

炼金术与人工智能： 休伯特·德雷福斯对人工智能发展的影响

陈自富

(上海交通大学 科学史与科学文化研究院, 上海 200240)

摘要：1965年12月，休伯特·德雷福斯以兰德公司顾问的身份，发表了编号为P-3244的《人工智能与炼金术》的研究报告，对兰德公司本身主导的人工智能（以下简称AI）研究提出了重大理论挑战，1972年德雷福斯以该报告为基础出版了《计算机不能做什么——人工智能的极限》，该书与1966年美国国家科学院的ALPAC报告，1973年英国科学研究理事会的LightHill报告一起，标志着AI发展历史上的第一次冬天，即使经历了上世纪80年代由于专家系统兴起的AI再次繁荣，以及90年代初AI的第二次冬天，AI的研究纲领已经变化甚多，但德雷福斯仍然坚持其基本观点，对隐含在AI研究纲领中的关于人类认知和问题解决能力的深层假设，从现象学和海德格尔哲学为核心的大陆哲学立场出发，始终进行批判性地思考和分析，无论AI科学家共同体对其观点是否认同，德雷福斯这些深刻的哲学思考，客观上推动了从AI研究早期基于知识主义、符号主义强纲领的盲目乐观，到目前对实现人类级别智能的智能机器建造的审慎态度，以及更加丰富的研究进路的转变。

关键词：德雷福斯；人工智能；人工智能冬天

中图分类号：TP18;TG115 **文献编码：**A **DOI:**10.3969/j.issn1003-8256.2015.04.008

1 德雷福斯生平及学术工作概述

休伯特·德雷福斯（Hubert Dreyfus），1929年出生于美国印第安纳的特雷霍特（Terre Haute），在哈佛大学接受本科和研究生教育，1964年在奎因的学生Dagfinn Føllesdal指导下，获得哈佛大学哲学博士学位，后者来自挪威，主要研究语言哲学、现象学、存在主义和解释学¹。

作为美国知名的现象学和存在主义哲学家，德雷福斯是欧洲大陆哲学的主要代表：胡塞尔、福柯、梅洛-庞蒂，尤其是海德格尔在美国顶尖的译者和诠释者，他的《在世存在：海德格尔〈存在与时间〉述评》第一部，被许多人认为是本科生学习海德格尔最重要的哲学思想的权威读本，他与Paul Rainbow合著了《米歇尔·福柯：超越结构主义和解释学》，还翻译了梅洛庞蒂的《意义与无意义》，但他最为人所知的工

作者简介：陈自富（1974-），湖南洪江人，上海交通大学科学史与科学文化研究院，博士研究生，研究方向为信息科学与互联网技术的历史、文化和传播研究。

作是从哲学角度对人工智能和认知科学的思考和批判，代表作主要是1972年的《计算机不能做什么——人工智能的极限》以及相关论文，1982年他与Harrison Hall合编的《胡塞尔、意向性与认知科学》也是研究胡塞尔现象学与认知科学关系的经典论文选集²。

1968年以后德雷福斯在加州大学伯克利分校度过了他主要的学术生涯，之前他先后任教于布兰迪斯大学（1957–1959），麻省理工学院（1960–1968），1998年荷兰伊拉姆斯大学因其“在人工智能领域杰出和有高度影响力的工作，以及对20世纪大陆哲学的分析和诠释所做的同样杰出的贡献”而授予其名誉博士，2001年当选为美国人文与科学院院士³。

德雷福斯桃李满天下，既包括提出了GOF AI（Good Old Fashioned Artificial Intelligence：有效的老式人工智能）概念的John Haugeland（1945–2010），还包括哈佛大学哲学系主任Dorrance Kelly，以及Taylor Carman、Iain Thomson、Mark Wrathall这些现象学和存在主义的中年哲学家，2010年音乐家和导演Tao Ruspoli在德雷福斯的影响和指导下，拍摄了反映海德格尔哲学思想的纪录片《在世存在》，采访了美国主要的存在主义哲学家，其中除德雷福斯本人之外，约一半学者均为其学生⁴。

