武汉理工大学

硕士学位论文

人工智能的哲学反思

姓名：周雅凡

申请学位级别：硕士

专业：科学技术哲学

指导教师：孙德忠

20081201

摘 要

人工智能（Artificial Intelligence）是20世纪50年代中期兴起的一个新型研究领域，既是计算机科学的一个分支，又是计算机科学、控制论、信息论、语言学、神经生理学、心理学、数学、哲学等多种学科相互渗透而发展起来的综合性学科。人工智能从其产生开始，就表现出强大的生命力，引发了思维变革，人脑不再被视为唯一的智能体。人工智能应用于生产与生活过程深刻地改变了人类生产和生活方式，延伸了人脑的功能，实现了脑力劳动的自动化，并极大地延展了人的本质力量，开拓了解放人类智能的道路。

但是，在人工智能带来经济腾飞与社会变革的同时，由于缺乏哲学深层反思的支撑，人工智能陷入了发展困境。然而，在哲学思维与科学机制的融合中，人工智能可能突破其当前的发展困境，成为人的本质力量延伸的最高体现，并将在新的联结模式下发展出新型智能集成体，成为高智能体系。

本文共分五个部分。第一章介绍人工智能的历史与现状，概括性总结了人工智能的内涵、发展与现状，指出人工智能发展遭遇到技术与哲学的问题。第二章分析当前人工智能的发展困境，首先从哲学向度针对阻碍人工智能发展的争论进行反思，肯定了人工智能具备智能的可能性。再从系统化极限、形式化极限、信息符号模式极限三个方面具体分析了人工智能的极限困境。第三章进一步分析人工智能蕴含的哲学问题，指出人工智能是哲学思维与科学机制的共同产物，哲学思维为人工智能的发展提供了认识前提，人工智能也推进了哲学思维的发展。第四章论述人工智能是人的本质力量的丰富和发展，以及智性的人是如何产生和人的意识是如何进化的。通过对智能人的本质力量的对象化、非对象化、自我对象化的分析，详尽地描述了其中的联系，论证了人工智能是人的本质力量延展的最高体现。第五章展望人类意识与人工智能意识的未来。立足于人类意识成长的历史和前景，对人工智能的意识实现进行可行性分析。笔者认为，在哲学思维的宏观指导和深度渗透下，通过将信息符号模式彻底转化为联结模式，由此而形成的集成式智能体使人工智能的意识实现具备了可能性。

关键词：人工智能，发展困境，哲学反思，人的本质力量

Abstract

Artificial Intelligence is in the Mid-1950s the rise of a new research field. It is not only a branch of computer science, but also a comprehensive subject developed from computer science, control theory, information theory, linguistics, neurophysiology, mathematics, philosophy and so on.Artificial Intelligence shows its great vitality from the beginning. It activates man's thinking, the human brain is no longer seen as the only agent. Artificial Intelligence is used in the production and artificial life. It profoundly changed the mode of production or life style, extended the funcations of human brains, made the mental automation come true, and greatly extend the strenght of huminine essence, opened up the path of liberation of human intelligence.

Artificial Intelligence make economic grow and social change, however, at the same time, due to the lack of deep philosophical reflection, the development of Artificial Intelligence gets into trouble which was showed from the analysis of philosophizing and scientific mechanism. Artificial Intelligence will become the maximum reflection of essence of the man's strength after getting out of predicament. It will be a new integrated smart system of intelligent system with the philosophy of science and technology in the link model.

This article is divided into five sections, the first chapter describe the history and current situation of Artificial Intelligence.It generally summed up the connotation, development and present status of Artificial Intelligence.Further more, it pointed out that the development of Artificial Intelligence encountered technical and philosophy problems.The second chapter analyzed the current difficulties of development of Artificial Intelligence.First of all, it reflected on the controversy about the obstacles to the development of Artificial Intelligence from a philosophical point of view.It also confirmed the possibility of Artificial Intelligence, cleared away obstamade progress.Then, it analyzed the ultimate predicament of Artificial Intelligence from three models of specific aspects, systematic limits, formal limits, information

symbols.Chapter third chapter contains further analysis of the philosophical problems of artificial intelligence. Artificial Intelligence that is the philosophy of science and mechanism of co-product of the philosophy of thinking for the development of Artificial Intelligence to provide a prerequisite for recognition, Artificial Intelligence has also promoted the philosophy of evolution. The fourth chapter on the strenght of huminine essence is an enrichment and development, as well as the intellectual who is who and how awareness of how evolution. By the people'sessential strength target, non-objectification, self-objectification analysis, a detailed description of the link which is derived artificial nature of force is the highest embodiment of the extension. The fifth chapter humanity prospect of Artificial Intelligence and sense of awareness of the future. Based on the human sense of the history and prospects for growth, achieve a sense of Artificial Intelligence to conduct a feasibility analysis. The author believes that the philosophy of the macro-guidance and the depth of penetration through the information into a complete model of symbolic link mode, the resulting integrated agent to make sense of artificial intelligence with the possibility to achieve.

Key words: Artificial Intelligence；The development of predicament；Reflections on Philosophy；The strenght of huminine essence

独创性声明

本人声明，所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得武汉理工大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：周雅凡　　日 期：2008.12.14

关于论文使用授权的说明

本人完全了解武汉理工大学有关保留、使用学位论文的规定，即学校有权保留、送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

签 名：周雅凡　　导师签名：孙恒忠　　日 期：2008.12.14

(注：此页内容装订在论文扉页)

引 言

1943-1946 年，在美国宾夕法尼亚大学研制成功世界上第一台程序控制的电子计算机 ENIAC（Electronic Numerical Integrator And Computer），开启了计算机时代。由于计算机科学大部分研究是基于“冯·诺依曼计算机”和“图灵机”的，所以它们绝大多数实际是机器的计算模型，研究的实质是逻辑物理的“思维法则”。思维法则赋予了这个科学机器生命力，试图赋予计算机活的灵魂，使之具有类人的智能性。1950 年，英国数学家和逻辑学家艾伦·麦席森·图灵（Alan Mathison Turing）就在《计算机器与智能》一文中设计了著名的“图灵测验”来证明机器具有智能。如果计算机通过了图灵测试，就证明了计算机和人类一样具有思维。

20 世纪，产业时代的需求也促进了计算机的发展，试图制造出可以像人类一样或能更好地完成智能行为的机器，代替人进行简单重复的产业劳动。而人工智能就是这种人造物的智能行为，成为计算机科学的终极目标。1956 年人工智能学科被美国科学家约翰·麦卡锡（John McCarthy）正式提出，引起了广大科学家与哲学家的重视。美国科学家尼尔森（Nilsson.N.J）指出，“人工智能的一个长期目标是发明出可以像人类一样或更好地完成包括知觉、推理、学习、交流和在复杂环境中行为的机器；另一个目标是理解这种智能行为是否存在于机器、人类或其他动物中。”[①]总的说来，人工智能的目的就是让计算机这台机器能够像人一样思考。那么，首先就必须知道什么是思考，更进一步讲就是什么是智慧。于是，人工智能除了是科学的产物，还是“爱智慧”的研究，是基于人的本质研究的学科，具有哲学属性。

在现实中，计算机逐步帮助人完成其它原来只属于人的工作，并以它的高速和准确提高人的工作效率。一开始，这种作用只局限在简单重复的计算和逻辑推理中，而人类越来越希望计算机不仅是高效的更是具有智能的，从而推进了人工智能的研究计划。1997 年 5 月，IBM 公司研制的深蓝（Deep Blue）计算机战胜了国际象棋大师卡斯帕罗夫（Kasparov）。这个消息令人振奋，虽然背后

---

① （美）尼尔森（Nilsson.N.J）. 人工智能[M]. 郑扣根, 译. 北京:机械工业出版社, 2000.1

隐藏了许多哲学问题，但人工智能研究者有理由相信人工智能是可以战胜人类智能的。随着人工智能的发展和广泛应用引起整个世界深刻而巨大的变化，智能产业化应用于生产和生活过程，深刻地改变了人们的生产方式和生活方式，人工智能部分代替和节省了人的脑力，并极大地延伸和扩展了人的脑力，实现了脑力劳动的自动化。当前，人工智能不仅成为计算机科学的前沿，同时也成为社会发展和人的本质力量延展的最高表现，计算机和人脑在目的和功能上渐趋一致。为了使这种延展能更好更快地发展下去，我们有必要对人工智能的困境进行反思，从而坚定人工智能研究的信念，开辟人工智能类人化的新途径。

事实上，随着各门自然科学对人脑机制和生命本质的不断深入了解，人工智能研究逐步摆脱经典逻辑计算的束缚，在新型联结模式下探索新的非经典计算途径，试图制造集成智能系统体系。研究人工智能的科学家与哲学家在这个智能体面前达成了基本共识，将哲学困境转化为科技发展的强大动力，这就为人工智能在更高阶段上的发展搭建了新的平台。

本文以国内外人工智能发展的事实为依据，首先考察计算机科学智能化的可能性及其进程；然后从理论和实证的角度出发，以认识论哲学理论为基础，深入剖析人工智能的困境，努力揭示人工智能的哲学内涵。全文以人为主体，以人类社会的当代发展为背景，以人类智能为参照，尽可能全面地展开对于人工智能的哲学反思，分析人工智能如何成为人的本质力量的延展，期望能够为人工智能和人类意识的合理化发展提供一种积极的思路。

第 1 章 人工智能的历史与现状

从古至今，人们从未中断过的思考之一便是创造出一种能够代替人的“机器”。例如，亚里士多德曾经想要建造一部能了解主人心意的机器代替奴隶，莱布尼茨也描述了一部会思考的机器如何运作的细节，中世纪西班牙哲学家拉蒙·鲁尔（Ramon Lull 约 1235—1315 年）认为有用的推理确实可以用机械装置完成。这些前瞻性的思想家认为，这个机器可以从机械角度来仿制，也可以从虚拟思维中分化。到了现代有人甚至认为可以从仿生学角度加以复制。随着计算机科学的诞生与发展，人工智能终于从想象变成现实，并迅速成长为现代科学和哲学的最前沿领域。

1.1 人工智能（AI）的定义

人工智能(Artificial Intelligence)，英文缩写为 AI。主要是研究、开发用于模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学技术的前沿科技领域，是自然科学进入信息时代的产物，它企图了解智能的认识实质，并生产出一种新型的、以类人智能模式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等具体科学。最终使智能机器具有通常与人类智能有关的功能，如判断、推理、证明、识别、感知、理解、设计、思考、规划、学习和问题求解等思维活动。

约翰·麦卡锡（John McCarthy，1927 年 9 月 4 日生于波士顿）作为人工智能的“教父”，毕生致力于研究人工智能。于 1948 年按时完成加州理工学院的学业，然后到普林斯顿大学研究生院深造。1948 年 9 月，他参加了一个“脑行为机制”的专题讨论会。会上，冯·诺伊曼发表了一篇关于自复制自动机的论文，提出了可以复制自身的机器的设想，这激起了麦卡锡的极大兴趣和好奇心，开始尝试在计算机上模拟人的智能。他发起并成功举办了成为人工智能起点的“达特茅斯会议”（Dartmouth Conference），提出了人工智能较早的定义：人工智能是指人造机器所表现出来的类似人的智能行为，即自主智能性。

在对计算机能智能化，即否思维、能否理解问题的研究过程中，J·R·塞尔在《心灵、大脑与程序》一文中首次将人工智能划分为“弱人工智能”和“强人工智能”。他认为，“就弱人工智能而言，计算机在心灵研究中的主要价值是为我们提供一个强有力的工具；就强人工智能而言，计算机不只是研究心灵的工具，更确切地说，带有正确程序的计算机其实就是一个心灵。”[①]随后发展成人工智能的两个学派，“弱人工智能观点认为机器不能思维，不可能制造出能真正地推理（Reasoning）和解决问题（Problem-solving）的智能机器，但电脑可以部分地模拟人脑的思维活动；强人工智能观点认为机器能够思维，能制造出真正地推理（Reasoning）和解决问题（Problem-solving）的智能机器，电脑可以完全模拟进而代替人脑的思维活动。”[②]

1.1.1 弱人工智能

弱人工智能学派以塞尔为代表，主要从心身问题入手。弱人工智能研究者认为，计算机程序将有助于理解人的心智，机器对符号的操作是一个形式系统的演算问题，这并不表明能对符号的语义形成独特的理解。不可能制造出能真正地推理（Reasoning）和解决问题（Problem solving）的智能机器，这些机器只不过看起来像是智能的，但是并不真正拥有智能，也不会有自主意识。认为可以在过程模式上“像人一样思考”，也可以在结果上“像人一样行动”。

芝加哥大学哲学系主任 John Haugeland，1985 年提出“像人一样思考”的理论：新的令人激动的努力，要使计算机能够思考……从字面上完整的意思是——有头脑的机器。试图片面的截取及“克隆”人的思维赋予机器。“像人一样思考”系统主要从两个不同的方面着手研究：第一，通过内省，捕捉我们自身的思维过程。第二，将关于思维的足够精确的理论通过计算机程序表达出来。艾伦·纽厄尔（Allen Newell）和赫伯特·西蒙（Herbert Simon）在 1961 年设立了“通用问题求解器”，对程序的推理步骤轨迹与人类个体求解同样的问题步骤轨迹进行比较，试图利用递归分析方法创立一种精确而且可检验的人类思维工作方式理论，为“像人一样思考”的认知模型提供了理论和现实基础。

计算机科学家Elaine A. Rich和Kevin Knight 1991年在《Artificial Intelligence》

[①](英) 玛格丽特·A·博登编著, 博登编. 人工智能哲学[M]. 刘希瑞, 译. 上海:上海译文出版社, 2006. 73
[②]杜文静. 人工智能的发展及其极限[J]. 重庆工学院学报, 2007(1):37

一书中提出：研究如何让计算机能够做到那些目前人比计算机做得更好的事情，使计算机“像人一样行动”。“像人一样行动”的类人行为系统以图灵测试为典型，它描述了一种假想的可实现通用计算的机器，例如：图灵机。使它表面上看着像人，能够通过图灵测试，并对人类思维本质提出疑问。

然而，人类并没有在机器能否思考的问题上停留很久，没有致力于研究如何通过图灵测试。因为人工智能的目标不是制造完全像人一样行动的机器人，来欺骗其他真实的人。只是致力于使机器具有类人行为，重视图灵测试的方法而非图灵测试的结论。

以目前的科学技术水平，包括神经科学、仿生学、计算机科学、认知科学的研究现状看，人工智能的基础研究主要是弱人工智能，研究路线以硬件技术为主，模拟原理主要为逻辑物理信息符号系统（即符号操作系统）假设和有限合理性原理。并且一般认为这一研究领域已经取得可观的成就。

1.1.2 强人工智能

强人工智能以赫伯特·西蒙（Herbert Simon）为代表，认为认知是一个信息加工的过程，是一个信息的输入、存储与加工、输出的心理发生过程。强人工智能研究者认为，人工智能可以模拟人的心智，有可能制造出真正能推理（Reasoning）和解决问题（Problem solving）的智能机器，并且，这样的机器能将被认为是有知觉的，有自我意识的。强人工智能主要有两类：

（1）类人的人工智能——“理性地思考”，即机器的思考和推理就像人的思维模式一样。建立在“像人一样思考”的认知模式之上，以形式化还原主义建造一种新型理性思维法则。表现为联结主义，联结仿生学和生理学等领域，其原理主要为神经网络及神经网络间的联结机制与学习算法。企图“创造”类人的理性思维。美国人工智能界的权威学者、麻省理工学院教授温斯顿(P.H.Winston) 1992 提出：人工智能是对使得知觉、推理和行动成为可能的计算机的研究。

（2）非类人的人工智能——“理性地行动”，即机器产生了和人完全不一样的知觉和意识，使用和人完全不一样的推理方式。表现为行为主义，以进化主义和控制论为中心，其工作原理为控制论及感知——动作型控制系统。

大卫·普尔（David Poole）是这类人工智能实用派的代表人物之一，他于1998年提出：计算智能是对设计智能化智能体的研究。使人工智能成为“理性地行

动的智能体系统”。智能体（Agent）就是某种能够行动的东西。不仅仅具有思维法则的逻辑思维模式，还具有类人的特性。不仅仅强调正确的推理，还要使正确推理有时也是理性智能体的部分功能。

综上可见，当前的人工智能研究形成了两大类别，即弱人工智能和强人工智能；四个代表性观点，即“像人一样思考”、“像人一样行动”、“理性地思考”和“理性地行动”。（如图 1-1 强人工智能与弱人工智能）在这四种形式系统中，达到“理性地行动”的智能体系统是最高级的境界，也成为目前人工智能研究的重心。

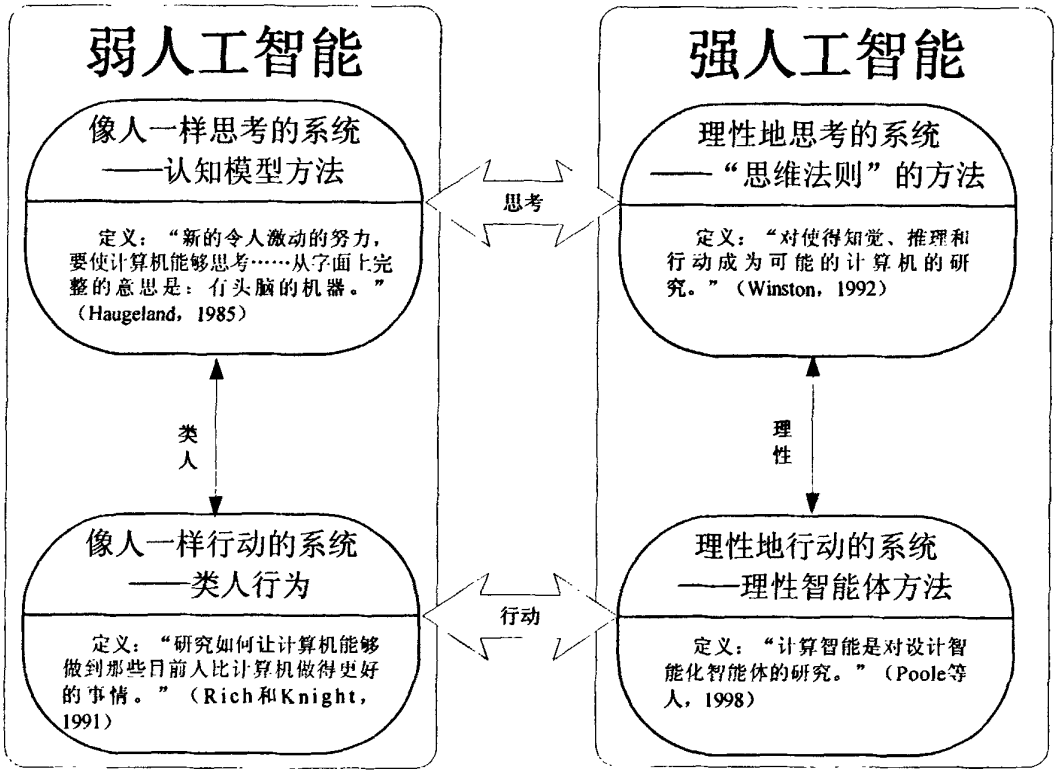

图 1-1 强人工智能与弱人工智能

笔者认为，无论是强人工智能与弱人工智能，对于人脑与计算机的区别都做出了肯定的回答。只是，目前强人工智能的联结主义认为，这种差别是可以通过技术的增强而逐步弥合的，是可以用科学的途径解决。过分地强调了信息科学的机械性，忽略了心灵的本质、神经生理学、心理学等方面人的属性问题。于是，我们看到强人工智能举步维艰。同时，弱人工智能也是片面的，他们一味追求人的“意向性”与计算机未来的“意向性”，从行为上、语义上、心理状态上尽量保持一致。那么弱人工智能也只能是机械的重复运动，失去了自我意

