



心智、语言和机器

——维特根斯坦哲学和人工智能科学的对话

顾名思义，“人工智能哲学”的任务，就是从哲学的角度，
对“人工智能”科学的观念前提和工作方法，进行反思性的研究……

徐英瑾 著

心智、语言和机器

——维特根斯坦哲学和人工智能科学的对话

徐英瑾 著

 人民出版社

责任编辑: 杜文丽

封面设计: 汪莹

版式设计: 周方亚

图书在版编目(CIP)数据

心智、语言和机器: 维特根斯坦哲学和人工智能科学的对话 / 徐英瑾 著.

—北京: 人民出版社, 2013.10

ISBN 978-7-01-012597-8

I. ①心… II. ①徐… III. ①人工智能—科学哲学 IV. ①TP18-02

中国版本图书馆CIP数据核字(2013)第226082号

心智、语言和机器

XINZHI YUYAN HE JIQI

——维特根斯坦哲学和人工智能科学的对话

徐英瑾 著

人民出版社 出版发行

(100706 北京市东城区隆福寺街99号)

北京中科印刷有限公司印刷 新华书店经销

2013年10月第1版 2013年10月北京第1次印刷

开本: 787毫米×1092毫米 1/16 印张: 29.75

字数: 760千字 印数: 0,001—3,000册

ISBN 978-7-01-012597-8 定价: 65.00元

邮购地址 100706 北京市东城区隆福寺街99号

人民东方图书销售中心 电话(010)65250042 65289539

版权所有·侵权必究

凡购买本社图书, 如有印制质量问题, 我社负责调换。

服务电话: (010) 65250042

鸣 谢

本研究得到如下基金的资助：

1. 复旦大学“985 工程”三期整体推进人文学科研究项目“对形式语义学之‘动态转向’的跨学科研究”(项目号 2011RWXKYB038)
2. 上海市教育委员会“曙光计划”项目“机器翻译中的哲学问题”(项目号 13SG11)
3. 国家社科基金重大项目“认知科学对当代哲学的挑战——心灵与认知哲学重大理论问题研究”(项目号 11&ZD187)
4. 全国优秀博士学位论文作者年专项资金资助项目“当代英美心灵哲学前沿问题研究”(项目号 200701)

全书文献引用规范说明

本书引用文献,均参照自然科学和工程技术类文献通用引用规范。即在需要引用文献处加入括号,括号内为文献作者的姓以及文献的出版年代。若引用出现直接引文,则在出版年代后标注文献的页码。该文献的版本信息统一在文后的“参考文献”中列出。“参考文献”中诸文献的排列次序以作者的姓的英文字母顺序排列,同一作者的文献按照出版年代先后排列。为节省篇幅,正文不再直接给出文献的版本信息。

举例说明。若正文中出现如下文字:

“但问题是,‘当系统变得越来越复杂的时候,那么理论对象之间可能的联系数量也会呈指数式爆炸,其带来的计算量,恐怕是任何计算机和有生命的认知系统都无能为力’ (Riegler 2002, p. 342)。”

上文括号内的信息,就表示这段引文来自一位姓为 Riegler 的作者于 2002 年发表的某部文献的页 342。查阅书后“参考文献”可知,这部文献的完整版本信息为:

Riegler, Alexander (2002), “When Is A Cognitive System Embodied?”, *Cognitive Systems Research* 3: 339–348.

