

让 AI 帮你“读书写书”

——访施普林格·自然全球图书业务总裁汤恩平

■本报记者 冯丽妃

高尔基说,书籍是人类进步的阶梯;莎士比亚说,书籍是人类知识的总结;列夫·托尔斯泰说,理想的书籍是智慧的钥匙……有没有想过,有一天你能根据阅读需要让一本书的内容随意“加长或缩短”?或者,如果你是一位作者,有没有想过让机器替你干活,助你写得更快、更好?

在 Niels Peter Thomas(汤恩平)看来,“人工智能(AI)时代,一切皆有可能”。近日,在其来华访问之际,《中国科学报》采访了这位施普林格·自然集团全球图书业务总裁。

《中国科学报》:今年4月,施普林格·自然出版了第一本由机器生成的书籍《锂离子电池》。请谈谈这本书的出版背景和读者反馈。

汤恩平:实际上,这是地球上第一本 AI 学术书籍,对我们来说也是第一次。它是大约两年前开始的一个项目。我和一个同事知道很多初创公司声称可以通过深度学习等技术处理自然语言,这让我们有些担心。如果某家科技巨头开始用 AI 出版图书,大概不太会关心版权、学术严谨性和同行评价,而只关心花哨的技术。我们只有比他们更快,才有机会解决这些出版方面的问题,所以我们秘密而不宣地做了这件事。

我们和德国法兰克福大学的研究人员合作,开发了一款叫作“Beta Writer”的先进算法,从施普林格·自然的内容平台上挑选、使用和处理锂电池领域的相关出版内容。一年半后,我们做到了。这本书只有序言是由人写的,目录以后的其他部分都是纯机器写的,我们没有改变一个字母和符号,所以书里有些句子可能有语法上的错误。但我们认为向世界展示这个算法能做什么更重要,所以整个过程是严格和透明的。

《中国科学报》:AI 书籍涉及的版权问题怎么办?如何避免 AI“抄袭”?

汤恩平:这本书是免费的,因此不涉及版权问题。版权非常复杂,各国的情况不尽相同。我们已找到了一种展示和出版的方法来应对这个问题。其实更有趣的是书里引用了一些内容,我们会像对待任何其他图书一样确保其学术严谨性,所以你可以看到各种参考文献、数字对象识别号以及很多脚注。我们确保了这本书所引用的作者原始观点都已注明出处。

因此这本书里没有抄袭,但要确保算法



汤恩平

不抄袭非常复杂。我们需要让算法阅读所有其他材料,然后得出新的结论,并以某种新的形式呈现。这本书当然是基于它读过的成千上万篇文章完成的,这就像研究者在写新书时会阅读大量材料,然后得出自己的结论,人类写的书籍总是基于其他人的作品,机器也是一样。但要教会机器遵守所有规则而不侵犯版权,这非常困难。但我认为我们已经成功地把它作为一个试验展示给科学界,就此进行讨论。目前还没有作者抱怨说他们不想在机器写的书籍中被提及。

《中国科学报》:你如何看待 AI 书籍的前景?

汤恩平:第一本 AI 书籍出版后,目前我们正在其他几个学科领域进行尝试。但我们没有计划在 1000 个主题上制作人工书籍,尽管这很容易做到,写这样一本书只需要半个小时。尽管我们花了两年时间创建算法,一旦有了算法,就可以把它用在其他各种主题上。

我确信 AI 会为图书出版带来有益贡献。当人类作者要写一本书时,我们可以生成一些文稿供其使用,帮助人类作者写得更快、更好。想象一下,如果你要写一本书,就必须阅读大量的材料,这样才能总结当前的情况并找出自己研究的重点,这实际上可以通过算法来实现。算法能帮你找到书中可引用的内容,总结已经出版的内容。未来,技术能为所

有作者带来帮助,让他们能更快地写书,由此促进科研过程。

今后它还可以帮助我们在每一本书的每个章节前创建一个概述,方便人们快速地浏览,而不再需要作者去完成。有一天,对于每一本书,你只需说出所需阅读时间,就会得到一个精确匹配的章节,算法可以缩短或加长它。或者,有读者想了解一本书,但对化学等知识知之甚少,我们可以用其他内容让一个句子变长以做出解释。所以,虽然它最初是为研究人员设计的,但也会让读者更轻松阅读一本书。

《中国科学报》:这本 AI 书同时也是一本开放获取(OA)的书籍。请问施普林格·自然目前的 OA 图书出版现状和趋势如何?与 OA 期刊有什么不同?