识的“自我”也只能简单地复述人类智能而放弃自主的人工智能。目前，两种人工智能都忽略了具有智能的人是存在与思维的统一体，无论是模拟过程还是模拟结果，都不是智能的“人”。

人是宇宙的精华，人的精神本质是社会科学几千年追问的难解之谜，人的生理构造同样也是困扰着自然科学的最高问题。因此，对于人脑的神经与心灵构造研究将无异于人类最伟大的工程，无法在短期内有更精确的回答。人工智能的研究者们，迫不及待地未等人的剖析完成之后便开始了计算机拟人化的研究。于是，只能摸索着逐步缩小计算机与人的差距，在现有的基础上尽力使之具有智性。然而哲学就是要以超验的思辨方法去解决自然科学不能解决的问题，只有在科学与哲学的相互辅助下，才能逐步弥合争论。在人工智能的研究领域内，哲学思维与计算机科学研究逐步融合，形成人工智能哲学。在这一点上，人工智能研究者们有了普遍的共同认识。

简而言之，人工智能是对人类智能的一种模拟和扩展，其核心是思维模拟，使之具备理性行动的能力。这也就奠定了人工智能必须具有哲学意义——具有人的本质力量，必须从人的哲学本质来改造人工智能，使机器成为智性的存在物并且在环境中展现，使之具有人类智能。

1.2 人工智能的历史和现状

19 世纪末数理逻辑的迅速发展和 1945 年第一台计算机被发明，正式开启了智能化研究。其发展主要经历了孕育时期、发展时期、联结时期、认知科学时期四个阶段：

1）1956 年前的孕育期

20 世纪 30 年代，丘奇(A.Church)、图灵等人提出了计算可定义算法和符号处理的理论概念。1948 年，美国数学家诺伯特·维纳出版了《控制论》创立了控制论；同年，美国数学家申农发表了《通讯的数学理论》创立了信息论，人工智能作为一门学科诞生的技术与理论条件已经基本成熟，并制造了一批数字计算机（实际上为数字计算器）。此时认为人工智能是技术专家的话语，他们用控制论和信息论的方法进行“可计算性与可定义性”(Computability and definability)操作，将计算系统引入了物理世界。

一般认为人工智能最早的工作是由两位神经生理学家 Warren McCulloch 和

Walter Pitts(1943)完成的，提出了一个人工神经网络模型（ANN，Artificial Neural Networks）。涉及了三种资源：基础生理学知识和脑神经元的功能；从伯特兰·罗素（Bertrand Russell）和怀特海德开始对命题逻辑的形式化分析；1950 年，图灵在论文《计算机器与智能》中提出了图灵测试、机器学习、遗传算法和增量学习理论，清晰地描述了人工智能的美好图景。1951 年，普林斯顿数学系的 2 名研究人员马文·明斯基和迪恩·埃德蒙建造了第一台神经元网络计算机，模拟了由 40 个神经元构成的网络。他们还扩展了数理逻辑，为人工智能的计算理论奠定了基础。

2）1956 年以后的发展时期

1956 年夏，约翰·麦卡锡在达特茅斯大学组织了数学、神经生理学、心理学、信息论、哲学、计算机科学、神经学的十名学者参加一个为期 2 个月的学术讨论班，提出了“人工智能”这一术语。他们以另一种创新性思维模式，开始了全方位地解读人工智能，将自然科学的各个领域联系在一起，认为人工智能不仅仅是一种哲学家的话语，也不仅仅是逻辑的符号，开始注重计算机与人的行为、智能方面的深刻联系。从此，观念开始转变，人们认为“人工智能不是关于计算机的研究，而是关于思想与行为中的智能的研究。计算机是它的工具，因为其理论被表达为使机器做人需要智力去做的事的计算机程序。”[①]当时出现了三个研究小组分别对人工智能进行细化研究：

（1）心理学小组

1956年，艾伦·纽厄尔（Allen Newell）、赫伯特·西蒙(Herbert Simon)和乔治·肖(George Shaw)展示了他们研制的“逻辑推理者”——“通用问题求解器”（General Problem Solver），用一个通用程式来模拟人类心理解决问题的方法。他们假定要解决的问题可以根据不同的状态来定义，然后用手段——目的分析法来判定一个问题的目前状态和期望状态或目标状态之间的差异，进而透过选择并应用不同的演算子集来达到目标状态，而演算子集合则定义了解决方案。

但是，由于该程式是基于一般逻辑的模拟，就必然导致产生无数个可能的演算子集，在解决复杂的实际问题时，将花费大量时间并且占用大量的记忆体，从而造成效率低下，导致该计划受阻。此后，虽然研究者们试图透过发明广义的搜寻机制来寻求问题的解决方案，这种经过简单“扩展”来解决庞大问题的

[①]（英）玛格丽特·A. 博登(Margaret A. Boden)编. 人工智能与自然人[M]. 1977. 56

方法仍然受到物理极限的限制，在现实的冲击下，只在“游戏问题”领域有效。

（2） IBM 小组

IBM 小组主要研究自组织、自学习、自适应能力的西洋跳棋程序。从 1952 年开始，阿瑟·萨缪尔（Arthur Samuel）就开始了西洋跳棋的研究，反驳了计算机只能做人想让它做的事情的观念，他的程序很快就学会下得比其创造者更好。这种跳棋程序于 1962 年战胜了美国一个州的跳棋大师，并于 1997 年的“人机大战”中战胜国际象棋大师卡斯帕罗夫。

这种被命名为专家系统的程式盛行于 20 世纪 70 年代初期至 80 年代中期，资讯科学家 Edward Feigenbaum 和 Bruce Buchanan 把基因领域诺贝尔奖获得者 Joshua Lederberg 的专业知识用电脑程式来演绎，使它达到人类专家的水准。主要借助于从人类专家那里提炼出的特定高品质规则，归结为经验定律（rules-of-thumb），运用形式的启发式研究方法，引发从知识匮乏以及弱方法到特定领域和高深知识技术的变革。

但启发式规则表现了知识的抽象形式，缺乏对领域的基本理解，只能显示用来解决问题的规则串列，无法进行相关性累积，这使得确认非正确性、不完全性和不一致性知识的任务相当困难，不具备从经验中学习的能力。一个专家系统不只是需要买一套推理系统或专家系统技术程式然后嵌入足够规则那么简单，它相对独立导致了无法快速扩展，在复杂的系统中同样面临极限瓶颈。因此也只在单调的象棋比赛中取得成功，而无法应用到其他领域。

（3）MIT 小组

1958 年约翰·麦卡锡从达特茅斯离开到麻省理工学院（MIT）任职，与马文·明斯基一起组建了世界上第一个人工智能实验室，做出了三项里程碑式的贡献。

第一，麦卡锡定义了高级语言 LISP，成为占统治地位的人工智能程序设计语言。使符号生动的具有逻辑生命力。

第二，提出了将计算机的批处理方式改造成为能同时允许数十甚至上百用户使用的分时方式（6me-sharing），解决了访问稀少而昂贵的计算机资源问题。

第三，在 1958 年，麦卡锡发表了题为《使用常识的程序》（Programs with Common Sense）的论文，在文中他描述了建议采纳者（Advice Taker），这个假想程序可以被视为第一个完整的人工智能系统。

这一阶段，随着计算机技术发展，发现了计算机高级语言的表示，物理符号算法也得到很大程度的提升。以麦卡锡为代表的科学家着重研究形式逻辑里

的表示和推理，开始探索理性的逻辑思维，逻辑符号主义得到了发展，但是又陷入了哥德尔定理的阴影中。同时，由于人脑的奥秘与思维模式这个复杂系统没有得到深入解读，并没有产生有成效的研究成果。由于在方法论上始终没能突破经典计算的藩篱，遭遇了形式化困境之后，弱人工智能对强人工智能进行了现实地批判。以英国科学家罗杰·彭罗斯（Roger Penrose）和美国分析哲学家J·R·赛尔（John R.Searle）为代表的各种反思和批驳不断涌现，使人工智能陷入技术困境。人工智能只是表面上取得了丰硕的技术和理论成果，实质却进入了低潮期（1965-1975）。

3）1986年以后的联结时期

20世纪80年代以后，物理学家约翰·霍普菲尔德（John Hopfield）使用统计力学的方法来分析网络的存储和优化特性，把节点集合当作原子集合处理，使人们对人工智能的注意力转向了神经元网络领域。美国认知心理学家鲁姆哈特（David Rumelhart），从描述语义网络中长时记忆的特点开始，进行关于记忆的神经元网络模型研究。真正的突破是在1986年，Rumelhart和McClelland在《Parallel Distributed Processing》一书中重新发明了向后传送学习演算法（BPN: Back Propagation Network），该演算法由Bryson和Ho于1969年首次发现的反向传播算法演化而来，当时至少四个不同的研究小组重新发明了此算法，并且应用于计算机科学和心理学中的学习问题。

这一时期的人工智能超越了符号模型和逻辑方法被称为联结主义的智能系统模型。“联结主义质疑了符号处理在认知的精神模型中是否真正有解释作用，这个问题还没有确切的答案，不过大部分人的观点认为联结主义方法和符号主义方法是互补的。”[①]近期，产生了新型的人工神经网络模式，根植于神经科学、心理学、数学和工程学等跨学科主题，比基于符号的推理模式提供更多的自然资料，可以学习，也可以脱离规则建立模式，并且可以处理模糊的不完整资讯。然而，它采用黑箱封闭的工作方式，缺乏解释功能，陷入繁复而耗时的困难阶段。

4）20世纪90年代后的认知科学时期

20世纪90年代，人工智能研究者开创非经典物理算法，并逐步从科学争论

---

① （美）Stuart Russell, Peter Norvig.人工智能——一种现代方法(第二版)[M].姜哲等译.北京:人民邮电出版社,2004.21

走向系统框架融合，尽力达成统一的科学范式。人们发现认知现象兼具结构的复杂性和运动的复杂性，人工智能无力对其进行独立的研究只能寻求跨学科、多层次的模糊研究方式，把智能研究的跨学科趋势进一步发展到更多更广的领域。人工智能的子系统已经取得了初步的进展，研究者们开始重新审视“完整智能体”的可能。艾伦·纽厄尔、约翰·拉尔德和保罗·罗伯逊在 SOAR 系统上构建了著名的完整智能体结构，使环境结束运动把目标瞄准了对嵌入真实环境的智能体的工作进行理解。“一般认为认知科学综合了心理学、计算机科学（人工智能）、哲学、神经科学、语言学、人类学六大领域共同对认知现象进行探讨，认知科学这种综合性研究方式出现的本身已经从一个侧面证明了人们承认人工智能的客观复杂性。”[①]认知科学就是一个极其复杂的大综合智能体，它涵盖了各学科领域，使各领域的争论和冲突被融合在一个完整的系统中。认知科学不仅发展了计算机科学深度的技术研究，也填补了边缘学科的空白。虽然联结模式的认知科学体系仍不完善，但人工智能在认知科学模式下达到了基本统一，使理性行动的智能体理论成为可能。

目前，国际上人工智能技术已经趋于成熟，也在目标上达到共识——制造综合智能体，尽可能地充分认识困境、避免陷入困境。在我国，计算机软件技术在理论上、方法上，已经达到发达国家水平，能够独立自主地开展重大问题的创新性研究，并取得了不俗的成绩。中国人工智能学会理事长钟义信举例说，中国科学家在模式识别领域创造性地提出仿生识别方法；我们提出了可拓学理论，较好地处理了过去在人工智能方面不处理矛盾的问题；中国用机器证明数学定理在全球可谓独树一帜。而在硬件、神经网络、认知科学等辅助学科方面，还未能引起足够重视，缺乏哲学思维的反思，与发达国家存在一定差距。一些智能型机器人还是在美国、日本、韩国等电子信息业发达国家产生。

[①]王哲. 人工智能的复杂性与信息科学纲领[J]. 系统科学学报, 2007, (02)

第 2 章 当前人工智能的发展困境

人工智能经历了 50 多年的发展，在计算机技术方面取得了长足的进步，解决了许多科学技术难题，缩短了各个领域科研项目的研究周期。但随着研究的深入，科学家及哲学家逐步发现人工智能的现实困境并引发了争论。人工智能想继续深入发展，就必须正视这些困境问题。

2.1 人工智能是否能进行真正意义上的思考与理解

作为一门科学，人工智能的终极目标是使机器能够具有人类智性，所以“机器能否思考”这个问题对人工智能学科至关重要。随着科学技术的迅速发展，人工智能认识问题的争论愈演愈烈，从初始的“图灵测试”哲学否定，到强人工智能与弱人工智能的分裂，到 20 世纪 80 年代以后强人工智能的哲学争议，再到 21 世纪的联结主义强人工智能争议等等。这个问题的争论主要在以下三个领域展开：

(1)哲学争论

在人工智能研究初期，科学家提出机器智能化有可能被实现的理论，一些哲学家试图证明人工智能是不可能的，即机器不可能智能的行动。图灵设计了这样一个游戏，称为“模仿者游戏”。假设游戏中有三个角色，一个是人，一个是有恰当程序的数字计算机，还有你。你可以通过一台电传打印机和他们俩交流，提问。你只能通过让他们回答你的问题，来判断哪个是人哪个是计算机。如果你不能正确的找出哪个是计算机，那么它就通过了图灵测试，也就证明了计算机和人类一样有思维。图灵测试拥护者认为，存在一种合适的软件将能够在硬件上运行，计算机可以创造出和人类一样的思考行为。而还有一部分哲学家将“图灵测试”解释为：在 1950 年，图灵在他著名的论文“计算机器与智能”（Computing Machinery and Intelligence）中提出问题，“机器是否能够通过关于行为的智能测试”，而不是问“机器能否思考”。这时争论的中心是站在哲学的立场上看计算机科学的问题，归根到底是哲学争论。一部分哲学家甚至认为人工智能研究亵渎了灵魂。

(2)技术争论

技术争论主要发生在强、弱人工智能两大学派之间。强人工智能研究者认为，人工智能可以模拟人的心智，能够将专家系统、模糊逻辑、感知机和机器学习等多种科学技术应用于现实中完全取代人类智能；弱人工智能研究者认为，计算机程序将有助于理解人的心智，机器对符号的操作是一个形式系统的演算问题，这并不表明能对符号的语义形成独特的理解，并不真正拥有智能，也不会有自主意识。

目前，人工智能研究的基础主要是计算机科学技术的弱人工智能，强人工智能的技术虽然取得了巨大成就，但仍然无法完整地应用于现实，还存在实际的技术困难。例如：强人工智能的“通用求解器”和“专家系统”技术，虽然推动了人工智能发展，能够快速的、专业的解决部分难解问题，取代人类局部专业工作，但在更广泛、更普遍的推广应用上不可避免的受到技术限制。强人工智能在思维模式中试图隔离若干系统，分门别类的系统探究，虽然意识到了心灵的问题，却始终没在技术上融入，无法取得实质性的进展。于是，弱人工智能学派认为强人工智能是无法实现的，用现实证明强人工智能的技术困境是无法跨越的。而强人工智能则始终认为舍弃了人性的智能不具有智性，却由于强人工智能没能在意识和认知方面取得突破性的进展，盲目的在技术上突破科学与哲学的隔阂，一直处于被动状况，无法提出确凿的证据证明强人工智能的可行性。“机器能否思考”的问题还是没能被实际的解决，只是陷入了计算机科学两种意见的相互对垒中。

(3)人工智能认识方式的争论

进入 21 世纪以后，科学技术得到全方面的发展，进入科技大融合时期，人工智能也已经全面进入联结主义时期，强人工智能成为主体。

塞尔设计了一个巧妙的思想实验证明强人工智能存在问题：“假定我被锁在一间屋子里面，并给了我一大批中文文本；而且假定我对中文一窍不通……之后，又给了我第二批中文脚本，并带有一套规则，使第二批与第一批发生联系。规则是用英文写的。我和其他以英文为母语的人一样是理解这些规则的。用这些规则，我可以把一组形式符号与另一组形式符号联系起来。……现在，假定又给了我第三批中文符号，同时还有一些指令，仍是英文的，这些指令使我可以把第三批元素同前两批联系起来，并指示我怎样送回某种特定形状的中文符号，作为对第三批中送给我的那些特定形状符号的相应。……第一批符号叫‘脚

本'，第二批符号叫'故事'，第三批符号叫'问题'，而且把我送回相应第三批文本的符号叫做'对问题的回答'，同时，把他们给我的那套英文规则叫'程序'。赛尔设想，屋外的人给我一些他能理解的英文故事并用英文问一些关于故事的问题，答案是肯定的，我能用英文回答。那么，又假定，一段时间后我擅长遵循指令来处理中文符号，同时程序员也变得擅长编写程序，导致从外部看，也就是据我被关在屋外的那些人来看，我对问题的回答与讲中文母语的人的回答毫无区别。"[①]

在塞尔的实验中，中文房间里的人可以给出回答，屋外的人也可能会认为屋内是一个中文母语者，换句话说，计算机是可以通过图灵测试的。但塞尔并不认为这就能证明计算机具有智能了。我们可以把塞尔的思想总结成一个三段论。大前提，如果强人工智能是真的，那么肯定存在一个关于中文的程序，并且不论在什么硬件上运行，这个程序都能产生理解；小前提，"我"可以运行一个关于中文的程序，但是这个程序不会因此而理解中文；结论，强人工智能的思想是错误的。于是赛尔否认了强人工智能"编程的计算机可以理解这些故事"和"这个程序在某种意义上解释了人类的理解"这两个论断。赛尔并不否认大脑是一台机器或者说这样的机器可以思考，也没有否认或许有一天科学家可以创造出人工智能产物。他只是要反驳强人工智能正确的输入输出就能产生心的理论。认为"编程计算机所理解的，正是汽车和加法机所理解的，就是说，恰恰什么都不理解。"[②]