其汉语意思为:亚历山大·里尔格尔的论文《一个认知系统何时是被具身化的?》,收录于杂志《认知系统研究》总第三卷,页 339–348。

导 论 哲学文化、人工智能科学以及人工智能哲学	1
0.1 为何说 AI 科学中的哲学维度是不可还原的?	2
0.2 哲学文化渗入 AI 的几个具体案例	4
0.22 案例二:德瑞福斯对主流 AI 的批判	6
0.23 案例三:布鲁克斯的具身化机器人方案	7
0.24 案例四:框架问题	8
0.3 不同国家的哲学文化对于 AI 科学发展的影响	9
0.31 美欧比较	9
0.32 美日比较	9
0.33 美苏比较	10
0.34 中美比较	11
0.4 人工智能哲学在国内外的研究现状概览	12
0.41 海外的人工智能哲学研究	12
0.42 国内的人工智能哲学研究	17
0.5 本书研究之路线图	18
0.51 我们需要一份怎样的人工智能哲学研究成果?	18
0.52 本书章节结构简介	20

第 1 篇

人工智能科学和人工智能哲学基础知识引论

第 1 章 人工智能核心技术面面观	30
1.1 人工智能学科中的传统派——符号人工智能	32
1.1.1 物理符号系统	33
1.1.2 “图灵机”概念和其概念衍生物	34
1.1.3 物理符号假设	37
1.1.4 简评“老而妙的人工智能”进路	41
1.2 新派人工智能之代表：“联结主义”或“人工神经网络”进路	43
1.2.1 自然神经元的工作方式	44
1.2.2 人工神经元的数学结构	46
1.2.3 人工神经元之间的联结方式	48
1.2.4 人工神经网络的学习规则	49
1.2.5 神经网络技术的运用：模式识别	50
1.2.6 对于神经网络技术的简评	53
1.3 遗传算法	54
1.4 贝叶斯网络技术	59
1.4.1 贝叶斯定理、全概率规则、链律	60
1.4.2 贝叶斯定理支配下的非单调性推理	62
1.4.3 在计算联合概论时直接使用链律所导致的实践困难	64
1.4.4 贝叶斯网络	65
1.4.5 从“司马光砸缸”看贝叶斯网络技术的局限性	69
第 2 章 人工智能科学在十七、十八世纪欧洲哲学中的思想先驱： 从笛卡尔到康德	75
2.1 人工智能科学和十七、十八世纪欧洲哲学关系刍议	77
2.2 笛卡尔和莱布尼茨：机器智能的反对者	80
2.3 霍布斯：符号人工智能之先祖	82
2.4 休谟：联结主义的哲学先驱	84
2.5 康德：“由上而下”进路和“自下而上”进路的整合者	86
2.6 康德的道德哲学和阿西莫夫的“机器人三定律”	92

第3章 机器智能在当代所遭遇到的哲学批判:从塞尔到彭罗斯..... 96

3.1 塞尔的“汉字屋论证”,以及它为何不成立.....	96
3.11 对于“汉字屋论证”之重述.....	97
3.12 郝泽对于塞尔论证之宏观逻辑结构的两种诊断模式.....	99
3.13 丹普尔对于汉字屋论证逻辑结构的两种诊断模式.....	101
3.14 笔者对于汉字屋论证逻辑结构的第五种诊断模式.....	103
3.15 对于“汉字屋论证”的深入检讨.....	104
3.2 彭罗斯对于机器智能的批判.....	107
3.21 基于“哥德尔不完备性定理”的反机器智能证明.....	108
3.22 人工智能需要关心大脑结构的量子层面吗?.....	119

第4章 维特根斯坦和人工智能科学关系刍议..... 122

4.1 为何要请出维特根斯坦?.....	122
4.11 维特根斯坦哲学的“综合”气质.....	123
4.12 维特根斯坦对于 AI 技术之哲学预设的深刻领悟.....	124
4.2 维特根斯坦哲学对于 AI 研究三大前沿领域的指导意义.....	127
4.21 中央推理系统的构建,以及“框架问题”的可能解决方案.....	127
4.22 中央推理系统和边缘感知模块之间的关系——以机器视觉的研究为例.....	130
4.23 统一的自然语言处理机制的设计.....	132
4.3 珊克教授关于“维特根斯坦—AI”关系的悲观论调,为何都站不住脚?.....	134
4.31 维氏本人曾说过“机器不可能思维”吗?.....	134
4.32 维氏意义上的“遵从规则”,就意味着无法以机械之方式遵从规则?.....	138
4.33 现代 AI 研究预设了所谓的“心理学主义”了吗?.....	140

