

情感机器 (湛庐文化机器人与人工智能书系)

作者：马文·明斯基

版权信息

本书纸版由浙江人民出版社于 2016年1月出版

作者授权湛庐文化 (Cheers Publishing) 作中国大陆 (地区) 电子版发行 (限简体中文)

版权所有·侵权必究

书名：情感机器

著者：(美) 明斯基 著

字数：358000

电子书定价：59.99美元

The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the

Future of the Human Mind by Marvin Minsky

Copyright © 2006 by Marvin Minsky

Published by arrangement with Simon & Schuster, Inc..

All rights reserved.

目录

[版权信息](#)

[编者按](#)

[机器人与人工智能，下一个产业新风口](#)

[机器人与人工智能书系](#)

[推荐序 情感机器 离我们有多远](#)

[引言 人类思维与人工智能的未来](#)

[引用和注释](#)

[第一部分 情感，人类特殊的思维方式](#)

[01 坠入爱河](#)

[“爱”的手提箱](#)

[精神奥秘之海](#)

[情绪与情感](#)

[本能机，让婴儿情感更好捉摸](#)

[云认知型思维](#)

[成人精神活动的6大层级](#)

[情感“瀑布”](#)

[思维维度的多样性](#)

[02 依恋与目标](#)

[沙子游戏：从叉子到勺子](#)

[依恋与目标](#)

[印刻者](#)

[依恋性学习模式](#)

[学习、快乐和信用赋能](#)

[价值体系的塑造](#)

[幼儿和动物的依恋](#)

[谁是我们的印刻者](#)

[自律，构建目标一致的自我模型](#)

[公众印刻](#)

[03 从疼痛到煎熬](#)

[疼痛之中](#)

[煎熬，大脑失去自由选择权](#)

[苦难机器](#)

[感觉、疼痛和煎熬](#)

[致命性的痛苦](#)

[悲伤](#)

[心智“批评家”：纠正性警告、外显抑制和内隐束缚](#)

[弗洛伊德的思维“三明治”](#)

[控制我们的情绪和性情](#)

[情感利用](#)

[第二部分 洞悉思维本质，创建情感机器的6大维度](#)

[04 意识](#)

[什么是意识](#)

[打开意识的手提箱](#)

[A脑、B脑和C脑](#)

[对意识的高估](#)

[如何开启意识](#)

[主观体验，心理学中的无解难题](#)

[自我模型与自我意识](#)

[笛卡儿剧场](#)

[不间断的意识流](#)

[05 精神活动层级](#)

[本能反应](#)

[后天反应](#)

[沉思](#)

[反思](#)

[自我反思](#)

[自我意识](#)

[想象](#)

[想象场景](#)

[预测机器](#)

[06 常识](#)

[什么是常识](#)

[常识性知识和推理](#)

[意图和目标](#)

[差异的“幻想”世界](#)

[在不确定性中，作出最优决策](#)

[相似推理](#)

[正面经验和负面经验的博弈](#)

[07 思维](#)

[是什么选择了我们思考的主题](#)

[批评家-选择器模型，思维跳跃之源](#)

[情感化思维](#)

[人类的19大思维方式](#)

[6大批评家，选择最合适的思维方式](#)

[先有情感，还是先有行为](#)

[庞加莱无意识过程的4大阶段](#)

[认知语境下的批评家选择](#)

[人类心理学的核心问题](#)

[08 智能](#)

[预估距离](#)

[平行类比](#)

[高效率学习的奥秘](#)

[信用赋能](#)

[创造力和天才](#)

[记忆与表征结构](#)

[表征等级](#)

[09 自我](#)

[多样的“自我”](#)

[人格特质](#)

[“自我”观念的魅力](#)

[为什么我们喜欢快乐](#)

[情感描述难题](#)

[发现感觉中独特的“质”](#)

[人类思维的组织方式](#)

[复杂的尊严](#)

[人类智能的3大时间跨度](#)

[致谢](#)

[译者后记](#)

