

一个哲学家的科学实验 —— 徐英瑾《心智、语言和机器——维特根斯坦哲学和人工智能科学的对话》评介

王培（美国天普大学计算机与信息科学系）

多年之后，“人工智能”（AI）重回公众视野，而徐英瑾的《心智、语言和机器——维特根斯坦哲学和人工智能科学的对话》（人民出版社，2013，以下简称《对话》）的出版可谓恰逢其会。

和很多哲学家站在人工智能之外对其评头论足不同，徐英瑾此书的最大特点就是真正进入到这个领域之中，对工作技术方案的哲学前提和后果进行分析。因此，对于希望对这个领域进行深入思考的读者，尤其对在相关问题上工作的研究人员来说，阅读此书应当会是一次既有挑战又有收获的思想历程。

因为我本人和徐英瑾有长期合作关系，也在本书的写作过程中参加过意见，这篇书评就不能是站在局外品评其成败之处，而更多的是对其的介绍和补充。当然，作为一个人工智能工作者，我的分析和结论和作为哲学工作者的徐英瑾不会完全相同。但正因为如此，希望这篇书评能够帮助读者加深对有关问题理解。

（一）人工智能的理论预设

徐英瑾的《对话》把人工智能和哲学史上的主要流派和代表人物在观念上的关系进行了一番详细的梳理。他的论题既有“不同国家的哲学文化对于AI科学发展的影响”之类的宏观概括，又有对笛卡尔、莱布尼茨、霍布斯、休谟、康德等人的思想和AI关系的介绍，尤其是对维特根斯坦和人工智能的关系进行了详细和独到的分析。有代表性的人工智能哲学研究也均有论及，并在此基础上总结出了他对这个领域中成果所应达到的标准。作为人工智能哲学的引论性著述，这本书已经相当完整。

在这里我要补充讨论的是人工智能工作者们的哲学观念，尽管这些观念常常不是用哲学术语表达的。正像《对话》（第1页）开头所说，很多人工智能工作者觉得哲学讨论都是空谈，殊不知他们的很多基本观点实际上属于哲学性的。一般来说，任何一个科学技术领域均有哲学观念的问题，因为每个理论都有其逻辑起点，包括基本预设、覆盖范围、直观背景等等。在这些问题上的争论往往不能完全在该领域之内得到解决，而需要在更大的尺度和范围中进行考虑，因而进入了哲学的范围。由于人工智能的研究对象涉及智能、认知、思维、心灵、意识等在哲学中被反复讨论和使用过的概念，和哲学的关系就比其它领域更密切和复杂。

在这一点上，人工智能和其它科学技术领域最大的不同是其中研究规范的庞杂。一个领域内有多个相互竞争的规范并不奇怪，但在其它领域中，主要的争议是对现象的解释和对问题的解决方法，而在人工智能领域中的争议首先是什么是要解释的现象和要解决的问题。“人工智能”直观意义很简单，可以说就是“让计算机像人脑那样工作”。但计算机毕竟不是人脑，也不可能所有方面都像人脑，因此一个人工智能系统只能是在某个方面像人脑，而在其它方面可能完全不像。在这一点上，人工智能主流的观点是把智能看成“能解决那些人脑能解决的问题”的能力。在这里，“问题”可

以是具体的实际应用，如“下围棋”和“开汽车”，也可以是抽象的认知功能，如“推理”和“学习”。(Russell & Norvig 2010)

这种把“智能”看作“解题能力”的观点实际上是一个关于智能和思维的哲学信念，尽管绝大多数人工智能工作者是将其作为一个自明的前提来接受的。造成这个现状主因是人工智能是作为计算机科学的一个子领域发展起来的，因此继承了其基础研究规范，而计算机科学的基础又是从数学中继承来的那些为解题过程提供规范的理论。(Feigenbaum & Feldman, 1995)

徐英瑾的《对话》(286页)中所介绍的戴维·玛尔的“三层次说”就是这一观点最为人知的表述之一。在玛尔看来，要解决某个问题，第一步是把该问题严格定义为对每个可接受的输入值指定输出值的“计算”或称“函数”，第二步是把该问题一个解法表达成明确地一步一步将输入变成输出的“算法”，第三步是把该算法实现在计算机中，而后计算机就可以依照此算法在这个问题的各个实例上完成这个计算了。这三个步骤也可以被看作对该解题过程的三个描述层次。