第2篇

知识表征系统和中央信念系统的构建

第5章 《逻辑哲学论》的遗留难题(一):“知识表征” 视角中的“颜色不相容问题”.....	146
5.1 《逻辑哲学论》和“知识表征”.....	146
5.11 从“知识表征”的角度重读《逻辑哲学论》.....	146
5.12 《逻辑哲学论》的“外延主义”预设.....	149

5.2 外延主义支柱下的定时炸弹:颜色不相容问题	151
5.21 “颜色不相容”何以成为“问题”?	151
5.22 深入讨论:为何颜色不相容问题无法在外延主义框架中得到消化?	153
5.23 如何在一种非外延主义的框架内安顿“颜色空间”?——维氏本人的启示	157
5.24 如何在一种非外延主义的框架内安顿“颜色空间”?——邱琪兰德的启示	158
5.25 邱琪兰德带给我们的启发及教训	161

第6章 《逻辑哲学论》的遗留难题(二):

概然性推理和归纳推理的表征	164
6.1 《逻辑哲学论》论概率	166
6.2 《逻辑哲学论》论归纳	170
6.3 插曲:《逻辑哲学论》对于概率概念的日常用法的说明	173
6.4 《维特根斯坦之声》和《大打字稿》关于概率和归纳的四段评论以及相关的注解	176
6.41 “引文一”及其注读	177
6.42 “引文二”及其注读	179
6.43 “引文三”及其注读	181
6.44 “引文四”及其注读	183

第7章 处理语义相关性的新技术平台:纳思系统

7.1 一种合格的“相关性”理论所应当满足的五大标准	192
7.2 为何相干逻辑不相关?	194
7.3 为何贝叶斯进路不相关?	196
7.4 为何描述逻辑依然不相关?	202
7.5 纳思系统引论	206
7.6 对于“相关性证据问题”的初步解决方案	210
7.7 对于“相关性证据问题”的一个更为精致的解决方案	213

第8章 一个维特根斯坦主义者眼中的“框架问题”

8.1 何为“框架问题”?	218
8.2 “关于惯性的常识法则”——解决框架问题的救命稻草?	222
8.3 让智能系统预先获得一个关于世界的物理学结构又如何?	225
8.4 来自维特根斯坦的启发	229
8.41 G.E. 摩尔对于维特根斯坦的刺激	230
8.42 为何摩尔的“基本常识命题”无法在形式系统中被全面地表征?	231
8.43 维特根斯坦对于信念网的动力学特征的描述	236
8.5 对于维特根斯坦哲学理想的工程学逼近	242

8.51 由纳思逻辑规则支持的动态语义网构建	243
8.52 纳思的长期记忆组织	248
8.53 纳思的实时任务管理	249

第 3 篇

人工视知觉模块的构建

第 9 章 “具身化挑战”和对视觉机制的人工模拟	254
--------------------------	-----

9.1 回应具身化挑战的第一个模式:联结主义	255
9.2 回应具身化挑战的第二个模式:(非符号化的)具身化认知	256
9.3 澄清:具身化不可以去符号化为代价	259
9.4 为何选择视觉?	261
9.5 人工视觉研究现状简介	262

第 10 章 维特根斯坦论“视—思”关系	265
----------------------	-----

10.1 维特根斯坦关于视觉问题的核心论点:“感知—语义连续论”	265
10.2 延伸性论点(甲):“低级知觉—高级知觉”之分, 并非“无语义—有语义”之分	269
10.3 延伸性论点(乙):感觉材料组织方式之“完型”特征, 乃是语义推理之灵活性的体现	271
10.4 延伸性论点(丙):三维视知觉对象的建立有赖于身体运动模式的参与	276
10.5 维特根斯坦的视觉理论和关于主动视觉的工程学研究	279
10.6 小结:从关于视觉的哲学讨论过渡到视觉科学	282

