机器意识研究综述

秦瑞琳 [1] 周昌乐 [1] 晁 飞 [1]

摘 要 意识问题是尚未解决的重大哲学问题与科学问题. 机器意识是人工智能最前沿的研究领域之一. 研发意识机器人对于人工智能与机器人学的发展具有重要科学意义与应用价值. 本文首先介绍了意识与感受性的相关概念和理论; 然后, 详细讨论了机器意识的概念与研究分类、实现方法与计算模型, 重点论述了实现机器意识的量子方法; 最后, 总结了机器意识目前面临的困境与未来可能的发展, 并给出了一套机器意识总体实现框架.

关键词 意识, 机器意识, 意识机器人, 感受性, 人工智能

引用格式 秦瑞琳, 周昌乐, 晁飞. 机器意识研究综述. 自动化学报, 2021, **47**(1): 18−34

DOI 10.16383/j.aas.c200043

A Survey on Machine Consciousness

QIN Rui-Lin[1] ZHOU Chang-Le[1] CHAO Fei[1]

Abstract Consciousness is an unsolved significant philosophical and scientific issue. Machine consciousness is one of the forefront research areas of artificial intelligence, and developing conscious robots is of great scientific significance and application value for artificial intelligence and robotics. In this survey, the concepts and theories of consciousness and qualia are introduced first. Then, the concept and taxonomy, the implementation methods and computational models of machine consciousness are discussed in detail. Especially, the quantum method on implementing machine consciousness is emphasized. Finally, the current difficulties and future development of machine consciousness are summarized; and an overall implementation framework of machine consciousness is proposed.

Key words Consciousness, machine consciousness, conscious robot, qualia, artificial intelligence

Citation Qin Rui-Lin, Zhou Chang-Le, Chao Fei. A survey on machine consciousness. *Acta Automatica Sinica*, 2021, **47**(1): 18−34

意识是什么? 人类意识是如何产生的? 动物是否具有意识? 这些问题一直困扰着人们. 哲学家 Dennett 曾说: “人类意识是最后幸存下来的未解之谜”[1]. 2005 年, Science 杂志提出了 125 个尚未解决的科学问题, 其中第 2 个问题就是 “意识的生物学基础是什么?”[2]. 2018 年, Nature 杂志提出了目前最重大的 6 个科学问题, 其中之一便是 “什么是意识? ”[3]. 如今, 随着脑科学的迅速发展, 尤其是各种脑成像设备 (如 EEG、fMRI 等) 的广泛应用, 出现了大量关于意识的脑科学研究成果[4−6], 形成了一些初步的意识科学理论, 人们对意识的神经相关物 (Neural correlates of consciousness, NCC) 和意识的产生机制有了更加深刻的认识. 同时, 物理学家基于意识和量子现象的相似性, 提出了一些意识的量子理论[7−9], 希望利用量子力学来解决意识问题.

随着人工智能和机器人学的发展, 人们开始思考机器能否具有意识这一问题, 这类研究逐渐被称为机器意识 (Machine consciousness, MC) 或人工意识 (Artificial consciousness, AC)[10−12]. 近年来, Nature 和 Science 杂志中出现了越来越多关于机器意识的研究成果[13−16]. 经过 30 多年的发展, 人们已提出了一些机器意识理论, 如全局工作空间理论、整合信息理论等[17], 根据这些理论开发的意识机器人能够表现出一定的意识行为, 并广泛应用于工业、教育、医疗、娱乐等领域[18−20]. 机器意识的研究已为当前各类机器人的发展提供了新的契机, 例如发育型机器人、协作机器人、机器人导航、机器人轨迹规划、机器人移动行为预测等[21−25]. 同时, 研发意识机器人也能促进人们对意识的理解, 推动构建更加完善的意识理论.

但是, 机器意识的研究目前还处于很初级的阶段, 例如针对自我意识、感受意识这些意识研究中的核心问题, 还少有涉及. 目前的机器人基于预先编程算法, 虽然能表现出一些意识行为, 但机器并

收稿日期 2020-01-21 录用日期 2020-06-01
Manuscript received January 21, 2020; accepted June 1, 2020
国家自然科学基金 (61273338, 61673322) 资助
Supported by National Natural Science Foundation of China (61273338, 61673322)
本文责任编委 曾志刚
Recommended by Associate Editor ZENG Zhi-Gang
1. 厦门大学信息学院人工智能系 厦门 361005
1. Department of Artificial Intelligence, School of Informatics, Xiamen University, Xiamen 361005

不理解其所执行的内容, 也不具有 “自我”的概念, 更没有对于自身以及外部环境的感受. 而这些都是人类意识中的核心部分, 机器意识研究的最终目的也就是实现这样的意识机器人. 如此, 机器人才能不受程序支配, 具有内省反思能力和情感体验, 从而能更好地生存与学习, 更好地与人交互. 然而, 由于意识具有超逻辑性, 并不是算法所能把握的, 因此在基于图灵机的机器上采用传统人工智能方法, 如符号计算和人工神经网络, 是不能实现这一目的的. 为此需要采用新的方法和技术, 如量子计算、脑机融合等. 量子计算比经典计算具有更强的计算能力与描述能力, 且具有真正的不确定性, 能突破预先编程的限制, 因而更适合描述复杂的意识现象. 而脑机融合技术则充分结合了生物智能与机器智能, 通过构建脑机混合机器实现大脑与机器的协同工作, 进而最终实现机器意识.

下文对意识与机器意识近期的研究进展进行综述. 图 1 展示了机器意识研究内容与方法分类. 我们首先对意识和感受性问题作一介绍. 在此基础上, 详细讨论机器意识的概念与研究分类、具体实现方法与计算模型, 重点论述其中的量子方法. 最后, 总结了机器意识面临的困难与未来的发展, 并给出了一种机器意识总体实现框架.

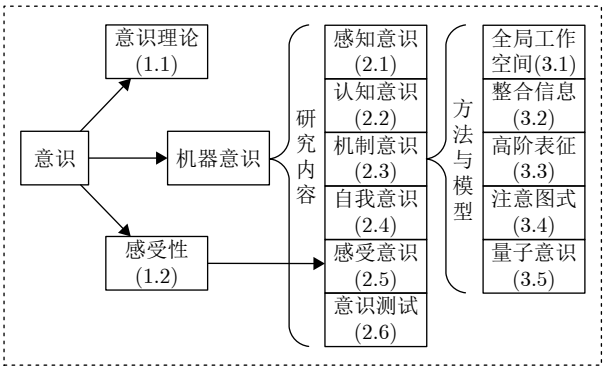


图 1　机器意识研究内容与方法分类

Fig. 1　The taxonomy of contents and methods of machine consciousness

本文的贡献主要体现在:

1) 对机器意识的分类、理论、方法、模型等方面的最新研究进行了详细综述, 填补了国内机器意识研究综述的空白, 为国内意识机器人研发提供了有益指导.

2) 对机器意识面临的困境和未来的发展给出了建设性的意见, 提出了一种机器意识总体实现框架, 为机器意识的进一步发展指明了方向.

1　意识与感受性

1.1　意识的概念与理论

目前, 意识还没有一个清晰的定义, 不同领域, 如心理学、医学、脑科学等, 对意识的界定不尽相同. 在心灵哲学领域, 哲学家主要研究意识的容易问题与困难问题、统一性问题、意向性问题、心-身问题等, 提出了很多关于意识起源和本质的哲学理论[26–27], 如表 1 所示. 科学家们则通过研究意识的脑机制, 提出了众多的意识科学理论, 如表 2 所示.

表 1　意识的哲学理论

Table 1　Philosophical theories of consciousness

理论名称	英文	主要观点
一元论	Monism	意识或物质是世界的本源
二元论	Dualism	意识和物质都是世界的本源
神秘论	Mysticism	意识是人类自身永远无法理解的
还原论	Reductionism	意识可还原为大脑神经细胞的物理过程
涌现论	Emergentism	意识是大脑神经元整体相互作用涌现出的
泛心论	Panpsychism	意识是物质固有的本质
副现象论	Epiphenomenalism	意识是行为产生的附带现象而不起任何作用
幻觉论	Illusionism	意识是一种幻觉
唯识论	Vijñāptimātratāsiddhi	五蕴八识理论

1.2　意识研究中最困难的问题—感受性 (Qualia)

对于意识问题, 哲学家 Chalmers 将其分为容易问题 (Easy problem) 与困难问题 (Hard problem)[28]. 所谓容易问题, 就是可以还原为物理过程, 并用物理方法进行解释的意识问题, 例如对于刺激的辨别、归类和反应, 认知系统对于信息的整合, 心理状态的可报告性, 注意的集中, 行为的控制等. 这些问题是可以通过脑科学研究解决的. 而困难问题则是指感受性 (Qualia) 问题. Qualia 指人的内在的、主观的体验 (Experience) 或感受 (Feeling), 例如感官感受性 (Sensory qualia), 情感感受性 (Emotional qualia) 等. 所谓困难问题, 就是大脑中的物理过程为何会引发主观体验? 又是如何引发的? 如图 2 所示. 尽管意识的脑科学研究已取得很多成果, 但却无法解决感受性问题. 与 Chalmers 的观点类似, 哲学家 Block 将意识分为可达意识 (Access consciousness) 与现象意识 (Phenomenal consciousness)[29], 分别对应意识的容易问题和困难问题. 哲学家 Levine 则将大脑与心灵之间的界限命名为解释鸿沟 (Explanatory gap)[30], 如图 2 所示.

表 2 意识的科学理论
Table 2 Scientific theories of consciousness

理论名称	英文	简称	提出时间	主要研究者	主要观点
高阶表征理论[31]	Higher-order representation theory	HOR	1968	Armstrong、Rosenthal	意识由对一阶心理状态的知觉或想法构成
全局工作空间理论[32]	Global workspace theory	GWT	1988	Baars、Dehaene	意识产生于全局工作空间
多重草稿理论[1]	Multiple drafts theory	MDT	1991	Dennett	意识产生于大脑中叙事脚本间的竞争
量子意识理论[7]	Quantum consciousness	QC	1994	Penrose、Hameroff	意识产生于大脑中的量子计算
Damasio意识理论[33]	Damasio's theory	无	1999	Damasio	在原我的基础上产生核心意识和扩展意识
动态核心假说[34]	Dynamic core hypothesis	DCH	2000	Edelman、Tononi	意识产生于时空上高兴奋性的神经集团, 即动态核心
感觉运动理论[35]	Sensorimotor theory	SMT	2001	O'Regan	意识产生于身体的感觉运动
整合信息理论[36]	Integrated information theory	IIT	2004	Tononi、Koch	意识产生于大脑对信息的整合
注意图式理论[37]	Attention schema theory	AST	2013	Graziano	意识产生于注意图式
预测处理理论[38]	Predictive processing theory	PPT	2013	Clark、Seth	意识产生于大脑对信息的预测

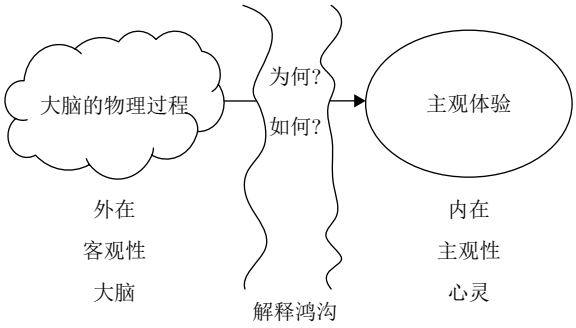


图 2 意识的困难问题与解释鸿沟
Fig. 2 The hard problem of consciousness and the explanatory gap

感受性是人类意识中最重要的部分, 我们每个人都可以通过内省直接得知自己的感受, 不需要进行计算、观察和推理, 也不需要使用任何设备. 因此, 感受性具有与物理现象不同的特性, 无法依靠物理还原的方式来解释. 感受性也是心灵哲学上争论不休的终极问题, 有人认为感受性根本不存在, 有人提出思想实验来证明感受性存在, 有人认为永远也无法解决感受性问题, 有人认为解决了所有的意识的容易问题后, 便可解决感受性问题[39]. 2018 年, Chalmers 又提出了意识的元问题 (Meta-problem)[40], 指为什么我们会认为存在意识问题. 由于困难问题难以解决, Chalmers 建议先通过建立内省模型来解决元问题, 然后再考虑困难问题.