2 兰德报告《炼金术与人工智能》的主要内容

1964年夏天，麻省理工学院的年轻教师休伯特·德雷福斯由其弟弟斯图亚特·德雷福推荐，进入兰德公司圣塔莫尼卡总部担任研究顾问，公司当时负责计算机科学部门的主管Paul Armer认为AI研究需要考虑哲学问题并能从中获益，而德雷福斯的简历和背景调查都不错，让他觉得德雷福斯可以为AI研究项目从哲学上提供公正的建议¹。

兰德公司在早期AI历史上具有特殊和重要的地位，当时其研究经费主要来自于美国空军等军方机构，对通用数字计算机的建造和研究情有独钟，不仅聘请冯·诺依曼担任顾问，还委托普林斯顿大学建造了JOHNNIAC，最早的程序存储结构（冯·诺依

曼）数字计算机之一。50年代他们聘请了Allen Newell从事北美防空指挥系统自动化的研究，Cliff Shaw为JOHNNIC上提供了最早的交互式解释型编程语言之一JOSS，Newell、Shaw与公司顾问，卡内基理工学院教师赫伯特·西蒙（Herbert Simon）及其部分学生一起，在推动早期人工智能研究中发挥着关键作用，1963年到加州大学伯克利分校商学院任教的Feigenbaum和Feldman选编了AI当时最重要的论文文集《计算机与思维》，书中收录的20篇论文中有6篇是兰德的研究报告²。

Paul Armer没有想到的是，德雷福斯提交的报告是如此尖锐，几乎摧毁了兰德正在进行中的AI研究的基础，他不得不延迟公布该报告，直至他认识到“仅仅是因为不喜欢不能成为不出版它的理由”，最终推迟了大约9个月于1965年12月出版，该报告后来成为兰德公司销量最高的报告之一³。

德雷福斯的报告标题是《炼金术与AI》，将AI与历史上的炼金术相提并论，意图说明当时进行的AI研究是没有基础的无用之功，而且在序言中，他针对Feigenbaum和Feldman在《计算机与思维》中所宣称的AI领域的显著进步是向终极目标的逐步接近的说法，提出了第二个尖酸的比喻：第一个爬上树的人可以声称这是飞往月球的显著进步⁴。

这份编号为P-3244的兰德公司报告当然不仅是因为标题和序言的激烈比喻而影响重大，关键还是报告中的观点确实对早期AI发展的成就提出了重大挑战。

报告共90页，分为以下四个部分⁴：

- （1）序言
- （2）Part I：人工智能领域的现状
- （3）Part II：当前困难的潜在意义
- （4）Part III：人工智能的未来

报告共引用了43篇文献，其中13篇来源于早期AI的核心文集《计算机与思维》中的论文，8篇来自于AI方面及其相关的专业期刊，8篇为兰德公司AI领域的研究报告和备忘录，余下14篇为相关专著、业内专家手稿、哲学和心理学文献，其中涉及哲学和心理学的文献是：

- （1）笛卡尔：《谈谈方法》

(2) 完形心理学创始人，德国心理学家Max Wertheimer：《创造性思维》

(3) 迈克·波兰尼：《经验和模式感知》，载于AI早期文集《心智建模》

(4) 维特根斯坦：《哲学研究》、《蓝皮书和褐皮书》

德雷福斯通过研究当时AI领域的主要学术论文，在序言和Part I对AI的进展进行了评估，他的主要结论是：

(1) 由于当时新闻媒体的宣传和AI科学家的乐观主张，一些分析哲学家（Putnam、Scriven、Ziff）和神学家、伦理学家对机器大脑的立场虽然相反，但均错误地以为高度智能的人造物已经或即将被科学家开发出来，而这个并不是事实，因此需要对AI的研究现状进行重新评估；

