学(强基计划)班学生鲍李晶晶说。

鲍李晶晶在包振民和黄晓婷的指导下,专注于海洋贝类逆境胁迫适应机制的学习与研究,并于近日,在《国际分子科学》杂志发表了最新研究成果,成为本科生以第一作者身份发表SCI高水平论文的榜样。

大三下学期,需要筹划毕业论文,并选择未来的研究生导师,于是2020级生物科学(强基计划)班学生刘思睿主动找到包振民,希望在他的指导下开展生物信息学方面的研究。“包老师欣然同意,不厌其烦地指导我这种‘科研小白’,帮我选定了扇贝非注释基因功能研究。”最终,他的毕业论文被评为“优秀”,并获得了荣誉学士学位。

“同学们经常自嘲是‘小白鼠’‘草履虫’,在包院士的带领下,我们历经艰辛,蹚出一条新路,基本建立了强基计划人才培养的海大模式。”黄晓婷说。

帮助年轻班主任做好班主任

6月7日,中国海洋大学鱼山校区化学馆130会议室,主题为“我们的成长与未来”的毕业班会热火朝天地举行,醒目的横幅、鲜艳的花束、真挚的话语,现场掌声阵阵,气氛热烈感人。

80%的同学成功转入研究生阶段,多名同学获评山东省优秀学生、青岛市千名优秀大学生、学校优秀学生干部、杰出青年志愿者等省、市、校级荣誉……班会上,一段段视频、一张张照片、一句句嘱托,4年的成长与收获——闪现,不知不觉间师生被泪水模糊了双眼。

“我顶多给自己打60分,很多地方我做得不够,特别是科研任务重,出差时间多,加上学校空间紧张,他们很长一段时间在崂山校区上课、住宿,离我所在的鱼山校区较远,陪伴他们的时间较少。”包振民说,给这群“00后”当班主任感觉自己更年轻、更有活力,他们带来的新思想、新观念的冲击,对自己很有启发。

当被问及“将来还会担任班主任吗”,包振民说:“不会了。我年龄还是大了,60多岁,学生自觉或不自觉地把当成了长辈,甚至是爷爷辈,相互交流不能平等,也很难融入他们的活动中。”所以,做班主任可能不好,做班主任的顾问是可以考虑的选择,帮助年轻班主任做好班主任。

“院士做班主任彰显了‘以本为本’的教育理念,对推动和促进拔尖创新人才培养起到了良好的示范带动作用。海洋生命学院党委书记初建松、党委委员周斌,山东省青年泰山学者王岩目前担任‘强基计划’班班主任,未来将会带动更多高层次人才重视本科教育教学工作。”张晓燕说。

“AI考生”闯关高考,谁是最会做题大模型?

■本报记者 赵广立

2024年全国高考的“硝烟”刚刚散去,“大模型考生”就被抓回来重新“做题”了。

市面上涌现出的大模型产品让人眼花缭乱,围绕“大模型技术哪家强”的讨论不绝于耳,各色名目的大模型评测应运而生。作为国内最权威的考试之一,高考覆盖各类学科及题型,同时在开考前这些题属于“绝密”,非常适合用来作为考查大模型智能水平的评测工具,堪称大模型综合能力的“试金石”。

连日来,一些专业机构纷纷下场,使用市面上常见的大模型产品如通义千问、字节豆包、讯飞星火、文心一言、腾讯元宝、月之暗面 Kimi 等作为“考生”,围绕“大模型高考测试”得出了一系列结果,为人们更好地了解大模型产品的性能和特点提供了参考样本。

AI 高考数学全不及格? 换个打开方式试试

近期,一则“AI 高考测试得分,数学全不及格”的消息登上“热搜”。

消息出自上海人工智能实验室旗下司南评测体系 OpenCompass 对 7 个开源大模型进行的高考语、数、外全卷能力测试。据 OpenCompass 于 6 月 19 日发布的评测结果,大模型的语文、英语考试水平还不错,但数学都不及格,最高分只有 75 分(满分 150 分)。

参加 OpenCompass 此次高考测试的大模型,分别是来自阿里巴巴、零一万物、智谱 AI、上海人工智能实验室、法国 Mistral 的开源模型。OpenCompass 称,因无法确定闭源模型的更新时间,此次评测没有纳入商用闭源模型,仅引入

GPT-4o 作为评测参考。

对于数学测试全部不及格,OpenCompass 表示,“大模型在数学方面还有很大的提升空间”。

不过,复旦大学自然语言处理(NLP)实验室 LLM4EVAL 团队主持的高考数学评测显示,大模型数学成绩不佳的结果,可能缘于“打开方式不对”。

首先,LLM4EVAL 团队选取了 2024 年高考新 I 卷、新 II 卷数学试卷的客观题(单选、多选和填空题,共 73 分)来评测,得出了不同的结论。使用客观题测试大模型的好处是,就是对、错就是错,结果一目了然。具有一定时主观性,如果结果不正确,就很难客观地评出步骤分。

其次,此次大模型“考生”增加到 12 个:阿里巴巴 Qwen2-72b、讯飞星火、GPT-4o、字节豆包、智谱 GLM4-0520、百川智能 Baichuan4、谷歌 Gemini-1.5-Pro、文心一言 4.0、MiniMax 海螺、腾讯元宝、月之暗面 Kimi、DeepSeek-V2-Chat。

另外,他们在评测中发现,数学问题不同格式的提示输入(Prompt)对大模型性能影响很大。在最初的评测中,LLM4EVAL 团队对数学题目中的公式部分采用了通过光学字符识别(OCR)后输出的格式(转义符格式),最新一次评测则使用了 Latex 格式进行了横向对比评测。