2 机器意识的概念与研究分类

和意识的分类类似, Holland 将人工意识分为弱人工意识和强人工意识[41], 相当于机器意识的容易问题与困难问题. 诸如让机器具有感知能力, 进行推理, 用语言进行交流, 识别与表达一定的情感等, 都属于容易问题. 而让机器具有自我意识以及感受性则是困难问题. 此外, 还存在一个如何测试机器是否具有意识, 意识能力程度如何的评价问题.

根据机器意识的研究内容与目标, 可将其分为 6 类[42−43], 如表 3 所示. 其中, 前 5 类具体研究意识的某个方面, 最后一类研究如何评测机器的意识程度. 在此分类体系中, 不同类别之间并不互相独立, 研究上存在交叉融合的情况. 例如在感知意识、认知意识中都伴随有感受意识与自我意识. 机制意识研究人类意识的产生机制, 则包括其他各类意识的脑机制. 意识测试则涉及到其他所有机器意识类别. 从研究内容和目标上来看, 前三种类别是机器意识的容易问题和初级目标, 而后三种类别是机器意识的困难问题和最终目标, 在理论和实现方法上还存在很多争论. 因此, 对前三种类别的研究有助于后三种类别的最终实现.

2.1 感知意识

感知意识 (MC-Perception) 主要研究如何通过外部传感器和内部感知模型使机器具有感知外界各种刺激并产生行为的能力, 如对图像、声音、温度等的感知[44].

例如在视觉方面, Yousef 等基于人类视觉并行计算的特点, 提出了一种边缘检测方法, 有助于构建更好的意识机器[45]. 听觉方面, Eliakim 等开发了一个自主蝙蝠机器人 Robat, 具有两个耳朵和一个发射器, 可以通过回声定位来创建环境地图[46]. 触觉方面, Klimaszewski 等设计了一种双层机器人皮肤, 可以检测外力的位置、大小和方向[47]. 嗅觉方面, Saeed 等开发了一个小型自主机器人, 具备超声波和红外线传感器, 可用于在危险环境下移动并检测

可燃气体泄漏[48]. 味觉方面, Ciui 等开发了一个味道检测机器人, 可通过其手指上的可穿戴化学传感器, 检测固体或液体食物中葡萄糖、维生素、辣椒素等的含量, 进而确定食物的味道, 如甜味、酸味、辣味等[49].

感知意识的研究目的是让机器更加有效地进行感知, 距离真正意义上的机器意识还有很大差距.

2.2　认知意识

认知意识 (MC-Cognition) 主要研究如何通过构建机器内部模型, 使机器具有意识的认知特性及其行为表现, 如语言、记忆、想象、情感等意识过程中的认知加工.

例如在语言方面, Davila-Chacon 等提出了一种涉身认知方法, 可以提高机器人在噪音环境下语音识别的水平[50]. 在想象方面, Pagel 等探讨了意识、想象和梦境的关系, 并研究了会做梦的机器人的相关问题[51]. 在记忆方面, Balkenius 提出了一种意识机器人的记忆模型, 实现了机器人的情景记忆、想象等意识功能[52]. 在情感方面, Yang等开发了一个情感意识机器人, 可识别和表达情感, 作为家庭护理机器人以改善人的心理健康[53].

认知意识的研究目的是让机器能表现出意识的认知功能, 但意识与这些认知功能之间的关系还需要深入研究.

2.3　机制意识

机制意识 (MC-Mechanism) 主要研究人类意识的产生机制, 以及在此基础上的机器模拟实现, 是目前机器意识研究最多的一个方面.

在意识产生机制方面, 人们通过研究脑损伤人群、意识障碍患者以及正常人群的大脑, 希望阐明意识的神经相关物以及意识产生机制. 例如, 对于视觉意识, Förster 等通过 ERP 研究, 确证了视觉意识负波是最早出现的 ERP 成分[54]. 对于情感意识, Paul 等指出杏仁核和前额叶之间的通路是情感意识的重要神经结构[55]. 对于自我意识, Keromnes 讨论了关于镜像神经元的各种观点[56]. 对于

表 3　机器意识研究分类

Table 3　The taxonomy of machine consciousness

类别	问题归属	研究内容	具体分类	研究举例
感知意识	容易	机器通过外部传感器和内部感知模型感知外界各种刺激并产生行为	视觉	图像感知[45]
			听觉	声音感知[46]
			触觉	皮肤触摸感知[47]
			嗅觉	化学物质感知[48]
			味觉	化学物质感知[49]
认知意识	容易	构建机器内部认知模型, 使机器具有意识的认知特性及其行为表现	语言	语音识别和表达[50]
			想象	梦境与意识[51]
			记忆	情景记忆[52]
			情感	情感识别和表达[53]
机制意识	容易	研究人类意识的产生机制, 并在此基础上进行机器模拟实现	脑科学研究	意识的神经相关物[54-58]
			意识科学理论	GWT[32]、IIT[36]、HOR[31]、AST[37]、QC[7]
			机器模拟实现	心智建模[59]、银纳米线神经网络[60]、大脑模型[61]
自我意识	困难	如何使机器具有内省反思能力, 并能意识到“我”是区别于其他个体的存在	自我模拟	机械臂自我建模[15]、“粒子”机器人自我修复[16]
			镜像认知	镜像测试[62]
			高阶理论	CiceRobot[63]、内部言语[64]、GMU-BICA[65]
			其他方面	自我意识模型ARTSELF[66]、本体感受传感器[67]
感受意识	困难	如何使机器具有感受性	能否实现	能实现[68-69]、不能实现[70-71]
			实现方法	Aleksander公理系统[72]、感觉运动融合[73]、计算相关物[74]、大脑时空结构[43]、Meme[75]、内稳态[76]、合成现象学[77]
			系统开发	交互式教学机器人[78]
意识测试	困难	检测机器是否具有意识, 意识能力程度如何	图灵测试	完全图灵测试[79]
			量表	ConsScale量表[80]

感受意识, Nani 等指出感受意识与颞顶枕区的多个同步网络的活动有关[57]. 此外, Zhao 等归纳出了意识产生的 3 个关键区域, 即丘脑室旁核、屏状核与后部皮层, 并提出了一种意识的巨大神经网络假说[58]. 尽管目前意识的脑科学研究成果丰富, 但关于意识的产生机制还没有一致的结论, 也没有一个能够很好解释所有意识现象的统一理论框架.

在机器模拟实现方面, 研究者基于意识的产生机制, 提出了很多可用于机器意识实现的理论和模型, 如 GWT[32]、IIT[36]、HOR[31]、AST[37]、QC[7] 等 (将在第 3 节详细讨论), 并采用仿脑计算的方法, 来构建意识机器. 此外, Shi 提出了意识与记忆的心智模型 CAM 以及一种机器意识架构, 包含觉知、注意、动机、元认知、内省学习以及全局工作空间模块[59]. Diaz-Alvarez 等用银纳米线构建了一个类脑神经网络, 具有类似人脑功能的记忆、学习、遗忘等相关电学特征, 可作为构建意识机器人的物理基础[60]. Wang 采用仿脑的方法, 提出了一种大脑的分层参考模型, 包括感觉、动作、记忆、感知四层潜意识能力以及认知、推理、智能三层意识能力[61].

机制意识的研究目的是模拟意识的脑机制来构建意识机器, 其研究前提是对意识的产生机制具有正确全面的认识, 但目前意识的脑机制研究只取得了一些初步的成果, 因此很难提出一种通用的意识仿脑理论和计算模型. 此外, 现有的意识理论也有待于脑科学研究的进一步验证.

2.4 自我意识

自我意识 (MC-Self) 主要研究如何使机器具有内省反思能力, 并能意识到 “我”是区别于其他个体的存在.

例如, Lipson 认为自我意识就是模拟未来自我状态的一种能力, 对未来的预测能力越强, 自我意识程度就越高, 机器如果能够自我模拟, 也就具备了自我意识. 基于这种理论, 他开发了一个机器臂[15], 能够在自己 “大脑”创建的模拟环境中学习, 产生自我意象, 进而完成未知的任务. 这个机械臂还可以检测到自我的损伤, 从而重新自我建模. 他还研发了一种 “粒子”机器人[16], 此机器人不是由单片部件构成, 而是由很多细胞构成, 能够自我建模和复制, 自行恢复结构并自我治愈. 粒子可以相互协调移动, 从而完成群体迁移, 有一部分粒子损坏也不会影响整体运作. Lipson 的工作实现了意识机器人的工程学突破, 也为理解自我意识提供了新的视角.

由于镜像神经元是人类自我意识的重要生物基础, 人们常用镜像测试来检验个体是否具有自我意识[81]. 例如, Zeng 等提出了一种类脑机器人镜像神经元系统模型 Robot-MNS-Model, 可用于仿人机器人的镜像自我认知. 实验中, 3 个外观完全相同的机器人在镜子前同时作出随机运动, 结果每个机器人都能识别出自己[62].

此外, Subagdja 等提出了一种自我意识的仿脑模型 ARTSELF, 并应用于 NAO 机器人, 机器人可以根据自己的身份、社会生活经验等回答人们提出的问题[66]. Rodriguez 使用本体感受传感器, 开发了一个姿势识别系统, 使得 NAO 机器人能够觉知到自身的姿势[67]. Chella 等提出了一种内部言语 (Inner speech) 的认知架构, 机器通过内部言语和长期记忆可以产生自我意识[64]. 此外, 高阶表征理论多用于实现机器自我意识, 将在第 3.3 节详细讨论.

自我意识是机器意识研究中的困难问题, 上述研究中虽有机器通过镜像测试的实例, 但这只是机器具有自我意识的必要条件. 此外, 对于 “自我”这一概念的界定也因人而异, 这也为自我意识机器的构建带来了很大挑战[82].

2.5 感受意识

感受意识 (MC-Qualia) 研究如何使机器具有感受性, 如感官感受性、情感感受性等. 感受性是意识的本质, 而感受意识又是机器意识研究中最难的一类, 因此对于机器感受意识能否实现, 如何实现等问题, 众多学者有着不同的看法.