其争论焦点在于：如果一台机器的唯一工作原理就是对编码数据进行转换，那么这台机器是不是有思维的？是否能理解？约翰·罗杰斯·希尔勒认为这是不可能的。他以中文房间为例来说明，如果机器仅仅是对数据进行转换，而数据本身是对某些事情的一种编码表现，那么在不理解这一编码和这些实际事情之间的对应关系的前提下，机器不可能对其处理的数据有任何理解。基于这一论点，希尔勒认为即使有机器通过了图灵测试，也仍然不算是真正的思考，也不一定说明机器就真的像人一样有思维和意识，也不具有"人性"的行为过程和结果。而只能算是对思考的模拟，是拥有普遍性人的思维，使其表面上看着像人，而并不真正具有自我意识。

---

① (英) 玛格丽特·A·博登编著,博登编.人工智能哲学[M].刘希瑞,译.上海:上海译文出版社,2006.75
② (英) 玛格丽特·A·博登编著,博登编.人工智能哲学[M].刘希瑞译.上海:上海译文出版社,2006.78

也有哲学家持不同的观点，丹尼尔·丹尼特（Daniel C.Dennett）在其著作《Consciousness Explained》里认为，“人也不过是一台有灵魂的机器而已，为什么我们认为人可以有智能而普通机器就不能呢？”[①]他认为像上述的数据转换机器是有可能有思维和意识的。有的哲学家还认为如果弱人工智能是可实现的，那么强人工智能也是可实现的。这一部分科学家和哲学家坚持当人类意识被完全的肢解，人工智能就能被完整的组合，还是具有一定的实现性。

无论是赛尔、希尔勒还是支持强人工智能的研究者，都没有否认强人工智能的可能性，普遍认为可以通过图灵测试。而现实是：强人工智能在认识理解上还是存在缺陷，它通过了图灵测试，仍然是表象地拥有智能，而没有真正地具有人一样的思维，在认识方式上还存在偏差。

那么，至少在突破了哲学困境、技术难题和认识方式困境之后，人工智能才能系统的发展，可能突破思维理解的困境。只有被自我肯定了，研究者们普遍统一了人工智能的发展方向，才能减缓人工智能发展的阻力，更迅速的投入到深层研究中去。这个困境是存在的，但不是永恒的，随着科学与哲学的融合，虽然不能完全解决这个困境，却可以被逐步消融，构成统一的科学共同体。

2.2 人工智能的极限

目前，人工智能的研究已经进入了强人工智能阶段，弱人工智能已经初步实现，研究者们也普遍认为强人工智能是可能实现的。然而，一部分科学家和哲学家在用逻辑的“思维法则”诠释系统时发现，人工智能的实现存在极限，强人工智能的逻辑系统企图用还原论的简单性科学纲领研究人类心智活动这种典型的复杂性现象，而这个逻辑还原论系统本身就存在着不完备性的极限。并且，这种还原论的对象是世界，复杂的世界在形式化还原的过程中同样存在不可跨越的极限。形式化的机器语言对世界的认识和把握只能用演绎推理和分析的方法，而人类还可以使用归纳推理、类比推理。人类心智还具有概括、综合、联想的能力，其复杂性远远超出强人工智能的现状。这也正是强人工智能可能实现而没能实现的重要原因。

[①] CONSCIOUSNESS EXPLAINED by Daniel Dennett.1991.By permission of Little,Brown & Company(Inc.),pp.3-7

2.2.1 系统化极限

美国逻辑学家，哲学家普特南（Hilary Putnam）如此评价道："20世纪早期逻辑领域出现的两位巨人哥德尔(Kurt Gödel)和杰克斯·赫伯德(Jacgues Her-brand)对于人工智能研究做出了重要的贡献。"[①]

1929 年，哥德尔解决了一阶逻辑的三大系统完全性，表明任何逻辑被蕴涵的语句都有一个有限的证明。打破了当时人们广泛意识——数学中的非有穷推理只是在能够靠有穷主义元数学来"解释"或"核正"的限度内才有意义，避免了把非有穷主义推理从元数学中排除。"哥德尔明确地把一阶逻辑的完全性问题当作纯粹数学问题求解，大胆使用选择公理、无穷引理和无穷域上的排中律，敢于向哲学上的偏见挑战，最终迈出了具有实质性意义的无穷推理这一步。"[②]

1930 年 9 月，他在哥尼斯堡举行的第二届精密科学认识论会议上正式宣布了第一个不完全性定理结果："假定古典数学是一致的，甚至能举出按内容为真但在古典数学形式系统中不可证明的例子。"会后把第一定理改为算术形式，同时建立了第二不完备定理。

哥德尔第一不完全性定理是指任何足以展开初等数论的数学形式系统，如果是一致的，就是不完全的，即其中必定存在不可判定命题。哥德尔第二不完全性定理是指任何足以展开初等数论的数学形式系统，如果是一致的，其一致性在系统内不可证。即任何足够丰富的数学形式系统，如果是一致的，那么它不能证明表达它自身一致性的命题是定理。哥德尔第一和第二不完全性区别在于："一个是要指出，在形式不可判定的算术命题存在的意义上算术是不完全的；另一个是要指出，算术的一致性不仅能够在算术理论自身中表达，而且表达算术一致性的命题本身就是一个特殊的形式不可判定命题。"[③]德国数学家外尔生动地道出了数学两难的困境："上帝是存在的，因为数学无疑是一致的；魔鬼也是存在的，因为我们不能证明这种一致性。"这种困境导致了："存在被蕴涵的语句，但是系统内没有有限的证明。针可能存在于隐喻的干草堆中，但是没有任何程序能够保证将它找出来。"[④]

[①] Hilary Putnam.Words and Life.[C]//Edited by James Co-nant.Cambridge,Mass:Harvard University Press,1994.391

[②] 刘晓力.理性的生命——哥德尔思想研究[M].长沙:湖南教育出版社,2000.67

[③] 刘晓力.理性的生命——哥德尔思想研究[M].长沙：湖南教育出版社,2000. 73

[④]（美）Stuart Russell, Peter Norvig.人工智能——一种现代方法(第二版)[M].姜哲等译.北京:人民邮电出版社,2004.225

简单说来，依照哥德尔手稿中的说法，数学的不可完全性可表达为如下几个等价命题："（1）数学是不可穷尽的。（2）任何一致的形式数学理论必定包含不可判定命题。（3）没有定理证明计算机（或程序）能够证明数学中所有的真命题。（4）没有数学形式理论既是一致的又是完全的。（5）数学是机械上（或算法上）不可穷尽的（或不可完全的）。"[①]

于是，哥德尔不完备定理能够证明："存在不能被证明的真值算术语句。"[②]简要的说就是，"对于任何能力强到足以描述算术的形式公理系统 F，都可能构造出一个所谓的哥德尔语句——G（F）。G（F）是 F 的一个语句，但是不能在 F 中被证明。如果 F 是一致的，那么 G（F）为真。哲学家 J.R.Lucas（1961）曾声称哥德尔定理表明机器在智力上是比人劣等的，因为机器是受到不完备性定理限制的形式系统——它们不能确定自己的哥德尔语句的真实性——而人类则没有这样的局限性。"[③]

人工智能模拟智能思维是通过逻辑思维法则算法的形式化系统实现，那么这个系统就必然存在不完备的极限。

2.2.2 形式化极限

生物学研究证明：人脑中的神经细胞对信息接受或不接受是由兴奋和抑制两种状态进行，并且人脑的神经脉冲传递也是按有和无的规律进行，这样我们可以采取 1 和 0 的二进制来模拟人类的思维活动。莱布尼茨发明的二进制是一种科学的数学方法，是构架模拟信息世界和现实世界的桥梁，使自然世界形式化成为了可能。从莱布尼茨的二进制到丘奇-图灵的物理算法再到逻辑算法，随着计算机技术的开发，让狂热的弱人工智能研究者看到了自然世界完全形式化的曙光。强人工智能也紧紧拥抱这根救命稻草，仍然将新联结模式认识框架建立在此基础之上。然而，在人工智能研究者那里如果只是把计算机科学作为我们探究心灵哲学的一个有力的工具，形式化极限的产生使强弱人工智能同时陷入了形式化困境中无法自拔。

[①]刘晓力.理性的生命——哥德尔思想研究[M].长沙：湖南教育出版社,2000. 54

[②]（美）Stuart Russell, Peter Norvig.人工智能——一种现代方法(第二版)[M].姜哲等译.北京:人民邮电出版社,2004. 225

[③]（美）Stuart Russell, Peter Norvig.人工智能——一种现代方法(第二版)[M].姜哲等译.北京:人民邮电出版社,2004. 731

波普尔将世界一分为三："第一，物理客体或物理状态的世界；第二，意识状态或精神状态的世界，或行为的动作发生倾向的世界；第三，思想的客观内容的世界，尤其是科学思想、诗的思想和艺术作品的世界。"[①]从实在的自然到外化的世界再到人文的灵魂，人工智能具有无法摆脱的形式化极限。

2.2.2.1 第一世界——本体极限

自然是博大的，世界也是运动发展的，在宇宙中还存在许多未解之谜，虽然这些谜是未知的，却并不是不可知的。随着科学技术的发展，这些谜团最终也将逐一攻破。如果我们把人类和自然定格，用黑箱方法静态的看待已知的世界，我们可以发现形式化的过程中，仍然存在着极限。

（1）物质运动着的联系无法被形式化

自然世界是运动发展的，遵循质量互变规律，是由量变引发质变，再由新的质变发生新的量变，运动发展的联系组成了整个自然世界。假设第一世界已经被完整的展现在人们的视野中，"世界变化有一个由量变到质变的过程，量变是相对的，质变是绝对的。"[②]而被计算机技术形式化后的信息却只能产生量变，也不可能产生互变的联系。没有质变的量变是有极限的，量变无法取代质变。也就是说，第一世界形式化是可能实现的，然而形式化后产生的信息模拟并不能完全取代第一世界。例如，水遇冷逐步固化能成为冰，遇热逐步汽化能成为气，气凝聚在一起成为水。它的本质仍然是 H_2O，这些量变过程就是无法被形式化的，量变中物质的联系也不能被形式化，质变的瞬间更不能被准确捕捉到。于是，能被形式化的仅仅是气、冰或水的特性。无法智能的判断三者其实是同一本质的结论，也无法在形式化领域感性描述其中的联系。甚至，哥德尔认为："有许多联系，今天的科学和正统的哲学对之还一无所知。"[③]由于，这些联系存在认识极限，同时存在被形式化的极限，因此形式化的信息并不能完全涵盖第一世界。

（2）第一世界形式化必然经过认识的过程才能被形式化，因而不能达到完全的形式化

第一世界是复杂的自然系统，不是凭靠简单的还原能实现的。物质是有存在的属性，包含存在的意义、存在的价值、存在的形式等。这种属性有两个方

[①]波普尔.科学知识进化论[M]. 纪树立,编译.北京:三联书店,1987.309
[②]周建.没有极限的科学[M].北京:北京理工大学出版社,2006.176
[③]刘晓力.理性的生命——哥德尔思想研究[M].长沙:湖南教育出版社,2000.9

面，一个是物质存在属性，一个是被认识的存在属性。在被认识的过程中，它就不仅仅是一个简单的存在物，它有存在的意义、价值、作用。不论是积极的还是消极的，它在不同的时间、空间，不同的事件中产生不同的认知。如此复杂的认识存在属性面临着形式化极限。

并且，物质存在属性是无法自我形式化的，它必然经过人为认识处理之后才能被形式化，那它也必然不能达到完全形式化的统一。因为，绝对的黑箱并不存在，黑箱被打开之后，物就不再是原来的物，必然受外界影响产生变化，形式化后也必然不是对原物的形式化。而被形式化的是认识，用规范认识物质存在，理性地描述感性认识的价值存在。那么形式化过程离不开认识的作用，思维与存在将不能在形式化过程中得到统一。

2.2.2.2 第二世界——现象的外化极限

人在对象化世界的同时，产生了意识。意识在实践中产生，也在实践中被表象外化出来，表现为知识、文化、社会等因素。它不是一成不变的，是运动发展的。随着意识的发展，形式化的意识——知识、社会也将是不确定的，即现象外化后也存在形式化极限。

（1）人工智能在知识来源上区别与人，存在知识极限

狭义的知识就是人类将认识形式化的过程和结果。由于历史是不断向前运动发展的，知识虽然具有一定的叠加性，但在形式化过程中，还存在超穷距离，不用超穷手段，不但填不满这一距离，甚至也没有希望去逼近。

在波普尔看来，第二世界是不断变化、不断进化的，之所以是进化的，是因为认识主体与环境在不断地发生相互作用。人是主动的，在不断的追问中自我认知、自我意识，因而认识也是进化的。人将原始自然对象化，再融入新的自然产生新的知识，如此反复，第二世界在认知主体与环境中不断相互作用，不断进化。“智能人之所以显现出其高智商，是因为智能人的智能是进化的，他可以通过学习、实践来改进自己的概念框架结构，不断地丰富与发展自己的聪明才智，提高自己的智能。”[①] 然而，从实用观点来看，人工智能是一门知识工程学：以知识为对象，研究知识的获取、知识的表示方法和知识的使用。到目前为止，人工智能还只是一个孤立的系统，“人工智能领域取得的最大成就莫过于‘专家系统’，但是这些系统说到底仅是高速的信息库搜索者”[②]，“机器需要的

[①]郑祥福,洪伟.认识论的自然化之后[M].上海:上海三联书店.2005.219

[②] Hilary Putnam.Words and Life.[C]//Edited by James Co-nant.Cambridge,Mass:Harvard University

大量信息只有通过自然语言加工才能弥补”[①]。那么，其发展的前提是智能信息知识库被充分形式化，并在现象外化的过程中受制于狭义的知识，于是必然存在知识的极限。

（2）个性外化极限

共性是物质的普遍性存在，个性则是不同的外化结果。人工智能的形式化，可以模拟数据复制物质存在属性，造成它存在的假象，甚至能让人产生感觉它确实存在。但这种形式化还是由人类的感觉认识转化成而来。普遍性的共性可以被普遍认识并模拟出来，是由于在形式化前已经进行过规则处理，而千差万别的个性将如何被形式化。我们不否认，存在那么一些规则和公约能将世界定性、定量分析，能将认识物质存在属性统一形式化，然而这项浩瀚的工程也将具有极限。“人工智能一直在设法摆脱这样一个事实：并没有任何优雅的、不费力的方法去获得这个巨大的知识库，至少开始时需要大量精力人工输入每一个断言。”[②]例如，一个孩子和一个青年人同拿一个苹果，苹果是真实存在的，它不会产生不同的物质属性，孩子和青年人都会认识到苹果是红色的共性，却产生偏大和偏小的认识差距。他们认识差异同样也是确实的，在产生统一的过程中，普遍认识的极限必然存在，个性认识的形式化极限也随之产生。

（3）认识论的自然化面临社会化极限

美国哲学家奎因（Willard Van Orman Quine）在1969年反对逻辑化认识论的过程中提出了自然化的认识论。认为必须从自然的或物理的角度去理解人类认识，把人的认识作为自然现象来研究。奎因认为，认识论应该是自然科学的一个分支，认识论是包含在自然科学中，而自然科学也包含在认识论中。在奎因看来，所有的理论都依赖于一种假定，即世界是存在的，我们的对象世界是实在的，但我们的认识并不存在一个阿基米德点，根本不存在独立于科学与认识的实在。认识论所要做的是找到一条与自然科学相一致的途径。

20世纪，“七八十年代以来，西方科学哲学的认识论研究基本上倾向于这种自然化的认识论研究。”[③]奎因的自然化认识论超越了“第一哲学”的困境，取消了哲学原有的优越感，把哲学与认识论科学化、自然化。使自然科学脱离哲

Press,1994.393

[①]Hilary Putnam.Words and Life.[C]//Edited by James Co-nant.Cambridge,Mass:Harvard University Press,1994.395

[②]Guha, Lenat .Cyc: A Midterm Report [J].AI Magzine,Fall,1990(33):33-59.

[③]洪伟. 科学认识论的五形态及其对人工智能的意义[J]. 自然辩证法研究 ,2004,(02)

学母体，彻底解放。将认知论自然化后，人类冲破了自然科学与哲学的界限，打破了传统认识论的逻辑形式，建立一种以科学为基础的新的自然主义的形而上学与认识论。让科学制约哲学，带来了信息科学，尤其是人工智能领域突飞猛进的发展。然而，自然主义的黑箱操作，始终无法将证明的信念自然化，隐含的知识不能形式化的被展现出来。另外，社会实践的意义也被排斥了。

笔者认为，奎因的自然化认识论完全抛弃了本体论，放弃了客观世界，排斥了实践的意义，将科学神化了，使认识论沦为了科学的工具，陷入了认识论自然化之后的困境。“认识论的自然化是人工智能模拟的前提，因为只有把认识论科学化，才能用物理符号对认识过程进行模拟。相反，对日常化认知的模拟是建立在自然化的基础上。”[①]然而，“日常认识论是生活实践的、相对的、整体的、协同的、非决定性的，遵循复杂性原则。”[②]自然认识论在对人工智能的考察上模糊了两者的区别，忽略了人具有社会性，是实践的产物，而机器却没有的重要事实。在人工智能研究过程中，部分人工智能专家也敏锐的发现了问题。1969 年，麦卡锡与海斯就详细表述了人工智能中面临处理常识知识的难题，并且提出了他们对常识知识表示的思路：“这对常识的研究就要求去形式化因果关系、能力和知识，一些也是哲学逻辑要考察的。”[③]脱离了社会环境的人工智能是机械的机器，这样的智能体不能说是有智性的，更不可能存在并运行在现实社会，必然成为认识论自然化后的社会形式化极限。

2.2.2.3 第三世界——灵魂表征极限

人工智能是人工和智能的融合，人类智能的展现是多方面的。那么，人类智能的人工化就必然存在以下问题：人文思想世界是否能完全形式化？人的心智活动能否完全形式化？形式化的意识是否具有自我意识？是否存在能够完全实现人类所有精神活动的计算机？

维特根斯坦早期持逻辑原子主义观点，认为心智表象的语言可以被严格的形式化。企图从哲学的角度建立“语言图灵机”。他在晚期指出：“一个概念之下的各种现象并不具有唯一一种或一组共同性质，他们只有彼此的相似，不存在所谓的共相。因此语言的完全形式化是一个不可能达到的幻影，只有学术意

[①]郑祥福,洪伟.认识论的自然化之后[M].上海:上海三联书店,2005.219

[②]王哲.人工智能的复杂性与信息科学纲领[J].系统科学学报,2007,(02)

[③]John McCarthy.Formalizing Commonsense: Papers by John McCarthy[M].Ablex Publishing Corporation, 1990. 22

义而无现实意义。”[①]

赛尔中文屋的实验也证明，一个完全不懂中文的机器人能够给在屋外的人正确的中文答案，这个程序在某种意义上解释了人类的理解，本身却完全不理解含义。在句法的层次上，塞尔所做的操作完全就是计算机所做的，它们运行操作的形式符号不可能有意思，它们甚至不是符号操作，因为这些符号毫无意义。用语言学的行话来说，它们只是句法的，而不是语义的。所以，计算机也不能理解中文。