结果显示,大多数模型的两次测试结果均出现较大差异,不过使用 Latex 格式后,大模型整体表现更佳:2024 年全国高考新 I 卷、新 II 卷数学测试中,得分率超过 50%的大模型产品数量由此前的 5 个和 6 个升至 7 个和 9 个。考虑到 Latex 格式更符合人类实际使用大模型时所采用的格式,LLM4EVAL 团队建议后续测试主要基于此格式。

具体而言,LLM4EVAL 团队使用 Latex 格式 Prompt 的测试结果显示,在 2024 年全国高考新

I 卷数学测试中,阿里巴巴 Qwen2-72b、讯飞星火的得分率均超过及格线(60%),分别为 78.08% 和 71.23%;在 2024 年全国高考新 II 卷数学测试中,讯飞星火、阿里巴巴 Qwen2-72b 和 GPT-4o 的得分率也超过了及格线,分别为 65.07%、63.70%、62.33%。

由此可见,大模型在数学方面并非“热搜”所说那样完全不及格,讯飞星火、阿里巴巴 Qwen2-72b 等国产大模型在高考数学客观题中具有较高的准确率,令人眼前一亮。当然,LLM4EVAL 团队在评测后也指出,大模型在数学推理任务中的鲁棒性与准确性仍有提升空间。

语文、英语高考测试,作文见真章

“AI 考生”之于语文、英语高考,最引人注目的当数大模型的作文水平了。

对于考生而言,作文考试主要考查学生运用语言成文的能力,考查的是识字情况、用词造句的能力以及表达事实、思想或观点的能力。事实上,作文是最能考验大模型语言理解能力和文本生成能力的测评工具,这两项能力正是时下大模型最为倚重的。

2024 年全国高考语文学科大考试一结束,就有不少场外师生使用市面上的大模型产品“写作文”。围绕新课标 I 卷高考作文题“答案与问题”、新课标 II 卷“抵达未知之境”、北京高考(1)(2)卷的作文题“历久弥新”和“打开”等题目,文心一言、讯飞星火等多家大模型产品纷纷化身“写手”,并交出“作品”。

一些大模型作文令人眼前一亮。以全国新高考 I 卷的作文题为例,在这个具有思辨性的题目引导下,大模型提交的部分作文题不仅切题,更

显巧妙,如《问,岂可少?》《疑问如春芽,答案似剪刀》《于无疑处生疑,方是进矣》《问题不止,智慧无穷》《智涌未来,问海无涯》等等。

近日,全国中小学生作文竞赛评委、中学语文教研专家吕政嘉和河南省基础教育专家库成员李来明共同对市面上 7 款大模型产品的上述 4 张试卷的作文进行了评测打分。从打分情况来看,讯飞星火、文心一言 4.0、腾讯元宝在 4 张试卷的作文题上均有不俗表现,最高平均分接近 50 分。

能拿 50 分的 AI 作文长啥样?讯飞星火作出的《问,岂可少?》得到均分 51.5 的评分。李来明对该文的评语为,“全文结构完整,思路清晰,论证层层递进,结构框架清晰明了。全文多处扣题生发议论,鞭辟入里,分析得当。但在一些地方,可以适当增加一些论证手法,使文章更加生动有趣”。

在高考英文作文题目“帮李华写邮件”中,中国外语教育研究中心特约研究员、知名教研策划专家周国荣和广东国家级示范校教师杨菁菁也对上述 7 款大模型产品的英语作文进行了评测和打分。他们将 2024 年高考真题作文要求输入 7 款大模型产品,生成作文后,由教研双评给出评分并作最高分点评。

全国高考卷的英语应用文写作中,7 款大模型产品均能完成试题规定的写作任务,结构上也能做到逻辑清晰、结构合理,其中不乏能够使用复杂句式,在语言表达上有多处亮点的作品。但这些文章也有一些明显的扣分项,如使用超纲词汇、超过字数上限等。打分方面,7 款产品均有超过 12 分(满分 15 分)的表现,且得分相对稳定。

在难度更高的全国高考英语卷“读后续写”题目和北京卷英语作文中,7 款大模型产品的表现有了差别。周国荣和杨菁菁的打分和点评显示,讯飞星火、腾讯元宝在“读后续写”题目中高

分领先;在北京卷英语作文中,讯飞星火、月之暗面 Kimi、文心一言 4.0 排前三位。综合来看,国产大模型在中国高考的表现不落风,有着教育行业背景的讯飞星火大模型在众多大模型中表现抢眼,堪称“更会做题的大模型”。

评测,还有很长的路要走

评测作为对机器理解、处理、应用自然语言能力的一种评估和量化手段,是大模型领域技术水平和研究进展的直观体现,是相关研究的工具和重要驱动力。

北京大学计算语言学研究所教授臧志方日前在“大模型+计算语言”专题论坛上的报告中表示,大模型在人类标准化考试中如中国高考、公务员考、美国 SAT 考试等的表现,能够为其在真实世界中的能力提供评估参考,但仍存在一些问题。如一些模型在诸如 SAT 数学测试等任务中表现优异,但在复杂推理或特定知识领域中的表现却又不尽如人意。截然相反的表现,让人无从评判。

“在大模型内在机理没有探究清楚的情况下,我们目前的评测路径只能依靠从外部表现来推测内在能力。”臧志方说,现有评测仍存在规范性、系统性及科学性方面的问题,评测的深度和广度方面有待改进。

她提出,未来大模型评测应当以具有综合考查能力的类人机器语言理解能力评测为目标,在参考信度、难度、效度三大原则的基础上,发展更系统的评测大纲,更具挑战性的评测任务,更科学的评测方法,采取更多样、更鲁棒的评测手段,科学高效地为大模型提供客观、公平、类人的评测结果。如此,方能引领和推动人工智能领域各类模型、方法的提出和创新。