按照这种信念，所谓“智能”无非是一组算法的总称，而人工智能和计算机科学的其它分支的差别只是由于它所研究的算法对应于人脑所表现出的问题解决能力。因此，尽管人工智能需要研究各个问题的具体解决过程，其基本理论需求并未超出现有的理论范围，其中主要是集合论、数理逻辑、概率论、可计算性理论、计算复杂性理论等。在一本流行的人工智能教科书中(Russell & Norvig 2010)，1987年被选为“人工智能开始采用科学方法”的标志年份(以Judea Pearl关于贝叶斯网络专著的出版为标志事件)，而这是因为“此后更常见的做法的是理论用现有的而非新造的，结论来自严格的定理或实验证据而非直觉，结果表现为现实应用而非简单例子。”在很多人看来，人工智能界主流是从早期的好高骛远变成后来的脚踏实地。但从另一方面来看，这也可以说是在理论上转向保守，将研究目标由通用系统转为专用系统，希望用技术手段解决理论问题，并在评价研究成果时强调近期应用而忽视远期潜力。尽管取得了一定的应用成果，人工智能的这一转向事实上抑制了对智能的基本问题的理论探索，以至于作为这个领域的奠基人之一的图斯基抱怨说“AI已经脑死亡了”。(Baard, 2003)

一些AI圈外人(如《对话》中提到的德瑞福斯、塞尔、彭罗斯等)，感到了人工智能在这方面的问题。他们正确地指出人的思维不能在人工智能现有的理论框架中得到解释，但错误地以此作为“真正的人工智能不可能实现”的理由。这些人工智能的批评者在这里和他们的批评对象一样，把计算机的一种特定的使用方式(遵循算法来解决问题)当做了其唯一可能的使用方式，因此不能为这个领域中的突破指引方向。(Wang 2007)

把“智能”看成一组算法的另一个后果就是人工智能领域的“身份危机”。由于“只有人脑能解决的问题”是一个随时间而变的概念，以其为目标导致人工智能说不清自己和计算机科学的其它领域的区别何在，更遑论建立一个统一的理论基础。一个常见的现象就是当一个以前只有人脑才能完成的甲任务被计算机完成后，总会有人视其为“人工智能的里程碑”，并期望乙任务、丙任务、丁任务等也会很快被完成。但过不了多久，大家就会发现完成甲任务的方法基本无法推广，而在其它任务被完成之前，即使甲任务的“完成”也是打了很多折扣的，且和传统的计算机应用相差不远。这时“真正的人工智能不可能实现”的论断又会重现，甚至先前的“里程碑”也会沦为其“技止此耳”的证据。这种“季节变化”在人工智能的历史中已经反复上演。很多人工智能工作者抱怨自己的工作得不到应有的承认，并试图使大家相信人工智能已经在我们身边，但往往说服力有限，因为人们靠直觉就可以感到所谓“智能系统”和人类思维的根本差别。

把“智能”看成一组算法也造成了人工智能领域的碎片化。在把每个问题独立定义和解决的过程中，各个认知功能之间的内在联系被割断了。其结果之一就是这些功能在很多所谓“人工智能”系

统中的表现和它们在人的思维活动中的完全不同。对于那些主要以人工智能作为研究人类智能或一般智能的途径的研究者来说,这种“分而治之”的办法是有根本缺陷的。即使完全从应用的需要来看,各个认知功能的协调运用在解决问题过程中也常常是不可或缺的。以自然语言理解为例,说“语言理解不需要思考”听上去明显有问题,但至今语言理解系统一般都没有多少推理能力。

正是对上述问题的思考导致了近年来“通用人工智能”思潮的兴起,其主要特征就是对智能的通用性和整体性的强调,以及对主流人工智能的基本预设的挑战。(Pei Wang & Ben Goertzel 2007) 要理解这一类研究工作和传统人工智能的区别,我们需要重新审视“智能到底是什么”和“怎样在计算机中实现智能”等基本问题。

(二) 纳思的理论预设

纳思(非公理化推理系统, Non-Axiomatic Reasoning System, NARS)是我设计的一个通用人工智能系统。徐英瑾在《对话》中在用通俗易懂的语言和实例介绍人工智能中的主要技术方案(符号进路、人工神经网络、遗传算法、贝叶斯网络)之后,重点介绍了纳思的主要成分,并以“相关性”为例分析了其理论意义。在书的后几章中,他更详细建议了用纳思处理视知觉和中文的方法。由于上述功能在纳思的规划中仍处于设计阶段,这些建议对今后的工作是有重要意义的。