第 11 章 再论“视—思”关系——玛尔和派利辛的视觉理论批判	284
---------------------------------	-----

11.1 玛尔的计算视觉进路	286
11.11 关于视觉机制描述的三层次说	286
11.12 对于玛尔工作的评论	292
11.2 派利辛的视觉索引论	294
11.21 派氏断裂论和玛尔的视觉表征论,尤其是佛笃的模块论之间的关系	294
11.22 派利辛用以支持断裂论的五个论证	297
11.23 派利辛对于“纯摹状词观点”的批判和他的“视觉索引论”	303

11.24 对于“视觉索引论”的评论	309
第 12 章 如何在纳思系统的技术平台上,构建一种新的人工视知觉机制?	315
12.1 为何选择纳思系统?	315
12.2 纳思系统视觉系统是如何获取客体的三维形状信息的?	322
12.21 彼德曼的“RBC 理论”	323
12.22 RBC 理论和纳思系统技术平台的互相融合	326

第 4 篇

自然语言处理模块的构建

第 13 章 从图灵到孔夫子	336
——人工智能系统的自然语言处理,及其七种哲学预设	
13.1 华生或斯金纳式的行为主义	337
13.2 外在论的语义学	338
13.3 莱布尼茨的“理想语言”假说	339
13.4 乔姆斯基的“深层句法”假说	340
13.5 休谟式的统计学进路	342
13.6 康德式的混合式进路	344
13.7 孔子式的基于实例的理解进路	345
第 14 章 维特根斯坦哲学视野中的自然语言理解	351
14.1 关于“语言游戏说”的第一组二律背反:关于行为主义	352
14.2 关于“语言游戏说”的第二组二律背反:关于统计学	354
14.3 关于“语言游戏说”的第三组二律背反,关于先天语法和思想语言	358
14.4 关于“语言游戏说”的第四组二律背反:关于实例在语言理解中的作用	362
14.5 对于四组二律背反的初步融贯化:寻找一条“既……又……”的中道	366
14.6 关于“最简先天论”的问答集	368
第 15 章 如何让计算机真正懂汉语?	383
——一种以许慎的“六书”理论为启发的汉语信息处理模型	
15.1 从词到字,再到《说文解字》	385

15.2	如何在纳思系统的平台上刻画“六书”以及相关的语法构建？	392
15.21	象形	396
15.22	指事	401
15.23	会意	404
15.24	形声	409
15.25	转注	410
15.26	假借	411
15.3	从字法转换到词法	413
15.31	多音的根词(从象形字到象声字的转变)	414
15.32	偏正式(会意字变种之一)	414
15.33	并列式(会意字变种之二)	416
15.34	重迭式(形声字变种)	417
15.35	前置成分和后置成分(形声字变种)	417
15.36	动宾式(两种会意结构的递归构造)	417
15.37	补充式(动宾式的变体之一)	422
15.38	主谓式(从词到句的过渡形式)	422
15.4	总结和答疑	424
尾 声 美丽的蓝图,以及蓝天中的两朵“乌云”		427
参考文献		435
专业术语索引		448
中文人名索引		454
外文人名索引		455
后 记		459