(2) 在博弈、问题求解、语言翻译和学习、模式识别这四个当时比较活跃，而且被公认为是人类智能才能胜任的领域中，AI研究都遇到了比较大的困难；包括国际象棋中的组合爆炸、启发式方法在机器定理证明中的停滞、10年来投入了1600万美元的机器翻译面临的上下文歧义问题、模式识别只能做到识别手写的摩尔斯电码（MIT林肯实验室）和英文字母的水平。

接下来，德雷福斯从哲学和心理学方面来评价这些困难的深层次意义，并且是以一个比较积极的态度，“倘若人们正视否定的结果，那么该结果也是有趣的。递减的进展代替了预测的加速成功，或许表明了未曾预料的现象。我们正在一个象速度那样的连续体上推进吗？当我们接近光速时，这样的进展会越来越困难，或者相反我们是在面对一个非连续体，就象那个爬树来登月的人？”，从这段原文可以看出，德雷福斯从一开始就未曾否定AI的意义，只是在质疑当时AI研究的主要假设和方法论。

德雷福斯主要是从人类与机器对信息加工形式的对比来阐述当前困难的深刻意义，他在Part II中列举了四种：

(1) 人类思维的边缘意识与AI的启发式搜索；

(2) 人类思维的本质/非本质区分与AI的试错法；

(3) 人类思维的模糊容忍度与AI的穷举；

(4) 人类思维基于上述三种信息加工形式的明晰组合（Perspicuous Grouping）能力。

由此，德雷福斯得出了人类能在下述困难逐步加大的条件下进行模式识别的结论：

(1) 模式可能歪斜、不完整、变形和在噪声环境中；

(2) 模式识别所需的特征虽然清晰甚至能形式化，但搜索难度会急剧加大（指数爆炸）；

(3) 特征可能依赖内外部上下文，从而不能从列表中隔离出来单独考虑；

(4) 可能没有公共特征，但“重叠的相似性的复杂网络”总能识别新的变化。

任何机器实现的模式识别能力，应与人类思维的能力等效，因此必须具备这些能力：对模式的特定实例把基本特征从非基本特征中区分出来、利用停留在意识边缘的线索（或暗示）、考虑上下文环境、把个体感知为典型，即把个体定位于一个范型实例。而目前AI在模式识别方面的困难，都给博弈、问题求解、语言翻译领域带来了巨大的困难。

但是，AI科学家共同体并不这么认为，德雷福斯指出：“尽管有这些巨大的困难，认知模拟和人工智能的工作者却并不气馁。事实上，他们还毫无理由地乐观。作为这种乐观的基础是人类信息加工必须按照象数字计算机那样的离散步骤来处理的信念，而且因为自然用这种加工形式产生了智能行为，那么合适的编程应该能够从机器导致这样的行为。”这种AI研究者中的教条主义必有其根源，这种根源就是心理学上的联想主义假设。

在Part III，德雷福斯认为没有经验和先验证据支持联结主义者的假定，因此也没有理由期待AI领域的持续进步，机械式信息加工有内在限制，而人类却没有这种限制。他指出了采用认知模拟进路的若干主要威胁：博弈中事实的无限性和无限“进行”（Progression）的威胁、问题求解中需求的不确定性和无限退行的威胁、上下文相互作用和循环性引用的威胁。

在以上基础上，德雷福斯把人类智能行为根据

AI实现的可能性分为四类：联结主义的、非形式化的、简单形式化的、复杂形式化的，认为联结主义和简单形式化的行为可以用计算机模拟，而非形式化的行为难以模拟，复杂形式化行为仅在很小程度上可以模拟。他希望：“所有智能行为可以映射到多维度连续体的假设鼓励AI工作者从两个有希望的领域中的成功，推广到另外两个尚无事实性成功期望的领域。”