汤恩平:我们第一本真正的 OA 书籍是在 2010 年出版的,现在总数近 1000 本。施普林格·自然每年出版新书约 1.3 万册,OA 图书目前比例仍较低,但重要的是其增长率非常高,目前年增长率为 40%。实际上,我们会首先决定是否出版一本书,然后再讨论它的商业模式,其中质量是唯一的决定因素。今年 OA 图书的增长率还可能更高。

这些 OA 图书涵盖了不同的学科领域。一个有趣和明显的趋势是,OA 期刊,我们看到的更多是生命科学和技术领域的;而 OA 书籍,则最多的是人文、社科领域,当然也有自然科学、工程、计算机、医学等学科。

OA 图书与 OA 期刊的最大差异其实就是图书与期刊之间的差异。学术期刊某种程度上更同质化,而学术图书更多样化,它有各种完全不同的形式,如教科书、专著、会议论文集、手册等。OA 图书并没有绿色开放路径。在互联网上免费上传 PDF 格式的手稿并不是开放获取书籍,真正的 OA 书籍出版涉及上百个环节,是一个非常复杂的过程,需要元数据、质量控制、图书管理员的标记记录等等。

《中国科学报》:OA 图书的业务模式是怎样的,由谁来支付出版费用?

汤恩平:OA 图书与 OA 期刊类似,需要收取文章处理费用(APC),这通常很少由作者来支付,多数情况下由科研资助机构或作者在大学从研究经费中支出,偶尔我们也会协助作者去寻找资助机构。需要说明的是,对于我们出版的非 OA

形式的传统书籍来说,作者不需要支付出版费用。相反,作者会得到版税,分享图书销售所得。之所以这样做是因为我们相信,如果一本书的质量好,我们就能通过国际市场销售,而不是从作者那里获得收入。这就是为什么我们只注重质量。

《中国科学报》:你在选择一本好书时有什么标准?

汤恩平:因为书籍类型是多样的,这意味着选择好书的标准会有所不同。拿专业书籍或专著来说,我们接受一本书的唯一原因是质量,但问题是质量作为一个概念并不容易衡量和定义。我们从图书出版史可以看到,有些书开始被忽视,10 年后才明白它实际上是一种里程碑式的成就。比如,日裔美籍物理学家中村修二曾在我们这里出版了一本书,并由此获得了诺贝尔物理学奖。这本书的使用和引用量都很少,它是如此高深,我们也不知道出版后有多少人能理解它,所以决策起来不轻松。我们要考虑“它是不是没有错误?是不是可靠的科学研究?有什么新的贡献吗?”这些是我们决策的标准。

《中国科学报》:施普林格·自然出版的由中国作者撰写的图书有什么样的趋势?

汤恩平:增长很快,需求也很大。我从 2011 年到 2014 年在北京工作过 3 年,当时,我们办公室人员还不到现在的 1/3,近 10 年间我们增加了很多本地员工。中国作者数量增长也很快,可谓“爆炸式”增长。当年我们出版来自中国作者的图书不足 100 本,如今每年已达 800 本左右。我们多数作者都曾以中文出版过图书或发表过论文,但我认为跨国的科研和学术交流非常重要,这有益于世界,也有益于中国研究人员,因为他们可以通过自己的书与国际读者进行更广泛的讨论。

目前,中国的研究在国际出版界并未得到充分展示。其中一个原因是语言障碍,另外一个原因是中国非常大,自己有较好的内部对话交流。若以欧洲或亚洲的小国家为例,有些领域只有很少的研究人员,他们必须走向国际,因为在本国没有人可以相互讨论。中国有这么多研究人员,有自己的学术会议,本土科研交流也很强,所以对他们来说压力没那么大。但达到一定程度,就有必要与国际同行进行交流,这会非常有益。

科学线人

全球科技政策新闻与解析

美限制警察搜索家谱 DNA 数据库



约瑟夫·德安基洛曾被指控为美国加州杀手,警方利用家谱数据库确定其身份。图片来源:NEWSCOM

美国司法部近日发布了新规定,对警察使用基因族谱学手段追踪犯罪嫌疑人的行为进行了规定,此前对业余人士使用家谱学数据的行为并无相应规定,此举旨在执法过程中平衡公共安全与隐私问题。

2018 年,约瑟夫·德安基洛被指控为加州杀手——涉嫌一系列强奸和谋杀案,执法人员将犯罪现场的 DNA 上传到公共基因族谱网站上,并使用传统族谱和其他手段缩小侦查范围,从而找到了犯罪嫌疑人。这一方法已使至少 60 起悬案中嫌疑人的身份被锁定。

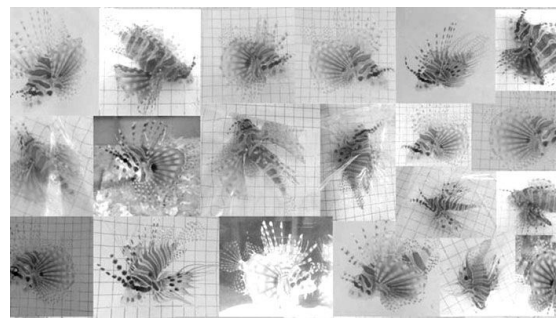
但这些搜索行为也引起了人们对隐私的担忧。即便不上传自己的 DNA 信息到数据库中,也可能成为警方的怀疑对象,而那些已经上传过数据的人,可能未经知情同意,其上传数据就被用于执法搜索。