本书技术插图信息汇总

图 1—1a	图灵机的加法运算(初始状态)·····	36
图 1—1b	图灵机的加法运算(停机状态)·····	36
图 1—2	人脑神经元结构简图(转自互联网)·····	45
图 1—3	阶跃函数的图像·····	47
图 1—4	S—曲线函数的图像·····	47
图 1—5	人工神经网络的结构和工作流程示意·····	48
图 1—6	形态各异的阿拉伯数字“6”的手写体·····	51
图 1—7	阿拉伯数字“6”之图像的数字矩阵·····	51
图 1—8	用色块表示的二维迷宫·····	55
图 1—9	二维迷宫的数字矩阵·····	56
图 1—10	父本决策和母本决策各自的染色体数码形式·····	57
图 1—11	两个子代决策方案的染色体数码形式·····	58
图 1—12	四事件之间因果关系示意(带数量未经压缩的数据槽)·····	65
图 1—13	四事件之间因果关系示意(不带数据版)·····	66
图 1—14	四事件之间因果关系示意(带数量已被压缩的数据槽,以及槽内数据)·····	67
图 1—15	司马光的朋友的问题解决路径的贝叶斯网络化·····	70
图 1—16	司马光的问题解决路径的贝叶斯网络化·····	73
图 2—1	寻找“齐桓公最有名的文臣”之匹配项的模拟推理流程·····	88
图 2—2	“照猫画虎”(Copycat)程序对于字符串之间的类比关系的搜索过程示意·····	90
图 2—3	“照猫画虎”(Copycat)程序各个部分之间的信息交换流程·····	92
图 5—1	维特根斯坦绘制的颜色概念双棱锥图·····	154
图 5—2	对“G(x)”解释空间的增扩方案·····	156
图 5—3	两层颞颞细胞所构成的詹—胡网·····	159
图 5—4	詹—胡网颞颞细胞的活动所构成的抽象数学空间, 以及在该空间完成的颜色语义编码·····	161
图 7—1	一个带数据槽的简单的贝叶斯网络·····	199

图 7—2	对于“尼克松菱形”的贝叶斯网络刻画方式	201
图 7—3	一个最简单的纳思语义网示例	207
图 7—4	由“图 7—3”所表示的纳思语义网扩展而成的新语义网	209
图 7—5	一个包含着隐含冲突知识的纳思语义网	209
图 7—6	对于正面证据的第一种纳思式表征图	211
图 7—7	对于负面证据的第一种纳思式表征图	211
图 7—8	对于正面证据的第二种纳思式表征图	211
图 7—9	对于负面证据的第二种纳思式表征图	212
图 7—10	对于不相干“证据”的纳思式表征图	212
图 7—11	对于“乌鸦悖论”的纳思式表征图	213
图 8—1	信念网所构成的经验知识的可能分布空间	237
图 8—2	经验命题在信念网中的“流动”	238
图 8—3	纳思系统的各构成要件的功能分布	242
图 8—4	一个带有连接权重的简化的纳思语义网	243
图 8—5	一个带有连接权重的纳思语义网(图 8—4 中语义网的增扩版)	243
图 8—6	一个收敛的纳思语义网	246
图 8—7	是非疑问句在纳思语义网中的表征	247
图 8—8	概念标签的覆盖范围和本地节点群的覆盖范围示意	249
图 9—1	人工神经元(联结主义)网络的工作流程,以及其和语义符号层的关系	256
图 9—2	纳思系统的表征形式和物理世界的接口设计问题	260
图 9—3	人工视觉研究的子课题分布情况一览	263
图 10—1	两类颜色判断的性质对比:内在关系 Vs 外在关系	266
图 10—2	兔—鸭图(此图原出自心理学家 Jastrow, 被维氏引用)	267
图 10—3	“看”与“看作”的区分——一种认知心理学的解读	271
图 10—4	雷哈的格式塔泡模型	274
图 10—5	在视觉空间中和在物理空间中的尺寸判定	277
图 11—1	玛尔的视觉信息加工三阶段说简图	288
图 11—2	一个圆柱的 2.5 维简图	290
图 11—3	对于复杂立体物的几何分解	290
图 11—4	各种广义锥	291
图 11—5	广义锥组合体的信息库分类简图	291
图 11—6	穆勒—莱尔错觉简图	298
图 11—7	“变形填充”(amodal completion)图	299
图 11—8	MOT 实验的四阶段	306
图 11—9	机器人视觉的三种表征方式:逻辑表征、索引词表征以及内部心理图像 (引自 Pylyshyn 2000)	309
图 12—1	一张简化的纳思语义网	316