德雷福斯在Part III中的结论并不是否定性的，他指出：

（1）短期来看需要考虑人类智能和机器智能的协作，只有从长期来看非数字化的自动机才能表现出在处理人类非形式化世界中关键的三种信息加工形式；

（2）目前的困难和停滞并不意味着之前对AI的投入完全浪费，而是应该调整到聚焦三种人类独特的信息加工形式上来。

在报告的最后，德雷福斯如此结尾：

如果炼金术士不再关注曲颈瓶和五角器皿，而把时间花在寻找问题的深层结构，如果人从树上下下来开始着手发明火与车轮，事情就会向一个更令人鼓舞的方向发展。毕竟，三百年后我们确实从铅提取了黄金（也登上了月球），但这只有在我们放弃了炼金术水平上的工作，达到化学水平甚至更深层次的原子水平上后才会发生。^[5]

3 《炼金术与人工智能》的影响和AI的第一次冬天

德雷福斯的这篇报告在当时由于其观点过于大胆，刚发布时反而没有引起AI学者内部的大规模反应，出于对报告内容的担忧，兰德公司1965年仅以最低级别的备忘录方式发布了油印版，1967年才发布了印刷版^[6]。

当时对于这份报告大致有三种不同的倾向和态度，大致如下^[7]：

首先是个别人的激烈反应，例如1968年1月麻省理工学院（MIT）的Seymour Papert发表了一份同样言

辞激烈的备忘录（AIM-154）：《休伯特·L·德雷福斯的人工智能：谬论种种》，其中他说道：“我得知反驳针对Simon的诽谤指控是不相干的，因为其他人确实已经对人工智能中的成就提出了错误的主张，也得知表明德雷福斯的解释混乱是浪费时间：即为什么机器能下跳棋但不能下国际象棋是因为计算机翻译俄语确实遇到了困难……我被告知他的观点必须作为拥有深度‘人文主义’内容的文学想象作品阅读。”

其次是虽然反对报告观点，但表面上态度比较绅士和含蓄，甚至置之不理。例如德雷福斯直接攻击的西蒙和艾伦·纽厄尔，还有Feigenbaum和马文·明斯基，这也是AI共同体中大多数人的态度。例如，西蒙和马文·明斯基都不认可德雷福斯提出AI研究陷入停滞的观点，他们并未在公共场合中表达此意见，而是在接受Pamela McCorduck的口述史采访中表达了这样的观点：西蒙在1957年做出的十年内AI在国际象棋和作曲达到人类水平的预言，是因为没有足够的人力和资金资源投入的问题，也就是这些问题的意义在AI中并不这么重大。明斯基回忆说：“也许西蒙低估了国际象棋有多难，但我想他真正的失误是另一个有趣的原因，因为他是管理学院院长。西蒙的重大失误在于他对有多少人从事国际象棋问题的估计上，当他做出预言时，我想他认为从那时起的三年或四年内，全世界有数百人非常努力地编写国际象棋程序……麦卡锡偶尔关注此问题，把好问题分配给差生，我的实验室则不鼓励这样。”

但是从多年后的口述来看，西蒙对德雷福斯这份报告的态度也是很激烈的：“……他在此前后与兰德公司没有任何联系，也根本没有相关技术背景，但他是兰德公司顾问这个事实马上给了他可信性……但我关于此事的不满只是兰德的名字附加在这个垃圾上，这才真正是错误的炫耀。”就连兰德公司的Paul Armer也不喜欢他，但同时也有一些人相当认可这份报告，Paul回忆说：“我认为这是糟糕的哲学……，我也不是哲学家，但我认为它不好。我与部门中喜欢此报告的其他人发生了激烈争吵，他们用我自己关于检查制度的声明将我置于尴尬之地……”。同样，Feigenbaum在多年后也如此说：“AI所需要的是好的

德雷福斯，AI中的概念问题确实很粗糙……但是德雷福斯用他没弄明白和过时的东西猛击我们的头……他提供给了我们什么作为代替？现象学！那是和绒线球、棉花糖一样无价值的东西！”。