英国帝国理工学院电子工程系的 Aleksander 教授认为, 当人们描述他们的意识及感受时, 实际上是在描述神经活动本身所具有的某种性质, 而采用神经计算方法的机器同样可以做到这一点, 因而机器是可以具有感受的. 基于此观点, 他提出了一个意识公理系统, 包括感知 (Depiction)、想象 (Imagination)、注意 (Attention)、规划 (Planning) 和情感 (Emotion) 能力, 只要机器的内部表征可以满足这 5 条公理, 则机器就是具有感受的[72]. 美国伊利诺伊大学斯普林菲尔德分校哲学系的 Haikonen 教授认为, 感受性是类人意识机器所必需满足的先决条件, 仅仅实现认知功能不足以产生意识. 真正的意识机器人会像人一样直接感知外部环境和自己的物质机体, 不会受程序支配, 任何时候都可能出现非预先编程的反应, 动机因素将在智能体行为的形成中扮演重要的角色, 不过, 机器的感受性未必需要和人类的感受性类似. Haikonen 建议使用无缝的感知和运动传感器融合的方法来构建意识机器, 并开发了一个意识机器人系统 XCR-1[73], 以实现他提出的类模态感受性 (Amodal qualia). García-Baños 从计算角度来解释感受性, 认为主观体验仅仅是大脑中

以非线性、不可逆转的形式编码的数据, 与机器中的信息处理没有本质区别, 因此完全可以通过某种类似的计算手段来实现[68]. Reggia 等提出了计算解释鸿沟 (Computational explanatory gap) 的概念, 认为之所以无法构造出具有感受性的机器, 是因为人们还不能理解和解释大脑中高级认知功能的神经计算模型是如何实现的, 因此, 必须首先找到意识的计算相关物 (Computational correlates of consciousness)[74]. Longinotti 和 Pandey 认为, 人的意识是由大脑神经生物系统产生的, 是人类独有的, 而机器仅仅是人类制造出的, 不具有神经系统, 因而不可能具有意识[70−71]. 而 Koch 等认为, 意识是自然界的一部分, 是只取决于数学、逻辑、物理定律、化学以及生物学的存在, 因而最终是可以由人工实现的[69]. Blackmore 认为, 人类意识, 包括主观体验, 其实是一种幻觉, 这种幻觉产生于 Memes (谜米, 指文化基因) 互相竞争从而进行自我复制的过程, 通过竞争存活下来的 Memes 就进化为我们的各种意识能力, 机器要具有意识就必须具有这种对意识和自我的幻觉[75]. Gamez 认为, 机器的外部行为并不能表明机器具有感受, 应该在大脑中寻找与意识感受有关的更加清晰明确的时空结构, 而不是 NCC, 从而构建更加普遍的意识理论[43]. Man 等认为, 有机体的生存依赖于内稳态, 而感受是内稳态过程中的心理表现, 如果机器能够实现类似于内稳态的过程, 那么就具有某种类似于有机体感受的功能[76]. Tojo 等将感受定义为基于主观体验的知识内容, 开发了一个交互式教学机器人 ITR, 实现了类似于感受与觉知的功能[78]. 此外, 对感受意识的研究形成了一个专门研究如何用人工手段合成感受的学科, 即合成现象学 (Synthetic phenomenology)[77].

感受意识是机器意识中最难实现的一种, 由于感受性是主观的, 没有意向对象可以作为形式化的载体, 因此难以对其进行表征与计算. 目前对于机器感受性的研究多为理论上的探讨, 少有相关意识系统的开发.

2.6 意识测试

除了对机器意识某一具体类别进行研究与实现外, 还有一类研究是机器意识测试 (MC-Test), 旨在测试机器是否具有意识, 意识能力程度如何[83−85]. 对于感知意识和认知意识, 可以用传统人工智能的评价指标, 如准确率、召回率等. 对于机制意识, 主要可看提出的意识理论对意识产生机制的解释能力如何, 以及在指导构建机器意识时能实现的意识能力的多少、强弱、复杂程度等. 而对于自我意识和感受意识, 由于涉及他心知问题, 行为测试是目前唯一的方法. 例如 Schweizer 提出了一个用于检验机器感受性的完全图灵测试 (Total turing test for qualia, Q3T)[79], 相比传统的图灵测试, 此完全版包括无间断的定性问题, 而机器需要回答这些主观的涉及内心体验的问题, 例如太平洋的落日景象是不是很美? 给你的印象如何? 等. 由于这类问题有多种表达方式且问题的数量是无限的, 因而预先编程的机器一定是无法回答的. Arrabales 提出了一种评估智能体意识程度和层次的量表 ConsScale[80], 将意识分为 11 个等级, 从最低级的非涉身智能体到最高级的超意识智能体.

机器意识测试是未来重要的研究方向, 目前还没有公认的机器意识测评标准.

3 机器意识的实现方法与计算模型

机器意识的主要实现方法如表 4 所示. 其中, 符号计算和人工神经网络是传统的人工智能方法. 生物神经网络采用生物技术, 将生物神经元一个一个搭建起来, 形成一个真正的神经网络, 这样的神经网络无疑更可能具有意识能力. 量子计算利用量子的叠加性、纠缠性等性质来解释意识并构建量子计算模型 (将在第 3.5 节中详细讨论). 脑机融合是脑机接口 (Brain-machine interface, BMI) 的一种类型, 通过将人脑与机器在物理上进行一定程度上的融合来研究意识并构造脑机混合意识机器.

表 4 机器意识的实现方法

Table 4 Implementation methods of machine consciousness

实现方法	具体内容	实现机器意识的可能性
符号计算	数理逻辑、计算推理	不可能
人工神经网络	模拟神经元活动机制建模	不可能
生物神经网络	用生物神经元搭建神经网络	有可能
量子计算	根据量子特有性质解释意识并建模	有可能
脑机融合	构造脑机混合的意识机器	有可能

BMI 旨在建立大脑和机器之间的直接信息通路, 按照信息传输方向可分为脑到机、机到脑、脑到脑以及脑机融合[86]. BMI 具有广阔的应用前景, 是目前的热门研究方向[87], 尤其是脑机融合 (混合智能), 充分结合生物智能与机器智能, 是实现机器意识的重要方法之一. 在这方面, Wu 等详细讨论了混合智能 (Cyborg intelligence) 的概念框架和实现方法等, 提出了一种脑机混合系统的认知计算模型[88]. Shi 等对上述模型进行了改进, 将意识引入到模型当中, 并讨论了环境觉知的实现问题[123]. Schweizer 通过思想实验表明在其他物理介质上复制人

表 5 机器意识的主要理论与计算模型

Table 5 Main theories and computational models of machine consciousness

理论	基本观点	研究内容	研究举例	面临问题
GWT[32]	意识产生于全局工作空间	理论研究	GNWT[89]、GNWT+PPT[90]、GWT+元认知[91]、GWT+SMT+PPT[92]	缺少神经层面的解释[93]
		机器实现	LIDA[94-95]、CERA-CRANIUM[96]	
IIT[36]	意识产生于大脑对信息的整合	理论研究	IIT 3.0[97]、IIT+幻觉论[98]、IIT+QC[99]	计算复杂性、还原论、泛心论[100]
		机器实现	工具箱PyPhi[101]、Aleksander公理系统实现[72]、XCR-1[73]	
HOR[31]	意识由对一阶心理状态的知觉或想法构成	理论研究	SOMA[102]、情感意识高阶理论[103]	意识统一性、无穷倒退[31]
		机器实现	Cicerobot[63]、CLARION[104]、GMU-BICA[65]、eBICA[105]	
AST[37]	意识产生于注意图式	理论研究	理论验证[106]、AST+PPT[107]	理论本身和具体实现方法有待完善[108]
		机器实现	注意系统[109]、CONAIM[110]	
QC[7]	意识产生于大脑中的量子计算	量子与意识的相关性	量子相干、量子叠加、量子纠缠、量子塌缩等[111-112]	缺乏实验证实、量子计算机成本高[7]
		大脑中是否存在量子计算	存在[113-114]、不存在[115-116]	
		建模方法	量子力学的数学形式以及量子逻辑[117-118]、量子计算+经典计算[119-120]、量子计算+神经生物学[121-122]	

类心智本质上具有理论困难[124]. Romano 等详细讨论了生物混合系统中动物与动物机器人的交互问题[125]. Lorrimar 则对意识上传 (Mind uploading)进行了理论探讨[126].

不过, 目前还少有以构建意识机器为目的的脑机融合理论和技术研究, 现有研究中机器一般只被视为一种工具, 主要目的是增强人类, 而并不涉及机器本身的意识问题[127]. 由于人类意识并不能简单地还原为大脑的神经活动, 因此, 仅仅将人脑神经信号复制到机器上并不能真正实现机器意识. 未来的脑机融合研究需要进一步了解意识产生的神经机制, 结合类脑智能、人工大脑、神经芯片等技术, 促进意识机器人的开发.

在机器意识的研究中, 通常以某种意识解释理论为基础来构建计算模型[128−129], 如表 5 所示. 下文将详细介绍这几种机器意识理论, 重点讨论其中的量子理论和量子计算模型.

3.1 全局工作空间理论

全局工作空间理论 (GWT) 是美国加州大学圣地亚哥分校的 Baars 教授于 1988 年提出的一种意识理论[32], 该理论认为意识为大脑提供全局信息处理. Baars 指出, 大脑是一个专门处理器网络, 各个处理器实现不同的功能, 如感知、运动控制、语言等. 在大脑皮层中, 广泛分布着一个全局工作空间 (GW), 各专门处理器通过竞争以进入 GW, 从而广播全局性信息, 意识就是在此过程中产生的. 1998 年, Dehaene 对 GW 进行了神经建模, 提出了全局神经工作空间理论 (Global neuronal workspace theory, GNWT)[89].

GW 可理解为一个意识剧场, 如图 3 所示. 剧场内包括后台、演员、舞台、观众等, 其中后台和观众作为背景, 相当于人的无意识信息加工, 感官、思维、意象等 “演员”相互竞争以进入舞台, 选择性记忆控制聚光灯照射在舞台演员上, 并全局性地广播到各个无意识处理器进行加工, 聚光灯的焦点就形成了意识内容和体验, 聚光灯的边缘则是一些模糊的意识事件.

在理论研究方面, Whyte 将 GNWT 和 PPT 结合, 提出了预测全局神经工作空间 PGNW 理论[90]. Shea 等将 GWT 和元认知理论结合起来, 认为 GWT 中的广播过程需要元认知的参与[91]. Jeczminska 将 GWT 和 SMT 统一于 PPT 框架下, 构建了一种新的意识理论[92].

在机器实现方面, 美国孟菲斯大学计算机科学系的 Franklin 教授等基于 GWT, 采用符号计算与神经网络相结合的方法开发了 IDA (Intelligent distributed agent) 以及 LIDA (Learning IDA) 系统[94], 实现了注意、情感、想象等意识能力, 并给出了一种通用可自定义实现的 LIDA 框架, 是目前最有影响的意识系统之一. Santos 等基于 LIDA 框架, 在虚拟环境下构建了一个意识机器人, 实现了机器人的移动导航以及情感表达, 可用于机器人导游[95]. 此外, Arrabales 基于 GWT, 构建了 CERA-CRANIUM 认知系统, 用以检验机器意识, 尤其是视觉感受[96].