图 12—2	一张包含“心像”的纳思语义网	316
图 12—3	六种典型几何离子的构成方式(转引自 Biederman 1987, 图 7)	325
图 12—4	三角形的不同殊相和三角形“标准相”之间的语义关系	327
图 12—5	对视觉对象的基本几何属性的侦测流程	328
图 12—6	对视觉对象的客观几何轮廓的推理和记录	329
图 12—7	一种不受推荐的对于空间关系词的纳思表征方式	331
图 12—8	两种几何离子的空间构成方式	331
图 12—9	纳思视觉系统对物理对象的空间属性进行侦测和标记的流程总图	333
图 15—1	一个简化的纳思语义网	397
图 15—2	以心像为词项的纳思网	398
图 15—3	新知觉对象在纳思语义网中的四种编码方式	400
图 15—4	象形字在纳思网中的构成	400
图 15—5	指事字在纳思网中的构成	403
图 15—6	会意字的意义融合机制的静态表征	407
图 15—7	会意字之间的意义推理关系	407
图 15—8	形声字在纳思系统中的构成	410
图 15—9	转注字在纳思系统中的构成	411
图 15—10	假借字在纳思系统中的构成	412
图 15—11	纳思系统在“一词多义”的情况下筛选出目标意义的推理过程	413
图 15—12	多音的根词在纳思系统中的构成	414
图 15—13	偏正词在纳思系统中的构成	415
图 15—14	并列词在纳思系统中的构成	416
图 15—15	重迭词在纳思系统中的构成	417
图 15—16	前置 / 后置成分词在纳思系统中的构成	418
图 15—17	纳思系统对于“吃”的意义的学习(一)	419
图 15—18	纳思系统对“吃”的意义的学习(二)	420
图 15—19	纳思系统对“说”的意义的学习	422
图 15—20	补充式语词在纳思系统中的表征	423
图 15—21	主谓式语词在纳思系统中的表征	423

哲学文化、人工智能科学以及人工智能哲学

读者现在打开的,乃是一本关于人工智能哲学(philosophy of Artificial Intelligence)的著作。顾名思义,“人工智能哲学”的任务,就是从哲学的角度,对“人工智能”(Artificial Intelligence,以下简称为 AI)科学的观念前提和工作方法进行反思性的研究。从学科分类上讲,该哲学分支应当属于广义的“特定科学的哲学”(philosophy of specific disciplines of natural science)之范畴,并和“物理学哲学”(philosophy of physics)、“生物学哲学”(philosophy of biology)等哲学分支相平级。但由于其和心灵哲学(philosophy of mind)之间明显的学术牵扯,西方学界更习惯视其为心灵哲学的一个下属研究领域。尽管在目前的英语世界,人工智能哲学的研究已经颇为繁荣,但在我国,它所受到的重视还很不够(相关文献概览详见后文)。其背后的原因,其实也不难想见。仅就国内的情况而言:一方面,不少 AI 学科领域内的学者们都视“天马行空”的哲学思辨为扎实的科学研究的巨大敌手,唯恐避之不及;另一方面,由于 AI 自身较高的学科门槛儿,人文学科训练出身的哲学学者也往往感到对其“难以下手”。于是乎,在汉语学术界,“人工智能哲学”便成为了哲学家和 AI 科学家都无法深入的“两不管”地带,任凭其间沃土抛荒,杂草丛生。

然而,抛荒的土地终归要被耕翻,被现代学术分工所隔离的研究工作,也必须被彼此贯通,否则,一些需要多学科方法协力研究的问题,也就难以得到真正的解决。而在人工智能科学中,这样的问题可谓俯首即拾。或说得更清楚一点:人工智能科学的研究需要哲学思维的参与,需要来自于哲学思维的帮助。

为了能够充分论证这个观点,本“导论”将对以下分论点进行逐一阐述:

