

情感智能与伦理

吕宝粮

摘要：情感智能指感知、理解、表达和调节情绪的能力，被认为是人类智能的重要组成部分。目前的人工智能尚未具备情感智能，无法解决人与机器自然交互的问题。未来的强人工智能将是逻辑智能与情感智能的有机结合。本文介绍构建机器情感的涌现理论，情感智能发展的四个阶段，霍金对人工智能的忧虑的现实性，情感智能的复杂性与多样性，情感智能研究中的伦理问题以及情感智能发展的现状。

关键词：人工智能 情感智能 强人工智能 伦理

一、引言

情绪与记忆、学习、决策同属于人的大脑的高级活动，情感交流是人与人之间交流的重要部分。情商反映一个人感受、理解、运用、表达、控制和调节自己情感的能力，以及处理自己与他人之间的情感关系的能力。在 20 世纪末，Salovey 和 Mayer 首次提出了情感智能的概念^①，认为除了逻辑智能，情感智能也是人类智能的重要部分。情感智能是指感知、理解、表达和调节情绪的能力。

近几年，人工智能领域最重要的进展之一，是 AlphaGo 围棋战胜人类围棋顶尖高手。从人工智能的角度，AlphaGo 围棋具有强大的逻辑智能，但它并不具有任何情感智能。在 2016 年 6 月夏季达沃斯论坛

^① P. Salovey, J. Mayer, "Emotional Intelligence", in *Imagination, Cognition and Personality*, vol. 9, 1990, pp. 185-211.

上，李世石曾几次提及相似的话题：“我再也不想跟 AlphaGo 下棋了。”李世石直言：跟 AlphaGo 下棋没有感情上的交流，困难超出想象。这正体现出情感智能的重要性。未来人工智能的研究，除了逻辑智能，我们必须关注情感智能。但是，目前无论是学术界还是工业界，大家还主要侧重于逻辑智能的研究与开发，而情感智能的研究尚处于萌芽阶段。随着人工智能的不断发展和在各行各业的应用，特别是那些需要人与机器自然交互的任务或人需要与机器协同作业的任务，情感智能将是关键的核心技术。例如，未来的家庭服务机器人，它们不同于现在汽车装配线上的工业机器人，不仅需要具有强大的逻辑智能，而且需要具有符合人类伦理、道德和生活习俗的情感智能。人们期望人工智能系统不仅要能“做事”，更要“懂人意”。举例说，老年人不会喜欢一个没有任何情感智能的家庭陪伴机器人或聊天机器人。不具有情感智能的弱人工智能无法满足人类社会的需求，未来的强人工智能（或称通用人工智能）应该是逻辑智能与情感智能的有机结合，是人工智能发展的必然趋势。

二、机器能具有人类的情感吗？

大脑是通用人工智能唯一的参照物。人类需要知道自己大脑的工作机理，从而构建机器的情感和强人工智能。美国马里兰大学的 Pessoa 使用行为和 fMRI 方法研究认知和情感以及认知和动机之间的相互作用。他认为大脑系统是高度互联的，而用于理解大脑的模块化方法从根本上是有问题的。尽管认知和情感 / 动机系统在很大程度上被认为是相互分离的，但他认为在一般信息处理架构中，认知和情感需要相互交织在一起，在构建智能机器人的过程中，情感和动机需要与所有的信息处理组件相结合。Pessoa 同样认为情感智能是智能机器人的必要成分，且情感功能不仅仅是机器人的一个功能模块，而应该

与学习、决策、行为、认知紧密耦合^①。

遗传算法之父约翰·亨利·霍兰德在 *Emergence: From Chaos to Order* 一书中提到整体协同耦合的系统远比各模块单独的行为复杂，并称之为涌现^②。涌现机制不是简单的功能堆积，其运算必须是非线性。大脑从单个神经元到高级功能正是涌现机制的一个例子。人工智能之父马文·明斯基在 *The Society of Mind* 一书中解释了如何利用无心智的小模块构建出心智的原理，即心智的涌现^③。他在 *The Emotion Machine* 一书中进一步解释人类大脑的运行方式，设计了能理解、会思考的机器，然后尝试将这种思维运用到理解人类自身和发展人工智能上^④。他认为机器的情感可以通过涌现机制由许多非情感功能构建，未来的机器将具有人类的情感。

三、霍金对人工智能的忧虑是杞人忧天吗？

人工智能追求的终极目标是让机器能像人一样学习、思考和具有情感。根据目前神经科学、认知科学、计算机科学和人工智能的发展水平，要实现像人一样的情感至少需要经历四个阶段。目前尚处于初始阶段，为纯逻辑智能，不具备任何情感智能；第一阶段，机器能精确识别人的情绪并进行情感反馈；第二阶段，机器具有自主学习能力并对客观世界有全面的感知；第三阶段，机器具有价值、意识和创造性，形成自我和价值观，从而具有与人类进行自然情感交互的能力。

我们需要指出的是不同于智商，情商的提升会降低机器伤害人类的风险。即机器的智商越低，人类越容易管理和控制机器。相反，在相同智商的情况下，机器情商越低，人类与机器交互的难度越大，机

① L. Pessoa, "Do Intelligent Robots Need Emotion?", in *Trends in Cognitive Sciences*, vol. 21, 2017, pp. 817-819.