关于纳思的详细技术性描述,可见我已经发表的论文和专著(Wang 1995, 2006a, 2013)、主页中收集的文章(<http://cis-linux1.temple.edu/~pwang/papers.html>),以及纳思网站(<https://github.com/opennars/opennars/wiki>)中收集的源代码、文档、实例等。在这里我主要介绍纳思的主要哲学背景和理论预设,尤其是其中和主流人工智能的不同点。

纳思和其它人工智能系统的不同很多都可以追溯到对智能的理解。在我看来,智能是“一个适应系统在知识和资源不足的条件下工作的能力”。所谓“知识和资源不足”,具体说来是下面几点:

- (1) 智能系统只能依赖于有限的信息加工资源。这里有限性主要指处理器的数目和速度以及存储空间的容量。对于智能机器人,有限性也包括能量储备和直接感知运动能力。
- (2) 智能系统必须实时工作。这就是说新任务可能在任何时刻出现,并且均有完成时间要求,如“五分钟之内”或“越快越好”等等,因此多个任务会争夺系统资源。
- (3) 智能系统必须对未来经验报完全开放的态度。这就意味着新知识可能和已有知识相冲突,新问题可能超出系统的知识范围,而这些情况均不应导致系统的瘫痪。

在这种条件下工作意味着系统中的所有来自过去经验的知识都可能被未来经验挑战,而且系统在解决一个问题时一般没时间考虑到所有的相关知识,因此无法保证所有结论都是绝对正确或最优的。这种情况下的理性只能是一种“相对理性”,即适应性,就是说系统只能用过去经验来尽力应对目前的新情况,用有限的资源来尽量满足当下的要求。

按照这种想法设计的系统和目前常见的主流人工智能系统有着根本性的不同。由于系统经验和资源需求随时间变化,系统对一个问题的解答往往不是固定的,因此“问题”和“解答”不能被看成经典意义下的“计算”或“函数”,并且解题过程也不遵循一个确定的“算法”,而是具体问题具体分析。即使是同一个问题实例,在不同的情形下所得到的处理也可能很不同。

这个意义下的人工智能不再是传统计算理论所能涵盖的。实际上这二者在前提上是互补的：前者研究知识和资源不足时的工作方式，而后者研究知识和资源充足时的工作方式。这就解释了纳思为什么不是基于现有理论基础之上的。在一个数学理论中，有关的知识一般都已经在公理和定义中，而解题时所用的资源是忽略不计的（只需是有限的）。这和我们日常的问题解决过程完全不同。

当现有的理论不能满足问题的要求时，正确的办法是尝试建立新理论，而不是削足适履，把问题限制在已有理论所能覆盖的范围。正像徐英瑾在《对话》（第276页）所指出的那样，建立一个正确的“战略”往往不像在“战术”上打主意见效快，但从长远看来却是更有效率的，因为战略上的错误是不能用巧妙的战术来弥补的。

纳思的上述理论预设并没有使其无法再现有的技术条件下实现。具体说来，纳思是在一个推理系统的框架里设计的，主要包括一个“逻辑”部分和一个“控制”部分。和数理逻辑致力于刻画从公理推出定理的证明过程不同，纳思所使用的是一个“非公理化逻辑”，其基本功能是根据系统的经验确定概念的意义和陈述的真值。它是“非公理化”的，因为其中没有一个来自经验的陈述具有“公理”的逻辑地位，即其真值不可能被未来的经验所挑战。尽管不再有传统意义下的“保真”性，非公理化逻辑的规范性仍然可以在相对理性的基础上被建立起来，而其中的推理规则的有效性则体现为其结论的真值恰当地衡量了前提所提供的证据。这样一来，传统逻辑中无法包容的归纳、类比等非演绎推理得到了在统一基础上的实现和辩护。

纳思的“控制”部分的任务是有效地分配系统的资源。在大量推理活动竞争有限资源的情况下，系统的时间和空间均是按竞争者（概念、任务、知识等）的优先程度分配的，其中综合了竞争者的自身特征、系统对其以往效用的评价、与当前系统目标的相关性等因素。尽管一个完整的问题解决过程不再遵循任何给定算法，其中的基本步骤仍各自遵循相应的算法，只不过这些“微算法”之间的组合依赖于系统的历史和环境中的很多因素，因此其整体效果一般是不可精确预测或严格再现的。