- 其一,笔者将在概念分析的层面上,揭示 AI 科学中的哲学维度的重要性和不可还原性;
- 其二,笔者将给出具体的技术案例,指出不同的哲学思想和相关 AI 的构建方案之间的深刻关联;
- 其三,笔者将从科学社会学的角度,揭示不同国家的哲学文化对于其 AI 研究所造成的积极的或消极的影响。

在作出这些分析的情况下,笔者将在本“导论”的余下篇幅中,对历史上已有的国内外“人工智能哲学”研究文献进行综述和评价。最后,笔者将对本书的总体研究计划进行概述,并对其核

心论点、理论终点和创新点进行提示。

0.1 为何说 AI 科学中的哲学维度是不可还原的？

在证明这一分论点之前，我们还是需要来看看哲学研究的基本任务为何。笔者将哲学工作的特征归结为以下三条：

第一，思考大问题，澄清基本概念。这里所说的“大问题”，即极具基础意义的问题。比如，数学哲学家追问数学家“数”的本性是什么，物理学哲学家追问物理学家“物质”、“能量”的本性为何，生物学哲学家追问生物学家“生命”的本性是什么。与哲学家相比较，一般的自然科学家往往只是在自己的研究中预设了相关问题的答案，却很少系统地反思这些答案的合法性。

第二，在不同学科的研究成果之间寻找会通点，而不受某一具体学科视野之局限。比如，科学哲学家往往喜欢追问这样的问题：如何会通生物学研究的成果和化学研究的成果？是不是所有的生物现象，都可以还原为更为微观的化学现象？而所有的化学现象，是否复又可被还原为更为微观的物理学现象？或者，存在着一种不同于“还原论”的会通方式？——相比较而言，职业科学家对于这些跨学科问题虽或偶有反思，但往往也不够系统和深入。

第三，重视论证和辩护，相对轻视证据的约束。这也就是说，评价哲学工作优劣的标准，主要是看一个哲学论证本身的合理性和常识的可接受性，却一般不用受到严格的科学证据的检测。而对于科学而言，合理的辩护程序却必须和实打实的经验证据相互匹配，否则得出的结论就无法被科学共同体所接受。这种差异，一方面固然使得哲学工作的自由度要远大于科学工作，但另一方面也使得哲学争议往往不如科学争议那样，容易取得学科共同体内部的一致意见。

综合以上三点我们不难发现，经过正规哲学训练的学者，在精神气质方面便很容易具备这样的特质：喜欢刨根问底，喜欢融会贯通，不受制于一门特殊经验科学的思维方式（或套用孔子在《论语·为政》中的训诫来说，“君子不器”），并倾向于对敌对的学术观点保持一种“绅士风度”，视哲学争议为正常。笔者将这些文化特质，统统归到“哲学文化”这个大的标签之下。但需要指出的是，在今天典型的科学训练中，上述这种“哲学文化”在一定程度上是受到排斥的。首先，对于处于“学徒期”的科学入门者而言，学会服从既定的研究范式乃是其第一要务，而对这些范式的“哲学式怀疑”则会导致其无法入门；其次，严格的一级、二级、三级学科分类导致学生们忙于如何熟悉特定领域内的研究规范，而无暇开拓视野，浮想联翩；最后，对于权威科学模式的服从，在一定程度上也压制了那些离经叛道的“异说”的话语权（与之相比较，在哲学界内部，对于“异说”的宽容度相对较高——只要你的论证符合一般的论证规范，任何古怪的观点都可以自由提出）。

但凡事都有例外。就对于哲学文化的宽容程度而言，AI 科学绝对算是个科学界内部的异数。从某种意义上说，该学科本身的诞生，就恰恰是“头脑风暴”般的哲学思辨的产物。

说到该学科的起源，就不能不谈到一篇经典论文和一个重要会议。1950 年 10 月，伟大的英国数学家、逻辑学家和计算机科学的理论奠基人阿兰·图灵（Alan Turing, 1912—1954）在英国哲