

---

# 具身智能如何进化？关键在于如何“试错”

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/32824.html>

*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

## 具身智能如何进化？关键在于如何“试错”

。人类智能的进化史，本质上是一部“试错文明史”。从石器时代的工具打磨到现代科技的精密设计，每一次认知跃迁都源于对未知的探索与修正。具身智能作为人工智能的下一代范式，其核心在于“具身性”——通过物理“身体”与环境的交互，在试错中积累经验、提升认知，最终实现智能涌现。

2025年政府工作报告首次将具身智能列为国家未来产业培育重点，其战略意义不仅在于技术突破，更在于构建一种“试错—认知—进化”的闭环逻辑，为智能体赋予适应物理世界复杂性的能力。

然而，正如生物智能体的试错需要长期不断探索、“量变引起质变”一样，具身智能的“试错—认知—进化”，也面临着“达尔文进化困境”。

### 物理试错的“达尔文进化困境”

自然界中，猎豹为提升捕猎成功率需经历数百次失败的追击，人类婴儿学会行走平均要跌倒2000次。对具身智能而言，这种在物理环境中试错的成本太高。

首先，硬件损耗与安全风险突出。机械臂碰撞、机器人跌落等意外可能造成精密传感器、驱动单元损坏，单次维修成本可达数万元，工业场景中更可能引发生产事故。

其次，环境复杂性导致试错效率低下。物理世界存在光照变化、动态障碍物等不可控变量，智能体需应对无限长尾场景，远超算法预设边界，如自动驾驶路测百万公里仍难覆盖极端天气。

再者，知识迁移存在“仿真鸿沟”。虚拟训练依赖理想化建模，与真实物理参数（摩擦系数、材质形变等）的细微偏差可能导致技能迁移失效，如家庭服务机器人常因地面材质差异导致避障失败。

此外，伦理与法律风险不容忽视。医疗、公共安全等场景中试错可能威胁人身安全，易引发责任归属争议。

硬件的高昂损耗与场景的不可逆性，共同构成具身智能物理试错的核心限制。然而，试错又是认知提升的必经之路。正如生物学中“自然选择”通过大量试错筛选最优解，具身智能的发展也需经历类似的迭代过程，只是需要寻找破解物理试错困境之道。

---

生成式人工智能能做什么？

近年来，以生成式人工智能作为试错“场景加速器”的研究已成为人工智能的热点之一。科技部“新一代人工智能”重大项目已将生成式人工智能列为具身智能核心支撑技术。

本质上，生成式人工智能通过创造高保真、高覆盖、高风险的虚拟场景，融合自主评估与智能试错机制，可有效突破物理试错的时空与成本局限。例如，利用扩散模型合成现实罕见的极端工况，弥补传统仿真无法覆盖的“认知盲区”；通过神经辐射场（NeRF）技术构建可微分物理环境，使虚拟试错能反向优化世界模型参数；基于GAN网络可在稀疏数据场景下合成高密度训练样本，支撑强化学习算法的快速收敛等等。特别是通过生成对抗网络与强化学习的结合，AI系统可主动探索人类预设规则之外的潜在场景，发现传统方法论难以捕捉的系统性风险。

不过，生成式人工智能虽能创造海量虚拟试错场景，但若缺乏对物理规律的深层理解，这些场景可能沦为无序的“数字游乐场”。

世界模型的价值正在于此——它为生成式AI的虚拟创造提供认知框架，将随机生成的场景转化为可解释、可迁移的“认知燃料”。

世界模型：试错的“认知罗盘”

以“世界模型”为认知框架的系统工程，有望拯救试错于“无序探索”的困境之中。世界模型基于人类对物理规律的抽象表征，为试错提供方向性约束与因果推理能力，进而将试错从“行为—结果”的简单表象关联，升级为“变量—机制—现象”的因果推演。正如人类依靠大脑前额叶预判行为后果，具身智能可通过世界模型预测动作链的影响。

试错机制与“世界模型”的协同演进，本质上是智能系统从“被动响应”向“主动认知”的范式转变。传统智能系统依赖预设规则或大数据训练，仅能被动响应已知场景。而具身智能通过试错主动探索环境边界，世界模型则将这些经验抽象为物理规律与因果关系的知识图谱，形成“预测—验证—校正”的认知闭环。

例如，自动驾驶系统在虚拟试错中遭遇AI生成的极端暴雨场景，世界模型不仅预测轮胎打滑概率，更反向推演路面摩擦系数与制动策略的关联，生成新的安全驾驶规则。这种协同突破了“数据喂养”的被动逻辑：试错机制负责在虚实场景中主动暴露认知盲区，世界模型则通过物理引擎与神经符号推理，将离散经验升华为可泛化的因果逻辑链，进而推动现有智能系统摆脱“刺激—反应”的被动链条，转向以世界模型为基座的“假设—验证”主动认知范式。

试错驱动认知跃迁：重构智能进化新范式

试错机制与生成式AI、世界模型的协同，标志着智能系统从“被动触发”到“主动认知”的范式跃迁。这种协同机制，本质上是对生物进化逻辑的数字化重构：生成式AI如同“基因突变”的加速器，不断拓展试错边界；世界模型则扮演“自然选择”的裁判角色，基于物理规律对试错结果进行因果评估与知识提炼。唯有二者耦合，虚拟试错才能超越单纯的数据堆砌，进化出指导现实行动的认知体系。

可喜的是，我国正以国家战略牵引这场认知革命：将具身智能作为认知革命的核心抓手，构建起以国家重大专项为引领、区域创新平台为支撑、产学研深度融合的新质生产力培育体系。

---

安徽大学具身智能研究院的成立正是对这一战略图景的落实举措。该院依托光电信息获取与防护技术国家重点实验室、自主无人系统技术教育部工程研究中心、安徽省无人系统与智能技术工程研究中心等，联合江淮前沿技术协同创新中心、科大讯飞、中科星驰等行业领军企事业单位，打造思维“立交桥模式”，有机组织各方力量，协同推进教育、科技、人才一体化合作，形成“锚定国家战略方向、突破高校平台瓶颈、加速产业协同创新”的产学研合作新模式。我们期待这种“立交桥式”产学研合作新模式，能为具身智能认知进化、推动技术转化应用、推动区域产业高质量发展贡献科技与人才力量。

（作者系安徽大学校长、中国人工智能自主无人系统专委会主任，本报记者赵广立整理）

作者：孙长银 来源：中国科学报

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发