在纳思中，推理系统的框架得到了逐步扩充，以统一实现各种（传统上不被看作推理的）认知功能，如学习、联想、规划、感知、决策、操作、通讯等。这些功能不是被彼此独立的模块分别完成的，而是同一个过程的不同侧面。当然，纳思内部也是由各种“部件”组成的，但其划分和目前人工智能诸子领域的划分完全不同，而且这些部件均不能被单独用来解决应用问题，因为它们之间有极强的相互依赖性。

由于纳思的智能不是直接体现在其解决问题的能力上，而是体现为一种“元能力”（即获得解决问题能力的的能力），它自然就是通用的，而其解决某一领域中的问题的能力完全来自于系统自身的经验，包括通过感知运动界面得到的直接经验和通过语言界面得到的间接经验。因此，这种对智能的理解导致计算机是在信息加工原则上像人类靠拢，而非复制人的个别解决问题能力。尽管让计算机解决各种问题自然是有其重大价值的，但这种工作并不一定加深我们对思维一般规律的理解。人工智能几十年的历史已经充分说明了这一点。

按照我这个智能定义，智能不是全知全能，智能系统也经常会犯错误。纳思需要通过大量的学习来达到在某个领域内有实用价值的水平。即使在达到这种水平之后，学习也不会停止，尤其是在环境不断变化的情况下更是如此。这种学习不仅限于知识的积累，也包括技能的习得和动机的演化。纳思的初始动机是由外界（设计者或用户）设定的，然后系统会从中生成派生动机，并维护整个动机体系的协调性和有效性。

纳思的概念设计和计算机实现已经基本完成，目前的工作已转向调试。尽管离实际应用还很遥远，目前的系统已经表现出很多与其它人工智能系统不同，而更像人类思维活动的特征。尽管纳思不是作为一个心理学模型来设计的，其智能定义和在此基础上建立的理论模型仍可以统一地解释很多心理现象，包括注意、遗忘、灵感、直觉、审美、情感等等。由于系统经验上的根本差别，在计算机中这些现象在具体表现上和人类很不一样，但在功能和原理上却足够接近，以至于可以用相同的概念来描述。

在纳思的设计过程中，对我影响最大的哲学家包括罗素、休谟、卡尔纳普、皮尔斯、库恩、拉卡托斯等人，大概是由于他们关注点和表达方式和我的理工背景和研究兴趣比较吻合（徐英瑾对这一点有所评论，见《对话》第9页）。对于维特根斯坦，在遇到徐英瑾之前我只是了解其语言理论，尤其是其“意义即用法”和我的“基于经验的语义学”可以说是有“家族相似”在其间的。经过徐英瑾的大力开掘，居然在维氏和纳思之间建立起密切的联系，以至于将纳思看作“对于维特根斯坦哲学理想的工程学逼近”（《对话》第242页），这自然是我所乐于得知的。

（三）人工智能近期发展分析

在徐英瑾的《对话》完稿后的两年中，人工智能又一次经历了由“春”到“夏”的季节变化。这些新发展是否挑战了书中的结论？这是本节要讨论的话题。

这次变化的主要动因是深度学习使得计算机在若干领域取得了出人意料的进展，导致公众对人工智能的成功可能性的估计大幅上升，以至于开始担忧其社会后果。

“深度学习”是人工神经网络模型中的一个新技术。（LeCun, Bengio & Hinton 2015）徐英瑾在书中已经对这个模型的主要想法进行了描述（第43页）。简单说来，每个人工神经元对应一个简单函数，其中输出值依赖于诸输入值的加权总和。人工神经网络是由很多人工神经元构成一个分层结构，其中每层的输出成为另一层的输入，在数学上表示为一个向量。从整体上看，这样一个网络就对应于一个复杂函数，由最底下的输入层的向量值确定最顶上的输出层的向量值，而这个函数的细节是主要由网络结构和其中神经元之间的联结强度（权重）所决定的。

说人工神经网络有学习功能，是因为一个网络所对应的函数不是在设计时定下来的，而是通过训练得到的。一个常见的训练样本是一个输入（如一张鸟的照片）和所期望的输出（如“鸟”这个词），二者均表示为向量。当网络的实际输出与正确输出不同时，其差值被一个学习算法用来调整有关的权重。当训练结束后，网络不但可以在见到一个已知输入时给出对应的输出，还可以为从未见过的输入根据其已知输入的相似性确定输出。