第三种态度就是对报告的基本立场持拥护立场，首先在兰德公司内部就有不少人对报告内容认同，他们也迫使Paul Armer最终将之公之于众，很快引起了AI共同体之外的广泛关注，在没有广告宣传的情况下，成为兰德公司卖得最好的报告之一，德雷福斯收到了来自世界各地的来信，1966年6月《纽约客》杂志还专门介绍了这篇报告，德雷福斯本人也因此报告被计算机科学共同体所接受，成为AI科学家不得不接受的一个事实。

接下来几年德雷福斯的报告逐步进入公众视野，其中一件事情是MIT的MacHack象棋程序击败了德雷福斯，对局结果印在了美国计算机学会（ACM）的人工智能专业组（SIGART）的公报上，西蒙对此写了一封题为“冷静！朋友！”的公开信作为对德雷福斯报告的正式公开回应，其中有一句这样的表达：“……这是一场真正扣人心弦的比赛，其中一个人的边缘无意识被另一个人彻底打败……”，边缘意识是德雷福斯批判早期AI的核心概念武器，西蒙的公开信最终还是表明了AI共同体与德雷福斯之间剑拔弩张的关系，德雷福斯回忆当时MIT从事AI的同事：“不敢被人看见与我一起午餐”^[8]。

《炼金术与人工智能》从事后来看对早期AI研究的影响是巨大的，一方面1972年德雷福斯在报告基础上进行扩充，出版了《计算机不能做什么：人工智能的极限》，这本书沿着报告的思路，不仅引用了更多的分析哲学和大陆哲学文献，而且也对明斯基1968年出版的AI论文集《语义信息加工》中反映的最新进展进行了研究，把报告中的联想主义心理学假设扩充为生物学、心理学、认识论和本体论四个方面的假设，从而涵盖了当时AI研究方法的全部基础。

德雷福斯从报告扩充到专著的最根本变化在于他引入了海德格尔、梅洛·庞蒂关于躯体在知识获取和应用中的作用，同时也用波兰尼、维特根斯坦的哲学来批驳AI研究的假设，使得其论点更加富于建设性。

《炼金术与人工智能》在另一方面的重大影响就是与1966年美国国家科学院的ALPAC（Automatic Language Processing Advisory Committee：自动语言处理顾问委员会）报告，1973年英国科学研究理事会的LightHill报告一起，成为了AI历史上第一次冬天（1974年–1980年）的主要标志。如果说德雷福斯的工作是因为其非专业出身而不被AI研究人员认可之外，而后两者的专业学术背景带来的影响力就直接导致了政府和军方大规模削减甚至终止对AI的研究资助^[9]。

ALPAC本身是由JALPG（Joint Automatic Language Processing Group：联合自动语言处理组）于1964年设立的，而后者是美国国防部、国家科学基金会、中央情报局三方为协调联邦层面的机器翻译研究而成立的组织，ALPAC的主席是当时贝尔实验室的电子和通讯专家约翰·皮尔士（PCM：脉冲编码调制的发明人）⁵。LightHill报告是英国科学研究理事会委托剑桥大学卢卡斯讲席教授、物理学家James LightHill爵士对当时英国以AI研究现状调查提交的一份独立报告⁶。

ALPAC报告的结论是机器翻译过于昂贵，比人工翻译慢而且不够准确，在较近的将来达不到人类翻译的品质，因此只对两个方面提供资助：将作为语言学分支的计算语言学当成一门纯科学研究，以及对人工翻译进行改进提高。LightHill报告的结论是仅支持对神经生理学和心理学过程的计算机模拟，而放弃对机器人和语言处理的资助，这导致了科学研究理事会终止对除爱丁堡、苏塞克斯、埃塞克斯三所大学之外的其他大学AI研究的支持。