美国司法部颁发的新政策将于 11 月 1 日生效,在其公开的新闻稿中提到,新规旨在“平衡国防部应对暴力犯罪和保护公共安全的不懈承诺及同等重要的公共利益之间的关系”,新规定指出,“法医基因族谱”通常仅用于谋杀和强奸等暴力犯罪,以及识别人类遗体(如族谱数据库政策允许此类搜索,则该政策也允许更广泛的使用)。警方应先用传统犯罪侦查方法,包括搜索原有的犯罪 DNA 数据库。

新规定还提到警察不能“钓鱼执法”:在没有被发现的情况下私自将假的个人资料上传到谱系网站,以找到犯罪嫌疑人的远亲。这些数据库网站也必须通知用户,执法机构可能会搜索这些数据。

新规禁止警察利用犯罪嫌疑人的 DNA 资料寻找与疾病风险或心理特征有关的基因,也禁止警察秘密地从犯罪嫌疑人的亲属处获取 DNA 样本,如从丢弃的杯子或人体组织中提取样本。除非获得搜查令,警察必须征得对方知情同意。(任芳言)

瑞典科学家涉嫌学术不端



发表在《生物学快报》的袁焘拼贴画。图片来源:《生物学快报》

近日,调查发现,2014 年瑞典乌普萨拉大学前海洋生物学家 Oona Lonnstedt 发表在《生物学快报》上的袁焘行为论文涉嫌造假,2016 年发表在《科学》的论文也可能存在造假行为。

去年,Lonnstedt 及其合著者在《生物学快报》网站上发表了一系列拼贴画,以此回复针对他们此前是否真的捕获了足够的袁焘以用于实验的质疑。但是批评者表示,这本五光十色的影像似乎只是同一条鱼的不同照片,甚至有的只是同一张照片精心篡改后的复制品。

2012 年,还在澳大利亚詹姆斯库克大学读书的 Lonnstedt 完成了有关袁焘行为的论文。但该研究遭受质疑——科学家怀疑 Lonnstedt 是否如实收集到了其声称数量的鱼群,并且如何在尚未录像的情况下记录海量的行为数据。

2017 年,乌普萨拉大学和瑞典一个国家级道德委员会共同敲定了大众对 Lonnstedt 同 Peter Eklov 合著的《科学》论文的质疑。该论文被撤稿,Lonnstedt 虽未被定罪,但失去了工作。

一位不愿具名的知情人透露,他在 Lonnstedt 的那篇《科学》论文发表前就向《生物学快报》提出了关于文章数据真实性的质疑。当《科学》论文被撤稿后,他又表示了对《生物学快报》论文也属造假的担忧。Lonnstedt 所报告的实验需要 86 条花斑斑鳍袁焘和 16 条触角袁焘。这些鱼群采集于澳大利亚大堡礁,在那里,所有研究人员都被要求登记捕鱼数量。该知情者查阅了在线记录后发现 Lonnstedt 只捕获了 12 条花斑斑鳍袁焘和 3 条触角袁焘。

2018 年 2 月,《生物学快报》表达了他们对该论文的担忧,并表示正在进行调查。2018 年 11 月,论文作者发表声明,一改此前说法,声称他们在研究中将鱼群所需数量降低到 40 条花斑斑鳍袁焘和 9 条触角袁焘。声明表示,为了提供“袁焘数量的证据”,作者在补充材料中提供了 50 幅鱼群图像的拼贴画。

不过科学家发现,拼贴画中同一张照片被用了两次。其中一张是经过镜像处理后的复制粘贴,对比度也有所调整。除此之外,不同的照片似乎拍摄的是同一条鱼。部分科学家重构了该组照片的时序,发现多张照片似乎是连续拍摄同一条鱼后生成的。

在《科学》丑闻发生后,詹姆斯库克大学承诺对事件进行调查。但是该大学发言人表示,具体调查过程还没有展开。向《生物学快报》投诉的知情人表示,期刊应立即将该论文撤稿。他认为期刊和大学对造假指控不进行严肃处理是“不道德的”,“这非常艰辛和沮丧”。(程唯伽)

科学七日

研究

接近中微子质量极限

德国物理学家对中微子的最大质量进行了迄今为止最精确的测量。近日发表的卡尔斯鲁厄中微子(KATRIN)实验的第一个结果显示,中微子的重量最多为 1.1 电子伏特(eV),而之前的测量上限为 2 eV。中微子是宇宙中最丰富的粒子之一,但它们往往会在未被探测到的情况下穿过物质,因此很难直接测量。KATRIN 实验监测了“重”氢同位素——氘的核衰变。在这个过程中,中子变成质子,并释放出电子和中微子,实验间接测量了看不见的中微子的能量范围,从而揭示了它们的质量。