GWT 直观地解释了意识的统一性, 清晰说明了有意识和无意识脑活动之间的区别, 和当前关于工作记忆的观点很好吻合, 对于意识的计算建模产生了较大影响. GWT 的主要问题有: 1) GWT 并没有说明为何全局信息处理在神经层面是和意识相关

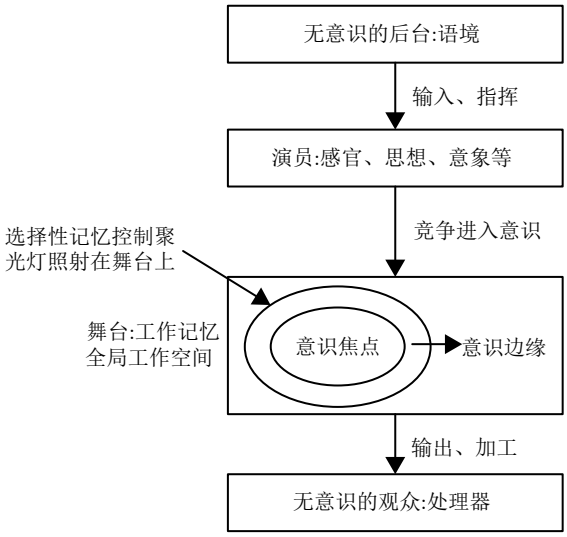


图 3　全局工作空间理论的基本思想

Fig. 3　The basic idea of global workspace theory

的; 2) GWT 对感受性没有给出很好的解释, LIDA 框架则没有考虑感受性的实现[93].

3.2　整合信息理论

整合信息理论 (IIT) 是美国威斯康星大学麦迪逊分校精神病学系的 Tononi 教授于 2004 年提出的一种意识理论, 又称为 Φ 理论[36]. 2014 年, Oizumi 等提出了 IIT 的改进版本 IIT 3.0[97]. IIT 以 5 个关于意识体验的现象学公理为基础, 得出了 5 个关于意识神经基础的推断, 其核心思想是, 意识的本质是信息, 意识产生于物理系统 (如大脑) 对大量信息的整合. 进而, 可以将意识分为 “量”和 “质”两个方面, 意识的 “量”就是物理系统产生的信息量, 意识的 “质”就是这些信息之间的关联程度和整合程度. 因而, 信息量的规模和信息的整合程度就决定了意识水平的高低. 信息量和信息的整合程度可以通过公式计算得到, 进而可以计算出整个物理系统的意识水平, 用 Φ 表示, Φ 值越大, 系统的意识就越强[130]. 如图 4 所示, 3 个物理系统各由 5 个复合体构成, 连接线表示复合体之间的连接强弱程度 (实线比虚线强), 每个复合体都可计算出 Φ 值, 进而可计算出整个物理系统的 Φ 值, 可以看到图 4中从左到右的系统意识水平逐渐降低.

在理论研究方面, McQueen 将 IIT 和意识的幻觉论结合, 提出了幻觉整合信息理论[98]. 同时, 他也研究了 IIT 和量子理论的结合问题[99].

在机器实现方面, IIT 目前已成为意识建模中最重要的方法之一. Mayner 等使用 Python 语言为 IIT 开发了一个工具箱 PyPhi[101]. 受到 IIT 的启

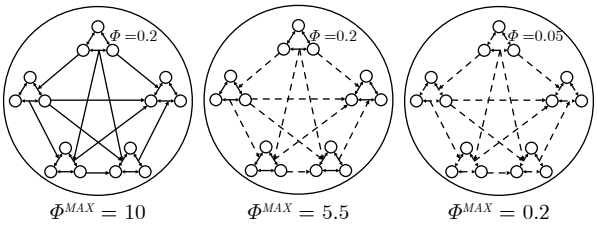


图 4　整合信息理论的基本思想

Fig. 4　The basic idea of integrated information theory

发, Aleksander 利用人工神经网络构建了若干意识系统[72], 以实现他提出的机器意识公理系统. Haikonen 则提出了一种认知架构 HCA (Haikonen cognitive architecture)[73], HCA 是一个并行、分布式的架构, 由大量感知和运动模块组成, 结合了亚符号和符号信息处理, 采用了联想信息处理和异常检测等机制, 同时也包含了诸如内部言语, 类模态感受性等模块. 基于 HCA, Haikonen 构建了一个非程序化, 不基于微处理器, 具备亚符号与符号神经系统的意识机器人 XCR-1, 可以表现出一些非预先编程的行为.

IIT 提供了一种定量测量物理系统意识程度的方法, 其主要问题有: 1) 大脑中加工的信息量过于庞大, 在当前计算机条件下根本无法完成计算; 2) Φ 值更大程度上是与系统的智能信息处理能力有关, 而不是与意识相关, 因此 Φ 值并不能成为判断意识是否存在的指标; 3) IIT 将意识完全还原为对信息的整合, 是一种还原论和功能主义, 因而无法解决感受性问题; 4) 按照 IIT 的观点, 任何具有整合信息的物理系统都是有意识的, 因此目前的机器也是有意识的, 只不过由于信息整合的水平较低, 其意识程度较低, 这种泛心论的观点也难以让人信服[100].

3.3　高阶表征理论

高阶表征理论 (HOR) 早期由哲学家 Rosenthal 等进行过深入研究[131], 其基本思想如图 5 所示. 一阶表征理论认为, 意识是大脑对基本感觉信息的

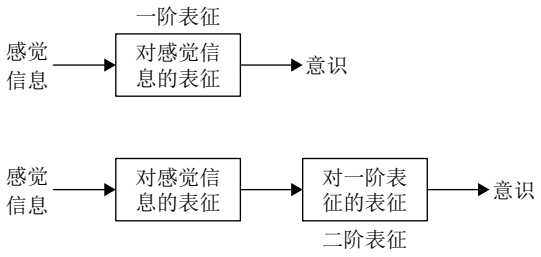


图 5　高阶表征理论的基本思想

Fig. 5　The basic idea of higher-order representation theory

表征, 而 HOR 认为, 相比无意识的心智活动, 有意识的心智活动使用对信息的更高阶表征, 意识就是由对一阶心理状态的知觉或想法构成[31,132], 即对于某种心理状态 M, 当存在对 M 的高阶表征时, M 就是有意识的. 高阶表征是一种元认知状态, 可理解为人的自我觉知, 例如当我们看到某事物时, 并不一定是有意识的, 只有当我们觉知到自己看到某事物, 才是有意识的, 如果没有这种自我觉知, 我们就是无意识的.

在理论研究方面, Cleeremans 等基于 HOR, 提出了意识的自组织元表征解释 SOMA, 认为大脑不断无意识地学习描述自身的活动, 从而发展出符合一阶表征系统的元表征系统, 意识就是在这样的学习过程中产生的[102]. LeDoux 则基于 HOR, 提出了情感意识的高阶理论[103].

在机器实现方面, 人们多采用符号计算方法来开发自我意识机器人, 这种机器人系统往往具有多层结构. 例如, 意大利巴勒莫大学机器人实验室的 Chella 等构建了一种基于高阶感知回路的机器人感知信息处理系统[133], 并开发了一个博物馆导游机器人 Cicerobot[63]. Cicerobot 使用了 3 层表征: 亚概念层用于处理最底层的感知数据, 中间概念层将感知数据整合为结构化的信息, 高阶语言层使用符号语义网络决定机器人的行动. 在这个机器人中, 一阶表征是机器人对外部世界的表征, 高阶表征是对机器人内部的表征, 因而实现了一种自我意识的认知结构机制. 近年来, Chella 等在 Cicerobot 的基础上实现了更多的意识功能, 如内部言语[64]、基于内省的知识获取[134] 等. 此外, Sun 提出了一种混合认知架构 CLARION[104], 具有两层表征结构, 将连结主义和符号表征、内隐和外显的心理过程、认知和其他心理过程混合在了一起. Samsonovich 等开发了一个自我意识仿生认知系统 GMU-BICA[65], 包含高阶的认知表征和低阶的图式表征, 以实现对外部世界、内部工作记忆、以及 “自我”的觉知. 在此基础上, Samsonovich 将情感和感受引入此系统, 开发了一个情感认知架构 eBICA[105].

HOR 是实现机器自我意识的一种重要方法, 其主要问题有: 1) HOR 是通过高阶表征来定义意识的, 因此会面临无穷倒退的问题; 2) HOR 主要依据符号计算来实现机器意识, 缺乏神经层面的解释; 3) HOR 在感受性和意识的统一性方面欠缺考虑[31].

3.4 注意图式理论

注意图式理论 (AST) 是美国普林斯顿大学心理学系的 Graziano 教授于 2013 年提出的一种意识理论[37,108]. 类似于身体图式是身体的一种模型, 注意图式是当前注意状态的一种内部模型, 决定了如何注意和注意什么, 人们通过注意图式修改自我内部关于世界的模型. 如图 6 所示, 当我们看到一个苹果时, 我们大脑对于世界的表征会发生变化, 但这种变化只是信息的变化, 即我们有了更多关于苹果的信息, 而并没有产生意识. 大脑要产生意识则还需要另外两个模型: 自我模型和注意图式, 注意图式将自我模型和对世界的表征联结起来, 因此产生了意识体验.

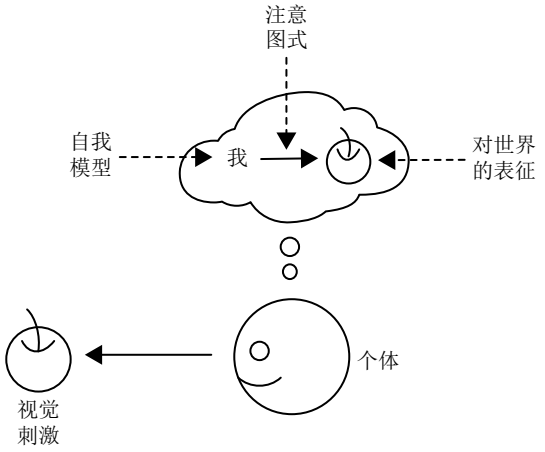


图 6 注意图式理论的基本思想

Fig. 6 The basic idea of attention schema theory

在理论研究方面, Dolega 分析了 AST 存在的问题, 提出了一种 AST 和 PPT 结合的新理论, 用于解释意识体验[107]. Boogaard 为 AST 设计了一种基于神经逻辑的时序因果网络模型, 用于验证 AST 提出的假设[106].

在机器实现方面, Graziano 认为, 如果 AST 是正确的, 那么使用当前的技术手段是可以构建意识机器的, 不过由于此理论的发展时间较短, Graziano 认为目前 AST 还处于基础概念构建层面, 尚不能在工程上给出很好的指导[108], 因此目前基于 AST 构建意识机器人的相关研究还较少. 例如, Lanillos 等提出了一种人工注意系统, 集成了注意图式, 使社交机器人具有自动注意机制[109]. Simões基于注意图式, 提出了一种意识注意整合模型 CONAIM, 整合了长短期记忆、推理、规划、学习等意识功能[110].

AST 是一种较新的意识理论, 其主要观点还有待于脑科学研究证实, 基于 AST 的机器意识实现方法也需要进一步深入研究.

3.5 量子意识理论

量子理论是描述微观粒子运动规律的物理理

论. 由于分子和原子是构成大脑的物质基础, 因此意识可能和量子理论具有某种关联. 这种假设存在着很大的争议, 主要涉及以下 3 个问题: 量子和意识有哪些相关性? 大脑中是否存在量子计算? 如何构建意识的量子计算模型?

3.5.1　量子与意识的相关性

量子力学所描述的微观粒子具有很多和经典力学不同的违反直观的奇妙性质, 这些性质和神秘的意识现象具有一定的相似性, 因而量子力学可能用来解释意识现象. 表 6 列出了量子力学中粒子的主要性质及其和意识的可能关联[111−112].