机器人与人工智能，下一个产业新风口

·湛庐文化“机器人与人工智能”书系重磅推出·

50多年来，人工智能经历了从爆发到寒冬再到野蛮生长的历程，伴随着人机交互、机器学习、模式识别等人工智能技术的提升，机器人与人工智能成了这一技术时代的新趋势。

2015年，被誉为智能机器人元年，从习近平主席工业4.0的“机器人革命”到李克强总理的“万众创新”；从国务院《关于积极推进“互联网+”行动的指导意见》中将人工智能列为“互联网+”11项重点推进领域之一，到十八届五中全会把“十三五”规划编制作为主要议题，将智能制造视作产业转型的主要抓手，人工智能掀起了新一轮技术创新浪潮。Gartner IT 2015年高管峰会预测，人类将在2020年迎来智能大爆炸；“互联网预言家”凯文·凯利提出，人工智能将是未来20年最重要的技术；而著名未来学家雷·库兹韦尔更预言，2030年，人类将成为混合式机器人，进入进化的新阶段。

国内外在人工智能领域的全球化布局一次次地证明了，人工智能将成为未来10年内的产业新风口。像200年前电力彻底颠覆人类世界一样，人工智能也必将掀起一场新的产业革命。

值此契机，湛庐文化联合中国人工智能学会共同启动“机器人与人工智能”书系的出版。我们将持续关注这一领域，打造目前国内首套最权威、最重磅、最系统、最实用的机器人与人工智能书系：

- 最权威，人工智能领域先锋人物领衔著作。该书系集合了人工智能之父马文·明斯基、奇点大学校长雷·库兹韦尔、普利策奖得主约翰·马尔科夫、图灵奖获得者莱斯利·瓦里安和脑机接口研究先驱米格尔·尼科莱利斯等10大专家的重磅力作。
- 最重磅，湛庐文化联合国内这一领域顶尖的中国人工智能学会，专门为“机器人与人工智能”书系成立了专家委员会。该专家委员会包括中国工程院院士李德毅、英特尔中国研究院院长吴甘沙、地平线机器人技术创始人余凯、IBM中国研究院院长沈晓卫、国际人工智能大会（IJCAI）常务理事杨强、科大讯飞研究院院长胡郁、中国人工智能学会秘书长王卫宁等专家学者。他们将以自身深厚的专业实力、卓越的洞察力和深远的影响力，对这些优秀图书进行深度点评。
- 最系统，从历史纵深到领域细分无所不包。该书系几乎涵盖了人工智能领域的所有维度，包括10本人工智能领域的重磅力作，从人工智能的历史开始，对人类思维的创建与运作进行了抽丝剥茧式的研究，并对智能增强、神经网络、算法、克隆、类脑计算、深度学习、人机交互、虚拟现实、伦理困境、未来趋势等进行了全方位的解读。
- 最实用，一手掌握驾驭机器人与人工智能时代的新技术与新趋势。你可以直击工业机器人、家用机器人、救援机器人、无人驾驶汽车、语音识别、虚拟现实等领域的国际前沿新技术，更可以应用其中提到的算法、技术和理念进行研究，并实现个人与行业的大发展。

在未来几年内，人工智能和机器人给世界带来的影响将远远超过个人计算和互联网在过去30年间已经对世界造成的改变。我们希望，“机器人与人工智能”书系能帮助你搭建人工智能的体系框架，并启迪你深入发掘它的力量所在，从而成功驾驭这一新风口。

机器人与人工智能书系

·专家委员会·

主席

李德毅

中国人工智能学会理事长，中国工程院院士

委员（按姓氏笔画排序）

吴甘沙

英特尔中国研究院院长

余凯

中国人工智能学会副秘书长，地平线机器人技术创始人

沈晓卫

IBM中国研究院院长

杨强

中国人工智能学会副理事长，国际人工智能大会（IJCAI）常务理事

腾讯微信事业群技术顾问，香港科技大学教授

胡郁

科大讯飞高级副总裁，科大讯飞研究院院长

秘书长

王卫宁

中国人工智能学会秘书长，北京邮电大学研究员

董寰

湛庐文化总编辑

推荐序
情感机器
离我们有多远

李德毅
中国人工智能学会理事长
中国工程院院士

我不真的凭我的眼睛来爱你，
在你身上我看见了千处错误；
但我的心却爱着眼睛所轻视的。
——莎士比亚

坠入爱河，本美妙无比，但莎士比亚的这句话可能要惊醒无数“梦中人”了。为什么体验如此美好，现实却如此残酷？我们的大脑为什么会自动忽略那“千处错误”，而“爱着眼睛所轻视的”？

事实上是，我们的大脑很容易会欺骗我们自己。马文·明斯基在《情感机器》中对人类思维的本质进行了深入的剖析：人类大脑包含复杂的机器装置，并由众多“资源”（resource）组成，而每一种主要的情感状态的转变，都是因为在激活一些资源的同时会关闭另外一些资源，这改变了大脑的运行方式。所以，愤怒用攻击代替了谨慎，用敌意代替了同情。