在1969年美国国会通过Mansfield修正案后，DARPA（Defense Advanced Research Projects Agency：美国国防部高级研究计划局）被要求资助面向任务的直接军事用途的研究，而不再支持间接的基础研究，其研究转向了诸如自动坦克、战役管理系统等具有明确目标的军事项目，到1974年已经很难从DARPA申请到经费资助。1971年DARPA雄心勃勃在BBN、IBM、卡耐基·梅隆大学、斯坦福研究院启动的语言理解识别（SUR：Speech Understanding Recognition）项目，至少包括1000个词汇，希望能够理解领航员的语音命令，但最终实现的系统只能听懂按特定顺序说出的词

句，DARPA对此很不满意，于1974年取消了每年300万美元的资助⁷。

无论德雷福斯的报告内容是否得到AI共同体认可，从导致AI第一个冬天的原因来看，他在1964年的工作相当准确地预见了当时在媒体上乐观情绪背后的深层次危机，这些危机在十年之后的爆发迫使AI共同体更加包容地对待哲学家对机器智能方面的思考和立场。

4 德雷福斯在《计算机不能做什么》后续版本的立场

1979年，《计算机不能做什么》第二次再版，德雷福斯持续跟踪了1972年之后的AI进展，并为此写了一个长达70页左右的第二版序言，其基本立场是认为他在第一版中的看法与之后的AI发展是相符合的^[10]。

德雷福斯把1967年到1977年之间的AI研究划分为两个阶段，第一个阶段是从1967年到1972年，称之为“控制微世界”，主要针对Terry Winograd用于处理积木世界的自然语言理解程序SHRDLU和斯坦福研究院的Shakey机器人项目提出了异议，认为认知过程的原理仍未取得重大进展，类似于积木这样的微世界，不能逐步逼近复杂的真实世界，Shakey项目反而导致了机器如何获得知识并予以表达和应用的难题，尤其是真实世界的常识知识问题。他说：“认为这种无尽头的、对人类其他实践活动的参照将会收敛起来，从而能对简单的微世界做相对孤立的研究，这种完全不合理的信念反映了一种要把在自然科学领域已取得成就的方向搬到AI中来的天真想法”。

第二个阶段从1972年到1977年，主要聚焦当时取得一定成功的专家系统，例如MYCIN，并分析了明斯基对知识表达的框架理论和香克的脚本理论的局限性，他最终得出的结论是人类智能难以在脱离其上下文环境及其局势的情况下，单独地表示出来和加以理解，并在AI中予以实现。

1992年，德雷福斯出版了《计算机不能做什么》的第三版，但书名改成了《计算机仍然不能做什么：人工理性批判》，新的版本并没有修改正文，但增加

了一篇约50页的新序言，在该序言中德雷福斯与时俱进跟踪并评估了四类不同的AI研究策略，第一类仍然是符号主义的，以1984年由Lenat启动的Cyc项目为代表，该项目试图组合人们日常生活中的常识知识的本体论和知识库，从而使计算机达到人类推理的智能水平，这是AI知识工程研究进路中最为宏大的项目，至今仍以OpenCyc的开源项目方式进行。其他三类分别是MIT人工智能实验室以海德格尔现象学为基础开发的交互式人工智能（海德格尔式AI）、80年代后复兴的神经网络建模和强化学习^[11]。

在上世纪90年代初，AI的研究进路已经出现了更加多元化的场景，德雷福斯承认AI这方面的进步，但仍然捍卫人类理性和智能的独特地位，这种独特性来自于躯体、社会和文化，“我们的需要、渴望和情感藉由关于我们行为适当性的感觉直接呈现给我们。如果这些需要、渴望和情感轮流依赖于被社会化到文化中的生物躯体的能力和弱点，那么即使强化学习策略仍有很长的路要走。”^[12]

5 德雷福斯对AI发展的主要影响

人工智能这个名词虽然是在1956年的达特茅斯会议上由麦卡锡提出，但关于这个学科的理论 and 实践探索至少可以追溯到二战前的图灵和维纳。在上世纪70年代之前，人工智能和认知科学基本不加区分，一直到1973年LightHill报告中才由英国认知科学家Longuet-Higgins提出“认知科学”这个名词，但当时人工智能与认知科学的区分仍不明显，这种情况要到1975年提出认知科学支撑学科的六边形模型后才改变^[13]。