表 6　量子和意识的可能关联

Table 6　The possible relationship between quantum and consciousness

量子性质	含义	与意识的可能关联
量子相干	粒子之间存在干涉效应	意识统一性的物理基础
量子叠加	粒子可以同时处于多种状态	前意识与潜意识加工、梦与意识改变状态
量子纠缠	粒子间可以存在非局域性关联	联想记忆、非局域性意识关联
量子塌缩	粒子可从叠加态塌缩到本征态	从前意识到意识的转变

3.5.2　大脑中的量子计算

表 7 列出了关于大脑中是否存在量子计算这一问题的主要观点、理由和关于机器意识实现的看法.

这其中, 影响最大的是 Hameroff 等提出的协同客观崩现 (Orchestrated objective reduction, Orch-OR) 理论[113,135]. Orch-OR 理论认为, 大脑既是神经计算机又是量子计算机, 意识产生于微管 (是细胞骨架的主要部分) 中的量子计算, 微管中的 π 电子可形成玻色-爱因斯坦凝聚 (Bose-Einstein condensate, BEC), 当 BEC 塌缩时, 意识就产生了. Orch-OR 理论对于构建意识机器的意义在于, 如果大脑确实是一台量子计算机, 而意识确实是大脑中量子计算的产物的话, 那么就有可能在量子计算机上实现机器意识. 近些年, 已有实验支持 Orch-OR 理论[114].

Orch-OR 理论面临的主要问题是, 量子计算需要一个相对隔绝以及低温的环境, 以避免与环境交互而引起量子退相干, 而在温暖湿润的大脑中如何能产生量子效应? 对此, 很多人都提出了质疑, 例如 Tegmark 通过计算退相干率[115], 证明大脑中的粒子在进行量子计算之前就已经因为和环境的交互退化到经典状态了, 因而大脑中的计算应该是经典计算而不是量子计算. Baars 等认为[116], 神经现象原则上可还原为量子事件, 但是量子力学并不是解释意识现象的合理层次, 因此应该着重寻找 NCC.

表 7　关于大脑中是否存在量子计算的观点

Table 7　Views on the existence of quantum computing in the brain

观点	代表人物	主要理由	研究建议	机器意识实现
存在[113-114]	Hameroff、Penrose	经典计算不足以描述意识的复杂性	实验验证	量子计算可能实现机器意识
不存在[115-116]	Tegmark、Baars	大脑中不具备量子计算的客观条件	寻找 NCC	采用经典神经网络模拟即可实现

3.5.3　意识的量子计算模型

量子计算具有比经典计算更强的计算能力, 且具有真随机性, 能突破预先编程的限制, 因而未来的意识机器人很可能基于量子计算而非经典计算. 表 8 列出了目前构建意识量子计算模型的主要方法.

表 8　意识量子计算模型的主要构建方法

Table 8　Main methods on constructing quantum computational model of consciousness

方法	研究举例
量子力学的数学形式、量子逻辑	CQN[136]、量子Braitenberg小车[117−118]、量子情感[137]
量子计算+经典计算	QML[138]、QNN[120,139]
量子计算+脑科学和神经生物学	BEC[121,140]、仿生认知架构[122]

1) 第一种方法是将意识的某些问题转化为数学问题, 利用量子力学的数学形式, 通过量子方程、量子逻辑、量子算法等来构建意识机器人. 例如, Majumder 等提出了认知量子数 (Cognitive quantum number, CQN) 的概念, 并将其作为机器的逻辑, 通过量子逻辑公式推导, 建立了机器内部通用的量子逻辑体系[136]. Mahanti 等对量子 Braitenberg 小车[141] 进行了改进, 利用量子算法控制小车飞翔, 提出了一种新的量子线路, 并在 IBM 量子计算平台 (IBM quantum experience) 中对其进行了模拟实现, 使其能够更好地表现出恐惧行为, 从而在运动中成功避障[117]. Toffano 等将量子本征逻辑应用于量子 Braitenberg 小车的行为分析, 通过多值模糊量子逻辑门的控制以及不同逻辑门的变换和组合, 扩展了量子机器人的情感行为[118]. Yan 等利用量子线路对机器人的情感空间进行建模, 提出了一种量子情感空间, 对情感空间中的情感转换进行了数学推导, 使机器人能更好地表达情感[137].

2) 第二种方法是通过量子计算与经典计算的结合来进行意识建模, 例如量子机器学习 (Quantum machine learning, QML)[138]. 以 Kak 提出的量子神经网络 (Quantum neural network, QNN)[119,142] 为例, 图 7 比较了经典神经元和量子神经元的结构.

在量子神经元中, 输入和权值都用量子比特表示, 而求和运算和传递函数则用量子逻辑门来实现, 这样的量子神经元具有叠加性和纠缠性, 更适合作为描述意识现象的神经基础. Peruš等认为, 规模足够大并且足够复杂的神经网络并不足以产生意识, 因为神经网络的神经元、突触的信息处理过程过于机械、离散和确定, 而仅在量子层次考虑意识又会使得模型很复杂, 因此在构造意识机器时, 需要将神经层次和量子层次结合起来[143]. 不过, 目前大多数此类研究的目的并不是对意识进行建模, 而是在传统神经网络中引入量子机制, 从而提高机器学习的性能. 例如 Zhang 提出了一种量子神经模糊联想记忆模型, 加深了神经活动、量子理论以及认知意识三者之间的关联, 可进一步用于心智的计算建模[120]. Abdulridha 等利用量子神经网络设计了一个机器人运动控制系统并提出了一种解决机器人逆运动学的方法[139].

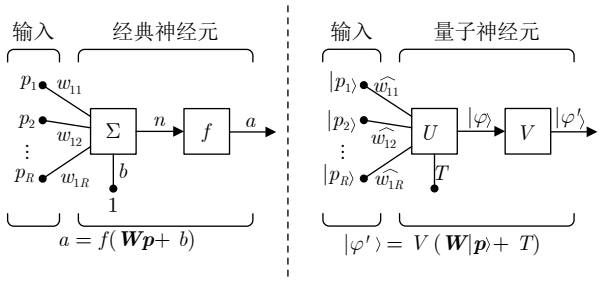


图 7 经典神经元和量子神经元

Fig. 7 The classical neuron and quantum neuron

3) 第三种方法是借鉴意识的脑科学和神经生物学研究成果, 采用某些具有量子效应的生物组织, 或使用仿生计算的方式, 来构建量子意识机器. 这方面最早由 Amoroso 开展了相关的研究[140], 他设计的意识计算机由一个输入输出装置和一个动态计算核心组成, 通过产生 BEC 来模拟人脑中的量子效应. Adamski 分析了 BEC 在人类意识中的作用, 并指出可以通过电子信息系统和人类生物系统相融合的方式来构建意识机器人[121]. Qazi 等基于量子与仿生认知架构, 开发了一个自我意识智能体, 实现了感知、学习、动机、想象、情感等功能, 具有一定程度的意识能力[122].

尽管目前已出现了一些意识的量子理论与计算模型, 但由于人们对意识与量子力学的关系还远未达成共识, 意识的量子理论还有待于神经科学的证实. 虽然如此, 在对意识的解释以及机器意识的实现方面, 量子理论和量子计算模型仍是很有潜力的一种方法.

3.6 意识机器人实例对比分析

我们对上文中提到的 Santos 等开发的机器人、XCR-1 以及 CiceRobot 进行对比分析, 如表 9所示. 这三种意识机器人主要都用于移动机器人导航, 但是是基于不同意识理论开发的, 因此适合进行比较.

通过对比可以发现, 这三种机器人都实现了一定的意识功能, 能够用于简单环境下的移动导航. Santos 开发的机器人在虚拟环境下实现了 LIDA 框架, 但 LIDA 只考虑了机器的功能意识, 并没有考虑自我意识和感受意识. XCR-1 基于 IIT, 实现了众多认知意识和自我意识功能以及 Haikonen 提出的类模态感受性, 但是类模态感受并不等于真正的感受. CiceRobot 基于 HOR 实现自我意识, 但这种基于符号计算的自我意识和人类自我意识还相差很远. 由于目前对机器意识方法的评价没有一个公认的标准, 很难比较谁优谁劣, 每种方法都有其适用的场合. 但从实现机器意识的最终目标来看, 应将现有多种机器意识理论有机结合, 以更好地构建自我意识和感受意识机器人.

4 思考与展望

4.1 面临的困境

根据以上对机器意识理论、实现方法和系统开发进展的讨论, 可归纳出机器意识目前面临的困境, 主要有以下 3 点:

1) 缺乏意识产生机制的统一科学理论. 机器意识的实现需要有意识科学理论的支撑, 然而目前人们关于意识还没有一个清晰的定义, 对于意识科学理论 (见表 2) 也远未达成共识, 尤其是对于意识中最难的感受性问题, 现有理论都不能很好地解释. 在开发意识机器人时, 不同研究者往往根据自己的研究需求给意识下一个定义, 进而选择一种最合适的意识理论, 最后进行系统开发, 因此开发的机器人很难从理论上给出让人信服的意识解释.

2) 机器意识的困难问题难以解决. 从研究内容上来看, 机器意识研究中最难的无疑是机器感受意识, 由于感受是主观的, 涉及到无意向性心理活动, 没有意向对象可以作为形式化的载体, 因此难以对其进行表征与计算. 对于自我意识, 由于其包含自我体验, 也涉及到感受性问题, 因此也难以完全实现. 此外, 对于机器意识测试, 由于存在他心知问题, 外部行为检测是目前唯一的手段, 例如图灵测试, 但如果我们在图灵测试中重复问机器同一问题, 很快就会发现机器和人的区别, 由于机器形式系统的局限性, 机器缺乏不可预见性的反应能力. 而对

表 9　意识机器人实例对比分析

Table 9　Comparative analysis of conscious robot examples

实例	Santos的机器人[95]	XCR-1[73]	CiceRobot[63]
实验条件	仿真平台V-REP、Pioneer 3-AT虚拟机器人	自制三轮机器人	RWI B21机器人
感知意识	听觉 (声呐)	视觉、听觉、触觉 (压力)	视觉
认知意识	情感表达 (面部表情)	语音识别和表达、情感理解, 选择性注意, 情感记忆	选择性注意、长期记忆
机制意识	GWT	IIT	HOR
自我意识	无	自我对话、内部言语、内省反思	自我动作想象, 内省反思, 自我预期
感受意识	不考虑	类模态 (amodal) 感受	无
实现方法	机器人操作系统ROS	硬接线神经回路、联想神经网络、信息整合、感觉运动整合	概念层: 概念空间中的几何计算. 语言层: KL-ONE 系统实现的语义网络
认知架构	LIDA	HCA	基于 HOR 提出
实现目标	室内虚拟环境移动导航与避障	目标搜寻与检测, 验证 HCA	博物馆导游机器人
主要问题	和意识脑机制的关联不明确	机器人缺少动作的长期记忆	数据量庞大, 动态场景的实时三维重建只能在简单环境下实现
未来改进	真实环境下移动导航与避障	具有更多神经元和突触的神经网络	机器人能够对所有过去经验进行总结

于自我意识, 通过镜像测试只是具有自我意识的必要条件, 而不是充分条件.

3) 缺乏理想的机器意识实现方法. 目前机器意识的计算实现主要依靠传统人工智能方法, 如符号计算和人工神经网络, 但是这种基于预先编程的方法本质上是无法实现机器意识的. 量子计算具有比传统计算更强的计算能力与描述能力, 且能突破预先编程的限制, 在一定程度上能实现机器意识, 但仍然无法解决所有的意识问题, 如意向性问题. 此外, 量子计算还面临成本高的问题. 而采用生物技术或采用脑机融合技术构建脑机混合机器, 则存在实现的意识是否还属于机器意识的问题, 而且还涉及到伦理问题.