而这人类思维的本质，正是我们研究人工智能、塑造最高级的“情感机器”的关键所在。

1956年，明斯基与约翰·麦卡锡一起发起并组织了达特茅斯会议，并首次提出“人工智能”的概念，而这场会议也成了人工智能的起点。期间，人工智能经历了两次上升、两次寒冬，终于在今日迸发出了野蛮生长的态势。而明斯基，在这段历史中一直是一位闪烁着耀眼光彩的杰出的人工智能权威，是当之无愧的人工智能之父。

随着历史车轮的迈进，人类社会已经进入了智能机器时代：工业机器人开始替代很多高重复性的人类劳动，甚至进入极端环境实施救援工作；轮式机器人进入了一个新阶段，将人类排除在外、实现交通零事故的目标将指日可待；被植入云端“大脑”的家庭机器人能听得懂人说的话，陪护机器人开始成为老人、儿童的新伙伴，陪他们度过了或孤独或患病的那些难熬的日子；而世界第一家机器人酒店也在日本开业，从前台到后勤的工作岗位全由机器人担任……机器人与人工智能的迅速发展，也给我们带来了更为严峻的问题：如果机器能够模拟人脑，具备意识、思维、自我观念等人类特质，会作出决断，是不是人类就能在无人驾驶和医疗这些关键领域实现质的飞跃？答案无疑是肯定的。

那么，大脑要进行何种变化才能改变我们的思维方式？机器应该如何演化，才能让它们具备人类的常识、常识性思考与反思能力？

在《情感机器》这本书中，明斯基为我们研究更高阶的人工智能——情感机器，提供了一幅详尽的路线图。他指出，情感是人类一种特殊的思维方式，并在洞悉思维本质的基础上，指出了人类思维的运行方式，提出了塑造未来机器的6大维度——意识、精神活动、常识、思维、智能、自我，揭示了人与机器根本性的不同，以及人之所以独一无二、足智多谋的原因，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。

毫无疑问的是，在未来几十年里，各国的研究者都将努力致力于更高阶人工智能的领域，但正如明斯基所言：只有当这些机器变得足够聪明，能够掩盖自己的种种缺点后，我们发明的系统才不会出现新的缺点。

情感机器，这一人工智能发展的终极答案，还将让人类上下求索。

引言

人类思维与人工智能的未来

诺拉·乔伊斯（Nora Joyce）对她的丈夫詹姆斯说：

“你为什么不写一部人人都能

看懂的书呢？”

我希望此书有益于以下读者：想要探究人类思维如何运作的读者，想要了解如何更好思考的读者，想要制造智能机器的读者以及想要学习人工智能知识的读者。此外，心理学家、神经学家和计算机专家也会对本书产生极大的兴趣，因为本书对这些人士一直奋力研究的课题提出了许多新观点和新看法。

我们都钦佩科技、艺术和人类学研究中取得的许多重大成就，却很少认可自己在日常生活中取得的成功。我们认识眼中看到的事物，理解耳中听到的词语，记得过去经历过的事情，因此可以利用这些经验应对以后将遇到的其他困难和机遇。

我们可以完成其他动物无法完成的非凡壮举：一旦无法采用常规思维方式思考，我们会反思；假如“反思”显示我们出现了错误，那么我们就重新塑造一些全新且有效的思维方式。但是，我们对大脑到底是如何成功完成这些事情的却所知甚少：想象如何发挥作用？意识缘何形成？什么是情感、感觉和想法？人类如何思考？

与此相反，人类在解决物理难题方面却取得了诸多进步。固体、液体和气体是什么？颜色是什么？声音和温度是什么？力、压力和张力是什么？能量的本质是什么？当前，区区几条定律就足以解释几乎所有这些神秘的事情，比如由物理学家牛顿、麦克斯韦和爱因斯坦以及薛定谔等人发现的方程式。

因此，心理学家自然也会想到模仿物理学家，即通过寻找一系列定律来解释大脑中发生的事情。然而，根本就不存在如此简单的一套定律，因为大脑拥有数千个部件，每一个部件都负责不同的特定工作：一些部件识别环境，一些部件促使肌肉执行行动，一些部件制订目标和计划，还有一些部件存储和使用大量的知识。尽管对大脑的运行方式不甚了解，但我们知道，大脑是在信息的基础上构建的，而信息又包含数以万计的遗传基因，因此大脑中每一个部位的运行方式都受不同的定律约束。