哥德尔也否认了灵魂的形式化。他认为，一方面，人心不能将全部数学直觉形式化，如果人心把某些数学直觉形式化了，这件事本身便要产生新的直觉知识（如该系统的一致性）；另一方面，不排除存在一台定理证明机器确实等价于数学直觉，但重要的在于，假定有这样的机器 M，由不完全性定理，我们不可能证明 M 确实能做到这点。

1960 年，大师麦卡锡研制出表处理语言 LISP（List Processing language）。这是一项划时代的贡献，它作为建造智能系统的重要语言工具在人工智能的各个领域得到了广泛的应用。LISP 语言实现了数据类型和句法结构简单化，甚至简单到具有唯一性和一致性；同时程序可作为数据被处理，数据也可作为程序来执行；LISP 一切功能由函数来实现，程序的运行就是求值，并且可以产生多个数值，给问题表示到来了方便；LISP 具有表的结构形式和规模的灵活性，不必预先设定；具有收集无用单元的功能。这种函数型语言为人工智能注入了新的生命力。可是，LISP 语言主要存在以下缺陷：LISP 的句法结构，特别是多层表的括号配对写法极为繁复；数据类型少(常用的只有表和原子)，使其表达能力受限；规范性差(无标准版本，不同的 LISP 文本间差别较大)；程序的可读性不及一般的高级语言，导致在表征中产生极限。

综上所述，从哲学角度、数学运算角度、计算机程序语言角度来看，可以实现的表征形式化都不同程度的存在缺陷，具有极限性。

2.2.3 信息符号模式极限

信息符号模式是人工智能的基础，同时也是认识论自然化的发展。主要以形式化算法研究为重心，以简单性科学研究纲领为指导，还原论是其核心的方

[①]王哲.人工智能的复杂性与信息科学纲领[J].系统科学学报,2007,(02)

法论。这种模式下造成的局限性主要表现在：

（1）实现方法方面，进行简单的还原。这是认识论自然化最直接的结果，将认识简单的还原、定义、使用。将普遍规律奉为科学的结论，忽略偶然性的发生。人工智能是复杂的跨学科多层次的综合体，人工智能对世界形式化还原的极限验证了对复杂表象无法彻底地简单还原。普特南明确指出："存在很多的认知心理学不是还原式的。事实上，我们没有任何理由做出下面的断言：对于人类认知的研究要求我们将认知要么还原为计算要么还原为大脑程序(brain process)"。[①]并且，"基于逻辑的机器、以纯算法的手段是不可能真正产生像'心'一样的东西"[②]。于是，认知时代的人工智能必然受到还原论的限制。

（2）实现环境方面，必须建立理解意义上的条件性假设。也就是说，信息符号模式下的人工智能不能在现时代人类社会存在，需要为其打造一个与之匹配的虚拟条件性假设环境。由于语句是特定的，语境环境是不确定的，语句只有在条件性假设的语境环境下才具有现实意义，不具有社会性。

（3）实现人工智能意识方面，使单一的机器人对人意识的模拟特定化。即不能让机器人做太多过于复杂的事情，必须使机器人处于相对单一的空间中进行智能型处理。机器人还不是智能共同体，只是各系统进行有目的、有针对性的分工协作。由于单一的机器无法承载过多不同方面的定义，只能进行特定化模拟，才能确切地处理事件，使人工智能成为了产业化的简单重复，成为机器工具，而非智能体，不具备真正的智能性。

信息符号模式实质上是对世界简单的形式化处理，于是不可避免的会陷入形式化极限。信息符号模式虽然引发了计算机科学革命，带动了计算机科学发展，也将在相当长的一段时间为人工智能等计算机科学提供原始的技术还原。但是我们必须清醒的认识到信息符号模式的局限性，简单的技术还原无法全面系统的达到形式化，信息符号模式下的人工智能也不能在现实环境中运行，人工智能也只是孤立地完成单一的任务，沦为工具。于是，寻找新型智能体模式成为必然。

---

① Hilary Putnam. Renewing Philosophy [M]. Cambridge,Mass:Harvard University Press,1992.18

② 郭艳华,周昌乐.2000b.汉语语句依存关系网协动生成方法.杭州电子工学院学报（4）

第 3 章　人工智能蕴含的哲学问题

哲学启蒙了人类思维，科学则将思维具体地实现了。在现代社会，科学引起发了人类文明质的飞跃，使人类物质生活与精神生活都得到了迅速发展。然而，科学不是万能的，它无法为“思维着的智性是什么？”这个问题做出合理解答。人工智能实质上是科学与哲学结合产生的新型学科，是人类对自我追思的技术发展。它从哲学追问中诞生，也蕴含着许多哲学问题。

3.1 人工智能的思维基础和科学机制

人工智能存在以上的困境和极限，乃是由于人们没有透过人工智能困境的现象看到本质，没有看到人工智能作为一门综合的自然科学也是需要哲学认识的支撑。人工智能虽然是计算机科学发展的高级阶段，建立在科学机制之上，但它同样也不能脱离哲学的思维基础。

3.1.1 哲学是否就是科学

进入现代社会后，科学技术的发展成为社会发展最重要的动力，大部分科学家甚至一部分哲学家都认为，哲学就是科学。

例如：马赫、维持根斯坦和皮亚杰等哲学家认为没有专门的哲学问题。马赫认为，“哲学家们并未打算解决宇宙之谜，他们仅仅带头消除妨碍科学探究的假问题（false Problems），而把其余的问题留给实证研究。”[①]维特根斯坦则认为，没有哲学问题，所有真正的问题都是科学问题。所谓的哲学问题都是假问题。受维特根斯坦的影响，很多分析哲学家认为哲学所能做的只是揭露和消除传统哲学所提出的语言之谜，而不是去解决哲学问题，因为他们认为哲学没有真正的问题。

在《我们关于外间世界的知识——作为哲学中科学方法的一个领域》一书中，罗素就写道：“自古以来，哲学就比其他任何学问分支做出了更大胆的断言，

[①]马赫著.认识与谬误[M].李醒民,译.北京:华夏出版社,1999.17

但获得了较少的成果。我相信，结束这种不能令人满意的状况的时机现在已经到了。”之后还断言，哲学“本质上是一门科学，它区别于特殊科学之处仅在于它所研究的问题的普遍性。”奎因也坚持：“新哲学认为，所有的知识都是科学知识，有待用科学的方法来确证和证明”，他们实质上消解了科学问题与哲学问题之间的区别。他们的观点与布伦坦诺和梅农的观点颇为相似，后两者认为，“哲学要么是科学，要么什么也不是。”[①]科学长足发展，疆域不断扩大，哲学自我分化也愈来愈严重，没有形成统一的体系，分散在各个领域。于是哲学逐步失去了立足之地，除去了哲学，就只剩下科学了。

于是，哲学被认为附属于科学，被具体科学解释。在人工智能领域，最初哲学也不被重视，用具体科学简单的还原，将哲学自然化、形式化表象出来。然而，这种哲学的自然化极限也已经被逐步显现出来。科学固然是发展的动力，它将科技成果产业化、科技化，放大了社会作用，却忽略了哲学是科学发展的根基，哲学为科学提供理论基础和方法论指导。

从与科学的区别来看，哲学是用超验思辨的方法，探讨科学尚未回答的问题时形成的学问。在科学突飞猛进的今天，哲学仍大有可为，原因在于，哲学始终关注并致力于解决如下科学所不能解决的问题：（1）实在性问题。哲学可以从总体上把握世界的物质客观实在性，意识作为物质的产物、人脑的机能和对客观物质的反映，也必须统一于客观实在性。而科学却脱离了实在，例如：网络虚拟于现实而存在，如果不用哲学的实在解释，科学也将是虚幻的。（2）综合性问题。哲学首先发现本原的总体综合问题，在逐步的研究中发展出分门别类的具体科学，并衔接产生边缘科学弥补研究漏洞和空白。（3）本原性问题。科学说明解决的是因果关系，而将原始的因抛弃，社会性本质也无法解释。（4）反思性问题。这也是科学的前提性问题，科学从不研究自己的方法，只有哲学才为科学提供有效的方法和坚实的理论基础。（5）价值性问题。科学的问题是：是什么？为什么？能怎么样的问题？而没有考虑过价值的问题。而科学的价值恰恰在于是否能引起广泛的哲学思考，成为发展的动力。

因此，哲学与科学之间是有距离的，哲学并不是科学，但哲学与科学又紧密相联。脱离了科学的哲学是传统的、保守的哲学；脱离了哲学的科学是盲目的、无系统的科学。科学离不开哲学，两者辨证统一的相互作用，倘若科学独

[①]张掌然.问题论[D].武汉大学, 2005

立于哲学，不被哲学认识，那它也只是无本之源，沦为改造世界的工具，不是能动的科学。人工智能就是要为僵化的科学注入生命本质，成为类人智能活的灵魂。

3.1.2 人工智能是哲学思维和科学机制的共同产物

人工智能不是单一的科学，而是哲学和科学联姻的结晶，它有着深刻的哲学思维基础：

（1）计算机科学不研究实在问题，将现实推向传统哲学。计算机科学面临现实实在的问题，却在虚拟空间用科学的方法解答，陷入了逻辑循环中。不回归现实本身，从哲学实在角度看待问题本身，人工智能只能是计算机科学的发展，也只能是物理逻辑下的理论，与现实彻底割裂开来，有着不可跨越的极限。在强人工智能模式下，科学家与哲学家一同构建了新型智能框架，哲学的思维基础被并入了智能体体系中，解救了建立在形式化虚拟模型基础上的人工智能。

（2）计算机科学将系统肢解为分门别类的研究，产生了计算机领域下的若干子学科，例如：数学、物理学、仿生学、心理学、计算机算法等。各领域的成果是显著的，而人工智能的成果几乎为零，我们无法找到智能的踪影。因为我们无法将这些成果有效的拼凑。只有用哲学的思维为边缘科学服务，构架一座座联结的桥梁，计算机科学才能成为一个复杂而庞大的体系，可以为人工智能研究输送源源不绝的理论数据和实践成果。

（3）计算机科学分析了形式的因和果，却忽略了哲学本质内涵，忽略了社会性和本源性本质。在计算机科学中，研究者们要达到的成果是制造智性的类人工智能，看到的因是替代人类部分工作，而哲学意义上的实质本因是要挑战人自身的思维意识。不追问人的思维本质，塑造的人工智能也只能是脱离现实社会的虚伪设想。

（4）计算机科学不具备反思性，无法反思自身，也无法形成独立于哲学之外的科学范式。“只有当实验和试探性理论相互连接在一起使之达成一致时，发现才会突现出来，理论才会变成范式。”[①]哲学是科学的科学，为科学提供方法，使科学共同体形成相对统一的科学范式。在美国科学史家，科学哲学家库恩（Thoms Samual. Kuhn）看来，范式是某些实际科学实践的公认范例，包括定

[①]库恩著.科学革命的结构[M].金吾伦,胡新和,译.北京:北京大学出版社,2003.49

律、理论、应用和仪器在一起，具有哲学反思特性，为特定连贯的科学传统研究提供模型。也是一种对本体论、认识论和方法论的基本承诺，是科学家集团所共同接受的一组假说、理论、准则和方法的总和，这些东西在心理上形成科学家的共同信念。目前的科学研究都是在科学范式的指导下进行的。

人工智能科学范式的产生，使哲学理论与科学实践有机的统一起来，使人工智能科学共同体建立在哲学思维与科学机制之上共同发展。从信息符号模式到联结模式再到认知模式，人工智能这个科学共同体，已经运行在哲学方法范式的科学指导之下，只是范式的选择和使用在不断进化的科学中调整。并且我们可以看到人工智能范式朝着系统化、综合化的方向发展，有利于全方位的实现机器智能。

（5）计算机科学是盲目的，没有价值目标和取向，只有哲学的导向才能使计算机科学更迅速的深层发展。价值的问题，科学家不能用事实的方法解决，科学能定性、定量的分析，却无法解释应该怎样的问题。

人工智能研究的价值不在于能研制出多么高明的智能体机器，而在于人类需要对智能的哲学解读，体现人的价值。人工智能研究并非要使人工智能超越人而凌驾于人之上，而是对人自我反思、自我追问、自我验证的现实证明，并使人工智能更好的体现人的价值。于是，人工智能不仅仅建立在科学机制，而是哲学思维和科学机制的产物。

3.2 人工智能与哲学的相互作用

人工智能不仅是在哲学作用下产生的，智能科学本身也蕴含着哲学认识问题，并且对传统哲学进行新的反思。爱因斯坦曾说过："认识论如果不同科学接触，就会成为空洞的图式。科学如果脱离认识论……，就成为粗俗的、混乱的东西。"[①]

3.2.1 哲学思维为人工智能的发展提供了认识前提

1）哲学思维对人工智能在产生和发展方面的启发作用

哲学是一种前科学状态的研究，在发展中逐步对若干问题进行分类研究，

[①]转引自张守刚，刘海波. 人工智能的认识论问题[M]. 北京：人民出版社，1984. 1

当某些问题的研究渐趋成熟时，它就从哲学中独立出来，成为有序的具体科学。分门别类独立出来的就是具体科学，而界限模糊的边缘问题，就成为新的哲学类型。此后，一个有了完满研究方案的问题已不需要哲学理论或某个明确的哲学概念了。罗素明确指出：一旦人们对某个对象有确定的知识，该对象就不再属于哲学的范围了，而通常是一门新的、独立的学科的形成。

由于，计算机科学的大部分研究是基于“冯·诺依曼计算机”和“图灵机”，它们实际研究的是机器的计算模型。主要是以人为模型核心进行研究，而人们对人这个研究对象暂时还不可能有明确的界定，并且人工智能技术也未完全成熟。所以，人工智能还不能摆脱哲学的范围，人工智能的产生就无可避免的带有哲学认识论的必然性。

哲学思维始终为人工智能提出价值性问题，并促进其发展。人工智能科学在分门别类的解答中逐步丰富自身。在亚里士多德那里，人类就开始为意识求解，莱布尼茨也试图制造一部没有灵魂会思考的机器。在人工智能这里，哲学理论部分成为了现实，认识被自然化了。意识的解读被科学化为心理学、仿生学、认知科学等具体学科；思维哲学被具体化为计算机语义结构、计算机仿记忆搜索算法等细微末节。这些科学是相对独立的，却无一例外的发源于哲学，哲学就是科学发生原始的因。

综上看来，人工智能科学是从哲学思想中启蒙，通过研究过程中分门别类独立出来的部分学科综合而产生的新型科学。在人工智能早期，简单的信息符号主义在哲学理论母体中诞生之后便企图彻底分离人工智能与哲学的关系，从事计算机科学单方面的研究。而事实证明人工智能始终无法脱离哲学的认识前提。进入了联结主义时期之后，才逐步弥合了两者的分裂斗争。

2）哲学认识论中的人工智能

近代，认识论已经成为哲学思维的首要问题，“认识论研究的是近代哲学的中心问题，也是现代哲学的思维对象和方式”[①]。它的基本任务在于提出和回答“人对对象的认识何以成为可能和如何达到科学”这个基本问题。

经验论和唯理论作为认识论的两个流派，是近代哲学和科学认识发展的产物，它们的分歧代表着科学与哲学的相互作用和融合。在人工智能领域体现为：哲学家强调通过最大机会的观察和计算，明确什么是潜在可知的；而人工智能

[①]欧阳康.哲学研究方法论[M].武汉:武汉大学出版社,1998.202

专家注重通过现有的观察和计算途径，弄清什么是可知的，用认识论研究知识及其形式和局限性。在实际情况中，“人工智能和认识论在本质上是互相交融和兼备的。”[①]主要表现在以下三个方面：

第一，认识起源的分歧

认识起源的分歧归根到底就是追问“知识的来源”问题，即经验论主义的“白板论”与理性主义的“天赋观念论”之间的争论，“这是经验论与唯理论分歧的一个焦点。”[②]而人工智能在面临形式化极限之前，需要解决的首要问题也是知识的来源问题。

经验论者认为，一切知识都起源于感觉经验，人心在获得感觉经验之前就是一块空无所有的“白板”，强调感觉的经验和反思的经验，视这两种观念为知识唯一的来源。提出人的心灵开始时就像一张白纸，而向它提供精神内容的是经验的观点。“我们的全部知识是建立在经验上面的；知识归根到底都是导源于经验的”。[③]然而，这种消极被动的“白板论”是唯物主义反映论，无法解释“如果心灵是‘白板’，则每个人的心灵是否就应该是一样而毫无差别的了？”莱布尼茨在肯定经验对于知识的基础作用的同时，同时指出了经验作用的局限性。他认为，感觉对于我们的一切现实认识虽然是必要的，但不足以向我们提供全部认识。然而印证一个一般真理的全部例子，不管数目怎样多，也不足以建立这个真理的普遍必然性。简而言之，莱布尼茨认为从感觉经验中得不到普遍必然性，“他在一定意义下认为一切观念都是天赋的，但它们只是作为潜在的禀赋或倾向而不是作为‘现实’的观念天赋于心中。”[④]

唯理论与经验论相反，以不同的形式肯定“天赋观念”。他们在肯定普遍必然性的同时，把“真理性”的认识界定为“天赋的”。在唯理论者看来，理性活动有逻辑的必然性，思维本身具有超越感官经验的先天认识原则，认为那种有普遍必然性的理性知识不能来自感觉经验，它的某些基本原则乃是人心中固有的或与生俱来的“天赋观念”。笛卡尔把“我思故我在”作为哲学的第一原则，把心灵实体作为最先确定的认知对象，把“天赋观念作”为知识的基础，构成了唯理论。而这种形而上学的“天赋观念”也自然地将唯理论推向了唯心主义

[①]董军,潘云鹤.人工智能的认识论问题.科学, 2002, 4: 34-36.
[②]赵敦华著.西方哲学简史[M].北京:北京大学出版社,2001.209
[③]北京大学哲学系外国哲学史教研室编译.十六——十八世纪西欧各国哲学:第二版,北京:商务印书馆,1975.366
[④]陈修斋主编.欧洲哲学史上的经验主义和理性主义[M].北京:人民出版社,2007.252

先验论，割裂了思维与存在的统一，使唯理论者不得不承认人的日常认识大多来自感觉经验，却仍然坚持那种具有普遍必然性的知识不是起源于感觉经验而是由理性固有的原则得来，是“现实”的“自明真理”。以至于笛卡尔也不确定的说道：“天赋观念”并非永远呈现而只是某种潜在功能。

因此，在对认识起源的理解中，我们既要承认一切思维和知识的内容都应当起源于感性的经验，又要从“形式”方面探讨思维的固有的能动作用。需要特别指出的是，人的认识不只是要形成关于经验对象的“表象”，更重要的是形成关于对象的“普遍必然性”的思想。“普遍必然性”的思想并不是直接地、现成地从感官经验的感性内容中归纳、概括、抽象出来的，而是在认识的来源上就必须肯定思维把握存在的规律即思维的能动作用。所以，在认识的来源问题上，我们既要超越习以为常的“经验”立场，又要挣脱唯理论者的“天赋观念论”，这就需要我们唯物地、辩证地理解认识的来源问题。