最初的人工神经网络只有一个输入层和输出层。这样的网络有很大局限性，有很多函数根本无法学会。后来，中间层被引入了，网络的能力因此有了极大提升。所谓“深度学习”就是发生于有很多中间层的网络。新发现的高效学习算法和大量的训练样本以及强有力的硬件相结合，造成了网络功能的大幅度提升。这项新技术被各大公司用在多个应用之中，并吸引了媒体和公众的广泛注意。

尽管深度学习的确有重大理论和实用价值，这项技术仍有某些根本局限性。我在（Wang 2006b）中列举了用人工神经网络构建通用智能系统的一些问题：

- 人工神经网络把知识表示成输入或输出向量，这对于通用系统来说往往是不方便的，因为知识往往涉及不同的抽象水平。如徐英瑾在《对话》中（第22页）所指出的，回避知识表

征无助于问题的解决。

- 智能系统所面临的问题往往不能被看作一个输入输出之间的函数关系，因为对一个给定输入，一般不是仅有一个唯一输出与其相对应。
- 智能系统中的学习应当表现在多个方面，而非仅仅能从具体例子中总结一般规律。
- 人工神经网络的学习和工作都遵循给定算法，因此对每个问题有确定的时空资源开销，所以无法满足实时系统的要求。

因为上述问题均来自人工神经网络的基本预设，它们不能被深度学习这种在这些预设之上构建的技术所解决。如徐英瑾所指出的，人工神经网络仍受限于其行为主义的智能观，即把智能看作特定的刺激反应关系。《对话》（第53页）“对于神经网络技术的简评”中对这项技术的局限性的结论并未被深度学习的出现所推翻。

同理，深度学习并不比其它现存人工智能技术或计算机技术更危险。实际上，对人工智能危险的警告从这个研究开始之日起就没有间断过，只是在大家不认为人工智能可能实现时，没多少人会认真对待这种警告。这一次，由于Nick Bostrom的《超级智能》（Bostrom, 2014）的发表和一些重量级人物（如霍金、马斯克、盖茨）的加盟，这场讨论一时席卷各大媒体。既然人类统治地球的最重要原因是其智能上的优势，计算机在这一点上超过人类的社会后果固然值得谨慎评估。有鉴于此，有人提出“安全研究先行”的观点，认为只有在能保证人工智能不可能危害人类的条件下，这项研究才应该被进行下去。另一方面，主流人工智能专家们表示人工智能早已经进入我们的生活，因此没什么可怕。人工智能固然有危险，但和其它计算机技术的危险并无根本差别。（Dietterich & Horvitz 2015）

在这一话题上我觉得双方的观点都有问题，而其核心仍是在对“智能”的理解上。被设计来解决某个或某类应用问题的系统的确不会有比其它技术更大的失控危险，因为其能力和方法均是基本上被人类设计者所决定的，尽管系统自身可以通过学习填补某些设计细节。在这一点上，主流人工智能专家们的回应是有根据的，但这个结论只适用于他们所造的“智能系统”，而并不适用于基于对智能的其它理解所可能开发出来的未来技术。而另一方面，对人工智能缺乏了解的人们的结论往往是基于他们所想象的“智能系统”，而和实际研究结果相去甚远。

在这篇文章中全面讨论人工智能的安全问题是不可能的，因此下面我仅以“动机”为例来看现有流行论点中的问题。一个智能系统的行为是被动机或目标所导向的。和生物与生俱来的本能性动机不同，人工智能的初始动机必定是由其设计者确定的，和其智能水平不具有必然的相关性。这就是说“高智能”既可用来做善，也可用来做恶。因此，初始动机的选择自然是至关重要的。传统系统只有单一动机，从中派生出的目标集也不会超出设计范围。但如果系统的智能很高，一个看似无害的目标也可能在多次派生后产生意想不到的后果。据Bostrom设想，如果一个超级智能系统以制造回形针为终极目标，那它有可能耗尽所有资源，以至于要消灭人类，只因为它相信这样可以制造更多的回形针。为避免这种灾难发生，Bostrom建议在人工智能的终极目标中包含全部人类道德和价值，并保证派生出的目标集与其一致。

