

THE EMOTION MACHINE

情感机器

[美] 马文·明斯基◎著 王文革 程玉婷 李小刚◎译

MARVIN MINSKY



中国人工智能学会·丛书·

THE EMOTION MACHINE

Commonsense Thinking,
Artificial Intelligence,
and the Future of the Human Mind



情感机器

[美] 马文·明斯基 (Marvin Minsky) ◎著

王文革 程玉婷 李小刚◎译



浙江人民出版社
ZHEJIANG PEOPLE'S PUBLISHING HOUSE

图书在版编目(CIP)数据

情感机器 / (美) 明斯基著; 王文革, 程玉婷, 李小刚译. —杭州: 浙江人民出版社, 2016.1

ISBN 978-7-213-06942-0

浙江省版权局
著作权合同登记章
图字: 11-2014-88 号

I. ①情… II. ①明… ②王… ③程… ④李… III. ①人工智能—研究
IV. ①TP18

中国版本图书馆 CIP 数据核字 (2015) 第 260538 号

上架指导: 科技 / 人工智能

版权所有, 侵权必究

本书法律顾问 北京市盈科律师事务所 崔爽律师
张雅琴律师

情感机器

作者: [美] 马文·明斯基 著

译者: 王文革 程玉婷 李小刚 译

出版发行: 浙江人民出版社 (杭州体育场路347号 邮编 310006)

市场部电话: (0571) 85061682 85176516

集团网址: 浙江出版联合集团 <http://www.zjcb.com>

责任编辑: 金 纪

责任校对: 张志疆 张彦能

印刷: 北京鹏润伟业印刷有限公司

开 本: 720 mm × 965 mm 1/16

印 张: 24.5

字 数: 35.8 万

插 页: 4

版 次: 2016 年 1 月第 1 版

印 次: 2016 年 1 月第 1 次印刷

书 号: ISBN 978-7-213-06942-0

定 价: 99.90 元

如发现印装质量问题, 影响阅读, 请与市场部联系调换。



ROBOT & ARTIFICIAL INTELLIGENCE SERIES

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

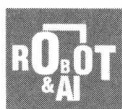
成为未来 10 年内的产业新风口。像 200 年前电力彻底颠覆人类世界一样，人工智能也必将掀起一场新的产业革命。

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。



ROBOT & ARTIFICIAL INTELLIGENCE SERIES

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

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。



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

而明斯基，在这段历史中一直是一位闪耀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。



目录

推荐序 /Ⅳ

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

引言 /001

人类思维与人工智能的未来

PART 第一部分

情感，人类特殊的思维方式

01

坠入爱河 /011

我们每一种主要的“情感状态”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。如果这种改变每次都会激活更多其他资源，那么最终将导致资源的大规模“级联”。

“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 /039

人类的一些目标是天生的本能，是由我们的基因决定的；另一些目标则是通过“尝试和错误”学习，来实现已有目标的次级目标；而高层次目标，则是由一种特殊的机器体系形成的。这种特殊的机器体系是指我们对身为依恋对象的父母、朋友或亲人的价值观的继承，这些价值观积极地响应了我们的需要，在我们体内产生了“自我意识”情感。

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 /069

任何疼痛都会激活“摆脱疼痛”这一目标，而这个目标的实现将有助于目标本身的消失。然而，如果疼痛强烈而又持久，就会激发其他大脑资源，进而压制其他目标。如果这种情况级联式地爆发下去，那么大脑

的大部分区域都会被痛苦占据。可见，在处于某种精神状态中时，我们也就失去了“选择的自由”。

疼痛之中

煎熬，大脑失去自由选择权

苦难机器

致命性的痛苦

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

PART 第二部分

洞悉思维本质，创建情感机器的6大维度

04

意识 /099

“意识”是一个“手提箱”式词汇，它被我们用来表示许多不同的精神活动。而这些精神活动并没有单一的原因或起源，当然，这也正是为何人们发现很难“理解意识是什么”的原因所在。心灵的每个阶段都是一个同时存在多种可能性的剧场，而意识则将这些可能性相互比较，通过注意力的强化和抑制作用，选择一些可能性、抑制其他可能性。

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 /135

我们的大脑是如何产生如此多新事物和新想法的？资源可以分为 6 种不同的层级——本能反应、后天反应、沉思、反思、自我反思、自我意识，以对想法和思维机制进行衡量。每一个层级模式都建立在下一个层级模式的基础之上，最上层的模式表现的是人们的最高理想和个人目标。

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 /169

我们所做的许多常识性事情和常识性推理，要比吸引更多关注、获得令人敬仰的专业技能复杂得多。你所“看到”的并不完全来自视觉，还来自这些视觉引发的其他知识。常识性知识的主体，即人类需要在文明世界中相处下去会涉及的许多问题，如我们所说的常识性问题，目标是什么以及它们是如何实现的，我们平常是如何通过类比来推理，以及我们如何猜测哪一项知识等，可能与我们的决策方式相关联。

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 /221

我们几乎从未认识到常识性思考所创造的奇迹。人人都有不同的思维方式。在众多的兴趣爱好当中，是什么选择了我们下一步将要思考的内容？每一种兴趣又会持续多久？批评家又是如何选择所使用的思维方式的？事实上，工作被隐藏在“脑后”，仍在继续运行。

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 /263

每个物种的个体智力都会从愚笨逐渐发展到优秀，即使最高级的人类思维也本应从这个过程发展而来。我们可以通过多种视角来观察事物，我们拥有快速进行视角转换的方法、拥有高效学习的特殊方式、拥有获得相关知识的有效方式并可以不断扩大思维方式的范围、拥有表征事物的多种方式。正是这种多样性造就了人类思维的多功能。

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 /305

是什么让人类变得独一无二？任何其他动物都无法像人类这样拥有各种各样的人格。其中一些性格是与生俱来的，而另一些性格则来自个人经验，但在每一种情况中，我们都具有各自的特征。每当想尝试理解自己时，我们都可能需要采取多种角度来看待自己。

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 /353
注释 /357
译者后记 /367



你不是一个人在读书！
扫码进入湛庐“趋势与科技”读者群，
与小伙伴“同读共进”！

诺拉·乔伊斯（Nora Joyce）对她的丈夫詹姆斯说：

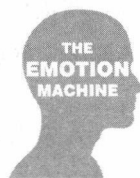
“你为什么不写一部人人都能
看懂的书呢？”

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，那么我们就能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



引言

人类思维与人工智能 的未来

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，我们就能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）。