认识的本质被揭露出来了，我们否定了纯粹的“天赋观念论”，也否定了纯粹的经验论。正如人工智能里计算机不能通过“天赋观念论”而摒弃感觉经验，因为“天赋观念”的心灵原则是未知的，可以通过经验来表象。计算机也不能用经验叠加突破思维理性，因为经验并不是知识唯一的基础，需要理性概括、判断、推理及假说的间接抽象。要达到真正的智能，只有正确的自我认识世界、认识信息、认识事物。能否理性的运用普遍必然性吸收和接纳感性的经验产生的思维和知识，能否产生自我认识、自我意识、自我追思是人工智能突破的关键。

第二，认识方法问题的分歧

经验论和唯理论在认识来源问题上的对立，已经蕴含着关于认识方法或逻辑的不同理解。这种不同理解，就是在个别与一般问题上的对立，即普遍必然性的认识能否从个别的感性经验中形成？人工智能主要经历了四个基本阶段，进行了几个模式的转变，同时也是认识方法的辨证发展过程。

经验论者强调经验中个别的、特殊的东西，认为“各种事物自身并没有普遍性，……各种事物的存在都是特殊的”[①]，偏执归纳和分析的方法，从个别中概括出一般的东西，或者把一切对象分解为个别的东西。近代经验论的奠基人弗兰西斯·培根从知识起源于经验这一原则出发，批判亚里士多德的三段论式的

---

① 洛克.人类理解论[M].北京:商务印书馆,1959.395

演绎逻辑，形成他的“新工具”的归纳逻辑。即由个别的感性经验上升为普遍必然性认识的“归纳法”。“归纳法是在观察实验的基础上，得到相应的感性经验材料；将感性材料整理和归类，区分为正面的例证、反面的例证和在不同条件下其性质有变化的例证，并对这三类例证进行比较研究；排斥掉非本质的东西，舍弃无用的材料，留下有用的材料；对留下的材料进行归纳，得到一般原理。”[①]培根指出，“正如现有的科学不能帮我们找出新事物，现有的逻辑也不能帮助我们找出新科学。”[②]这是因为，在培根看来，只有归纳的方法，才能使认识从经验事实开始而上升为一般原理，这种归纳法能够使人们循序渐进地揭露出不同等级的自然规律，防止人们超越经验作无谓的猜想。“归纳法可以从有限的事例中寻找出有共性的东西，比较容易把握，而归纳法的短处则是试图以局部来说明整体，这在逻辑上是有疑问的。”[③]出此，归纳法只适用于认识内部连续的事件，对个别事物的归纳不足以证明一般命题真理的普遍性和必然性，从过去发生的事件不能推导出将来也必然会发生同样的事件，适用于认识的初级阶段。

与经验论相反，唯理论者强调一般的、普遍必然的东西，断定普遍的必然性原理不能从个别的感性经验中得来，只能来自“天赋的”心灵自身。侧重演绎和综合的方法，从普遍必然出发，推导出特殊个别，并致力于把特殊个别综合为整体。近代唯理论的奠基人笛卡儿认为，在追求真理的出发点上，必须首先探求出一种无可怀疑的原则，并在这个原则的基础上去形成普遍性的思想。这就是笛卡儿的新演绎逻辑。从理性主义的知识观出发，笛卡尔认为人类可以达到理性范围内的一切形而上学真理，“笛卡尔将直觉和演绎当作获得科学知识的根本方法，并进而用普遍怀疑对之进行修正和补充。”[④]并在《谈谈方法》一书中，笛卡尔确定了著名的四条认识规则：(1)“对于我没有清楚认识为真的东西，我决不把它当作真的接受。也就是说，要小心避免判断的仓促和偏颇，并且只接受其中清楚明白呈现给我的心灵、以致我不能有任何怀疑的判断”；(2)“根据可能和必要，将我所考察的每一难题分成多个部分，以便尽可能地将这个难题妥善解决……”；(3)“按照恰当的次序进行我的思考，从最简单、最容易

[①]王天成.创造思维理论[M].长春:吉林教育出版,1989. 7
[②]培根.新工具[M].北京:商务印书馆,1984. 10
[③]周建.没有极限的科学[M].北京:北京理工大学出版社,2006.176
[④]周晓亮.直觉与演绎：笛卡尔的方法论选择及其困境[J].云南大学学报,社科版,2005 年,第 1 期

理解的对象开始，以便一点一点或逐渐上升到认识最复杂的对象……”；(4)“在一切情况下都进行完全的列举和普遍的审查，以致我确信没有任何事情遗漏”。[①]莱布尼茨也从一位友人送给他的中国“易图”（八卦）里受到启发，最终悟出了二进制数之真谛，发展了数理逻辑，为计算机的现代发展奠定了坚实的基础。

笛卡儿的新的演绎逻辑，是一种关于思想自身发展的内涵逻辑，是思想从单纯上升到复杂、从抽象上升到具体的内涵逻辑。虽然是不被广泛认可的唯心思想，却带有辨证发展的内涵，在德国古典哲学集大成者黑格尔那里，构成了概念辩证发展的关于人类思想运动的逻辑。但是，演绎法同样具有认识的局限，“演绎法必须要有一个适当的逻辑基础来作为该逻辑体系演绎的逻辑起点，虽然演绎体系本身是逻辑的，但逻辑系统的建立和演绎系统的更替却是非逻辑的，并不存在一个现成的道路来引导我们如何建立新的逻辑系统。”[②]适用于认识的高级阶段。

经验论与唯理论最终都无法否认对方方法论的积极作用。并且，是否所有的感性经验都能转化完全为普遍认识？普遍必然性是否能演绎出个别的、特殊的东西？包括情感、欲望、语言、思想是否都能完全演绎为形式化？这些对经验论和唯理论的反思同样造成了新的人工智能困境。于是，在黑格尔第一次揭示了个别和一般、归纳和演绎、分析和综合的辩证关系之后，经验论的归纳法和唯理论的演绎法逐步被统一于认识的发展过程中。在人工智能领域，归纳经验是智能学习发展的必要条件，而辨证上升联合的逻辑演绎模式也是智能化统一发展的充分要件。人工智能是用归纳的方法使计算机科学上升到理性思维高度，用演绎推理也将计算机科学的成果以具体的形式表现出来。

第三，认识可靠性问题的分歧

认识的可靠性问题，即感觉经验与理性知识何者为核心。经验论者认为，归纳感觉经验是人的认识唯一可行的可靠方法，而演绎理性知识如果不脱离实际的无聊游戏，顶多也只是把归纳所发现的真理阐述出来的形式技巧。“错误或虚假倒不是在感官里，感官并不主动，它只是接受影像，……错误或虚假是在判断里，或是在心灵里；判断或心灵没有给予应有的周密细致地对待，没有注意到离得远的东西只是由于离得远或由于别的原因，而应该比它们离我们较近

[①]笛卡尔.谈谈方法.载于 The Philosophical Works of Descartes, volume I,92.
[②]周建.没有极限的科学[M].北京:北京理工大学出版社,2006.176

时显得小和模糊；在别的情况下也是这样。”[①]这就把“错误”或“虚假”归咎为“判断”或“心灵”，也就是归咎为人的“理智”。认为理性知识才是虚构出来迷惑心灵的，感觉经验的个别性、特殊性归纳才是现实而可靠的。

与此相反，唯理论者则认为感觉经验是“骗人的”，把归纳看成不值得信任的“意见”，只相信清楚明白的理性直观及其推理是最可靠的。笛卡儿说：“因为我曾经多次观察到：塔远看像是圆的，近看却是方的，竖在这些塔顶上的巨像在底下看却像是些小雕像；像这样，在无数其他的场合中，我都发现外部感官的判断有错误”[②]。“荷兰唯物主义者、唯理论者斯宾诺莎继承笛卡儿的认识论思想，把知识分为三类：第一类是感性经验知识，包括传闻的知识和泛泛的经验；第二类是推理知识；第三类是理性的直观。他说：‘只有第一类知识是错误的原因，第二类和第三类知识是必然真实的。’可见斯宾诺莎也排斥感性经验，只信赖理性知识的可靠性。”[③]

经验论和唯理论的争论在近代哲学发展过程中，逐步的从彻底分裂极端的第一阶段走向相互承认感觉经验或理性知识各自合理性的第二阶段，却一直把感性与理性割裂开来。其实，人类哲学思想从它开始产生的第一天起，就在自身之内包含着一个深刻的矛盾：它来自于经验，但又是超验经验的结果；它是理性思维、范畴和概念的运动，但又只有经验才能推动它。经验论和唯理论都竭力把对方的方法结合、消融在自己的方法中，却始终无法真正的结合。正如在康德那里，“思维、知性仍保持其为一个特殊的东西，感性也仍然是一个特殊的东西，两者只是在外在的、表面的方式下联合着，就像一根绳子把一块木块缠在腿上那样。”[④]黑格尔则在哲学史上第一次提出了辩证统一问题，“经验论和唯理论的方法论的错误都是由于违背了这一辨证原则而走向形而上学的片面性的结果”[⑤]，经验论与唯理论这一艰难探索过程恰好暗示了一个思维方法的内在规律：归纳和演绎是辨证统一的。“没有一种归纳可以是没有预先的假设而盲目进行的，……而演绎的形式原则本身也是在人类认识亿万次重复的实践活动中形成的起来，离开了经验认识和归纳，其可靠性是没有依据的。”[⑥]

[①]伽桑狄.对笛卡儿沉思的诘难.[M]北京:商务印书馆,1963.75
[②]北京大学哲学系外国哲学史教研室编译.十六——十八世纪西欧各国哲学.北京:商务印书馆,1975.179
[③]外国哲学史研究集刊:第5辑,上海:上海人民出版社,1982.14
[④]黑格尔.哲学史讲演录:第4卷[M].北京:商务印书馆,1983.271
[⑤]陈修斋主编.欧洲哲学史上的经验主义和理性主义[M].北京:人民出版社,2007.285
[⑥]陈修斋主编.欧洲哲学史上的经验主义和理性主义[M].北京:人民出版社,2007.284-285

在人工智能的研究中，弱人工智能认为理性演绎运用的抽象观念是缺乏实在性的，只有经验论的归纳方法才可靠，它试图尽力搜集所有个别且特殊的东西，归纳概括出一般的普遍必然性，并赋予给计算机，结论是无法等价于或超越人类智能；早期的强人工智能则注意到演绎的理性逻辑，却由于未将感性的经验归纳融入，始终无法达到人工经验与智能理性的统一。长此以往，人工智能要么陷入虚构世界，欺骗虚构他们的心灵；要么陷入理性循环，不断的证伪可能是虚构的事实。要想达到真正的智能，只有两者相互被承认，又同时被对方怀疑，成为螺旋式上升的辨证过程。也只有在赋予计算机感性经验认识的同时，添加理性逻辑分析，才能可靠地完成真正的“人工”和“智能”。

3.2.2 人工智能的产生促进了哲学思维的发展

哲学是一种对元问题的追问。哲学思考什么问题是可研究的，可解决的，什么问题是不可研究的，不可解决的。在人工智能领域，哲学的元问题追问，也就是人工智能的发展方向，即何种人工智能思维方式有解。此时，哲学思考为人工智能提出问题和方法，人工智能科学去解决可解问题。然而运用哲学方法解决人工智能问题的同时又可以发现新问题，再对新问题进行哲学反思。于是两者都得以继续循环发展。

1)人工智能对元哲学的诘难——促进哲学理论的发展

元哲学是以哲学为自身对象，而不以具体的哲学问题为对象，注重对语言的形式化研究，从而分析、检验和判断以往哲学的合理性和有效性。

对这个“哲学究竟应该研究什么?”的哲学本原问题，古典实用主义者詹姆斯的回答是：“所有的问题，包括存在着什么事物的问题，都简化为什么事物将有助于创造一个更加美好的世界的问题。”[①]杜威的回答是：哲学应该学会把关于来世这些让人迷惑不解的问题的给宗教，把有关认识过程这类棘手难题让给心理学。莱布尼茨则是运用哲学的方法和数学的基础发明了二进制，随后科学家与哲学家一致将哲学难题推向了未知的人工智能。莱布尼茨通过对元哲学的反思，力图寻求一种可以获得知识和创造发明的普遍方法。这种努力导致许多数学的发现，并运用对中国的科学、文化和哲学思想的总结，系统地阐述了二进制计数法，为计算机的现代发展奠定了坚实的基础。于是，对元哲学的诘难

[①]转引自罗蒂.后形而上学希望[M].张国清,译.上海:上海译文出版社,2003.178

造就了人工智能的基础——二进制。而人工智能本身是哲学思维发展的产物，具有哲学内涵，人工智能的元哲学反思也使人类的目光从具体哲学回归到元哲学本身，回归到哲学家们花费两千多年时间试图理解并解答宇宙中的两大疑难问题：人类是如何思考；人类之外的物体是否有思维。早期，“人工智能试图借助于符号操作来帮助人类理解人的智能过程，因此它提出了在什么条件下我们才能有理由将心的状态归之于一个纯粹的物理系统问题。它对于认知科学和心的哲学发展做出了很大贡献”[①]，打破元哲学的传统范围，重新审视和判断了以往哲学的合理性和有效性，促进新型哲学的发展，注入了新的动力。

2)人工智能对传统哲学的诘难——促使科技哲学的产生

哲学是问题之学，其基本职责是在批判中发现问题和提出问题，而且是本原性和前提性的问题。在罗蒂看来，问题也有发现出来的问题和制作出来的问题之分，前者是自然地真实存在的，是自然科学技术问题；后者是人为地虚构制造出来的，是传统哲学问题。后者可以通过语言转换而被清除掉，前者则未必。真正的哲学是应当用来解决真实存在的哲学问题的。而现在确实存在着“与令我们的祖先困惑问题不同的问题要去解决。”这个所谓的不同问题既指哲学对象性问题，也包括元哲学问题，即传统哲学问题的性质的问题。

也就是说，现代，人们企图解决的不是传统哲学问题，而是自然科学的哲学问题，不能纯粹使用传统哲学，因为它只是可以被语言转换而清除掉的问题。是既包括哲学对象——具体科学，又包括元哲学的问题，进入科学哲学领域。

正如美国哲学家普特南指出：“总的来说，我认为科学哲学领域中将持续讨论的问题，是那些我们认为几乎与科学哲学完全无关的问题，那些完全可以、甚至更加恰当地被看作是属于思维哲学或语言哲学领域的问题，诸如模拟人脑计算机模式的可能性和重要性问题；语言的计算机模式的重要性问题；关于真理理论、真理性质、真理检验的性质问题；以及即便不存在一种严格的科学方法，科学怎么能够成为客观的等问题。”[②]因此，人工智能是一个不同于传统哲学的问题，结合科学技术的方法排斥传统哲学的阻碍，解决现实存在的哲学性问题。

[①]尼古拉斯·布宁编.西方哲学英汉对照辞典[M].余纪元,译.北京:人民出版社,2001. 80

[②]麦基编.思想家[M].周穗明,等译.北京:生活·读书·新知三联书店,1987.365-366

3.3 人工智能的社会功能与社会后果

人工智能经过50多年曲折的发展，在“人类活动的所有各方面——包括工作的性质，学习的方法，政府的结构，战争的手段，艺术创作以及人类对自我的概念等——产生深远的影响”[①]，产生了积极的社会功能和社会后果，主要包括以下三个方面：

第一，人工智能促进了科学、技术、哲学的融合

人工智能是多学科多领域的共同作用的产物，人工智能也导致了许多新兴学科的发展和融合。例如，融合了认知科学、神经网络科学、信息科学、计算机科学、信息论、控制论。人工智能的研究成果表明，虽然人工智能尚未解开任何古老的心灵认识之谜，但是它为我们提供了规范和拓宽哲学想象力的新方法。然而，科学一般也会存在不能解决的问题，需要借助哲学的方法论。并且，尽管科学是从哲学中逐步演化而来，哲学同样也不能替代科学解答具体科学的问题。两者相互交叉，相互作用，科学有哲学的外延部分，哲学有科学的理论基础。

第二，人工智能使科学更贴近人本身

计算机心理学家认为，人工智能可以帮助我们认识心灵——它是什么，它是怎样工作。人工智能再将这些被认识到的基本常识性知识形式化（用某种形式语言，如用编程语言，表示为形式得当的公式），并且将其公理化（这里的公理表示的是物理世界和社会现实中的普遍真理），继而把它推广到常识性思维中去。

人工智能也使人更仔细的关注大脑，包括：脑部内部结构，利用建立计算机模型揭示一个已知的（单细胞的或多细胞的）神经解剖学结构，解读怎样使大脑能够加工某些特定类型的信息；同时，进行外部功能模仿，使表象更与人接近，促使更深层地研究人的本质，更全面地探究认知科学。

第三，人工智能代替人完成了许多工作

在哲学对社会的需求分析下，人工智能取得了比较丰硕的成果。例如：人工智能改变了传授计算机科学和实施软件研发的方式；人工智能使语音识别系统、库存控制系统、监视系统、机器人和搜索引擎等具体应用成为可能；人工

[①]（美）库兹韦尔.灵魂机器时代——当计算机超过人类智能时[M].沈志彦,祁阿红,王晓东,译.上海:上海译文出版社,2006.6

智能改变了人类的生活，改变了人类工作和娱乐的真正本质，使人类从单一的计算和重复的思维中解放出来，开始真正的智性思维；人工智能的发展颠覆了人类对传统智能、意识和人类未来命运的构想，进入的新的信息智能时代。人工智能也完成许多人类无法完成的物理极限，在航空和水下探测等领域，都有人无法到达的物理极限，人类研究出小型智能机器人代替人登月，在人为设置的程序下，采集了信息和数据，带回了部分有用的实体。

第4章 人工智能是人的本质力量的丰富和发展

人工智能研究的基础是人，而人又是什么？“人是一种对自己的存在不断进行自我认识、自我探究的存在物。”[①]从人类开始思维世界的同时也在认识人本身，主要从物质和精神两个方面追问。史料证明，世界演化的过程中形成了物种，猿人通过对自然世界的改造从众多物种中进化出来形成了智能的人，又因为人的智性意识使人的进化从自然转向了人本身，产生了形式化的人的本质力量。人类现在又试图用人工智能的科学方式将人的本质力量这一种无形的能力展现出来，并将人的本质力量发展壮大。

4.1 自然界必然产生思维的智性人

人是自然中具有自我意识的智能体，随着自然进化而来，具有历史必然性。主要经历了形体进化的初级阶段和意识进化的高级阶段，最终形成自然界中高等生物——智性的人。

1)人形体的产生

地球，从形成至今大体上已有四十五亿年到四十六亿年之间。在它形成之后约十几亿年的时间内，地球上只有机械运动、物理运动和化学运动的无机物，当时既没有生物，更没有人类，因而也就不存在意识现象。这段时间是生命的史前期。在自然界长期发展的过程中，大约在三十三亿年左右，才产生了生命物质，以后发展成两大类：植物和动物。