② J. H. Holland, *Emergence: From Chaos to Order* (Oxford: Oxford University Press, 1998) .

③ M. Minsky, *The Society of Mind* (New York: Simon & Schuster, 1986) .

④ M. Minsky, *The Emotion Machine* (New York: Pantheon, 2006) .

器伤害人类的风险也越大。看看我们人类的社会，就不难理解智商与情商的这种关系。最近上海和广东接连发生的两起惊天命案，无一不是人的情商出现了严重的问题。

如果我们用数值来对智商进行量化的话，那么普通人的智商可以被看作是从 70 到 140 的一个数值，据说爱因斯坦可能会高达 160。如果同样用这样简单的数值来衡量情商的话，可能需要用从 -160 到 +160 的数值来表示。把符合人类伦理、道德、文化和生活习俗的情商定义为“正”情商，用 +1 到 +160 的数值表示，而把那些无法对自身的情绪进行管理直至危害社会公德的情商定义为“负”情商，用 -1 到 -160 的数值表示。实际上，不同于智商，“正”情商和“负”情商的最大值和最小值很难界定。

2015 年 3 月 24 日，德国之翼航空公司一架编号为 4U9525 的空中客车 A320 型客机在法国东南部的阿尔卑斯山脉南麓坠毁，导致机上载有 144 名乘客和 6 名机组人员的全部身亡的航空惨事。事后的调查结果显示，该机坠毁是因为副驾驶患有严重的精神疾病，他蓄意将飞机撞向阿尔卑斯山山体。对于这种因为精神疾病的原因，导致了“负”情商的行为，我们如何用一个数值来衡量呢？通过这个例子，我们想说的是，相对于智商，情商更难界定和量化。

1940 年科幻作家阿西莫夫提出机器人三原则：第一条：机器人不得伤害人类，或看到人类受到伤害而袖手旁观；第二条：机器人必须服从人类的命令，除非这条命令与第一条相矛盾；第三条：机器人必须保护自己，除非这种保护与以上两条相矛盾。一旦机器人具有了人类一样的情感，如何保证让机器人具有“正”情商并遵守上述三原则是一个非常具有挑战性的问题。

迅速发展的人工智能引发了关于强人工智能与伦理的讨论。著名理论物理学家霍金多次表达了对人工智能的忧虑。他认为：“强大的人工智能的崛起，要么是人类历史上最好的事，要么是最糟的；是好是坏我们仍不确定。但我们应该竭尽所能，确保其未来发展对我们和

我们的环境有利。我们别无选择。”

霍金对人工智能的忧虑是杞人忧天吗？如果我们赞同马文·明斯基关于未来机器会具有人类情感的理论，那么霍金的忧虑是现实的，有道理的，而并非杞人忧天。我们也不能因为霍金是理论物理学家，就对他的关于人工智能的深思嗤之以鼻。不管你是否喜欢强人工智能的提法，抑或反对研究强人工智能。但是，人类对于自然的好奇心和对美好生活的追求，探索自身大脑奥秘的脚步不会停止，探索强人工智能的努力也不会停止。

四、为什么情感智能需要考虑伦理问题？

在我们现实社会里，成功人士不仅需要高智商，更需要高情商，缺一不可。人从小到大的成长过程，是智商和情商同时提高的过程。与智商相比，情商的培养和教育往往更加困难和复杂。作为老师对这一点深有体会，我们教学生如何写论文相对比较容易，但是要教会学生如何做人却是一件非常难的事情。如果只注重智商的培养，而偏废了情商的提高，那么这样的人在社会上很难与别人融洽相处，也难以发挥出应有的作用。

我们研究情感智能的目标之一，是要建立一种新的机器学习框架，使机器能像人一样在提高智商的同时提高情商。也就是说，让机器逻辑智能与情感智能同时学习。人类的例子已经告诉我们，为什么智商与情商需要同时培养和提高。如果一个孩子，小的时候被父母娇生惯养，一点不懂规矩，一点不讲礼貌，等长大了再教育，十有八九难以扭转。在第二节里我们介绍过 Pessoa 对人脑的认知和情感功能偶联性的研究结果，他认为情感智能不能仅仅被当成是智能机器人的一个功能模块来使用，需要的时候调用这个模块，不需要的时候放到一边去。情感智能应该是与学习、决策、行为、认知的紧密耦合。

从人类智能的角度看，情感智能要比逻辑智能更加复杂和具有多

样性，因此更难以学习和构建。举例来说，如果把训练家庭服务机器人如何回避障碍物并找出最佳路径，把一杯饮料送到主人面前，看作是服务机器人的一项逻辑智能。那么，在一个三世同堂的大家庭里，这台服务机器人如何按照家庭的风俗习惯来给家里人送饮料就必须具有情感智能。如果考虑尊敬长辈，那就应该先送年长的。如果还有位感冒在家休息的年轻妈妈，那该如何操作呢？这种情况可能没有标准答案，要根据这个家庭平时的生活习惯以及他们对女性和病人的态度来决定。

