

# 50年前的一个创新“公案”

■黄四民

2019年,NASA(美国国家航空航天局)总部的总工程师扎卡里·皮尔特勒博士与合作者发表了一篇题为《创新从何而来》的文章。这篇文章已经被人遗忘了50多年、著名的创新来源案例研究重新回到人们视线中,即由美国国防部内部开展的被称为Project Hindsight(后知项目)的项目。

该项目早在1963年8月就已经启动,其主要目的是针对二战结束至1962年期间开发的20种主要武器系统,衡量其科技投入的成本效益并提出改善科研管理方法的建议。

项目由国防部国防研究与工程办公室副主任谢尔温·查尔默斯推动,项目规划方面由经验丰富的工程师雷蒙德·伊森森主要负责。每种武器系统的研究团队大约5人,整个项目成员既有军方内部工程师,也包括来自西北大学和MIT(麻省理工学院)的管理领域学者。项目的中期报告于1966年6月发布,最终版于1969年10月发布,巧的是这一年7月NASA的阿波罗11号登月舱在月球表面成功着陆。

在艾森豪威尔总统和肯尼迪总统的时代,政府的研发财政支出年增长率远超过GDP增长率。在艾森豪威尔执政时期,研发支出平均每年增长15.1%,在肯尼迪执政时期为16.6%。而到了约翰逊执政时期(始于1963年11月)仅为3%,甚至低于生活成本的上升。可以想象美国国防部需要“学会过紧日子”,这应该就是Project Hindsight的大背景。

该项目从中期报告到最终报告历时三年多的时间,原因是该项目中期报告的结论掀起了一场激烈争论。其中期报告的结论是,基础科学和应用科学对创新的贡献甚微。

具体来讲,研究团队对每种武器系统进行分解并识别其创新,向后追溯20年时间,确定其每一创新的“关键”研究和探索性开发(简称RXD)事件,并将RXD事件分为科学、技术和工程三大类。在所有确定的710件RXD事件中,9%的

贡献归于科学类事件,91%归于技术类事件。而上述9%的科学类事件中,6.7%为国防需求驱动,归为应用科学;2%为商业(非国防)需求驱动,归为应用科学;只有0.3%为非直接指向的科学。

这一结论对在艾森豪威尔和肯尼迪时代“大手大脚”花钱的科技界无疑是一声惊雷,尤其是在转向“过紧日子”的关键时期。更具争议的是该结论挑战了当时人们对创新源头的认知,即二战结束以来为科技界奉为圭臬的万尼瓦尔·布什创新性线性模型。

于是,上世纪60年代中后期,美国政府、学界和工业界针对基础科学对创新的影响、科学与技术的关系以及创新流程等深层问题开展了一系列大规模调查,其中包括美国国会、国家研究理事会、国家科学基金会、空军科学研究办公室等资助与监管机构,也包括IEEE(电气和电子工程师协会)、哈佛大学、MIT、斯坦福大学、伊利诺伊理工大学研究院以及西电公司、陶氏化学公司、GE(通用电气公司)等学术机构和行业龙头企业。

同期,英国也开展了类似的大规模调查研究。针对美国上述调查的不足之处,英国的调查扩大了规模并兼顾了创新的成功与失败案例。例如英国科学政策委员会支持的WFK(Wealth From Knowledge)小组针对1966年和1967年共84项英国企业女王荣获奖项开展技术创新案例分析,英国科学政策委员会设立SAPPHO项目,对比分析两期共72对成功创新与非成功创新案例。正是在这72年里,许多西方国家迈出了对科学进行规划的第一步。

在Project Hindsight引发的关于创新源头的大调查和大辩论中,美国科学界与国防部和国家有着截然不同的态度。科学界的最初反应自然是质疑,其中最强烈的批评来自美国国家科学基金会(NSF)。

1967年,NSF委托伊利诺伊理工大学研究院开展了TRACES(Technology in Retrospect

And Critical Events in Science)项目,对Project Hindsight的中期报告结论进行反击,得出了相反的结论,但是得出结论的方式遭受质疑。

其后,NSF委托Battelle(巴特利)实验室进行调查,其结论却介于Project Hindsight和TRACES之间。

总而言之,科学界不喜欢Project Hindsight的结论。截至2019年底,查尔默斯和伊森森1967年在《科学》上发表的项目总结文章被引用了270多次,而批评其结论的一篇文章被引用了近1800次。

尽管迫于压力,1969年10月发布的Project Hindsight最终报告修改了措辞,但是美国国会于1969年11月通过了由参议员迈克·曼斯菲尔德提出的曼斯菲尔德修正案,要求国防部仅对那些“直接与具体军队职能或行动有明显关系”的项目提供支持。

