

人工智能未来——发现、理解与创造

戴琼海

人类社会的发展就是不断发现、理解与创造的过程。原始社会利用工具解决生活问题，发现现象并理解现象背后的规律，进而改造甚至创造这个世界，这就是人类社会发展的脉络。

信息时代，艾克特 25 岁带领团队做出了第一个计算机系统。冯·诺依曼给出了现代计算机系统的新架构，并沿用至今。计算机推动了世界的数字化，包含两个历程，一是符号化，二是模型化。数字是表达现象，模型化是对现象的理解过程。

1946 年至今不到百年，计算机的发展为人类带来了巨大的价值。

人工智能参与理解与改造世界

人工智能时代出现了三位深度学习的奠基人，也是 2018 年度的图灵奖获得者。第一是 Geoffrey Hinton（杰弗里·辛顿），反向传播算法的代表人物；第二是 Yann LeCun（杨立昆），卷积神经网络的代表人物；第三是 Yoshua Bengio（约书亚·本吉奥），序列概率模型的代表人物。

那么人工智能将如何参与理解和改造世界？王恩东院士曾有一问，人类怎么预测蛋白质的作用？诸如此类的复杂问题，靠人类的预测是无法做到的。以深度学习为代表的人工智能推动了例如科技、医疗、电子、金融等行业的快速发展，正如总书记说的，人工智能“具有溢出带动性很强的‘头雁’效应”。

同时，我们更应看到人工智能的局限性。

其一，自动驾驶领域中人工智能的相关应用，已经凸显了其在鲁棒性、迁移性及能效比等方面的问题；

其二，在医疗领域中的应用凸显了人工智能算

戴琼海，中国工程院院士、中国人工智能学会理事长、清华大学教授。



法自适应能力的局限，清华大学跟 301 医院合作，用 20 万条男性 50 岁的脑卒数据做训练，但在做预测时发现对女性脑卒疾病的预测准确度并不高；

其三，人工智能可解决一定的问题，但其工作原理还没有明确的可解释性。

还有一个经典案例——莫拉维克悖论（Meravec's paradox）。一台波士顿动力的机器人能翻跟头、跳舞和干很多复杂的事，但让它把一个物体放到有障碍物的桌子上去，它做不到，这就是人工智能的问题所在——难以理解场景与对象间的关系，人工智能能干成年人干的活，但理解能力不如一岁的孩子。

由此可见，人工智能还有很多瓶颈问题要解决。怎么解决呢，就需要追根溯源。我们发现，深度网络的发展很大程度上受到了脑科学的启发，仅仅是视觉听觉部分脑功能的发现，就极大推进了人工智能的发展。如果有机会了解全脑，那会为人工智能带来多大的变化？

我们来看深度学习和人类视觉的不同：生物视觉是宏观和微观回环交互的。比如画画首先画轮廓，再画细节；人眼看东西也是，先看全场景，再聚焦某个小场景、某个小目标，是一个回环交互的过程，

即高级视觉的抽象和初级视觉的边缘检测存在回环交互。但计算机视觉只能从微观到宏观，不能从宏观到微观，这就给人工智能造成了困惑。举个例子，一头熊照片的碎片，人眼看到一点点边缘时，就知道这是一头熊；而人工智能因为缺乏宏观与微观的交互，只能通过不断的学习才能知道这是一头熊。

可见，对场景当中复杂关系的理解，是人工智能非常重要的部分。以前的人工智能针对小场景、少对象、简单关系，用微观图像训练一个模型，设计一个算法，而让它去理解大场景的时候就无能为力了。未来的人工智能应该具备对大场景、多对象、复杂关系的精准理解，这样才能够弥补现有人工智能的不足。

脑科学与新一代人工智能发展

脑科学研究人类如何思考，图灵奖关注的是机

器如何思考，这是两条平行线。人工智能专家往往将脑科学的部分现象和模型应用到人工智能方面。这就给了我们一个启发：能不能在脑科学和人工智能之间架起一个桥梁，这个桥梁就是认知计算。支撑这个桥梁的两个桥墩，是多模态回路观测以及多层次认知模型。通过解决微观、宏观的回环交互问题，来创造新一代的人工智能。

该怎样去做这些研究呢？通过观测可以看到，小鼠有亿级的神经元，恒河猴有百亿级的神经元，人类有千亿级的神经元，神经元的多少代表了聪明的程度。斑马鱼虽然只有 80-100 万个神经元，它也可以工作。MIT2020 年的研究构建了一款只有 18 个神经元的机器人，就能够实现自动驾驶。

2016 年美国 IARPA 做了 MICrONS 项目，称为阿波罗脑计划，花费一亿美金，研究一立方毫米大脑皮层 10 万个神经元是怎么连接的，参与项目的有 CMU 的 Tai Sing Lee、哈佛大学的 David Cox、贝勒医学院的教授，这是个交叉的大项目。

