# **人工智能如何自我进化**

刘逸川 2019.12.17

设计一种能够改进自身的人工智能是困难的，因为改进所需要的智能水平很高，或者说人工智能建立在复杂的基础上。无论是数学、程序还是各种智能模型都超出人工智能的理解能力，而要改进自身几乎不可能避开这些。那么可以从建立在简单基础上的低水平人工智能开始，使基础和智能水平相适应，从而能够不断自我进化达到很高水平。这不是新思路，只是缺乏理论指导，本文旨在提出一种初步理论。

自我进化的人工智能的关键在于正确设计初始智能和进化规则。强化学习也是如此，但它是复杂的人工智能，进化到一定程度会停止，需要人类来改进。简单的人工智能将类似人工生命，一开始是非常简单的，所有的复杂结构都是进化而来，这样能够避免自身超出自身的改进能力。初始智能和进化规则虽然是简单的，但是并不容易设计，需要符合一些原理。

初始智能的进化肯定会受到阻碍，这些阻碍足以导致进化失败，只有将阻碍转化为动力的进化方式才可能成功。比如初始智能可能会犯一些错误，但是正是从错误中学习才更加深刻，没有错误反而不好。闪电、火山、宇宙射线等对生命有害的因素对早期生命而言却是必不可少的，早期生命将劣势充分转化为优势不是偶然，如果生命不具有这种特性将无法诞生。初始智能能够进化得越来越复杂在于一个改变引起其他改变，这些改变又会引起更多改变，像滚雪球一样。这是一个递归的过程，初始智能和规则都需要体现递归的原理。智能体内部在任何时期都需要是递归的，进化这种时间上的递归和智能体这种空间上的递归是统一的。智能体必须是一个严密系统，各部分合理组织、紧密联系，同时既保持一定稳定性又保持一定可变性，保持稳定性是为了更好地变化，保持可变性是为了更好地稳定。智能体在很长时间内都不是在理解自身的基础上改进自身，而只是根据一些简单原则和方法。这些原则和方法需要不断更新，这是由智能体自身控制的，需要实现每次更新都是正确的并且促进下一次更新。智能体不一定需要自我复制，所有改变都可以发生在一个智能体上。规则之间的联系应该是非常紧密的，而且规则是统一的，各个部分采用同样的规则，这其实也是递归。因为当两个看似不相关的部分相遇时，它们能够很好地发生作用，就像遇到同类一样。规则还需要体现自相似这种递归，即简单规则组成的宏观规则和简单规则是相似的，只是层次不同而已。随着智能体的进化，规则本身也会进化，但是初始规则是不变的，新规则是初始规则的组合，那么初始规则必须是完备的，能够组成任何可能需要的规则。

初始智能处在一个虚拟环境中，与之构成一个系统共同进化。环境也可以理解为智能体的一部分，相当于它用于改进自身的工具。自我进化产生的智能必然是通用智能，因为只有通用智能才能解决自身遇到的各种问题。由于在虚拟环境中进化，它只是具备了智能的基本原理，就像婴儿大脑一样，有一个适应真实世界的过程，需要人类的引导和训练。

自我进化的人工智能是不可简约的，即要知道结果只有按照初始智能和规则一步步演化，没有更简单的方法可以准确预测。同时由于它是一个复杂的整体，除非深刻理解它，否则来自外部的修改可能都达不到预期效果，甚至适得其反，也就是说人类对进化过程是很难正确干预的。人类所做的只是设计初始智能和规则，进化过程可以为了研究而修改但是最终成品的进化过程是不应该修改的，因为容易出现无法预料的错误。研究进化过程时生命是一个重要参考，但是模拟生命进化是很困难的，因为生命建立在复杂的物质基础上。即使原理是简单的，还要充分理解物质性质才能正确理解和模拟生命，不如模拟更简单且更合适的。智能也是递归的，因为智能能够创造智能，其原理具有很强递归特性，从而非常适合自我进化。

人类为什么很难理解自我进化的原理？因为我们处在一个已经演化得非常复杂的世界，和初始智能的世界有巨大差异。我们遇到阻碍不一定需要转化为动力，有时忽视反而更好，有很多因素能够使我们突破阻碍，但是对于初始智能而言这是唯一能够成功的方式。我们的很多事物不需要是自我调节的，无论什么坏了都可以由人类来维修，但是对于初始智能而言没有任何高级智能能够帮助它。

既然人类可以直接设计人工智能，为什么还需要使其自我进化？因为有些是无法教给人工智能的，需要它自己去领悟。我认为技术奇点会由自我进化的人工智能达到，因为这样的智能才是完备的，直接设计的人工智能难免会有考虑不周之处，即使能够自我改进也会在未来遇到无法解决的问题。

目前主流的人工智能是将智能转化为计算，自我进化的人工智能则是将计算转化为智能。人类掌握了巨大的计算力却没有充分发挥出来，因为需要人类设计和改进算法。自我进化的人工智能其实是一种创造算法的算法，能够不断地将计算力转化为算法。人类学会创造新的智能是一个飞跃，学会创造自我进化的智能也是一个飞跃。