恩格斯在《自然辩证法》中谈到：“拉普拉斯以一种至今还没有人超过的方式详细地证明了，一个太阳系如何从一个单独的气团中发展起来；以后的科学愈来愈证明他是正确的。”[②]在太阳系的行星上最初是热的那种物质运动形式占优势。卫星、小行星和流星逐步的冷却下来，随着冷却，相互转化的各种物理运动形式的交替就愈来愈显著地出现，化学亲和性开始起作用，化学元素得以区分，获得了化学性质。随着气温的下降，当行星上有了一层硬壳的时候，一

[①]夏甄陶.人是什么[M].北京:商务印书馆,2000.1

[②]恩格斯.自然辩证法[M].北京:人民出版社, 1984:26

部分气态物质向液态、然后又向固态过渡，形成了积水。最后温度降低到至少在相当大一部分地面上不越过蛋白质能在其中生存的限度，有生命的原生质便形成了。经过了许多万年，这种没有定形的蛋白质能够由于核和膜的形成而产生了第一个细胞。有了整个有机界形态发展的基础，形成了无数种无细胞膜的和有细胞膜的原生生物。一些逐渐进化为最初的植物，另一些逐渐进化为最初的动物。动物中，进一步进化发展出无数的纲、目、科、属、种，最后进化出神经系统获得最充分发展的那种形态——脊椎动物，然而其中的一种在自然界获得了自己的意识——这就是人。

2)人类意识的产生

意识是自然界长期发展的产物，它的产生，经历了一个长期的、复杂的物质运动过程：从无机物质的反应特性到有机低等生物的刺激感应性；从低等生物的刺激感应性到高等动物的感觉和心理；从动物心理到人的意识。

（1）动物心理

1500 万年以前，动物界中出现了人猿——森林古猿。同所有低等生物一样具有刺激感应性意识，并开始逐步与低等动物区分开来，具有高等动物的感觉和心理。

晚期的猿人已经具备两足直立行走的身姿，迈出了从猿到人具有决定意义的一步。由粗壮体型向纤细的体型转变，更有利于直立行走，前肢被逐步解放出来，手的专门化、灵巧化意味着工具的出现，开始了生产劳动。进入石器时代后，创造了大量坚固耐用、工艺精致的石器工具。利用工具开拓了洞穴，自卫，甚至猎取其他动物，一方面开始了杂食生活，摄取了更丰富的蛋白营养；另一方面增加了动作的多样性、灵巧性和准确性，在劳动力富余的情况下开始了分工，迈开了从猿到人的转变的历史性步伐。在一切大自然的困难面前，都显示出智慧的光芒。

（2）人类意识

森林古猿的不断进化，到距今三百万年左右，地球上才出现了人类——早期智人，成为有思维意识的群体。

早期智人，在劳动对象、劳动工具、劳动行为本身建立了比较稳定的联系，其中认知能力和实践能力有了明显的长进。婚姻和社会制度都得到了完善。人类语言也有了长足的进步。由此，人类形成史上，早期智能人是基本形成的人。

晚期智人，发达的语言能力已经形成，学会了复杂石器制造之外的骨器磨

制，开始有选择的制造生产工具和装饰用品。用磨尖的鹿角狩猎、捕鱼，用兽皮兽骨、木桩、茅草、树叶构建房屋和帐篷。随着生产工具水平的提高，生产技术和生产力得到了提高。装饰用品和雕像图腾的制造，代表着已经开始形成了意识，包括：审美意识、宗教意识。实行族外婚，悼念死者的社会制约代表着人类已经形成了人伦道德观念。到此，我们已经可以看到文明的雏形。

（3）自我意识

远古蒙昧时代的原始人，并没有把自己同自然界的动物区分开来。但“当人开始把自己同其他动物加以比较时，就意味着人已经萌芽了关于自己存在的自我意识。”[①]自我意识的觉醒标志着智能人的诞生。

随着工具的使用区分了狭义的动物和人。狭义的动物也有工具，而这只是它们身躯的肢体，它们的生产对周围自然界的作用，在自然界面前只等于零。人却用手通过劳动实践给自然打上自己的印记。随着手的发展头脑也一步一步地发展起来，语言在手的劳动中和劳动一起生产出来，开始有意识和意识的交流理解，猿脑就逐渐地过渡到人脑。在进一步的发育中，感觉器官也伴随着人脑发展起来。形成了现代有对自然界改造的劳动和通过感觉器官对自然界认识的思维映像的人。现在有一种信念牢固地确立了：物质按照规律在其永恒的循环中运动，它在一定的阶段上——时而在这里，时而在那里——必然地在有机存在物中产生出思维着的精神。自然界运用运动不灭原理无限发展，必然产生思维的智性这个地球上“最美丽的花朵”——具有智能的人。

4.2 人的本质力量存在的形式

人的本质力量就是人认识、改造、顺应、协调自然的智慧与力量，由活动（运动）的形式转化为物质存在形式，以对象化的形式展现出来。人的本质力量主要存在对象化、非对象化、自我对象化三种形式。从与人工智能的关系来看，人的本质力量对象化意味着人工智能产生的可能性；人的本质力量非对象化形式展现了人工智能产生的必然性；人的本质力量的自我对象化则进一步验证了人工智能发展的现实性。并且，人工智能也成为人的本质力量延伸的最高体现。

---

① 夏甄陶.人是什么[M].北京:商务印书馆,2000.3

4.2.1 人的本质力量对象化——人工智能产生和发展的可能性

人的本质力量对象化是指人通过实践使自己的本质力量转化为对象物。“人的本质力量包括人的自然力、智力、意志、情感、创造力等的总和。人的感觉、感觉的人性，都是由于它的对象的存在，由于人化的自然界，才产生出来。”[①]

马克思告诉我们，任何事物要显示自己的存在，都必须依靠周围对象并通过它所作用的对象来予以确证。换言之，人要显示自身的存在，要确证自身的本质力量，就必须把这种力量外化于对象，使这种力量在对象身上得到实现。因此，人的本质力量必然还有一种通过对象显现的现实性存在形态。

在人进入劳动的历史时空后，人类利用一切资源，包括肢体、工具改造自然，使自然界人化，提高人类生存能力。人的本质力量，就是作为主体的人的属性，它是人在物质的或精神的劳动实践中形成的，它以内蕴的形式具体地存在于各个时代具体的人那里。对象对他来说如何成为他的对象，这取决于对象的性质以及与之相适应的本质力量的性质，因为正是这种关系的规定性形成一种特殊的、现实的肯定方式。人的眼睛对对象的感觉不同于耳朵，眼睛的对象不同于耳朵的对象。这是由人的内在本质力量的独特本质决定的。在人的生命历程中，它渐渐积淀为一种随时可以外化的潜能。因此，可以说，这是人的本质力量的一种可能性存在形态。

自然界的人化就是人的本质力量的自然对象化过程。以一种物质的力量作用于自然界，改造自然界，这一过程就是人类通过实践创造对象世界，占有自然界的过程，其结果则产生了获得人的本质的自然界。这种获得了人的主观性、人的本质的自然界，就是所谓的“人化自然”，为人工智能实现提供了可能。

因而，人工智能也是人的本质力量对象化的一种形式，试图将人的本质力量（包括人的自然力、智力、意志、情感、创造力等的总和）形式化为统一的逻辑符号形式，通过形式化对象显示虚拟性的存在形态。于是，人的本质力量对象化为人工智能形式化提供了哲学理论支撑，使人工智能成为可能。

4.2.2 人的本质力量非对象化——人工智能产生和发展的必然性

1）人类意识产生之前——非对象性的存在物是非存在物

[①] (德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.87

在远古，人这种生物很长一段时间是受自然支配的，自然限制了人类活动的范围，制约了人类的生存能力。人只是从自然中演化出来的一种复杂的生物，依赖于自然而存在。

此时，古猿还不能称之为真正意义上的人，它还只是无意识的低等动物，是不被认识的非存在物。“假定一种存在物本身既不是对象，又没有对象。这样的存在首先将是一个惟一的存在物，在它之外没有任何存在物存在，它孤零零地独自存在着。”[①]猿人就是孤零零的独自存在，在它之外没有任何存在物存在，或者说其它物质存在只是毫无意义的存在。于是，要使自然具有存在的意义，智性思维的产生是充分必要条件。

2）人类意识产生之后——人直接是自然存在物

早期猿人所具有的“意识”，基本上依然是动物式的意识，“它们自身的本能还没有清晰地被意识到，还不可能处于自觉意志控制之中”[②]。于是，人是被自然改造的。此时的自然是相对非对象化的，而不是绝对的。因为，“猿人还是自然的奴隶，偶然性的奴隶。”[③]然而，人既是主体又是实体，是从自然界中演化而来，之后又具有社会属性。具有社会属性的人，是人类社会发生史的第二次飞跃，代表着人类已经进入了有理性、有意识、有思维的历史范畴。但是人的自然化痕迹仍然没能被抹去，人的发展还是必须符合事物发展的客观规律，在这个前提下，不断的超越自我。

人是自然的产物，同时也在自然中进化。不被人类认识及改造的局部自然就具有人的本质力量的非对象性。因为它是神秘的，是未知的，虽然它并非是不可知的，但起码现在，它未展现在人类的文明领域，不为人类所认识。那么，我们不能否认它的存在，只能说它是孤零零地独自存在，它具有人的本质力量非对象性。

比如，在风雨雷电自然现象未被解密之前，人类受到自然灾害的影响，是自然直接作用于人，改造人类生活，却不被人认识。此时，这种自然客观规律就是非对象性的。人类通过科学的途径认识到了其形成和发生过程之后。很长一段时间，这种客观规律仍然主导了人类生活，影响着人类的生产劳动、生活，人有将其对象化的意志，却没能改变其非对象化的属性。随后，人类完全解密

[①] (德)马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.106
[②]胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.58
[③]胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.34

了这一客观规律，在不违背自然客观规律的条件下，对其进行改造，利用科学技术手段制造了人造雨解决农田久旱问题，奥运会期间根据天气情况和赛事要求安排人工消云减雨作业保证赛事的顺利进行。此时，这一自然气候现象已经是人的本质力量对象化的结果。

由此可见，人的本质力量非对象化是暂时的必经过程，人的本质力量对象化才是最终的必然结果。那么，面向未知的人工智能科学，人的本质力量非对象化阻碍了人工智能发展，推动了人类对未知世界解密，促使针对性地研究人工智能困境，将人的本质力量对象化显示出来。

4.2.3 人的本质力量自我对象化——人工智能产生和发展的现实性

马克思认为，“人不仅仅是自然存在物，而且是人的自然存在物，就是说，是自为地存在着的存在物，因而是类存在物。他必须既在自己的存在中也在自己的知识中确证并表现自身。”[①]也就是说在对象化自然的同时也对象化包含在自然存在物中的自身。简而言之，就是“人作为主体，作为能动因素，要求工具适应自己、强化自身；而作为被工具加强的对象和受动因素，人又必须适应工具和环境、必须直接适应制造和使用工具的活动。”[②]人将自然对象化为工具的现实性存在形式，同时又运用工具将自然存在物——人本身通过对象显现现实性存在形态。

我们可以整理出一条线索，自然必然诞生智性的人，并对人进行不断改造，持续到人类消亡的那一刻，人是受动者。然而在自然进入到社会的历史领域下，人也是社会的主动者，此时的自然已经不再是原始的自然，而是人化的自然，它仍然对智性的人产生作用，同时也承受着人的反作用。

套用马克思的话说，“人是一个特殊的个体，并且正是他的特殊性使他成为一个个体，成为现实的、单个的社会存在物，同样，他也是总体，观念的总体，被思考和感知的社会的自为的主体存在，正如他在现实中即作为对社会存在的直观和现实享受而存在，又作为人的生命表现的总体而存在一样。”[③]那么，人在对象化人化自然的同时又再对象化自我，人不仅仅是主体，同时也是人化自然——社会的一部分，是人的本质力量对象化的受体。

[①](德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.107
[②]胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.22
[③](德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.84

综上看来，实践不仅是自然界人化、人的本质力量对象化的过程，而且同时也是人的自然化过程。在相互做用的过程中，人类开始关注人的本质力量，认识到了人的自我意识。

马克思在《1844年经济学哲学手稿》中写道："人的本质，人，在黑格尔看来=自我意识。因此，人的本质的全部异化不过是自我意识的异化。……因此，对异化了的对象性本质的全部重新占有，都表现为把这种本质合并于自我意识；掌握了自己本质的人，仅仅是掌握了对象性本质的自我意识。因此，对象自我的复归就是对象的重新占有。"[①]

自我意识的产生是自我反思、自我确证、自我对象化的表现。自我意识通过自己的外化所能设定的只是物性，即只是抽象物、抽象的物，而不是现实的物。"这种自我反思、自我认识之所以可能，只是因为人已经作为类存在物现实地在地球上存在了，并且人们已经或至少已经初步地意识到了这种存在，即有了关于这种存在的自我意识。也就是说，先有了人的存在并有了关于这种存在的自我意识，然后才有人关于自己起源和产生的自我反思、自我认识。"[②]在如此循环反复中，人的价值得到了升华，工具的使用价值得到了展现，人的本质力量外化得以更完美的实现。

事实证明，随着人类的智慧与理性逐步发展，人的意识逐步升级到人的自我意识，是理性的思维，开始注重灵魂的重塑。成熟社会中的人们经常遇到或提出"我是什么"，"我应该是什么"，"我如何去达到我应该达到的目的"等关于自我意识的一类问题。这就是人的本质力量对象化目标转移的证据，对象已经从现实性存在状态转移到人的思维、意识、灵魂上来了。

人进入历史后，随着对象性的现实在社会中对人来说到处成为人的本质力量的现实，成为人的现实，因而成为人自己的本质力量的现实，一切对象对他来说也就成为他自身的对象化，成为确证和实现他的个性的对象，成为他的对象，就是说，对象成为他自身。马克思举例说：人的眼睛这个本质力量的独特性不同于耳朵，这个独特性就是本质力量的独特的本质，因而也是它的对象化的独特方式。"因此，人不仅通过思维，而且以全部感觉在对象世界中肯定自己。"[③]人的自我对象化是通过对象化现实存在物后，利用自我意识抽象为第三存在物

[①](德)马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.103
[②]夏甄陶.人是什么[M].北京:商务印书馆,2000.12-13
[③](德)马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.87

反对象化自身的过程。

马克思对人的本质力量做出了如下结论："一个存在物如果在自身之外没有自己的自然界，就不是自然存在物，就不能参加自然界的生活。一个存在物如果在自身之外没有对象，就不是对象性的存在物。一个存在物如果本身不是第三存在物的对象，就没有任何存在无作为自己的对象，就是说，它没有对象性的关系，它的存在就不是对象性的存在。"①

因而，人是自然存在物，并且是对象性的存在物，同样是第三存在物的对象，通过人的本质力量对象化自然而对象化存在。对象是现实的存在，第三存在物是自我意识抽象的物。（如图 4-1 人的本质力量关系图）

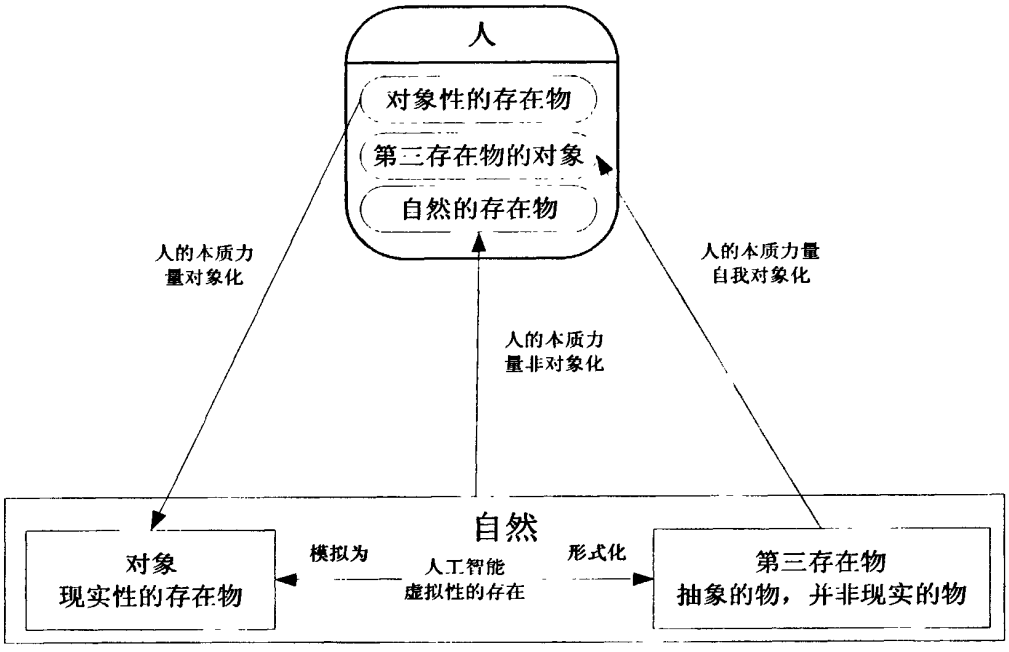

图 4-1 人的本质力量关系图

人工智能实际是一种虚拟性的存在，连接现实存在与抽象存在，将现实存在形式化为抽象存在，并反作用模拟成现实存在应用于自然社会中。从某种意义上来说，人工智能的虚拟存在是人的本质力量对象化的结果，这种虚拟存在以抽象物的形式存在，反而将人的本质力量自我对象化，为发展注入新的活力。于是，人工智能的产生和发展具备一定的现实意义。

①(德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.106

4.3 人工智能是人的本质力量延伸的最高体现

人之所以为人，成为自然界的高等生物，在于意识的产生。意识是自然界长期发展的产物，它的产生，经历了一个漫长的、复杂的物质运动过程，这个过程与整个自然界的发展过程相适应。在劳动实践中将人的本质力量对象化为物质现实性存在物，大部分现实性的存在物异化为抽象的物又反作用于人，并重新生成新的现实性存在物。以人工智能为代表的工业形式又将人的本质力量延展，使人具有无机的身体，发挥更强的社会效应。

4.3.1 劳动工具使智人和动物真正区别开来

在自然的演化中，有一部分猿人开始了简单的劳动，劳动使人具有了人性，劳动使人脱离的简单的存在，产生了思维，逐步形成了智人。于是使用工具的劳动将人与动物区别开来。马克思说："动物和自己的生命活动是直接同一的。它就是自己的生命活动。人则使自己的生命活动本身变成自己意志的和自己意识的对象。他具有有意识的生命活动。"[①]人的本质已经得到了显现，开始独立于生命存在而改造"命运"。