前两位科学家都来过清华，专门讨论 10 万个神经元活动到底怎么构建，讨论神经元活动连接和机器学习算法如何建立关系。

清华开发了 RUSH-II（多维多尺度高分辨光学显微成像系统）来观察小鼠、大鼠全脑神经元的连接和它行为的变化。这是世界上最大视场、数据通量最高的高分辨率光学显微镜，视场大小是一个平方厘米，分辨率达到 0.4 个微米，就是 400 纳米，意味着不仅是神经元，用来在神经元之间传递信息的轴突和树突的连接都可以看清楚。目前正在观测猕猴的全脑神经元活动，这是更接近人类的灵长类动物。现在正在开展新的工作，将脑机和观测仪器相结合，不但看神经元连接，还要看放电过程，这样能够揭示意识是如何转移的。

新一代人工智能按照这样的路径，不



现有人工智能的局限性

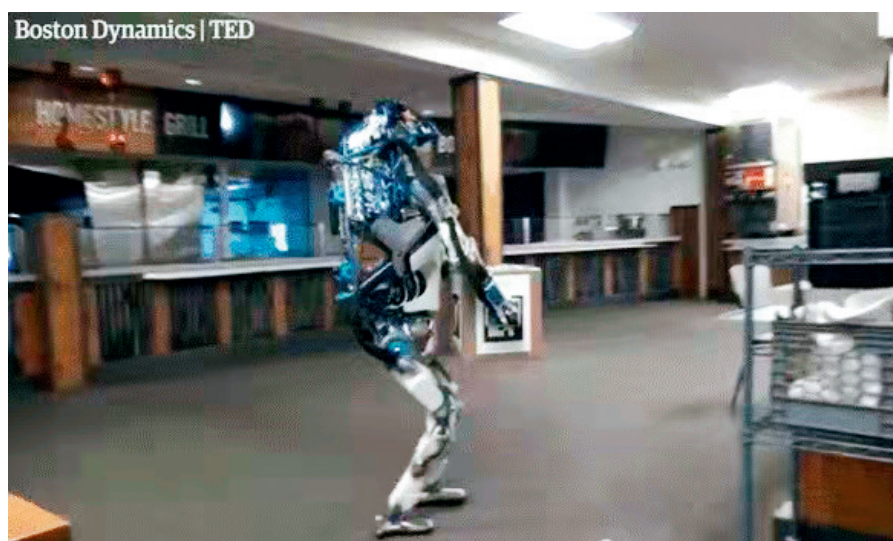
仅要做微观观测，还要做宏观观测。将 CT 核磁共振、全脑高分辨率光学观测、多模态观测结合起来，才能理解神经元之间传递了什么信息。我们也在跟数学家讨论，并构建新型的网络模型，包含了记忆环路、生物机制、物理的熵平衡原理，来推导新型的神经网络模型。我们需要考虑生物化学机制的发觉，数学物理机制的约束，

短期记忆、长期记忆的过程，新一代神经网络推理的自增强等等很多方面，这是我们从脑科学到人工智能做的事情。

大场景多对象智能理解

人工智能能够拓展人类发现、理解与创造的能力。人眼的感知能力会受到空间、时间、波长等多个维度的限制。人工智能可以具备超越人类的感知能力，利用仿生光学复现鹰眼、猫眼、果蝇等的感知能力；同时大量的信息凸显了人类自身的处理能力不足，必须交给机器来继续感知。

人眼感知视觉的像素数不到 6 亿，但是机器视觉可以达 10 亿甚至百亿像素，感知能力更强，带来更丰富的信息。对大场景范围中多动态目标之间复杂关系的理解，就需要构建新一代人工智能模型，让它看得全、看得清、看得准，要做一个大场景多对象数据平台。但宽视场和高分辨的矛盾难以解决，这是物理上面临的挑战。因此我们提出了非结构化的概念和原型系统，很多相机长得不一样，可以自调整，“鲁棒性”（robust）非常高。以此为基础构建数据平台，是 10 亿像素的大场景多对象数据视频平台 PANDA。比如，在清华主楼的门口，非常



波士顿动力的 Atlas 机器人

多的人在迈步走，我们可以对这些对象实时识别和理解；又如，马拉松比赛中的万人人脸识别，就是在处理和解决大场景、多对象、复杂关系。目前在 CVPR、ICCV 还有“2021 全球人工智能技术大会”上做了数据的公开并且比赛，有 6 千多支队伍参加了这场比赛。我们也在讨论是不是可以建一个大的数据平台，来解决复杂场景中复杂关系的理解问题。

从感知智能走向认知智能，第一个要解决模型问题，第二个要解决数据问题。认知智能能够促进大范围动态场景时空关联建模分析，支撑数字城市构建与理解，可以构建物理世界的孪生数字城市。目前我们正在杭州开展工作，做一个数字孪生的城市形态，希望为智慧城市做出贡献。

我们要从脑科学出发，来构建新一代人工智能的理论、方法和技术，同时构建一个大的数据平台来验证理论和模型的可行性。在未来，新一代人工智能需要大场景、多对象的数据平台，大到能够从物理城市构建孪生的数字城市，最后构建智慧城市。这样一来，我们新一代的人工智能理论、模型和算法就初具雏形了。

（本文为作者在 2021 人工智能计算大会上的演讲）