但是到今天为止，对于AI的定义，仍然存在广义和狭义的区别，在《MIT认知科学百科全书》的序言中，将传统AI当成计算智能：

“有两种互为补充的人工智能观：一种是作为关注智能机器建造的工程学科，另一种是关注对人类智能进行计算建模的经验科学。在该学科早期，两种观点很少得到区分。其后根本性的分歧出现了，前者主导着现代人工智能，后者是许多现代认知科学的特征。基于这个原因，我们采用更为中性的术语‘计算

智能’作为本文标题——两个共同体都使用计算术语来处理智能理解的问题。^[14]”

美国人工智能学会（AAAI）对AI的定义也小心翼翼地维持两种人工智能观的平衡：“对作为思维和智能行为为基础的机制的科学理解及它们在机器中的具身实现⁸”，但大多数AI学者还是从狭义来理解AI，例如麦卡锡的定义是：“制造智能机器的科学和工程，特别是智能的计算机程序，它与利用计算机来理解人类智能的类似任务有关，但不必自我限制于生物学上可观察的方法。”⁹”

在目前流行的AI教科书《人工智能——一种现代方法》对AI的定义是：“理性智能体的设计过程……着重讨论理性智能体的通用原则以及构造此类智能体所需的组成部分”^[15]，这里强调的是一种工程和实践意义上的人工智能观。

为什么对AI的描述仍然存在这些分歧，这里无疑和早期AI的研究纲领的单薄和宏大抱负、乐观预言的失败有关，德雷福斯的兰德报告起到了导火索作用，使得在70年代初之后AI在美国和英国的声望空前下降。

德雷福斯不仅促使公众、政府机构等利益相关者重新审视AI，而且始终运用德-法传统的大陆哲学思想来为最新的AI研究开方抓药，包括：

（1）胡塞尔现象学中的意向性理论；

（2）波兰尼、庞蒂关于躯体在知识获取和人类行为中如何发挥作用的理论；

（3）海德格尔关于在世存在、上手、预备上手等一系列存在论现象学的观点；

对于早期AI中过于简单的认识论假设，对人类智能行为的不恰当的模拟，德雷福斯也引用了维特根斯坦的“家族相似性”和其语言哲学方面的观念，他的这些工作至少从两个方面对AI的历史发展起到了推动作用：

（1）影响了AI工作者的研究实践，特瑞·温诺格拉德（Terry Winograd）不仅把海德格尔哲学引入到斯坦福大学计算机系的课堂上，甚至于在麻省理工学院的人工智能实验室，80年代后也转向了德雷福斯提出的研究方向，“对那些追随人工智能历史的人来说，

具有讽刺意味的是麻省理工学院的实验室会成为‘海德格尔式的人工智能’的摇篮。正是在麻省理工学院德雷福斯首次明确表达了他的批评，而20年来人工智能实验室的智识氛围（intellectual atmosphere）对承认他言论中的含义明显是敌对的。尽管如此，该实验室现在完成的某些工作似乎受到了海德格尔和德雷福斯的影响。^[16]”

（2）对于认知科学从符号-计算-表征为代表的早期研究纲领向当前以涉身-交互-情景为代表的研究纲领的转变，起到了重要的引导作用。德雷福斯当然不是推动这种转变的唯一代表，作为现象学大家和人工智能哲学研究的旗手，他开启了一个批判的潮流，在其之后，塞尔、Haugeland、丹尼特等人不断挑战GOFAI的形式化、认识论上人与世界的分离观点、事实的离散性和原子性等核心观念。