劳动使人具有了本质力量，对象就是自然。在对象化的过程中，人使用了工具征服自然，先是猿人用肢体充当工具进行最原始劳动，导致了人类发生史的第一次飞跃——获得了人的属性。然而生产对周围自然界的作用在自然界面前只等于零，所以"猿人还是自然的奴隶，偶然性的奴隶。"[②]再是借助天然工具劳动，拓展赖以生存的无机界范围，扩大与动物的差距。在对天然工具的不断试误的过程中，积累经验，使用成熟，对天然工具有选择性的使用。在智人晚期阶段还开始制造工具，进入了石器时代，开始分工劳动，导致了人类发生史的第二次飞跃——具有社会性。人类在思维中逐步产生了智慧，而"工具在自身内在张力的推动下比较快的发展着。它由于有人的智慧、经验参与其中，所以它能够在形态学方面突破人的种属尺度和人的机体组织的局限，而按照非人的种属尺度的原则被制造出来。"[③]

从本质上讲，一切工具都是人体器官在质和量上的延长，这种延长也就是

[①] (德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.57
[②]胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.34
[③]胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.22

人的体外进化。马克思对此作过经典的表述："植物、动物、石头、空气、光等等，……从实践领域说来，这些东西也是人的生活和人的活动的一部分。人在肉体上只有靠这些自然产品才能生活，不管这些产品是以食物、燃料、衣着的形式还是以住房等等的形式表现出来。在实践上，人的普遍性正表现在把整个自然界——首先作为人的直接的生活资料，其次作为人的生命活动的材料、对象和工具——变成人的无机的身体。自然界，就它本身不是人的身体而言，是人的无机的身体。人靠自然生活。这就是说，自然界是人为不致死亡而必须与之不断交往的人的身体。所谓人的肉体生活和精神生活同自然界相联系，也就等于说自然界同自身相联系，因为人是自然界的一部分。"[①]在这里，马克思把自然界看作人类的机体，但他认为，由于异化劳动，自然界与人发生了分裂，因而自然界就不属于人了。工具则使人的能动作用扩大了，工具成为了人的无机的身体，延展了人的本质力量。

4.3.2 人工智能成为人的本质力量延展的最高境界

马克思在《1844年经济学哲学手稿》中指出："工业的历史和工业的已经生成的对象性的存在，是一本打开了的关于人的本质力量的书。全部人的活动迄今为止都是劳动，也就是工业，就是同自身相异化的活动。人的对象化的本质力量以感性的、异己的、有用的对象的形式，以异化的形式呈现在我们面前。"[②]这就是说，通过工业的劳动形成的自然界，是真正的、人本学的自然界。只有把工业看成人的本质力量的公开的展示，才能理解自然界人的本质，或者人的自然本质。

工业导致了三次革命，使手工业、工场手工业转移到重工业，再转移到以信息产业为核心的高新技术产业。从历史上看，工业革命起源于劳动工具的机器，"在使用机器生产商品达到一定规模之后，利用机器生产机器本身的需要变得明显起来，这是很自然的。"[③]在一系列的变革中，机器从工具机升级为智能机，人的本质力量也逐渐得到越来越强的延伸和拓展。

1）机器——工具机

能够制造和使用工具使人类和其他动物根本区分开来，随着生产力的发展，

---

① (德)马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.56

② (德)马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.88

③中共中央马克思恩格斯列宁斯大林著作编译局.马克思恩格斯全集：第47卷[M].北京:人民出版社,1979.413

生产关系的变革，“人与动物的根本区别在于人会制造和使用机器。”[①]机器——工具机的出现，标志着人类进入了工业时代，“同一工作机同时使用的工具的数量，一开始就摆脱了工人的手工工具所受的器官的限制。”[②]所以，手工工具只不过是改变了形态的自然物，是人的肢体延长或“器官投影”；而机器则是改变为工业过程的自然过程，“是人们在实践中形成了某种内在性能要求的观念，并以此为功能目标设计制作出来的一种超自然的人工物”。[③]

第二次工业革命是新型机器的产生和发展阶段，“工具由人的机体的工具变为机械装置即工具机的工具以后，发动机才获得了独立的、完全摆脱人力限制的形式。”[④]在生产劳动中，从事生产的劳动者失去了人的主体地位，再一次的沦为“自然”的奴隶，作为工具机的守护者，为其服务。

在工场手工业和手工业中，是人利用工具，在工厂中，是工人服侍机器。在前一种场合，劳动资料的运动从工人出发，后一种场合，则是工人跟随劳动资料的运动。在工场手工业中，工人是一个活机构的肢体。在工厂中，死机构独立于工人而存在，工人被当作活的附属物并入死机构。

随着大工业的兴起，从事生产劳动的人，失去了人的本质力量，回归到原始的生产劳动中，这是“历史的一种退步”。虽然生产得到了发展，生产水平得到了提高，劳动成果有了质的飞跃，但失去了人的本质力量。在机器生产过程中，劳动产品不具备意义，是独立于社会、人之外孤立的存在。

这种工具机也不是万能的，它脱离了人独立存在，却无法突破第三世界的束缚。人是有智性的，生产是有目的性的改造，而文化、意识、思维的生产是工具机无法做到的。“大工业发展到一定阶段，也在技术上同自己的手工业以及工场手工业基础发生冲突。”[⑤]工具机突破了工具的人身限制，却失去了人身的本性。

2）智能机

进入第三次工业革命后，带来了契机，产业也变成了一种哲学。自然科学已经发展到了较发达的水平，为我们带来了物质基础，却对哲学的本性始终是疏远的，它们互相摒弃对方。正如马克思说的“过去将它们暂时的结合起来，

[①]乔瑞金.马克思技术哲学纲要[M].北京:人民出版社，2002.165
[②] (德)马克思.资本论:第一卷[M].北京:人民出版社,1975.411
[③]万长松.曾国屏.机器、产业与人的本质力量[J].科学技术与辩证法,2006,(02)
[④] (德)马克思.资本论:第一卷[M].北京:人民出版社,1975.413
[⑤] (德)马克思.资本论:第一卷[M].北京:人民出版社,1975.420

不过是离奇的幻想。存在着结合的意志，但缺少结合的能力。”[①]

近 2 个世纪，无论是自然科学还是人文科学都得到了迅猛发展，尤其是计算机科学、意识的起源追问和生命科学都有了突破性的进展。为自然科学与哲学的结合提供了可能。出现了立足于哲学的科技产业，例如：“20 世纪 80 年代初，先进制造技术以信息集成为核心的计算机集成制造系统（CIMS，Computer Integrated Manufacturing System）开始得到实施；80 年代末，以过程集成为核心的并行工程（CE，Concurrent Engineering）技术进一步提高了制造水平；进入 90 年代，先进制造技术进一步向更高水平发展，出现了虚拟制造（VM,Vir-tual Manufacturing）、精益生产（LP，Lean Production）、敏捷制造（AM，Agile Manufacturing）、虚拟企业（VE，Virtual Enterprise）等新概念。”[②]

由于科学技术已经成为第一生产力，以信息产业为核心的高新技术产业成为现代工业的主导力量。人们把重心转移到用高科技手段制造智能型的工具，即“利用机器制造机器本身”。用智能机贴近人类生活，成为更高级的工具。工业已经是人类发展必不可少的物质基础，而大工业使人重新沦为奴隶，将自然科学与哲学结合的需求凸显出来。人工智能的发展就是为了把人从机器机械化生产中解放出来。将机器从人的身体限制中解放出来，使机器智能化，具备人性；使思维产业化，赋予智能机，人工智能取代旧工业成为必然趋势。工业是一本打开了人的本质力量的书，现代工业——一种结合了哲学的科技产业（例如人工智能），则是更好的延展了人的本质力量。马克思也在《资本论》中谈到：“现代工业通过机器、化学过程和其他方法，使人的职能和劳动过程的社会结合不断地随着生产的技术基础发生变革。”[③]现在，人工智能的研究已经证实了这一观点，成为人的本质力量延展的最高境界。

---

① (德)马克思.1844 年经济学哲学手稿[M].北京:人民出版社,2000.89

②万长松,曾国屏.机器、产业与人的本质力量[J].科学技术与辩证法,2006,(02)

③ (德)马克思.资本论:第一卷[M].北京:人民出版社,1975.533

第 5 章 人类意识与人工智能意识的未来

意识是人存在的核心，没有意识的人在医学上被宣称脑死亡，没有意识的机器也只能是冷冰冰的机器，不能构成人工智能。“意识是人类特定存在的中心事实，因为如果意识不存在，那么我们生存所凭借的全部其他人类的特定方面——语言、爱、幽默等等，都将不可能存在。”[①]无论是人类智能还是人工智能，意识的发展是关键因素。人的意识是无限发展的，而人工智能也通过各种途径使机器具备类人意识，使人工智能的智能性成为可能。

5.1 人类意识的无限发展

进入 19 世纪以来，自然科学各领域都取得了长足的进步，人类在对自然深入认识的同时，也进一步反思自身。从物种起源到文化起源再到意识起源，人类越来越关注人是什么，从哪里来，是如何发展的，人脑工作原理等问题。神经科学、心理学、生命工程、生物学等学科朝着大综合的方向发展，致力于集中探讨人脑精神现象之谜。从生理——心理——哲学为一体的角度综合探讨人类意识的起源与发展。

诚如上文探讨的，人是自然的产物，是从自然客观规律演化中形成的高等生物。我们不能说人是最高等的生物，但人是具有极高智性的存在物，不仅仅作为单纯生物而存在，并且是在实践活动中生成和发展着的群体性社会存在物。意识则是人的必备属性，随着人的出生而积累，随着消亡而停止。人类种群的意识还以知识、文明、文化等形式传承积累下来。意识蕴含在社会生产劳动实践中，是自然及生产劳动实践的“自然”创造的，派生于实践、派生于社会。那么只要人这个种群还具有社会性，还是存在与思维的统一，意识就一直存在，并且以螺旋上升的模式循环发展。

人由自然中来，产生原始的意识，将人与动物区别开来。意识以人的本质力量形式对象化自然，被对象化的自然又成为实践劳动中的生产资料，反作用

[①]塞尔.心、脑和科学[M]. 扬音莱,译.上海:上海译文出版社,1991.8-9

于人本身，产生高等的意识。如此循环反复，自然在更迭中发展，社会在进步，意识在螺旋上升。

物质按照规律在其永恒的循环中运动发展，人的意识在认识物质的过程中完善，意识通过人的本质力量对象化物质及人本身。生生不息，运动不灭，意识发展不止。

5.2 人工智能意识的可能性展望

人工智能意识是人类意识的模拟、物化和延长，是将人的本质力量对象化与自我对象化过程中产生的虚拟存在意识。它不是本来意义上的智能，而是合理条件下虚拟的无限智能。随着科技发展，人工智能在某些方面会超越人的意识，但在总体上仍将以人类意识为本。美国科学家库兹韦尔在《灵魂机器的时代》一书中大胆的预测：信息技术、生物工程、纳米材料已经成为现在科技三大前沿，在21世纪，科学技术的融合将带来更迅猛的发展，三大技术将合力打造出新的智能机器，重塑人类的大脑和躯体。于是，人工智能研究者致力于研究新型人工智能意识，用人工智能的科学技术将意识更好的展现。

20世纪初，计算机运算速度实现了每三年翻一番，50-60年代达到每两年翻一番，现在的速度提高到每一年就翻番了。人工智能发展中却存在一个值得思考的奇怪现象：对诸如下棋、解码之类与环境相对隔离的、看似很困难的任务而言，计算机系统可以超过受专门训练的人。然而，一些最通常的、通过长期进化而形成的认知功能，比如视觉和听觉，经过几十年努力发展的人工智能系统还不如婴儿的能力。这些现实证实了人工智能的认识论自然化的困境，体现人工智能意识存在缺陷，显示出认识论日常化研究的必要。历史地看，人工智能的发展不时地陷入没有预想到的深层困境，这提醒我们不仅应当从人工智能发展的技术问题方面专项研究，而且还须从人工智能的最根本概念和理论上去寻找原因，人工智能需要更为宏观的方法论指导。

人工智能要想冲破这层层困境，必须加强与哲学的联系，“事实上，新近的哲学进展基本上没给科学带来任何冲击。”[①]哲学并不是要消灭自然科学，只是使自然科学更贴近现实的人，令人工智能意识尽可能的类似于人的自我意识。

[①]Russell S，Norvig P. Artificial Intelligence: Amodern Approach. NJ：Prentice Hall，1995. 817

而且，不管自然科学家采取什么样的态度，他们还是得受哲学的支配。问题只在于："他们是愿意受某种坏的时髦的哲学的支配，还是愿意受一种建立在通晓思维的历史和成就的基础上的理论思维的支配。"[①]

今后，人工智能的目标主要向以下几个方面发展：模糊处理、并行化、神经网络和机器情感。力图弥补生硬的自然科学缺陷，使人工智能具有真正的智性、具备人工智能意识，并且使用某些未知的方法消除各种认识极限。目前，人工智能的推理功能已获突破，学习及联想功能正在研究之中，下一步就是模仿人类右脑的模糊处理功能和整个大脑的并行化处理功能。人工神经网络是未来人工智能应用的新领域，未来智能计算机的构成，可能就是作为主机的冯·诺依曼机与作为智能外围的人工神经网络的结合。研究表明：情感意识对于计算机与人的自然交往至关重要，是智能的一部分，而不与智能相分离，因此人工智能领域的下一个突破可能在于赋予计算机情感意识能力。

人类意识是看不见摸不着的信息流，但仍然可以通过生理——心理——哲学大综合为一体的科学进行解读。生理方面，通过医学解剖、神经科学等手段对复杂的人脑进行研究，可以逐步解读人的大脑构造，解释意识行为的发生。心理方面，通过思维模式的分析和研究，人类已经逐步掌握人类心理的普遍规律，用表象的思维归结意识的本质。哲学方面，人类在不断的追问中思考，思考人的本质、人的价值、人的意识。从黑箱的表象观察推测到灰箱的系统化研究，人工智能系统在突破了上述困境之后，也将迎来希望。

虽然，人类目前揭开的只是冰山一角，并且研究进行的同时，事物仍然依照事物的客观规律向前发展，具有无法突破的极限。但人工智能依然顽强的展现其生命力，在对人类意识解读的同时完善自身，成果是卓然的，主要体现在以下三个方面：

1）人们对人工智能的潜在感觉（即意识）进行了认真推测，从自主意识入手，使算法具有类人的特性。

人工智能的计算，在经历了传统经典算法后，结合非经典计算理论和技术，通过量子计算方法与神经网络方法；基于生物计算机制进一步完善遗传演化方法；以及基于自然界的自组织机制，将群体计算方法加以完善，借以实现人类神经集群相互作用的心智自涌现机制，以联结模式的手段并行运算，可以打造

[①]恩格斯.自然辩证法[M].北京:人民出版社, 1984. 187

高智能性的类人。事实证明计算算法的能力是强大的，它的处理速度和效率也以几何次方的速度增长，我们有理由相信，将会设计出更接近人类思维的算法出现，使人工智能成为可能。20 世纪 70 年代早期至 80 年代中叶“专家系统”的发现、20 世纪 70 年代早期开始的“演化计算”与 20 世纪 80 年代后期进行的“用文字计算”研究，使计算机算法成为面向机器智慧的新哲学。

它具备类人的思维模式，能复制和制造类人的意识，具有人的本质力量，那么将它放在特定的社会中，它能独立存在，它具备人工智能的特性。缺陷是，模糊演算算法与用文字计算——典范转移算法还处于发展的初级阶段，传统算法受形式化极限的影响，部分感情、思维是无法进行计算的，没能突破人类思维的限制达到理性思维。

2）人工智能的主观体验成为人们研究的课题，直接截取人类意识，使行为符合人的特征通过图灵测试。

随着人体器官组织的人工培育生物技术的不断成熟，完全“可以按照各种需要人工培育人类和动物的大脑皮层组织，并直接与数字芯片相衔接用于控制机器行为。”[①]

人类现在已经能够片段地截取意识信息，并且将其形式化，虽然具有极限性，但仍然可以趋近。目前，生物技术上的克隆也已经不再遥不可及，可以通过基因工程放大或缩小人的部分功能，与信息技术有机结合。20 世纪 80-90 年代，出现了一些虚拟实境（Virtual Reality）的研究，由著名数学哲学家希拉里·帕特南提出“瓶中脑”（brains in vats）的案例，假说有一个邪恶的科学家将我的脑由我的身体取出，并把它养在一个灌满营养液的瓶中，通过电流的连接输入感觉内容，究竟会怎样呢？“我们可以想象所有的人类都是瓶中的脑子或者在瓶中的神经系统，假设自动机械具有给我们一种集体幻觉，而不是许多分开且无关联的幻觉的程式。这样，当我说话时，由我的脑子将刺激神经意识传导到电脑，造成我去‘听’我自己发出的声音并‘感觉’我的舌头在动，也造成你‘听’我说的话并‘看’到我在说话的错觉。你并没有真的嘴，没有真的耳。”[②]试验已经证明，通过信号的输入，你感觉到你是在和一个有生命特征的人在交流，表面上看它是智性的，它表征着对方的灵魂。然而，这种感觉终究是自己的幻觉，并不是

[①]周吕乐.心脑计算举要[M].北京:清华大学出版社,2003.214

[②] The Case of the Brains in a Vat by Hilary Putnam.From Reason,Truth,and History ,1981,pp.5-7.Rerinted with the permission of Cambridge University Press.