这种建议虽然听上去很有道理，但实际上基本行不通。首先，确定一个能代表人类道德和价值的终极目标显然不是一个很容易取得共识的任务。即使这样一个目标可以被明确表达，它和要造回形针这样的具体任务之间的逻辑关系也不是直截了当的，往往要通过若干步中间结果。由于这个原因，同时由于知识和资源限制，实际上我们往往无法准确预知一个行动的所有后果。比如说，那些主张

人工智能研究应当缓行的人实际上无法证明这样做可以的确使人类更安全，因为有些未来的灾难可能恰好是靠人工智能才能避免的呢！

在（Wang 2012）中，我以纳思为例讨论了通用智能系统中的动机问题。其主要结论是：这样一个系统一般同时有多个目标，而它们之间常常彼此竞争和冲突。一个操作的被执行不是单个目标所决定的，而是多个目标相互平衡的结果。而恰恰是这种相互制约使一个系统避免了极端化的行为。如果一个人工智能系统想通过消灭人类来实现世界和平，它肯定不是真的有智能。此外，智能系统中的“当前动机”不完全被其“初始动机”所决定，而在很大程度上被其经验所决定，就好像一个成年人的动机大部分不是先天决定的，而是逐渐形成的。这些因素不仅没有被Bostrom等人考虑到，也尚未进入人工智能主流视野。

基于上述理由，我对人工智能的安全问题的基本看法是：像科学史上其它重大课题一样，人工智能对人类既是个挑战又是个机会。因此，对其社会后果抱谨慎态度是绝对必要的，尤其是要尽量防止仓促地把理论成果投入实际应用。但另一方面，对人工智能的安全问题的研究必须以对相关理论和技术问题的深入了解为前提。具体说来，由于智能系统必须是有适应性的，其行为不仅是被其（先天）设计所决定的，而更取决于其（后天）教育、训练、使用等环节，而后者往往是在有关的讨论中被忽视的。

（四）结论

由于人工智能的领域特征，其中有大量的哲学问题。我们甚至可以说只有在哲学上正确的方案才存在技术上成功的可能。但不幸的是，很多人工智能工作者缺乏在哲学层面上检讨其理论预设的兴趣和能力，而哲学工作者又常常缺乏深入了解技术问题的勇气和基础。

在这种情形下，徐英瑾的《心智、语言和机器——维特根斯坦哲学和人工智能科学的对话》代表了一种难能可贵的眼光、决心、和才能的结合。尤其是其不跟风、不怕难、不惧做少数派的态度，更是当前学术界所需要的。

我希望这本书的出版会对人工智能的哲学研究起到促进和示范的作用。在这个领域，人工智能工作者和哲学工作者相互理解的努力是会使双方受益的。

参考文献

Mark Baard, AI founder blasts modern research. *Wired News*, May 13, 2003.

Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press (Oxford), 2014.

Tom Dietterich & Eric Horvitz, Benefits and Risks of Artificial Intelligence, Blog at <https://medium.com/@tdietterich/benefits-and-risks-of-artificial-intelligence-460d288cccf3#>, 2015

- Edward A. Feigenbaum & Julian Feldman (editors), *Computers and Thought*, MIT Press (Cambridge, MA), 1995.
- Yann LeCun, Yoshua Bengio & Geoffrey Hinton, Deep learning, *Nature* 521:436-444, 2015.
- Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd edition, Prentice Hall (Upper Saddle River, NJ), 2010.
- Pei Wang, *Non-Axiomatic Reasoning System: Exploring the essence of intelligence*, Dissertation for the degree of Doctor of Philosophy in Computer Science and Cognitive Science, Indiana University, 1995.
- Pei Wang, *Rigid Flexibility: The Logic of Intelligence*, Springer (Dordrecht) 2006a.
- Pei Wang, Artificial general intelligence and classical neural network, *Proceedings of the IEEE International Conference on Granular Computing*, 130-135, Atlanta, USA, 2006b.
- Pei Wang, Three fundamental misconceptions of artificial intelligence, *Journal of Experimental & Theoretical Artificial Intelligence*, 19(3):249-268, 2007.
- Pei Wang & Ben Goertzel, Introduction: Aspects of artificial general intelligence, In *Advance of Artificial General Intelligence*, 1-16, IOS Press (Amsterdam), 2007.
- Pei Wang, Motivation management in AGI systems, *Proceedings of AGI-12*, 352-361, Oxford, UK, 2012.
- Pei Wang, *Non-Axiomatic Logic: A Model of Intelligent Reasoning*, World Scientific Publishing (Singapore), 2013.