人类大脑包含复杂的机器装置，一旦认识到这个事实，我们就需要采取与物理学家完全不同的做法，即寻找更为复杂的方式来解释我们最为熟悉的精神活动，而不仅仅满足于寻找简单的方式。诸如“感觉”“情感”或“意识”等词语的意思对我们来说如此自然、清楚和直白，我们根本不知道该如何对它们的含义进行深思。然而，本书认为，这些常用的心理学词汇并不描述任何单一、确定的过程，相反，它们都在试图描述大脑中复杂过程的影响和效果，例如第4章将讨论的“意识”就指代20多种不同的过程！

使简单的问题变得复杂，这种方式表面上看起来会使事情变得更糟。然而，从大的方面来说，增强复杂性有助于简化工作，因为一旦把未知的事物分割成细小的部分，我们就用更小的问题来取代更大的问题；虽然这些更小的问题看起来可能仍然难于解决，但却不再是难解之谜。另外，第9章提出，把人类自身当成复杂机器的行为并没有伤害人类的自尊心，相反却能够增强人类的责任感。

在开始将大问题分割成小问题之前，本书认为，人类的大脑是由很多被我们称为“资源”的部分组成的（见图0-1）。

□

图0-1 大脑由不同“资源”组成

我们使用图0-1来说明一系列大脑资源是如何控制和影响人类思想的，从而来解释人类的一些精神活动（比如愤怒、爱或尴尬）。例如，被我们称作“愤怒”的状态能够激发一些资源，使其能够快速而有力地作出反应，同时压制一些让计划和行动都变得更为谨慎的资源。因此，愤怒用攻击代替了谨慎，用敌意代替了同情。与之相似，“恐惧”的状态会激发让你萌生退意的资源。

大众：有时我会觉得周围的一切是那么明亮和有趣，而有时（尽管周围没有发生任何变化）我又觉得周围的一切是那么黯淡和无聊，朋友们也觉得我有些“沮丧”和“压抑”。为什么会有如此的心态、情绪、情感或性情？为什么情感会引发这些奇怪的效果？

人们普遍认为“这些变化是大脑中某些化学物质、压力或消极的想法造成的”，但是这个答案却没有指出这些过程的实际运作方式，而选择“资源集”的想法却指出了思维变化的具体方式。例如，第1章伊始就是对以下常见现象的思考：

当一位朋友爱上了某人，他就像变了一个人似的成为一个有目标、有规划、懂得换位思考的人，就像一个刚刚被打开的开关或刚刚开始运行的程序。

大脑内部究竟发生了什么，会使其思维方式发生如此大的变化？以下是本书使用的研究方法：

每一种主要的“情感状态”（emotional states）都是因为激活了一些资源，同时关闭了另外一些资源——大脑的运行方式由此改变了。

但是，是什么激发了这样的资源集呢？我们将在以后的章节中提到，大脑也必须具有我们称之为“批评家”（Critics）的资源。每种“批评家”资源会专门识别某种情况，然后激活其他特定的资源集。一些“批评家”资源是与生俱来的，意在激发我们的一些本能反应，如愤怒、饥饿、恐惧和口渴，其发展、演化有利于人类祖先们的生存。因此，愤怒和恐惧是为了抵御和保护，而饥饿和口渴是为了给人类的进化提供所需要的营养物质（见图0-2）。

图0-2 不同“批评家”可以激发出不同的本能反应

然而，随着人的不断学习和成长，我们也不断发展出激活另一些资源集的方法，这种发展过程使得人们形成了一种情感状态，我们称之为“理性”而非“感性”。例如，在生活中遇到难题时，大脑能在不同的思维方式间进行转换和思考，从而选择不同的资源集来把问题细分为更小的部分，或通过类比法，或从回忆中寻找解决方法，或向他人寻求帮助。

这种方法将有助于我们发现独特的人类智能。

每一种主要的思维方式都是因为激活了一些资源，同时关闭了另外一些资源——大脑的运行方式由此改变了。

例如，本书的前几章试图从情感状态如何使用资源的角度出发，说明这种方法如何解释诸如爱、依恋、痛苦和沮丧等情感状态。之后的章节也会以同样的方法分析思维更加具有“智能”的层面。