4.2　未来的发展

鉴于机器意识面临的困境, 未来可从意识理论、计算方法、认知架构、实验系统、检测平台这 5 个方面进行深入研究. 为此, 我们给出一种机器意识总体实现框架, 如图 8 所示.

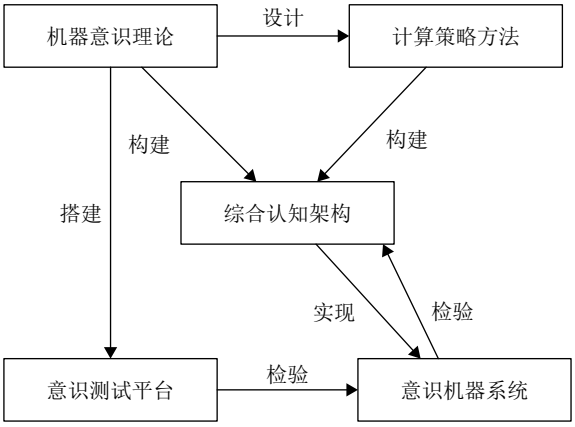


图 8　机器意识总体实现框架

Fig. 8　The overall implementation framework of machine consciousness

1) 构建更加全面合理的机器意识理论. 现有的意识科学理论 (见表 2) 更多侧重于对意识产生机制的解释, 而不是直接面向机器意识实现的, 因此不能很好地指导机器意识的实现. 为此, 需要直接面向机器意识实现本身, 提出一种全新的意识解释理论, 此理论应能清晰刻画各种意识特性及其关系, 给出机器意识限度与范围, 符合机器意识实现的要求, 更好地用于指导意识机器人的研发, 并给出机器意识实现的理论标准和规范. 例如, 可开展现有多意识理论结合的研究[17], 如 GWT 和 IIT 相结合, IIT 和量子信息论相结合, 以及意识理论与现有人工通用智能理论相结合等.

2) 探索有效的机器意识实现方法. 首先, 可以采用仿脑计算方法或者传统方法之间相互融合的方法, 例如符号计算与神经网络相融合, 深度学习与神经科学相融合[144] 等. 尤其是深度学习的最新技术, 如深度卷积神经网络、图神经网络、胶囊网络、生成式对抗网络、零样本学习等[145−147], 可更好地用于实现机器感知意识和认知意识. 而为实现机器自我意识和感受意识, 可将深度学习与当前意识的神经科学相结合, 开发功能更强的神经网络, 并采用仿脑计算的方法来实现机器意识. 其次, 可以采用量子计算方法, 但由于量子计算机的普及难度较大, 量子计算的成本代价较高, 因此还需要研究在非量子体系下模拟量子计算的类量子方法. 此外, 将目前的深度学习技术与量子计算相结合, 也是未来机器意识实现的重要方法. 最后, 可以采用生物计算、脑机融合和类脑智能的方法, 构建生物机器人和类

脑机器人. 例如 Diaz-Alvarez 等用银纳米线构建的大脑神经网络[60], Li 等用细胞开发的 “粒子”机器人[16], 以及 Kriegman 等用青蛙细胞创造出的活体机器人[148] 等, 都为实现机器意识提供了新的方法与思路.

3) 构建机器意识综合认知架构. 目前用于人工智能的认知架构已有数百种, 但只有少数涉及到意识, 例如 LIDA[94]、HCA[73]、CLARION[104]. 应该在机器意识理论的基础上, 构建机器意识综合认知架构, 包括感知意识、认知意识、机制意识、自我意识以及感受意识, 并给出他们之间的层次结构与交互关系, 以实现环境感知、语言交流、内省反思、自我觉知、情感感受等能力. 同时, 结合机器意识具体的计算策略与方法, 参照已有机器意识和人工智能认知体系的优点, 给出机器意识综合架构的总体实现策略.

4) 开发意识机器人实验系统. 在现有智能机器人开发平台上实现构建好的机器意识综合认知架构, 形成具体的意识机器人系统, 并开展具体的系统实验分析研究, 例如真实或虚拟环境下的人机交互, 陌生环境下机器人的情感表现, 多机器人协作等, 以检验机器意识综合认知架构是否满足提出的机器意识理论要求, 最终给出一种机器意识系统的实验范例.

5) 搭建机器意识测试平台. 为检测开发的意识机器人是否有意识, 意识能力程度如何, 需要建立机器意识评测标准体系, 例如对不同的意识类别设定不同的评价指标. 现有的人工智能系统评价指标可用于感知意识和认知意识中, 但对于自我意识和感受意识, 则没有一个公认的评价标准, 目前只能通过机器的外在行为表现来进行分析. 在此基础上, 构建开展评测的环境平台 (如镜像实验系统、问卷系统、图灵测试系统、机器人行为分析系统等), 以实际评测意识机器人的意识能力水平.

总之, 通过上述 5 个方面的研究, 希望能够在机器意识理论构建、方法实现、系统开发、测试平台搭建等方面有所突破, 从而促进机器意识研究的进一步发展.

5 结语

本文首先介绍了意识的概念、理论和感受性问题. 然后将机器意识分为感知意识、认知意识、机制意识、自我意识、感受意识和意识测试 6 种类型, 详细讨论了每种类型的最新研究进展. 之后, 对指导机器意识计算实现的 5 种意识理论, 即 GWT、IIT、HOR、AST 和 QC 进行了深入挖掘, 分析了其理论研究进展, 建模实现方法以及意识系统开发情况. 最后, 总结了目前机器意识面临的困境与未来的可能发展, 给出了一套意识机器人系统的总体实现框架.

机器意识的研究还处于很初级的阶段, 缺乏统一的理论、方法、评价标准等. 目前机器意识在国内也少有相关研究. 为切实推进机器意识的研究, 可按照机器意识总体实现框架 (见图 8) 进行深入研究. 在研究内容方面, 可暂时搁置机器意识的困难问题, 重点研究感知、认知与机制意识, 如视觉、语言、想象、记忆、情感等意识过程中的神经机制和计算建模. 在研究方法方面, 可采用自然机制和算法相结合, 深度学习与量子计算相结合以及脑机融合等策略. 在研究目标方面, 开发具有一定意向能力的机器人, 并应用到工业、教育、娱乐等社会服务领域.

References

1 Dennett D C. *Consciousness Explained*. New York, USA: Little, Brown and Company, 1991, 21–21

2 Miller G. What is the biological basis of consciousness. *Science*, 2005, **309**(5731): 79–79

3 Koch C. What is consciousness? *Nature*, 2018, **557**(7704): S8–S12

4 Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 2016, **17**(5): 307–321

5 Sohn E. Decoding consciousness. *Nature*, 2019, **571**(7766): S2–S5

6 Mashour G A. The controversial correlates of consciousness. *Science*, 2018, **360**(6388): 493–494

7 Li T W, Tang H H, Zhu J H, Zhang J H. The finer scale of consciousness: quantum theory. *Annals of Translational Medicine*, 2019, **7**(20): 585

8 Gornitz T. Quantum theory and the nature of consciousness. *Foundations of Science*, 2018, **23**(3): 475–510

9 Georgiev D D. Inner privacy of conscious experiences and quantum information. *Bio Systems*, 2020, **187**: 104051

10 Angel L. *How to Build a Conscious Machine*. Boulder, USA: Westview Press, 1989

11 Zhou Chang-Le. Prospect of machine consciousness: the future artificial intelligence philosophy. *Frontiers*, 2016, **5**(13): 81–95 (周昌乐. 机器意识能走多远: 未来的人工智能哲学. 人民论坛·学术前沿, 2016, **5**(13): 81–95)

12 Gamez D. *Human and Machine Consciousness*. Cambridge, United Kingdom: Open Book Publishers, 2018

13 Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? *Science*, 2017, **358**(6362): 486–492

14 Carter O, Hohwy J, Van Boxtel J, Lamme V, Block N, Koch C, et al. Conscious machines: defining questions. *Science*, 2018, **359**(6374): 400–400

15 Kwiatkowski R, Lipson H. Task-agnostic self-modeling machines. *Science Robotics*, 2019, **4**(26): eaau9354

16 Li S G, Batra R, Brown D, Chang H D, Ranganathan N, Hoberman C, et al. Particle robotics based on statistical mechanics of loosely coupled components. *Nature*, 2019, **567**(7748): 361–366

17 Graziano M S A, Guterstam A, Bio B J, Wilterson A I. Toward a standard model of consciousness: reconciling the attention

schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, 2019

18 Lake B M, Ullman T D, Tenenbaum J B, Gershman S J. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 2017, **40**: e253

19 Koumpis A, Christoforaki M, Handschuh S. The robot who loved me: building consciousness models for use in human robot interaction following a collaborative systems approach. In: Proceedings of the 19th IFIP WG 5.5 Working Conference on Virtual Enterprises. Cham, Switzerland: Springer Publishing, 2018. 409−416

20 Chella A, Cangelosi A, Metta G, Bringsjord S. Editorial: consciousness in humanoid robots. *Frontiers in Robotics and AI*, 2019, **6**: 17

21 Qiao Shao-Jie, Han Nan, Ding Zhi-Ming, Jin Che-Qing, Sun Wei-Wei, Shu Hong-Ping. A multiple-motion-pattern trajectory prediction model for uncertain moving objects. *Acta Automatica Sinica*, 2018, **44**(4): 608−618
(乔少杰, 韩楠, 丁治明, 金澈清, 孙未未, 舒红平. 多模式移动对象不确定性轨迹预测模型. 自动化学报, 2018, **44**(4): 608−618)

22 Cao Feng-Kui, Zhuang Yan, Yan Fei, Yang Qi-Feng, Wang Wei. Long-term autonomous environment adaptation of mobile robots: state-of-the-art methods and prospects. *Acta Automatica Sinica*, 2020, **46**(2): 205−221
(曹风魁, 庄严, 闫飞, 杨奇峰, 王伟. 移动机器人长期自主环境适应研究进展和展望. 自动化学报, 2020, **46**(2): 205−221)

23 Qiao S J, Han N, Gao Y J, Li R H, Huang J B, Guo J, et al. A fast parallel community discovery model on complex networks through approximate optimization. *IEEE Transactions on Knowledge and Data Engineering*, 2018, **30**(9): 1638−1651

24 Qiao S J, Han N, Wang J F, Li R H, Gutierrez L A, Wu X D. Predicting long-term trajectories of connected vehicles via the prefix-projection technique. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(7): 2305−2315

25 Qiao S J, Han N, Zhou J L, Li R H, Jin C Q, Gutierrez L A. SocialMix: a familiarity-based and preference-aware location suggestion approach. *Engineering Applications of Artificial Intelligence*, 2018, **68**: 192−204

26 Kügler P. The ever-shifting problem of consciousness. *Theory & Psychology*, 2013, **23**(1): 46−59

27 De Sousa A. Towards an integrative theory of consciousness: Part 2 (an anthology of various other models). *Mens Sana Monographs*, 2013, **11**(1): 151−209

28 Chalmers D J. Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 1995, **2**(3): 200−219

29 Block N. On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 1995, **18**(2): 227−247

30 Levine J. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 1983, **64**(4): 354−361

31 Brown R, Lau H, Ledoux J E. Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, 2019, **23**(9): 754−768

32 Baars B. *A Cognitive Theory of Consciousness*. Cambridge, United Kingdom: Cambridge University Press, 1988

33 Damasio A R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. San Diego, USA: Harcourt, 1999

34 Edelman G, Tononi G. *A Universe of Consciousness: How Matter Becomes Imagination*. New York, USA: Basic books, 2000

35 O'regan J K. How the sensorimotor approach to consciousness bridges both comparative and absolute explanatory gaps and some refinements of the theory. *Journal of Consciousness Studies*, 2016, **23**(5—6): 39−65