6 结论

作为一名没有受过控制论、信息论和计算机专业训练的哲学家，德雷福斯用这篇报告闯入了AI学术共同体的核心领地，并提出了振聋发聩的建议，在哲学的认识论转向、语言转向之后，也影响了哲学的心灵转向，某种程度上弱化了人们对认知科学将要消解传统认识论的担心。同时，对于人工智能和认知科学这样研究人类自身心灵和思维能力的工程学科和交叉学科，为传统哲学与这些新兴尖端学科的对话提供了一个正面示例，因为德雷福斯并不是象伦理学家、神学家那样捍卫传统的人类中心论的人文价值，而是从认识论、形而上学方面来挑战AI的研究方法基础，不管AI科学家共同体中的部分成员从一开始如何封闭和敌对，最终还是从外部打开了一扇科学家与人文学者交流的窗户。

注释：

1 https://en.wikipedia.org/wiki/Hubert_Dreyfus.

2 https://en.wikipedia.org/wiki/Hubert_Dreyfus.

3 https://en.wikipedia.org/wiki/Hubert_Dreyfus.

4 https://en.wikipedia.org/wiki/Being_in_the_World.

- 5 http://www.nap.edu/openbook.php?record_id=9547.
6 http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm.
7 https://en.wikipedia.org/wiki/AI_winter.
8 <http://aitopics.org/topic/ai-overview>.
9 <http://aitopics.org/topic/ai-overview>.

参考文献:

- [1][3][6][7] Pamela McCorduck. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*[M]. San Francisco: W.H.Freeman And Company, 1979.
[2] Philip Klahr, Donald A. Waterman. *Artificial Intelligence: A Rand Perspective* [J], *AI Magazine*, 1986(7):54–64.
[4][5] Hubert L.Dreyfus. *Alchemy and Artificial Intelligence* [M]. Santa Monica: RAND Corporation, 1965.
[8][9] Daniel Crevier. *AI: The Tumultuous History of the Search for Artificial Intelligence* [M]. New York: Basic Books, 1993.
[10] 休伯特·德雷福斯. *计算机不能做什么:人工智能的极限*[M]. 北京:三联书店,1986.
[11][12] [16] Hubert L.Dreyfus. *What Computers Still Can't Do: A Critique of Artificial Reason* [M]. Cambridge: MIT Press, 1992.
[13] Howard E.Gardner. *The Mind's New Science: A History of the Cognitive Revolution* [M]. New York: Basic Books, 1987.
[14] Wilson, Keil. *MIT认知科学百科全书*[M]. 上海:上海外语教育出版社,2000.
[15] 斯图尔特·罗素,诺文. *人工智能:一种现代方法*[M]. 北京:人民邮电出版社,2010.

(责任编辑:王保宁)

Alchemy and Artificial Intelligence: The Influence to Artificial Influence Impacted by Hubert Dreyfus

CHEN Zifu

(School of History and Culture of Science, ShangHai JiaoTong University, ShangHai 200240)

Abstract: In 1965, Huber Dreyfus as a consultant in Rand Corporation, presented a proposal titled artificial intelligence and alchemy, which numbered as P-3244 challenged research on Artificial Intelligence (AI) sponsored by Rand Corporation himself in theory greatly. In 1972, Dreyfus published 《What computers can't do: The Limits of Artificial Intelligence》 based on this proposal, which book and other two reports: ALPAC report presented by National Academy of Sciences, LightHill report presented by Science Research Council, are the beginnings of first AI winter. Even after the AI revival in 1980s because of emergence of Expert System, and the second AI winter in early 1990s, the research programme of AI had changed greatly, Dreyfus still stand his ground firmly, he always thinks and analyzes the assumptions on human recognition and ability of problem-solving behind the AI research programme implicitly in a critical way, his in-depth review about AI comes from the standpoint of Phenomenology and Heidegger's which is the core of continental philosophy. No matter what AI scientific community agree with him or not, the meaningful thoughts of Dreyfus really had driven such changes in AI field: from the blind optimism based on intellectualism and Symbolism strong programme in early stage to the prudent view about building intelligent machines at the human-level intelligence at present and more diversified research programme today.

Keywords: Hubert Dreyfus; Artificial Intelligence; AI Winter