上世纪70年代中期,美国国防部高级研究计划局(DARPA)局长乔治·海耳迈耶启用“海耳迈耶问题”决定研究经费的分配,并一直为DARPA沿用。

2004年,美国陆军小范围再次开展Project Hindsight调查,覆盖了四个陆军武器系统;艾布拉姆斯主战坦克、阿帕奇攻击直升机、毒刺防空导弹与标枪反坦克导弹。

项目由曾任陆军研究实验室主任和国家标准与技术研究院院长的约翰·莱昂斯教授、曾任陆军装备司令部首席科学家及陆军研究和实验室管理部主任的理查德·夏泰教授等负责。其主要结论与1969年10月发布的Project Hindsight最终报告的结论大致相同。

但是这一次的调查没有像上世纪60年代那样成为各利益相关者的关注焦点,因为经过40年的时间,人们已经在什么是基础研究、什么是创新以及它们之间的关系上有了更深入的认识。

<http://blog.sciencenet.cn/u/Siminhuang>

## 本科·科研入门

中学物理中我们便学习过压强的基本概念及其国际单位帕斯卡。在流体压强性质的认识过程中,科学家帕斯卡做出了杰出的贡献,继承并进一步发展了托里拆利等人的工作,得到了今天物理学课本中所熟知的不同高度下压强的计算公式。

1971年,在国际计量大会上,科学家们通过投票,将帕斯卡规定为压强的国际单位,即1个标准大气压等于101325帕斯卡。不仅如此,帕斯卡还对科学研究方法本身做了很多深入的思考,他的一些思想被后续学者逐渐发展为检验一个论断是否是一个科学论断的核心标准,即“可证伪性”。

有关压强的研究历程对于本科生了解科学研究的详细过程很有帮助,“可证伪性”的概念也让本科生更为明确地区分科学和非科学。

1644年,托里拆利实验已经明确证实了真空的存在。当时的人们已经认识到了大气压的一些基本性质,但是,关于压强的诸多问题仍然有待解决。例如,在不同的测量高度下,流体中的压强会发生怎样的变化?压强又跟哪些因素有密切关系?

1647年,帕斯卡接触到了托里拆利采用水银做的真空存在性验证实验,并进一步地获取了托里拆利相关实验的详细信息和主要结论。托里拆利论文中的结论强烈地激发了帕斯卡的研究兴趣,并将这种兴趣贯穿其一生之中。

首先,帕斯卡重复了托里拆利的实验,搭建了大气压测量的相关实验台,基本熟悉了大气压实验的操作要点。值得一提的是,在本科生开展科研的过程中,尤其是在科研的初期阶段,本科生也可尝试性地围绕近期发表的高水平成果进行跟进,这个经历可以让本科生对相关实验平台和原理较为熟悉,为后续的创新工作打下一定的基础。

其次,帕斯卡更换了气压计中的液体并进行了多次实验,发现了一些规律。1647年,帕斯卡在一篇论文中更为系统地讲述了空气压强对于各类液体所能支撑的高度以及气压计中真空的本质。1648年,帕斯卡又发表了另一篇论文讲述流体静力学相关理论。这些初期的科研工作为帕斯卡对于压强的性质更加了解。

这也提醒了本科生,对于一些好的想法,要敢于尝试,否则它们就只会一直停留在想法的阶段。

后续,通过不断地思考,帕斯卡意识到真正决定压强大小的另外一个因素是测量大气压时所在位置的高度。他对这个想法坚信不疑。

帕斯卡居住地附近的法国多姆山高度为1460米左右,是一个很好的测试地点。遗憾的是,帕斯卡因身体虚弱无法爬山,他便拜托他的姐夫代劳。在帕斯卡连续几个月不厌其烦的劝说下,他的姐夫皮埃尔于1648年9月19日携带着水银气压计与其他两个朋友一起攀登了多姆山。

结果是令人鼓舞的。在山脚下,他们用水

银气压计测量了大气压力。到了山顶,当他们再一次进行大气压的测量时,发现水银气压计的刻度线显著地下降了,表明山顶处的大气压明显地低于山脚下的大气压。该实验重复了五次之后,结果依然如此。估计皮埃尔没有想到一次爬山便可以验证在后来几百年的几乎所有物理学课本中都会提到的公式——否则他便不会耽搁几个月的功夫了。

帕斯卡仍不满足,他跑到巴黎的圣雅各伯屠羊场教堂的高约50米的钟楼继续测试,结果显示压强也降低了些许。通过上述实验验证工作,帕斯卡最终确认影响大气压强大小另一个核心因素是测量位置所处的高度。

实际上,帕斯卡的上述科学发现具有广泛的影响力,深刻地揭示了压强的本质特征。浮力原理最早是由阿基米德在公元前3世纪发现的,并被广泛应用。但是,浮力原理的数学证明却需要引入帕斯卡的理论才能够完备。另外,日常生活中使用的咖啡壶以及青藏高原考察所使用的浮空艇均与帕斯卡的贡献有关。