大众：使用同一种方法来描述情感和常规性思维的做法看起来很奇怪，但是思维基本上是理性的、超然的、独立而有逻辑的，而情感则因为其非理性的感觉和偏见，从而活跃了思维方法。

传统上认为，情感为简单明了的思维增添了额外的特征，正如艺术家借用颜色为黑白的图画增添色彩一样。然而本书认为，我们的许多情感状态是由思维方式压制某些资源而形成的。例如，第1章提出的“依恋”是在压制一些资源的条件下形成的，被压制的资源则被用来识别其他资源所犯的错误。

大众：我依然认为你关于情感的观点忽略了太多内容。例如，当我们的胸部或内脏感到不适、心悸或是感到眩晕、颤抖或出汗时，诸如恐惧或恶心的情感状态就会影响我们的身体和大脑。

我承认这种观点可能太过极端，但有时候，为了探索新的想法，我们必须摒弃以往陈旧的思想，至少暂时需要如此。例如，当前人们普遍认为，人们的情感与身体状况紧密地联系在一起，然而本书第7章则将持相反的观点，认为身体部位是大脑用以改变（或保持）精神状态的资源，例如，我们可以通过持续保持某种面部表情来表示自己支持某个计划。

因此，尽管本书名为“情感机器”，但我们仍然认为情感状态与人们所认为的“思考”过程并无大异，相

反，情感是人们用以增强智能的思维方式。这就是说，当我们的热情没有高涨到对自己有害的程度时，不同的思维方式就成为被人们称作“智能”（英文为intelligence或resourcefulness）的重要组成部分，这个过程不仅适用于情感状态，也适用于我们所有的精神活动。

如果你仅以一种方式“理解”某件事情，那么你可能根本无法理解它，一旦走进死胡同，你便无路可逃。但是，如果同时使用多种方式来表征某物时，一旦你遭受挫败，便可以转换到其他思维方式，直到找到一种适合自己的方法为止。

以此类推，当设计模拟人脑的机器即创建人工智能时，我们需要确保这种机器的多样性。

如果一个程序仅有一种运行方法，那么一旦这种方法不再适用于该程序，则该程序很快就会陷入停滞。相反，如果一个程序同时有几种运行方法，一旦其中一种方法出现问题，程序会很快转而使用其他方法或者寻找合适的替代方法。

以上就是本书的中心思想，它与认为每个人都有中心核、一种看不见的精神或自我，这种精神或自我又产生了多种多样精神活动的主流观点完全不同。这种主流观点看起来降低了人类的身份，即我们所有的美德都是间接获得的，我们所有的成就都是不值得称颂的，因为这些成就是其他途径附送的礼物。相反，我认为人类的尊严来自人类自身的构造：处理多种情况和困境的不计其数的方法。多样性正是人和动物以及与过去所制造的机器的主要区别所在。本书的每一章都会讨论人类特殊智能的来源途径：

第1章 我们天生就拥有多种精神资源；

第2章 我们能从人与人的关系中学习；

第3章 情感是不同的思维方式；

第4章 我们能反思我们自己最近的想法；

第5章 我们试着学习多维度思考；

第6章 我们能够积累大量的常识知识；

第7章 我们可以在不同思维方式之间转换；

第8章 我们发现了很多表征事物的方式；

第9章 我们可以塑造多样的自我。

几百年来，心理学家一直在寻找能够解释人类大脑活动的方法，但是到目前为止，仍然有很多思想家认为思维的本质非常神秘。许多人认为大脑是由一种只存在于生命体内的物质成分组成，机器不能感觉或思考，不会担忧自身的变化，甚至不会感受到自己的存在，更不会产生促成伟大的画作或交响乐作品的思维。

本书有如下目标：解释人类大脑的运行方式，设计出能理解、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。

引用和注释

本书所引用部分源自具体的人物，我们会将其写在注释里。

马塞尔·普鲁斯特（Marcel Proust, 1927）：每一位读者都只能读出自己已有的知识。一本书仅仅是一种光学仪器，作家在这个光学仪器里让读者发现自己，这种发现必须借助书本。

另一些引用则为假想读者的评论：

大众：如果日常的思维如此复杂，那为什么它在我们看起来却如此直白？

欢迎访问：电子书学习和下载网站 (<https://www.shgis.cn>)

文档名称：《情感机器（湛庐文化机器人与人工智能书系）》马文·明斯基 著.epub

请登录 <https://shgis.cn/post/728.html> 下载完整文档。

手机端请扫码查看：