将灵魂形式化反馈给你的，而是制造了你期待看到的表象，足以迷惑你的心智，让你产生幻想的触发事件而已。形式化的极限只是用一种虚幻的表象弥补了，无法自组织的完成意识的产生、处理、运行，片段的记忆无法形成系统的智能体，只能说它是一个低智能的类人行为机。

虽然这种感觉是虚拟的，也可能被不真实幻觉蒙蔽。但仍然证明了信息数据已经可以影响人类的意识，添加记忆片段。那么制造一个符合人类需求的智能机也许并不遥远，它能够自主的完成指令。

3）寻求自我意识产生的基本方法，模拟智性的思维模式，人工智能使人类智能得到越来越大的提高。

在对人类记忆存取模式解读的过程中，人们发现人类大脑思维是智性的提取，记忆是以片段的形式存留，从记忆库中搜索提取，是通过突发触点的关键词检索到的。这种思维模式是并行且模糊的，人工智能的思维模式却是精确的框架式运行。目前，科学家已经充分注意到这些现实差距，并致力于开发并行的联结思维模式。不再进行简单的还原，重新对人类智能的本质进行反思，提供智性思维模式直到人工智能发展，为人工智能的成功又开辟了新的途径。

近年来，针对以上缺陷，一些基于神经网路技术的方法被用来为模糊规则寻找数值资料，因为自适应或神经模糊系统可以发现新的模糊规则，或者根据所提供的资料改良并调整现存的规则，换句话说是输入资料——输出规则，或者输入经验——输出一般规律；模糊规则也被用来弥补人工神经网络的缺陷，因为人工神经网络必须繁复地归结经验定律、规则，并反复输入到需要这些规则的地方，否则程式将无法继续运行。这种新型思维模式是具有自学习性质的，规则被自动发现并调节整个人工神经网络模型，将模糊逻辑和神经计算相统一的专家系统提升了基于知识系统的适应性、容错性和可执行速度。目前，专家网络神经和模糊系统已经诞生并已经广泛地应用于解决不同的意识问题，主要试行在工程，医疗，财经，商业和管理等领域。

5.3 人工智能等价于或超越于人类智能的问题

美国科普作家霍根（John Horgan）在《电脑比我们聪明吗？我们究竟是谁？》一文中，明确指出：“西洋棋是为电脑量身定做的游戏。深蓝威力非凡，能够在一秒钟之内计算出几亿个一步棋所有可能的结果。假如这种矽晶怪物必须如此

费力才能胜过一个人，那么我们如何能期望电脑可以替外交官进行限武谈判？”[①]人脑之所以与电脑不同，原因之一是电脑有一个单独的、可信赖的中央处理器（或是个数字处理器），运作起来非常快速；而人脑中的处理器虽然是较慢且较模糊的，但却有数千亿个处理器，而且每个处理器与其他几千个处理器相连接，这使得人脑能立即辨认出复杂的外在事物。“电脑则必须一次一步地处理每个琐碎的细节。”[②]那么，人工智能能否等价于或超越于人类智能？

5.3.1 黑箱、灰箱、白箱 ：人类智能的祛魅化进程

在人类认识世界、从事科学研究的时候，常常会遇到这样的情况，有一些我们所要认识或控制的客体，由于种种条件的限制，其内部的结构一时不能够被我们直接观测到，仿佛是一个既不透明且又密封的箱子，其复杂的结构和神妙的机理，珍藏其中，人们无法从外部或无法打开来直接探察其内部的奥秘。对于这样一种内部结构尚不能直接观测，只能从外部去认识的客体，控制论创始人维纳起先称它为“闭盒”(Closed Box)，后来艾什比和维纳又称它为“黑箱”(Black Box)。

人类意识的产生标志着人开始产生与其他生物不同的思维，认识世界的同时，也认识世界的存在物——人本身。由于认识能力和认识环境的限制，只能通过表象认识，利用外部观测，考察对象与周围环境的相互联系来了解黑箱的特征和功能，猜测其内部构造和机理。对于新鲜事物，黑箱的研究方法是人类认识必经的初级阶段。人脑是个极其复杂的系统，人们还不能直接从外部，或通过生理解剖窥视其内部构造细节和思维机理。而人工智能的研究已经开始了，“随着仿真技术的发展，数字化模型技术的发展，激励人们试图用‘黑箱’技术，信息论、控制论技术和计算机技术探测人脑这个令人最着迷的‘黑箱’，并模拟人脑的意识、思维功能与机制，移用于电脑乃至机器人的研制和改进过程中。这当然地引发了人们对人脑思维乃至整个意识现象的研究兴趣与不懈进取。”[③]在人工智能发展初期，随着科学技术化发展的深入，研究者将这一领域划分成若干子学科，逐个进行系统定性研究。黑箱方法一般包括三个原则和步骤：第一，

---

① Smarter Than Us?Whos Us?by John Horgan,Section4,p.15,Column 1,Editorial Desk,May 4,1997,Copyright© 1997 by The New York Times.Reprinted by permission.

② “Can a computer be conscious?” by Steven Pinker.Mysteries of Science Section,August 18,1997/August 25,1997.Copyright,August18/ August 25,1997,U.S.News & World Report.

③胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.2

以相对孤立的原则，确认黑箱。第二，以观测和主动试验的方法，考察黑箱。第三，建立模型，阐明黑箱。因此，具有机械性，没有摆脱大工业时代机器的阴影，尚未使智能成为体系，缺乏关联性，并且将研究者沦为“活的附属物并入死机构。”

随着科学的大融合，各学科齐头并进，结合内部解刨学和外部客观认识人脑的反映，人脑的工作机制和思维模式逐步被认知，为灰箱的系统方法提供了可能。而弱人工智能研究者也在现实中明白，整体不是部分的简单叠加。要突破量变到质变的飞跃，必须在已被认识的半框架结构中逐步完善。虽然人类智能系统仍未完全明了，但开放式的联结模式让研究者看到了希望，开始有目的、有联合的灰箱研究，成为发展的必然阶段。

白箱就是机制和结构完全明了的系统，是一个较高阶段。虽然这将是一个漫长的过程，但仍将促使研究者们通过不懈的努力逐步解开未解之谜。当事物已经被赤裸裸的展现在人类面前，我们还有什么理由不去将它对象化，使其扩大人的本质力量。

人类智能是人类追问了千年的命题，人工智能则使人类智能丰富和发展，在自然科学和人文科学的作用下，曲折前进。黑箱、灰箱、白箱的科学方法见证了人工智能研究的过程，黑箱的方法是纯自然科学的纵向深入研究。灰箱的方法将研究领域扩大，横向的将触角深入到各领域，将自然科学与人文科学联系起来，虽然进展缓慢，但还是产生了质的飞跃。白箱将是终极阶段，更好的融合自然科学与人文科学，产生更利于发展的新科学技术哲学。只有进入了白箱阶段，才能系统的、完整的使对象化的世界处于领先地位。只有在黑箱研究的技术已经成熟、灰箱研究的模式框架已经基本建立的前提下，努力进入白箱阶段，人工智能才能等价于或部分超越于人类智能，更好的成为人的本质力量的延伸。

5.3.2 将信息符号模式彻底转化为联结模式

目前，人工智能的研究在信息符号模式下已经进入了认识论的自然化之后的困境。而新型联结模式的认知，使单一走向集成，将计算、认知与情感全方面的融合，令计算机具有更接近大脑的功能。虽然联结模式还不够完善，但集成化趋势逐步弥合了信息符号模式僵硬的叠加。

虽然自然科学下的认识论让我看到了信息科学的先进性，发现了数字信息

的神奇力量，使工业得到突飞猛进的发展，人工智能的实现成为可能。但在详尽的系统分析之后，我们也清晰的看到了其局限性。看到了在信息符号模式框架下，日常感情、自我追思、自我意识等日常化认识被拦在了自然科学的大门之外。失去了自我意识的人是植物人，不具有人的特性。然而，缺乏自主意识和情感的人工智能充其量也只能算是高级的工具。信息符号模式的人工智能，给人类带来新的效率，它能快速完成人类传统计算几十年、几百年也算不清的计算题，它能精确完成数据的定位及提取，它能无障碍的跨越各数据库之间完成数据的检索、搜集、分析。迅速、准确、便捷成为信息符号模式人工智能的典型特点。但它始终是人操作的机器，在运行的同时将人并入到死机构中去，简单将实物以数据的形式描述，机械重复地计算统计、分析储存，实行量变。

人是一个极其复杂的机构系统，联结模式同样也是新型的复杂人工智能理论框架，即广义的仿生学。试图将人的构造、人的本质力量、人的思维模式、计算机技术等科技领域有机的组合起来，使计算机这个中央处理器不再做无用功。在我国，钱学森教授也主张采用“从定性到定量的综合集成技术”，把人的思维和思维的成果、人的知识和智慧以及各种信息和资料统统集合起来，形成一个工程领域，称之为“大成智慧工程”（Metasynthetic Engineering）,使之成为集中人类智慧的工作体。

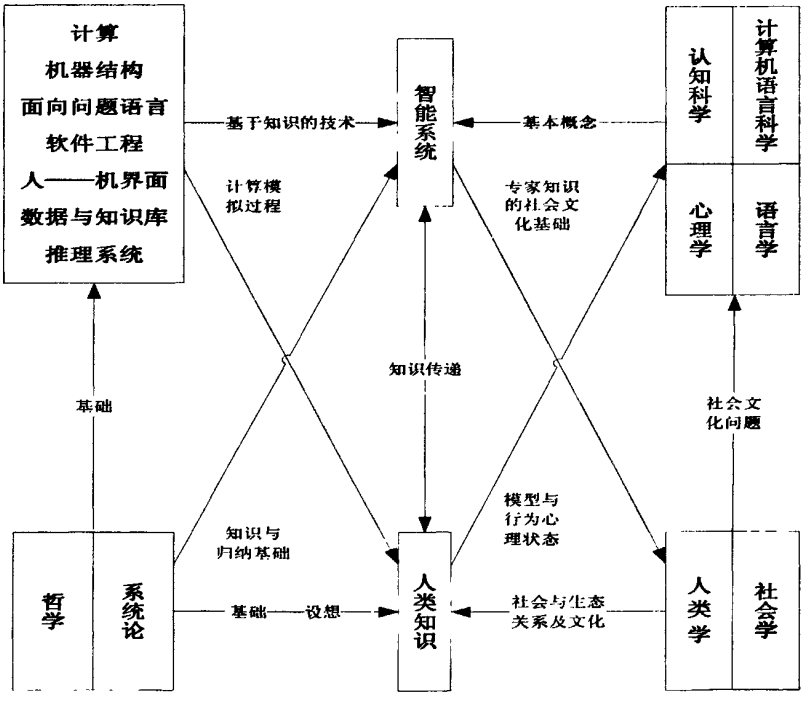

图 5-1 集成智能系统的相关学科关系图

从这个集成智能系统的相关学科关系图（图 5-1 集成智能系统的相关学科关系图）中我们可以看出，智能系统不仅仅是计算机科学内部的事，还是哲学的运用和联系心理学、语言学、现代计算技术、认知科学和计算机语言学，系统的概括社会学、人类学、系统科学、人文科学。计算技术是联系所有学科的脉络，联结了各学科形成智能系统。各学科都是相互联系、相互辅助统一的，同时也是各自深化发展的子系统。

在新型人工智能模式的组合下，把人脑的智能水平与计算机处理信息的高速、准确以及共享数据资源的能力等内在优势有机地结合起来，人工智能等价于类人智能成为可能，人工智能的意识也随智能体的发展而进化。但由于极限的存在，人工智能将不可能完全取代人类智能，只能在部分环境作用下片面的超越人类智能，总体上不断趋近人类智能。然而，假设人工智能真的能取代人类智能，那么人类还将有什么存在的价值？这也将与人工智能的研究目的背道而驰。

结束语

人工智能经历了50多年的发展，在各种争论中曲折前进，虽然人工智能要达到真正的智能体，还需要相当长的一段时间，但笔者仍然乐观的认为，当人工智能突破以上层层困境，充分认识到哲学内涵，冲破系统化极限、形式化极限、信息符号模式极限的束缚，从根本上解决人工智能发展的局限，运用更成熟的计算机软件应用技术和更完善的开放式联结集成模式，就能更好的解决人工智能的复杂性和广泛性问题，人工智能就能成为人之外的智能体具有理性思维。

人工智能是伴随人类发展的必然产物，最终将融入人类，随着人工智能应用领域的不断扩大，人工智能不仅仅局限于工业、医疗、国防和计算机等领域，已经广泛地运用于交通、农业、航空、通信、气象、文化、教育和海洋工程、管理与决策、博弈与竞技、情报检索等部门。人工智能最终将成为人的本质力量的延展，成为“人的无机身体”，并更好的发展人类社会。人工智能研究已经是确定的事实，成为无法倒退的历史，一些人工智能反对者也不由自主的享受着人工智能带来的便利。目前，我们应该做的不是思考人工智能研究的合理性，而是要统一认识，寻求一种有效的方法帮助进一步实现机器智能化。

美国著名的人工智能专家爱德华·费根鲍姆（Edward Albert Feigenbaum）也对人工智能的发展作了一个历史性的回顾与总结，指出：“能推理的动物已经（也许是不可避免地）制成了能推理的机器，尽管这种大胆的（有些人认为是鲁莽的）投资有着种种明显的风险，但是，无论如何我们已经开始干了。……不管阴影有多么黑，险恶有多么大，我们切不可被吓住而不敢走向光明的未来。”

参考文献

[1]（美）尼尔森（Nilsson.N.J）.人工智能[M]. 郑扣根,等译.北京:机械工业出版社,2000.9

[2]（英）玛格丽特•A•博登编著,博登编.人工智能哲学[M].刘希瑞,译.上海:上海译文出版社,2001.7

[3] 杜文静.人工智能的发展及其极限[J].重庆工学院学报,2007(1)

[4]（英）玛格丽特 • A • 博登(Margaret A.Boden)编.人工智能与自然人[M].1977.6

[5]余纪元译,尼古拉斯•布宁编.西方哲学英汉对照辞典[M].北京:人民出版社,2001.2

[6]（美）Stuart Russell, Peter Norvig.人工智能——一种现代方法:第二版[M].姜哲,等译.北京:人民邮电出版社,2004.6

[7] 陈修斋主编.欧洲哲学史上的经验主义和理性主义[M].北京:人民出版社,2007.4

[8] 王哲.人工智能的复杂性与信息科学纲领[J].系统科学学报,2007,(02)

[9] 刘晓力.理性的生命——哥德尔思想研究[M].长沙:湖南教育出版社,2000.9

[10] 波普尔.科学知识进化论[M].纪树立,编译.北京:三联书店,1987.11

[11] 周建.没有极限的科学[M].北京:北京理工大学出版社,2006.4

[12] 郑祥福,洪伟.认识论的自然化之后[M].上海:上海三联书店.2005.8

[13] 洪伟.科学认识论的五形态及其对人工智能的意义[J].自然辩证法研究,2004,(02)

[14] 郭艳华,周昌乐.2000b.汉语语句依存关系网协动生成方法.杭州电子工学院学报(4)

[15]（德）恩斯特·马赫.认识与谬误[M].李醒民,译.北京:华夏出版社,2000.1

[16] 张掌然.问题论[D].武汉大学, 2005

[17] 库恩著,金吾伦、胡新和译.科学革命的结构[M].北京:北京大学出版社,2003.1

[18] 欧阳康.哲学研究方法论[M].武汉:武汉大学出版社,1998.7

[19] 董军,潘云鹤.人工智能的认识论问题.科学, 2002, 4

[20] 赵敦华著.西方哲学简史[M].北京:北京大学出版社,2001.1

[21] 北京大学哲学系外国哲学史教研室编译.十六——十八世纪西欧各国哲学:第二版[M].北京:商务印书馆,1975.7

[22] 王天成.创造思维理论——德国古典哲学创造思维理论的精华[M].长春:吉林教育出版,1989.7

[23] 培根.新工具[M].北京:商务印书馆,1984.10

[24] 周晓亮.直觉与演绎：笛卡尔的方法论选择及其困境[J].《云南大学学报》,社科版,2005年(第1期)

[25] 笛卡尔.谈谈方法.载于 The Philosophical Works of Descartes, volume I,

[26] 伽桑狄.对笛卡儿〈沉思〉的诘难[M].北京:商务印书馆,1963.75

[27] 张守刚,刘海波.人工智能的认识论问题[M].北京:人民出版社,1984.8

[28] 外国哲学史研究集刊:第5辑.上海:上海人民出版社,1982.4

[29] 黑格尔.哲学史讲演录:第4卷[M].北京:商务印书馆,1983.10

[30] 罗蒂.后形而上学希望[M].张国清,译.上海:上海译文出版社,2003.7

[31] 麦基编.思想家[M]. 周穗明,等译.北京:生活•读书•新知三联书店,1987.7

[32] 夏甄陶.人是什么[M].北京:商务印书馆,2000.1

[33] 恩格斯.自然辩证法[M].北京:人民出版社, 1984.8

[34] 胡潇.意识的起源与结构[M].北京:中国社会科学出版社,2004.12

[35]（德）马克思.1844年经济学哲学手稿[M].北京:人民出版社,2000.5

[36] 乔瑞金.马克思技术哲学纲要[M].北京:人民出版社,2002.2

[37]（德）马克思.资本论:第一卷[M].北京:人民出版社,1975.6

[38] 万长松.曾国屏.机器、产业与人的本质力量[J].科学技术与辩证法,2006,(02)

[39] 塞尔,扬音莱译.心、脑和科学[M].上海:上海译文出版社,1991.6

[40] 周昌乐.心脑计算举要[M].北京:清华大学出版社,2003.3

[41] 洛克.人类理解论[M].北京:商务印书馆,1959.39

[42]（美）库兹韦尔.灵魂机器时代——当计算机超过人类智能时[M].沈志彦,祁阿红,王晓东,译.上海:上海译文出版社,2006.4

[43] 中共中央马克思恩格斯列宁斯大林著作编译局.马克思恩格斯全集:第47卷[M].北京:人民出版社,1979.7

[44] John McCarthy.Formalizing Commonsense: Papers by John McCarthy[M].Ablex Publishing Corporation, 1990.

[45] The Case of the Brains in a Vat by Hilary Putnam.From Reason，Truth，and History ,1981,pp.5-7.Rerinted with the permission of Cambridge University Press.

[46] CONSCIOUSNESS EXPLAINED by Daniel Dennett.1991.By permission of Little,Brown&Company(Inc.)

[47] Hilary Putnam. Renewing Philosophy[M].Cambridge,Mass:Harvard University Press,1992.

[48] Russell S，Norvig P. Artificial Intelligence: Amodern Approach. NJ:Prentice Hall，1995.

[49] Smarter Than Us ? Whos Us ? by John Horgan,Section4,p.15,Column 1,Editorial Desk, May 4,1997,Copyright©c1997 by The New York Times. Reprinted by permission.

[50] "Can a computer be conscious ?"by Steven Pinker. Mysteries of Science Section, August 18,1997/August 25,1997.Copyright,August18/ August 25,1997,U.S.News & World Report

[51] Hilary Putnam.Words and Life.[C]//Edited by James Co-nant.Cambridge,Mass:Harvard University Press,1994.

[52] Guha, Lenat .Cyc: A Midterm Report [J].AI Magzine,Fall,1990(33)

致 谢

转眼间，三年的研究生生活即将结束，在求学和论文写作过程中，许多老师和同学给予我很大地帮助，在此，我要向他们表示诚挚地谢意！

衷心感谢我的导师孙德忠副教授，是他为我打开了学术研究的大门，带我走进了科技哲学的领域，这三年来我每一个进步无不凝聚着导师的心血。本文是在导师的悉心指导下完成的，在论文写作过程中，孙老师多次亲自指导，指点迷津。孙老师渊博的学识、严谨的治学态度、宽以待人的师表风范和孜孜不倦的敬业作风使我受益匪浅。在此，谨向尊敬的孙老师表示诚挚的谢意！

感谢政治与行政学院全体老师，他们的启迪和教诲，让我成功的走到了今天。在专业数据收集期间，计算机学院的彭德巍老师讲授的人工智能专业课为我提供了技术理论，在此向他表示感谢！

非常感谢师兄弟姐妹们以及同窗们在论文资料收集过程中和三年的学习生活中给予了我无私的帮助与关心！特别感谢清华大学王海威同学帮助下载及传输相关方面课件资料！

衷心感谢一直默默关心、支持我学业的父母及家人！

最后，谨向曾经关心和支持我的师长和朋友们致以诚挚的谢意！

攻读硕士期间发表的论文

[1] 周雅凡，自然科学与唯物主义的相互作用，2008，知识与创新，刊号：ISSN 1814-3458

[2] 周雅凡，人的本质问题，2007，人文社科研究，刊号：ISBN 978-7-5419-9793-8