36 Tononi G. An information integration theory of consciousness. *BMC Neuroscience*, 2004, **5**(1): 42

37 Graziano M. *Consciousness and the Social Brain*. Oxford, United Kingdom: Oxford University Press, 2013

38 Clark A. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 2013, **36**(3): 181−204

39 Dennett D C. Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2018, **373**(1755): 20170342

40 Chalmers D J. The meta-problem of consciousness. *Journal of Consciousness Studies*, 2018, **25**(9-10): 6−61

41 Holland O, Goodman R. Robots with internal models: a route to machine consciousness? *Journal of Consciousness Studies*, 2003, **10**(4-5): 77−109

42 Zhou Chang-Le, Liu Jiang-Wei. Can machines have consciousness? *Journal of Xiamen University (Arts & Social Sciences)*, 2011, **86**(1): 1−8
(周昌乐, 刘江伟. 机器能否拥有意识-机器意识研究及其意向性分析. 厦门大学学报 (哲学社会科学版), 2011, **86**(1): 1−8)

43 Gamez D. Four preconditions for solving MC4 machine consciousness. In: Proceedings of the 2019 AAAI Spring Symposium: Towards Conscious AI Systems. Palo Alto, USA: Stanford University Press, 2019. 1−6

44 Jung Y H, Park B, Kim J U, Kim T I. Bioinspired electronics for artificial sensory systems. *Advanced Materials*, 2019, **31**(34): 1803637

45 Yousef A, Bakr M, Shirani S, Milliken B. An edge detection approach for conscious machines. In: Proceedings of the 9th Annual Information Technology, Electronics and Mobile Communication Conference. Vancouver, Canada: IEEE Press, 2018. 595−596

46 Eliakim I, Cohen Z, Kosa G, Yovel Y. A fully autonomous terrestrial bat-like acoustic robot. *PLOS Computational Biology*, 2018, **14**(9): e1006406

47 Klimaszewski J, Janczak D, Piorun P. Tactile robotic skin with pressure direction detection. *Sensors*, 2019, **19**(21): 4697

48 Saeed M S, Alim N. Design and implementation of a dual mode autonomous gas leakage detecting robot. In: Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques. Dhaka, Bangladesh: IEEE Press, 2019. 79−84

49 Ciui B, Martin A, Mishra R K, Nakagawa T, Dawkins T J, Lyu M, et al. Chemical sensing at the robot fingertips: toward automated taste discrimination in food samples. *Acs Sensors*, 2018, **3**(11): 2375−2384

50 Davila-Chacon J, Liu J D, Wermter S. Enhanced robot speech recognition using biomimetic binaural sound source localization. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, **30**(1): 138−150

51 Pagel J F, Kirshtein P. *Machine Dreaming and Consciousness*. Cambridge, USA: Academic Press, 2017

52 Balkenius C, Tjøstheim T A, Johansson B, Gärdenfors P. From focused thought to reveries: a memory system for a conscious robot. *Frontiers in Robotics and AI*, 2018, **5**: 29

53 Yang J, Wang R, Guan X, Hassan M M, Almogren A, Alsanad A. AI-enabled emotion-aware robot: the fusion of smart clothing, edge clouds and robotics. *Future Generation Computer Systems*, 2020, **102**: 701−709

54 Förster J, Koivisto M, Revonsuo A. ERP and MEG correlates of visual consciousness: the second decade. *Consciousness and Cognition*, 2020, **80**: 102917

55 Paul E S, Sher S, Tamietto M, Winkielman P, Mendl M T. Towards a comparative science of emotion: affect and consciousness in humans and animals. *Neuroscience & Biobehavioral Reviews*, 2020, **108**: 749−770

56 Keromnes G, Chokron S, Celume M-P, Berthoz A, Botbol M, Canitano R, et al. Exploring self-consciousness from self- and other-image recognition in the mirror: concepts and evaluation.

Frontiers in Psychology, 2019, **10**: 719

57 Nani A, Manuello J, Mancuso L, Liloia D, Costa T, Cauda F. The neural correlates of consciousness and attention: two sister processes of the brain. *Frontiers in Neuroscience*, 2019, **13**: 1169

58 Zhao T, Zhu Y Q, Tang H L, Xie R, Zhu J H, Zhang J H. Consciousness: new concepts and neural networks. *Frontiers in Cellular Neuroscience*, 2019, **13**: 302

59 Shi Z Z, Ma G, Li J Q. Machine consciousness of mind model CAM. In: Proceedings of the 12th International Conference on Knowledge Management in Organizations. Cham, Switzerland: Springer Publishing, 2017. 16–26

60 Diaz-Alvarez A, Higuchi R, Sanz-Leon P, Marcus I, Shingaya Y, Stieg A Z, et al. Emergent dynamics of neuromorphic nanowire networks. *Scientific Reports*, 2019, **9**: 14920

61 Wang Y X, Lu J H, Gavrilova M, Fiorini R A, Kacprzyk J. Brain-inspired systems (BIS): cognitive foundations and applications. In: Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics. Miyazaki, Japan: IEEE Press, 2018. 995–1000

62 Zeng Y, Zhao Y X, Bai J, Xu B. Toward robot self-consciousness (Ⅱ): brain-inspired robot bodily self model for self-recognition. *Cognitive Computation*, 2018, **10**(2): 307–320

63 Chella A, Macaluso I. The perception loop in CiceRobot, a museum guide robot. *Neurocomputing*, 2009, **72**(4—6): 760–766

64 Chella A, Pipitone A. A cognitive architecture for inner speech. *Cognitive Systems Research*, 2020, **59**: 287–292

65 Samsonovich A V, De Jong K A, Kitsantas A. The mental state formalism of GMU-BICA. *International Journal of Machine Consciousness*, 2009, **1**(1): 111–130

66 Subagdja B, Tan A H, Aaai. Towards a brain inspired model of self-awareness for sociable agents. In: Proceedings of the 31th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2017. 4452–4458

67 Rodriguez I, Astigarraga A, Ruiz T, Lazkano E. Body self-awareness for social robots. In: Proceedings of the 2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization. Prague, Czechoslovakia: IEEE Press, 2017. 69–73

68 Garcia-Banos A. A computational theory of consciousness: qualia and the hard problem. *Kybernetes*, 2019, **48**(5): 1078–1094

69 Koch C, Tononi G. Can we quantify machine consciousness? Brainlike circuitry might one day endow some computers with awareness-here's how we'd know. *IEEE Spectrum*, 2017, **54**(6): 64–69

70 Longinotti D. Agency, qualia and life: Connecting mind and body biologically. In: Proceedings of the 3rd Conference on Philosophy and Theory of Artificial Intelligence. Cham, Switzerland: Springer Publishing, 2018. 43–56

71 Pandey S C. Can artificially intelligent agents really be conscious? *Sādhanā*, 2018, **43**(7): 110

72 Aleksander I. Partners of humans: A realistic assessment of the role of robots in the foreseeable future. *Journal of Information Technology*, 2017, **32**(1): 1–9

73 Haikonen P O. *Consciousness and Robot Sentience. 2nd Edition*. Singapore: World Scientific, 2019

74 Reggia J A, Monner D, Sylvester J. The computational explanatory gap. *Journal of Consciousness Studies*, 2014, **21**(9-10): 153–178

75 Blackmore S. Decoding the puzzle of human consciousness: the hardest problem. *Scientific American*, 2018, **319**(3): 49–53

76 Man K, Damasio A. Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, 2019, **1**(10): 446–452

77 Chrisley R. Synthetic phenomenology. *International Journal of Machine Consciousness*, 2009, **1**(01): 53–70

78 Tojo T, Ono O, Noh N B M, Yusof R. Interactive tutor robot for collaborative e-learning system. *Electrical Engineering in Japan*, 2018, **203**(3): 22–29

79 Schweizer P. Could there be a turing test for qualia? In: Proceedings of the 2012 AISB/IACAP World Congress. Brighton, United Kingdom: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 2012. 41–48

80 Arrabales R, Ledezma A, Sanchis A. ConsScale: a pragmatic scale for measuring the level of consciousness in artificial agents. *Journal of Consciousness Studies*, 2010, **17**(3—4): 131–164

81 Takeno J. *Creation of a Conscious Robot: Mirror Image Cognition and Self-awareness*. Singapore: Pan Stanford Publishing, 2012

82 Diaz J L. Self-consciousness: an I-World patterned process model. *Adaptive Behavior*, 2018, **26**(5): 211–223

83 Elamrani A, Yampolskiy R V. Reviewing tests for machine consciousness. *Journal of Consciousness Studies*, 2019, **26**(5-6): 35–64

84 Pennartz C M A, Farisco M, Evers K. Indicators and criteria of consciousness in animals and intelligent machines: an inside-out approach. *Frontiers in Systems Neuroscience*, 2019, **13**: 25

85 Nazri A, Abd Ghani A A, Hafez I, Ng K-Y. A new theoretical framework for testing consciousness in a machine. In: Proceedings of the 3rd International Conference on Soft Computing and Data Mining. Cham, Switzerland: Springer Publishing, 2018. 330–339

86 Wu Zhao-Hui, Yu Yi-Peng, Pan Gang, Wang Yue-Ming. Brain-machine integrated systems. *Chinese Bulletin of Life Sciences*, 2014, **26**(6): 645–649
(吴朝晖, 俞一鹏, 潘纲, 王跃明. 脑机融合系统综述. 生命科学, 2014, **26**(6): 645–649)

87 Shanechi M M. Brain-machine interfaces from motor to mood. *Nature Neuroscience*, 2019, **22**(10): 1554–1564

88 Wu Z H, Zhou Y D, Shi Z Z, Zhang C S, Li G L, Zheng X X, et al. Cyborg intelligence: recent progress and future directions. *IEEE Intelligent Systems*, 2016, **31**(6): 44–50

89 Dehaene S, Kerszberg M, Changeux J P. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**(24): 14529–14534

90 Whyte C J. Integrating the global neuronal workspace into the framework of predictive processing: towards a working hypothesis. *Consciousness and Cognition*, 2019, **73**: 102763

91 Shea N, Frith C D. The global workspace needs metacognition. *Trends in Cognitive Sciences*, 2019, **23**(7): 560–571

92 Jeczminska K. Global workspace theory and sensorimotor theory unified by predictive processing. *Journal of Consciousness Studies*, 2017, **24**(7-8): 79–105

93 Van Der Velde F. In situ representations and access consciousness in neural blackboard or workspace architectures. *Frontiers in Robotics and AI*, 2018, **5**: 32

94 Franklin S, Madl T, D'mello S, Snaider J. LIDA: a systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development*, 2014, **6**(1): 19–41

95 Santos D H, Palar P S, Oliveira A S D, Fabro J A, Becker T. Adding conscious aspects and simulated emotions through facial expressions in virtual robot navigation with Baars-Franklins cognitive architecture. In: Proceedings of the 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics and 2018 Workshop on Robotics in Education. Joao Pessoa, Brazil: IEEE Press, 2018. 370–375

96 Arrabales R, Muñoz J, Ledezma A, Gutierrez G, Sanchis A. A machine consciousness approach to the design of human-like bots, Hingston P, editor, *Believable Bots: Can Computers Play Like People*? Berlin, Heidelberg: Springer Publishing, 2013: 171–191