虽然关于真空和压强性质的科学实验已经大体完成了,但真正与已有理论的论战才刚刚开始。从这场论战中,可以看到一项伟大的科学学说的建立所可能遇到的阻力。

论战的另一方是亚里士多德学派。在上述论战的过程中,帕斯卡逐渐明晰了科学论断所应具有的特征,即“可证伪性”。

亚里士多德学派认为,在处于真空状态的管道之中,存在着某种“不可观测”的物质。但这个论断是不可被证伪的。如果观测到某种物质,说明真空不存在。如果没有看到物质,便可以用存在某种“不可观测”的物质进行说明。因此,帕斯卡断言亚里士多德学派的论断并不科学。

帕斯卡的一些思想被后续哲学家所继承并发扬,形成了证伪主义。证伪主义突出了经验在检验科学理论方面的重要作用,其虽然是论战的对象,但可以为判断科学理论的错误性提供证据。上述思想在科学哲学领域占据着重要地位。

事实上,在开展科学研究以前,科学方法的积累对于本科生的长期发展和科学素质提升都是很有必要的。

<http://blog.sciencenet.cn/u/upflyzhang>

## 帕斯卡的实验与证伪主义

■张宇宁

## “变异”的紫薇花

■陆仲绩

台风过境,刮了一夜的风,下了一夜的雨,还没有停下来。

早上起来,看窗外那盆新插的紫薇花,细长枝条垂垂在别的花盆缝隙处,花朵挂着雨滴,湿淋淋地依旧绽放,花蕾还挂在花朵下,柔柔弱弱得惹人怜。

新插的紫薇花,母株来自老家的庭院。夏日骄阳下,紫薇花的盛花期,推开门就能看到一簇簇的花球,每根枝条前都顶着金黄色细米粒的花蕊,在微风中摇曳。江南乡间一小屋,堂前绿色丛中一团火似的艳红,真是蓬华也生辉。

久居都市,唯独这盆栽的紫薇常伴身边。起初是阿爸在调理,常开常盛,煞是喜人。自从由我接手后,总觉得原来的繁茂可言。给予最好的位置最细心的养护,总有些不如心意,于是暗下决心,要让其重振辉煌。

去年夏末秋初期间,腾出一只较大的泥盆作为其新家,盆底翻新,底部铺上欲腐烂的部分作为肥料,剪老株的粗状枝条,分段插入,浇足水,上面用暗黑色塑料纸遮阴,冬天来临时,用透明塑料袋罩住保暖。那段时候,还不时擦开去看看,浇浇水。春风吹来,看枝条处已经萌发出细细绿芽,该是再换盆的时候了。花与人一样,不能太挤,住房面积大些,日子过得宽敞些,灿烂窗台、烂漫景色,只待花开时节。

可能是老盆的枝条被修剪得有点过分,叶子



密密匝匝、郁郁葱葱,就是不见花蕾。老盆里的花尚未开,新的却已经悄然绽放,虽然稀少,只有二匝,但也按捺不住高兴,拍了照片。远在北京的四叔来信了:老家的紫薇花是纯红的,你这扦插而成的竟然镶上银边了,变异了,应该说更漂亮了,好好培育。

果不其然,变异了。原先的纯红,犹如一团火,热烈,激情四溢,就像是远方的召唤,永远是游子心中激荡着的一种牵挂。如今,远离故土,又没学得阿爸的家传手艺,纯真的鲜红却镶嵌了银色的思念,娇艳、可人。人有时候也是蛮奇怪的,要做一件事,事情做了,却“变异”出另一个结

果。“花解怜人花亦愁,隔帘消息风空透”,变了,人到了这个年龄,也是这样变的吗?变得多了一份淡淡的幽思。

风,有时会搅起人一点点起伏的思绪,而雨,则又引得人一点点对先人幽幽的思念。

江南,每年都有一场又一场的台风,每到那个时候,不管我们在哪里,都会牵挂东海边这座小小的院落。枝叶摇曳“晴方好”,庭院氤氲“雨亦奇”。那时候,雨打枝头,堂前紫薇花瓣撒落一地,留在枝条上的也蹒跚着,牵拉下来,有的挂在残损的围墙上淋湿后的紫薇花,依然浓郁、奔放。

门外布满青苔的石板,偶尔会发出“砰砰”的声响,那是过路人在跳起来摘花了。阿娘(宁波人对祖母的称呼)急得在屋里大声叫唤:“不要把我的墙给攀倒了!”下雨天,陈年的老墙更容易被推倒。随即传来一阵阵年轻人朗朗的笑声和说话声,渐渐远去,窄窄的巷子又恢复了宁静。