政策

研究“安全”

数月来,美国政府因所谓的安全漏洞而不公平地排斥外国出生的研究人员,此举引发了公众的强烈抗议。美国白宫科学顾问 Kelvin Droegemeier 在近日的一封信中说,美国国家科学技术委员会正在努力制定政府层面的要求,包括研究人员公开哪些信息才能获得联邦研究资助。他写道,5 月份由国家科学技术委员会组成的小组正在收集“研究型企业受益或利用或损害的一些例子”。Droegemeier 表示,现在公开联邦机构调查的人数还为时过早。

流感疫苗

美国政府正在启动一项使流感疫苗开发现代化的计划。美国总统于 9 月 19 日签署了一项行政命令,指示卫生与公众服务部和国防部在 120 天内提交一项计划和预算。其目标是提高该国为未来流感大流行做好准备的能力,并开发出更好的疫苗以预防季节性流感暴发。该项目将支持研发一种通用流感疫苗,以取代每年开发的季节性疫苗。此外,政府还将制定一项战略,研究更快生产季节性流感疫苗的方法。目前,大多数流感疫苗都是用鸡蛋生产的,这个过程至少需要 6 个月。

环境

重视自然

BBC 电视主持人 David Attenborough 将



▲ KATRIN 实验监测氘同位素——氘的核衰变。图片来源:Michael Zacher



► 2011 年 3 月大地震和海啸后的福岛第一核电站。图片来源:DigitalGlobe/Getty

成为英国政府委托进行的生物多样性经济价值评估的公众代表。该评估报告将由英国剑桥大学经济学家 Partha Dasgupta 担任主席,以研究生态系统服务(如授粉和森林固碳)对全球经济规模和增长的贡献。据估计,自然环境每年可带来价值数万亿美元的利润,审查报告希望能准确量化这一数字,以迫使各国在允许开发绿色空间之前三思而后行。审查小组预计将在明年 10 月于中国举行的《联合国生物多样性公约》缔约方大会上发表报告。

能源

核灾害

9 月 19 日,东京一家法院裁定,3 名电力公司高管与 2011 年福岛核灾难有关的刑事过失罪名不成立。Tsunehisa Katsumata, Sakae Muto 和 Ichiro Takekuro 曾在东京电力公司工作,负责运营福岛第一核电站。2011 年 3 月 11 日,日本东北部发生里氏 9 级地震,随后发生海啸,福岛第一核电站的 6 个反应堆中有 3 个发生了熔毁。这场灾难导致大约 15 万人撤

科研诚信

大脑研究

德国科学基金会(DFG)日前发现,图宾根大学神经科学家 Niels Birbaumer 在研究中有科学不端行为,他声称已经开发出一种能够读懂瘫痪患者想法的大脑监控技术。调查发现,Birbaumer 的两篇论文中数据不完整,科学分析存在缺陷。9 月 19 日,DFG 宣布将禁止 Birbaumer 在 5 年内申请资助,并建议撤回这两篇论文。“DFG 发现了我的科学不端行为,并对我实施了制裁。我必须承认,我无法反驳针对我的指控。”Birbaumer 表示他会坚持自己的研究。

奖励撤销

西班牙奥维耶多大学分子生物学家 Carlos López-Otin 获得的 2017 年杰出导师奖将被《自然》杂志取消。2017 年《自然》杰出导师奖宣布后,《自然》杂志收到了相关报道,报道指出,Carlos López-Otin 联合撰写的一篇文章评估发现了图像完整性问题。撤销该奖项反映了评委们的观点,即实验室的高质量指导不应导致如此重大问题。

趋势观察

研究人员报告说,随着肉类产量的增加,印度等国家畜用常用抗菌药物的耐药性越来越强。9 月 19 日发表在《科学》上的一项关于亚洲、非洲和南美洲牲畜抗菌素耐药性的研究,肯尼亚、乌拉圭和巴西也出现了耐药性热点。

2000 年以来,由于养殖方式更为集约化,包括在动物身上使用抗生素,这些地区的肉类产量大幅上升。研究人员分析了在发展中国家进行的 901 项流行病学研究,这些研究关注 4 种常见细菌——沙门氏菌、弯曲杆菌、葡萄球菌和大肠杆菌。他们利用这些信息绘制了多药耐药性存在和开始出现的地区。研究结果表明,农场动物中最常用的增重抗菌素类药物的耐药率非常高。从 2000 年到 2018 年,细菌抗药性的比例在鸡和猪身上几乎增加了两倍,在牛身上增加了 1 倍。(辛雨)