一旦我们掌握了为机器构建情感智能的技术，那将带来各式各样的伦理问题。我们不会无聊地给微波炉和冰箱构建上情感，让它们彼此相爱。但是，对于家庭老人陪伴机器人和医院护理机器人这类服务机器人，如何构建它们的情感智能就必须考虑人类社会的伦理、道德和文化习俗。因此，如何将伦理、道德和风俗习惯形式化并作为常识让机器自主学习是情感智能和机器学习的挑战性课题。

五、情感智能离我们的生活还有多远？

目前，心理学对情绪这个词本身还没有统一、严格的定义，大多数实验还局限在理想化的环境；精神疾病的诊断停留在主观量表，缺乏客观定量的指标；神经科学在动物模型上对情绪的研究主要针对恐惧状态；商业应用的情感交互主要依靠图像、语音、文本和容易采集的生理信号，但是这些信号难以准确地区分人的复杂的情绪。

日本是机器人生产和使用的发达国家，早在30年前工业机器人已经广泛应用于工业生产线。日本政府和产业界的一个强烈愿望是想开发服务机器人，希望能像工业机器人那样，让服务机器人走进广大家庭，以解决日本老龄化社会问题。但是，直到目前为止，这个愿望依然未能实现，其中的一个重要原因是无法让这些服务机器人具有情感。

2015年日本软银、阿里巴巴和富士康宣布，联合出资开发情感机

机器人 Pepper。但是三年已经过去了，Pepper 的研发并不顺利。有些日本媒体甚至把 Pepper 说成是胸带平板电脑的机器人，并没有把它看作是一台情感机器人，因为它目前的功能确实算不上具有情感。

2016年9月《科学美国人》杂志提出了关于人类未来的20个大问题，第3和第15个问题分别是“我们是否可能理解意识的本质”以及“我们是否能用可穿戴技术探测自己的情绪”。麻省理工学院的 Picard 回答了第15问。她指出：“情绪与生物化学和电信号有关，这些信号会到达我们体内的每个器官，例如，压力影响我们的身体和心理健康。可穿戴技术让我们能量化这些信号在较长时间内的模式。在未来十年，可穿戴设备可以为我们的健康提供个性化的预报。比如，基于用户最近的压力、睡眠、社交情感活动，让我们的健康和幸福指数在未来一周提高80%。未来20年，可穿戴式设备和从其获得的分析数据，能显著降低精神和神经疾病的发病几率。”

笔者的实验室长期从事基于脑电、眼电和眼动信号的多模态情绪识别、疲劳驾驶检测和睡眠质量评估的研究工作，提出了脑电与眼动信号融合的情绪识别方法，发现了脑电信号对于三类情绪识别的稳定性以及脑电与眼动信号对于情绪识别的互补特性^①；开发了基于前额眼电和脑电的疲劳驾驶检测模型，在实验室模拟驾驶平台和实际驾驶环境的评测结果表明，该模型可以有效地检测驾驶员的疲劳状态^②；提出了利用5—10分钟脑电信号对被试前夜睡眠质量评估的便捷性方法^③。我们最新的研究结果表明，可以利用多模态脑机交互技术将抑郁症诊

① Yi-Fei Lu, Wei-Long Zheng, Bin-Bin Li, Bao-Liang Lu, “Combining Eye Movements and EEG to Enhance Emotion Recognition”, Proc. Of IJCAI-2015, 2015; Wei-Long Zheng, and Bao-Liang Lu, “Identifying Stable Patterns over Time for Emotion Recognition from EEG”, IEEE Transactions on Affective Computing, 2017.

② Wei-Long Zheng and Bao-Liang Lu, “A Multimodal Approach to Estimating Vigilance using EEG and Forehead EOG”, in *Journal of Neural Engineering*, vol. 14, 2017.

③ Jia-Jun Tong, Yun Luo, Bo-Qun Ma, Wei-Long Zheng, Bao-Liang Lu, Xiao-Qi Song, Shi-Wei Ma, “Sleep Quality Estimation with Adversarial Domain Adaptation: From Laboratory to Real Scenario”, Proc. of IEEE IJCNN-2018, 2018.

断的哈密尔顿量表转化成一系列脑机交互实验，从而实现基于多模态脑机交互的抑郁症辅助诊断系统，为抑郁症的诊断提供客观的量化指标和自动诊断技术。

六、结语与展望

尽管情感智能的研究尚处于萌芽阶段，情感机器人的开发也举步维艰，但随着机器学习和可穿戴设备的迅速发展，情感智能的研究已经从纯逻辑智能的初级阶段迈入了情绪识别与情感反馈的第一阶段。未来几年，有望通过可穿戴脑机交互系统实现人和动物在实际场景和自然环境下的情绪识别与情绪调控，有助于揭示自闭症、抑郁症等精神疾病的机理，开发早期诊断与治疗的方法和技术。对于特殊岗位的工作人员，如飞行员、高铁司机和危险品运输司机，实现实时的极端情绪状态识别与疲劳驾驶检测。