窗外的雨水沿着屋檐滴滴地下,大风卷起细嫩的枝条呼呼地吹,屋内阿娘捧着茶壶坐在藤椅上眯着眼睛……朗朗的笑声也会感染人,随着年轻人的远去,屋里也发出一阵“如释重负”的笑声。漂亮的花总是惹人喜爱的,越多人喜爱自己种的花就越有成就感,只是“不要把我的墙给攀倒了”……

泥盆藏乾坤,紫荆有渊源;故乡由此远,天涯寄余生。

<http://blog.sciencenet.cn/u/lzj6189>

## 它们生活在北极

■高登义



▲北极兔

海豹憨厚可爱,但近数十年,“见北极熊容易,见北极海豹难”。何故?人类之中的贪婪者大量捕杀海豹牟利。在此情况下,虽然我去过北极19次,但拍摄到清晰的海豹全身照片只有两次,此张是外国朋友赠送。



海象,喜群居,有时下水游泳也是集体行动。海象最美的是它们游泳时浮在水上的头部或面部。偶尔,海象也在地面翻跟头玩耍。

<http://blog.sciencenet.cn/u/dengyigao>

## 从杭州失踪案谈人工智能的短板

■张军平

最近杭州的失踪案在经过十几天网民各种“福尔摩斯式”的推测后,终于被破破了。毋庸置疑,大数据对案件的侦破起到了关键性的作用。通过对小区和邻近路段的监控视频的分析,最终推断出来某某没有出过小区,然后将嫌疑人锁定。

遗憾的是,从公开的报道来看,人工智能似乎并没有发挥其应有的能力。公安人员在分析视频时,是通过人工检查,24小时的视频要看一天甚至更长时间,以免漏掉潜在的目标。一如在世界人工智能大会上,复旦大学附属华山医院感染科主任张文讲“新冠疫情在中国得到成功的控制,也不是得益于人工智能,而是传统医疗和有效的城市管理”。

为什么人工智能在此次案件的侦破中没有发挥应有的作用呢?按道理来说,利用人工智能的快速检索能力,通过对视频监控数据进行摘要提取,可以很快地把嫌疑犯找出来。更何况,人工智能在海量搜索和计算速度这两方面都有人无法比拟的优势。

然而,事实并非如此。

其中的原因可以从判断人工智能算法优劣的两个指标说起,即误检率和漏检率(常称为召

回率)。误检率是指本来没有发生的,如未发生的交通违法行为,没有感染新冠病毒的人被误判为新冠肺炎患者。而漏检率是指发生了却未被检出的,就是说明有嫌疑人出现了却未检出,有新冠肺炎症状却没被检出。

误检率相对来说损失小一些,但仍然会浪费时间和增加成本。比如在交通违法行为上,疑似违法者需要花时间去与交通管理部门进行交涉后方能取消。在刑事案件中,可能导致公安部门不必要的出警,极端情况会产生“狼来了”的不利影响。而在医疗诊断中,对于心理素质不强的人来说,误诊为癌症对身心的打击会很强烈。所以,基于人工智能的模型应该尽量避免这个指标过高。

而漏检率的危害就大多了。举例来说,在视频监控的快速检索中,如果把出现嫌疑人的视频漏掉了,那么有可能会让整个案件无法形成闭环的证据链。而新冠患者如果漏检,则会致病毒快速扩散,对疫情的防控、相关地区各行业的正常运营都产生巨大的影响。

那么,人工智能有没有可能把误检率做到零,同时也把漏检率降到零呢?

很遗憾,答案是肯定的。如果在文献中看到有这种情况出现,一种可能是数据集本身的规模

太小,并没有涵盖更多未知样本的情况。

单个指标或两个指标达不到零的原因有不少,一是有可能数据本身是有噪的,如不可约简掉的噪声;二是数据的分布,如果对数据做了某种分布假设,那就无法保证不满足该假设的样本能被正确分类或预测。

除此以外,当数据缺乏明显的标签时,就更难保证了。比如这起失踪案,如果连嫌疑犯是谁都未知,人工智能就无法建成有效的预测模型。

另外,有信息表示本次失踪案的监控视频中,有一帧是只有半个下半身影子的图像。依赖现有的人工智能算法,这种图像很容易被忽略掉,因为它可能达不到形成响应的阈值。其后果却严重地影响对案件的分析。此时,快的搜索就不如人类慢的、缜密的思维模式有效了。

尽管我们看到人工智能在众多刑事案件侦破中起到了重要作用,但也要认识到,仍然有一些情况是不能完全依赖人工智能的,尤其是需要用到强大因果推断和直觉思维能力的场合。

未来也许需要人和机器实现混合智能增强,才能把各自的优势有效结合,形成更强大的智能计算。

<http://blog.sciencenet.cn/u/heruspex>