吴飞，浙江大学求是特聘教授，主要研究领域为人工智能、多媒体分析与检索。现任浙江大学本科生院院长，国家人工智能教材基地执行主任，国务院学位委员会智能科学与技术学科评议组成员，教育部计算机101计划核心课程《人工智能引论》负责人，曾获教育部、浙江省、中国人工智能学会和中国电子学会科技进步一等奖和第二届全国教材建设奖一等奖。

卷首语 Foreword

深度必藏于表象之下：还原论的困境与复杂系统的涌现

吴飞

万物之始，大道至简，简而生繁，衍化无穷。以简驭繁、化繁为简，历来是人类探索世界的核心科学范式。在物理学发展史上，还原论(reductionism)长期占据这一范式的主导地位，其核心主张是将复杂系统的各类现象归结为基本组成单元及其相互作用规律，再从简单规律出发重建复杂系统的全貌。从分子到原子、从原子核到夸克，科学家通过层层还原探索物质本质，一直试图将自然界四大基本力纳入统一的理论框架——这正是爱因斯坦晚年倾力追寻的“统一场论”理想，这一研究思路深刻塑造了现代物理学的发展方向。

然而，以还原论研究复杂系统时，却遭遇了根本性挑战。1972年，诺贝尔物理学奖得主安德森在《科学》杂志发表“多者异也”一文，深刻指出：还原论假说从来都不意味着建构论(constructionist)假说。将所有事物还原为简单的基本定律的能力，并不天然具备从那些基本定律出发并重建整个宇宙的能力。事实上，随着粒子物理学不断揭示更深层的基本定律，这些定律与其他学科问题乃至社会问题的关联反而越来越疏远。在面对尺度和复杂性的孪生难题时，以还原论为基础的建构论假定就完全崩溃了。大量基本粒子构成的巨大复杂集体的行为，并不能依据少数粒子的性质做简单外推就能理解。取而代之的是，在每一个复杂性的发展层次之中，都会呈现出全新的物理概念、物理定律和物理原理。要理解这些新行为所需要做的研究，就其基础性而言，与其他研究相比毫不逊色。

这正是“涌现”的核心，即单体简单行为在相互作用中使得系统整体层面突然产生的新质，无法还原为部分之和所得结果。系统科学将这种整体独有、孤立部分不具备的性质，称为整体涌现性。涌现性就是系统组成成分按照特定结构方式相互作用、相互补充、相互制约而催生的全新整体特征，是一种典型的系统结构效应。

2021年诺贝尔物理学奖授予复杂系统研究领域的真锅淑郎、哈塞尔曼和帕里西(其研究揭示了气候、物理系统的复杂涌现规律)，进一步2024年该奖项又授予了在深度神经网络领域做出奠基性贡献的专家霍普菲尔德和辛顿。这一系列颁奖清晰地昭示：物理学范式正在从追求井然有序的方程规律，向理解复杂系统转变，虽然用来解决问题的手段和方法，未必能从基本定律推导出来。

当前生成式人工智能正是这一范式的典型体现。信息通过层层递进、逐级抽象的非线性映射方式被处理。这种信息处理方法体现了复杂系统所呈现的非线性特点：首先，网络的整体行为无法通过简单分解各组成部分来理解，系统会表现出混沌、分岔或涌现等非线性动力学特征；其次，深度神经网络对输入信息的理解是亚符号化(sub-symbolic)——知识以分布式方式编码在连接权重中，无法直接提取出类似“如果–那么”的显式规则，这使得模型的理解过程难以像符号系统那样进行清晰的逻辑推理解释。这一特性既反映了动态系统的内在复杂性，也揭示了人类认知在理解此类系统时面临的固有局限。

但这并不意味着深度神经网络完全不可理解。只是需要我们认识到，在复杂系统面前，不能一味强求物理学的简约公式，不应期待一切复杂现象都能用寥寥数语确定性概括，而应接受基于近似性或概率性的解释框架，这需要数学、物理、脑科学与认知科学的协同进化。

正如奥地利诗人霍夫曼斯塔尔所言：“深度必藏于何处？确定性恰在表象之下。”深度神经网络的本质价值，不在于层数的堆砌，而在于通过多层非线性变换，穿透数据表层噪声，提取出支配现象的内在非线性隐式模式。