

# 驾虚卸实：气象界发展“trustworthy” AI技术

■ 贾朋群 张萌



刚刚过去的2021年，全球气象界的一个热点，是以“机器学习”（machine learning, ML）和“神经网络”（neural network, NN）等为代表的AI技术，从几年前有“炒作”嫌疑和爆炸式的登场，到开始入驻气象研究和业务各领域并产生效果和积极影响的过渡年。伴随这样的过渡，“trustworthy”一词被很多学者加在AI之前，组成“可信的AI”。强调AI技术可信，既与AI技术本身的神奇有关，还是现代气象科学本身“物理求真”的传统使然。

2021年10月，欧洲中期天气预报中心（ECMWF）和欧空局（ESA）共同主持召开“地球系统观测和预报中的机器学习研讨会”。会上来自ECMWF的学者介绍了ML在该机构的发展情况。未来10年，ML会在全业务链得到应用，按照应用从弱到强分布的领域为：观测筛选(observation screening)、后处理应用、模式输出特征研判(feature detection)、同化偏差纠正、参数化方案仿真、模式的观测学习分量(learn model components from observation)和学习运动方程(learn equations of motion)。这些应用领域，如果说前面几项针对预报业务，还仅仅是改进意义上的发展，ML强介入的最后三项则具备颠覆性的意义：即以经典力学和热力学为支撑的预报系统中的参数化方案、模式分量乃至运动(动力)方程，都可以通过ML技术，基于已有预报系统的预报和观测结果重新打造。

报告人随后介绍了“自下到上”和“从上到下”两种打造方法，前者强调了基于物理的ML、可信的AI、混合模式和不确定性量化等科学概念和领域。这里，学者指出针对天气和气候应用，需要建设定制的ML解决方案，而方案中强调“trustworthy”，无疑是针对AI技术的“灰色”甚至“黑色”的本质，画出了底线。

与欧洲学者相呼应，美国国家科学基金(NSF)在其最新AI战略布局中，以强强联合方式成立了“天气、气候和沿海海洋可信的AI研究所”(AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography, AI2ES)。AI2ES汇集了来自俄克拉何马大学等7所高校，谷歌、IBM、NVIDIA等4家高技术企业和NCAR及NOAA等联邦机构的研究实体，扛起了在环境科学领域提倡“trustworthy”AI的大旗。2021年，NCAR学者在一次学术会上解释AI2ES要发展可信的AI技术的宗旨时，先提出了ML是否是面对未来气象挑战的“高招”(Silver Bullet)的问题。他问道：为什么可解释的AI非常重要？回答这个问题时，NCAR学者引用了系列科幻小说《银河系漫游指南》作者道格拉斯·亚当斯的回答：42<sup>①</sup>，即关于生命、宇宙和一切事情的终极问题的答案(“The ultimate answer to life, the universe and everything is...42!”)。

学者引用科幻作家的终极答案，实际上并没有多少调侃的意味，他实际上也在负责任地提出AI2ES需要面对的：what are the limitations/ boundaries of this silver bullet?(这一高招的局限和边界在哪里?)。完整回答这个问题，“可信的”要素完全不可丧失，这也是为什么在AI2ES的LOGO(附图)中，与人工智能和环境科学并列的第三个关键词，选择了Risk Communication(风险沟通)，在更现实的意义上，点出了环境学者在引入AI时要有的意识。

ML和NN等AI技术，已经成为气象科技发展重要因素和变数，要发展与气象预报这一与国计民生密切相关领域的、融合了AI的新技术，“trustworthy”或许是我们选择、改进和应用各种AI技术的前提。坚持这样的前提，不可或缺的做法之一，就是更加准确、科学地定义或再定义融入了各种AI技术的新术语，例如：De-biasing Techniques(消偏技术)、Fast linearised models for DA(快速线性化DA模式)、Fast Emulation of Parameterisations(参数化快速仿真)、Data driven Parameterisations(数据驱动参数化)和Non-linear Ensemble Averaging(非线性集合平均)等。

① 在亚当斯的经典科幻小说《银河系漫游指南》中，一台名为“深思”(Deep Thought)的超级计算机经过700万年的思考，得出了关于“生命、宇宙和万事万物终极问题”的答案，这个答案就是“42”。巧合的是，42在数学领域具有许多特殊的意义。其中之一便是2019年由MIT的Andrew Sutherland和布里斯托大学的Andrew Booker领导的团队解开了著名的数学难题——丢番图方程(Diophantine Equation)  $x^3+y^3+z^3=k$  中最难以琢磨的数字  $k=42$  的解。这个由三个立方组成的难题于1954年在剑桥大学被首次提出。对于较小的数字，这类方程比较容易求解：例如，29可以写成  $3^3+1^3+1^3$ ，而32是不可解的。在1~100之中，除了33和42之外，所有问题最终都通过各种技术和超级计算机解决，或者被证明是无法解决的。Booker设计了一个巧妙的算法用布里斯托大学的超算花了几个星期的时间算出了33的一个解决方案。但对于求解42他发现所需的计算量要高一个数量级，可能超出了超级计算机的能力。在尝试破解33时，搜索界限是  $10^{16}$ ，但对于破解42来说需要将搜索界限设置为  $10^{17}$ 。通过UK-based Charity Engine的帮助，他们从遍布世界各地的40多万名志愿者的家用电脑中获得计算能力，最终得到了  $x^3+y^3+z^3=42$  的第一种解决方案： $42 = (-80538738812075974)^3 + 80435758145817515^3 + 12602123297335631^3$ 。目前，在101~1000中还有10个数字尚未被破解；另外，对于  $k=3$  是否存在更多解，也是人们关注的一个方向。