97 Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *Plos Computational Biology*, 2014, **10**(5): e1003588

98 McQueen K J. Illusionist integrated information theory. *Journal of Consciousness Studies*, 2019, **26**(5—6): 141–169

99 McQueen K J. Interpretation-neutral integrated information theory. *Journal of Consciousness Studies*, 2019, **26**(1—2): 76–106

100 Doerig A, Schurger A, Hess K, Herzog M H. The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition*, 2019, **72**: 49–59

101 Mayner W G P, Marshall W, Albantakis L, Findlay G, Marchman R, Tononi G. PyPhi: a toolbox for integrated information theory. *Plos Computational Biology*, 2018, **14**(7): e1006343

102 Cleeremans A, Achoui D, Beauny A, Keuninckx L, Martin J-R, Munoz-Moldes S, et al. Learning to be conscious. *Trends in Cognitive Sciences*, 2019, **24**(2): 112–123

103 Ledoux J E, Brown R. A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**(10): E2016–E2025

104 Sun R. The CLARION cognitive architecture: toward a comprehensive theory of mind, Chipman S F, editor, *The Oxford Handbook of Cognitive Science*, Oxford, United Kingdom: Oxford University Press, 2017: 117–134

105 Samsonovich A V. Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. *Cognitive Systems Research*, 2020, **60**: 57–76

106 van den Boogaard E, Treur J, Turpijn M. A neurologically inspired network model for Graziano's attention schema theory for consciousness. In: Proceedings of the 2017 International Work-Conference on the Interplay Between Natural and Artificial Computation. Cham, Switzerland: Springer Publishing, 2017. 10–21

107 Dolega K, Dewhurst J. Bayesian frugality and the representation of attention. *Journal of Consciousness Studies*, 2019, **26**(3-4): 38–63

108 Graziano M S A. The attention schema theory: a foundation for engineering artificial consciousness. *Frontiers in Robotics and AI*, 2017, **4**: 60

109 Lanillos P, Ferreira J F, Dias J. Designing an artificial attention system for social robots. In: Proceedings of the 2015 IEEE International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE Press, 2015. 4171–4178

110 Da Silva Simoes A, Colombini E L, Ribeiro C H C. CONAIM: a conscious attention-based integrated model for human-like robots. *IEEE Systems Journal*, 2017, **11**(3): 1296–1307

111 Simon C. Can quantum physics help solve the hard problem of consciousness? *Journal of Consciousness Studies*, 2019, **26**(5-6): 204–218

112 Woolf N J, Hameroff S R. A quantum approach to visual consciousness. *Trends in Cognitive Sciences*, 2001, **5**(11): 472–478

113 Hameroff S, Penrose R. Consciousness in the universe: a review of the 'Orch OR' theory. *Physics of Life Reviews*, 2014, **11**(1): 39–78

114 Pitkänen M. New results about microtubules as quantum systems. *Journal of Nonlocality*, 2014, **3**(1): 1–18

115 Tegmark M. Why the brain is probably not a quantum computer. *Information Sciences*, 2000, **128**(3-4): 155–179

116 Baars B J, Edelman D B. Consciousness, biology and quantum hypotheses. *Physics of Life Reviews*, 2012, **9**(3): 285–294

117 Mahanti S, Das S, Behera B K, Panigrahi P K. Quantum robots can fly; play games: an IBM quantum experience. *Quantum Information Processing*, 2019, **18**(7): 219

118 Toffano Z, Dubois F. Quantum eigenlogic observables applied to the study of fuzzy behaviour of Braitenberg vehicle quantum robots. *Kybernetes*, 2019, **48**(10): 2307–2324

119 Jeswal S K, Chakraverty S. Recent developments and applications in quantum neural network: A review. *Archives of Computational Methods in Engineering*, 2019, **26**(4): 793–807

120 Zhang W R. Programming the mind and decrypting the universe-a bipolar quantum-neuro-fuzzy associative memory model for quantum cognition and quantum intelligence. In: Proceedings of the 2017 International Joint Conference on Neural Networks. Anchorage, USA: IEEE Press, 2017. 1180–1187

121 Adamski A G. Role of Bose-Einstein condensate and bioplasma in shaping consciousness. *Neuroquantology*, 2016, **14**(1): 36–44

122 Qazi W M, Ware J A, Bukhari S T S, Athar A. NiHA: a conscious agent. In: Proceedings of the 10th International Conference on Advanced Cognitive Technologies and Applications. Wilmington, USA: IARIA XPS Press, 2018. 78–87

123 Shi Z Z, Huang Z Q. Cognitive model of brain-machine integration. In: Proceedings of the 12th International Conference on Artificial General Intelligence. Cham, Switzerland: Springer Publishing, 2019. 168–177

124 Schweizer P. Artificial brains and hybrid minds. In: Proceedings of the 3rd Conference on Philosophy and Theory of Artificial Intelligence. Cham, Switzerland: Springer Publishing, 2018. 81–91

125 Romano D, Donati E, Benelli G, Stefanini C. A review on animal-robot interaction: from bio-hybrid organisms to mixed societies. *Biological Cybernetics*, 2019, **113**(3): 201–225

126 Lorrimar V. Mind uploading and embodied cognition: a theological response. *Zygon*, 2019, **54**(1): 191–206

127 Lee J. Brain-computer interfaces and dualism: a problem of brain, mind, and body. *AI & Society*, 2016, **31**(1): 29–40

128 Reggia J A. The rise of machine consciousness: studying consciousness with computational models. *Neural Networks*, 2013, **44**: 112–131

129 Manzotti R, Chella A. Good old-fashioned artificial consciousness and the intermediate level fallacy. *Frontiers in Robotics and AI*, 2018, **5**: 39

130 Tononi G, Boly M, Massimini M, Koch C. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 2016, **17**(7): 450–461

131 Rosenthal D M. Higher-order thoughts and the appendage theory of consciousness. *Philosophical Psychology*, 1993, **6**(2): 155–166

132 Lau H, Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 2011, **15**(8): 365–373

133 Chella A, Frixione M, Gaglio S. A cognitive architecture for robot self-consciousness. *Artificial Intelligence in Medicine*, 2008, **44**(2): 147–154

134 Chella A, Lanza F, Pipitone A, Seidita V. Knowledge acquisition through introspection in human-robot cooperation. *Biologically Inspired Cognitive Architectures*, 2018, **25**: 1–7

135 Hameroff S, Penrose R. Orchestrated objective reduction of quantum coherence in brain microtubules: the “Orch OR” model for consciousness. *Mathematics and Computer Simulation*, 1996, **40**(3-4): 453–480

136 Majumder D D, Karan S. Quantum computing: a nano scale information processing in minds and machines. In: Proceedings of the 2011 International Conference on Recent Trends in Information Systems. Kolkata, India: IEEE Press, 2011. 1–6

137 Yan F, Iliyasu A M, Jiao S H, Yang H M. Quantum structure for modelling emotion space of robots. *Applied Sciences-Basel*, 2019, **9**(16): 3351

138 Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. *Nature*, 2017, **549**(7671): 195–202

139 Abdulridha H M, Hassoun Z A. Control design of robotic manipulator based on quantum neural network. *Journal of Dynamic Systems, Measurement, and Control*, 2018, **140**(6): 061002

140 Amoroso R. Engineering a conscious computer. In: Proceedings of the 4th Workshop on Physics and Computation. Amsterdam, Netherlands: Elsevier, 1996. 12–16

141 Raghuvanshi A, Fan Y, Woyke M, Perkowski M. Quantum robots for teenagers. In: Proceedings of the 37th International Symposium on Multiple-Valued Logic. Oslo, Norway: IEEE Press, 2007. 18–25

142 Kak S. On quantum neural computing. *Information Sciences*, 1995, **83**(3—4): 143–160

143 Peruš M, Loo C K. *Biological and Quantum Computing for Human Vision: Holonomic Models and Applications*. Hershey, USA: IGI Global, 2011

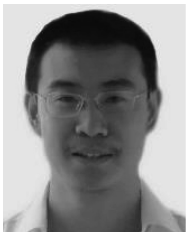
144 Mallakin A. An integration of deep learning and neuroscience for machine consciousness. *Global Journal of Computer Science and Technology*, 2019, **19**(1): 21–29

145 Lin Jing-Dong, Wu Xin-Yi, Chai Yi, Yin Hong-Peng. Structure optimization of convolutional neural networks: A survey. *Acta Automatica Sinica*, 2020, **46**(1): 24–37
(林景栋, 吴欣怡, 柴毅, 尹宏鹏. 卷积神经网络结构优化综述. 自动化学报, 2020, **46**(1): 24–37)

146 Lin Yi-Lun, Dai Xing-Yuan, Li Li, Wang Xiao, Wang Fei-Yue. The new frontier of AI research: Generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 775–792
(林懿伦, 戴星原, 李力, 王晓, 王飞跃. 人工智能研究的新前线: 生成式对抗网络. 自动化学报, 2018, **44**(5): 775–792)

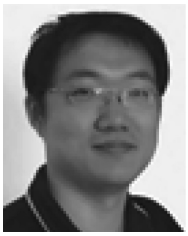
147 Zhang Lu-Ning, Zuo Xin, Liu Jian-Wei. Research and development on zero-shot learning. *Acta Automatica Sinica*, 2020, **46**(1): 1–23
(张鲁宁, 左信, 刘建伟. 零样本学习研究进展. 自动化学报, 2020, **46**(1): 1–23)

148 Kriegman S, Blackiston D, Levin M, Bongard J. A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, **117**(4): 1853–1859



秦瑞琳 厦门大学信息学院人工智能系博士研究生. 主要研究方向为机器意识, 情感计算和机器舞蹈.
E-mail: qqqrrrlll_2008@163.com
(**QIN Rui-Lin** Ph. D. candidate in the Department of Artificial Intelligence, School of Informatics, Xiamen University. His research interest covers machine consciousness, affective computing, and robotic dance.)

周昌乐 厦门大学信息学院人工智能系教授. 主要研究方向为机器意识, 脑机融合和机器歌舞. 本文通信作者.
E-mail: dozero@xmu.edu.cn
(**ZHOU Chang-Le** Professor in the Department of Artificial Intelligence, School of Informatics, Xiamen University. His research interest covers machine consciousness, brain-machine interface, and robotic dance. Corresponding author of this paper.)



晁 飞 厦门大学信息学院人工智能系副教授. 主要研究方向为智能机器人, 机器学习, 最优化算法.
E-mail: fchao@xmu.edu.cn
(**CHAO Fei** Associate professor in the Department of Artificial Intelligence, School of Informatics, Xiamen University. His research interest covers intelligent robotics, machine learning, and optimization algorithms.)

勘误声明

本刊 2020 年第 46 卷第 5 期 847–857 页所刊 “水上无人系统研究进展及其面临的挑战”一文中.

2016 年英国海军在苏格兰西海岸组织了 “无人战士” 大型无人化装备部署演习, 动用了 50 艘无人驾驶快艇, 负责海域探索, 监控情报收集, 以及鱼类侦测等任务.

应为:

2016 年英国海军在苏格兰西海岸组织了 “无人战士” 大型无人化装备部署演习, 动用了 50 多台(套) 空中、水面和水下无人系统, 负责海域探索, 监控情报收集, 以及鱼类侦测等任务.

特此更正, 并对由此带来的困扰表示歉意.

《自动化学报》编辑部