

LAW SCIENCE OF  
ARTIFICIAL INTELLIGENCE

No.1

# 人工智能法学 研究

2 0 1 8 年 第 1 期 总 第 1 期

主编 岳彩申 侯东德 副主编 张建文



社会科学文献出版社  
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)



集人文社科之思 刊专业学术之声

刊 名：人工智能法学研究

主办单位：西南政法大学人工智能法律研究院

主 编：岳彩申 侯东德

副 主 编：张建文

## LAW SCIENCE OF ARTIFICIAL INTELLIGENCE No.1

### 学术顾问委员会（以姓名首字母排序）

主 任：岳彩申

副主任：侯东德

委 员：黄 辉（香港中文大学法学院）

卢代富（西南政法大学经济法学院）

Mathias Siems（英国杜伦大学法学院，剑桥大学商业法研究中心）

梅传强（西南政法大学法学院）

任惠华（西南政法大学刑事侦查学院）

王国胤（重庆邮电大学大数据智能研究院）

熊明辉（中山大学哲学系和法学院）

熊 榆（英国诺森比亚大学纽卡斯尔商学院）

杨 东（中国人民大学法学院）

赵万一（西南政法大学民商法学院）

张 渡（澳门科技大学资讯科技学院）

张凌燕（西南政法大学司法鉴定与证据科学研究院）

### 编辑部

主 任：叶 明

编 辑：黄金华 李 媛 赵 吟 姚万勤 郑志峰

本刊编辑部地址：西南政法大学敬业楼 5070 室

电 话：023-67258286

投稿邮箱：rgznflyjy@swupl.edu.cn xnzfrgzn@163.com

### 第 1 期

集刊序列号：PIJ-2018-266

中国集刊网：<http://www.jikan.com.cn/>

集刊投约稿平台：<http://iedol.ssap.com.cn/>

LAW SCIENCE OF  
ARTIFICIAL INTELLIGENCE

No.1

# 人工智能法学 研究

2 0 1 8 年 第 1 期 总 第 1 期

主编 岳彩申 侯东德 副主编 张建文



社会科学文献出版社  
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)



# 发刊词

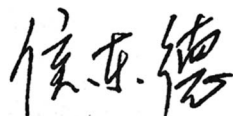
21 世纪是一个繁荣的世纪，也是一个科技引领时尚的世纪。以互联网为依托，我们实现了全球范围内的信息交流与资源共享。在大数据的支持下，我们得以通过海量数据的计算收获有意义的结果。寄寓于人工智能，我们希望机器在可能的应用场景中替代人类实现高效率的识别认知和分析决策。无论是基于符号主义认为人工智能源于数学逻辑，还是基于连接主义认为人工智能源于仿生学，实践已经表明，人工智能在计算机视觉、语音工程、自然语言处理、决策系统、大数据应用等方面给经济社会生活带来了全新的体验。然而，技术是一把双刃剑，在解放人类体力和智力的同时，也隐藏着巨大的风险。世界人工智能百强公司的领导们经常担忧发展过于迅速致使我们面临来自机器人的威胁，甚至因反应迟钝最终被机器所淘汰。不容否认，机器通过深度学习越来越聪明，可能超出设计者最初的预想及控制范围。当人工智能技术的载体联合作用形成复杂生态系统时，原因与结果之间将不再是简单的线性关系，细微的变化可以引起爆炸性的反馈回路，也许会出现科幻小说和电影中恐惧的画面。

技术本身没有错，人工智能本身也并不可怕。担心机器人会替代人类成为地球的主宰者，无异于当初担心发明原子能用于取代矿物能源会毁灭人类，事实并非如此。技术的中立揭示问题的根本在于人类如何运用自己掌握的技术，如何乘借科技之风遨游多维空间。这离不开制度规则的约束，离不开上层建筑行之有效的方向引导和实操调控。科学的制度规则不仅能够确保人工智能技术在既定的轨道上运行，服务于实体经济的长足发展，而且能够促进人工智能技术本身的进步，平衡安全与效率的社会价值。在依法治国的大背景下，法律人如何面对人工智能的发展，如何应对新型的挑战，如何跳出学科界限的藩篱从全局的视角发现问题、解决问题，将是一项任重而道远的事业。

为了提供聚焦于人工智能法学研究的交流平台，吸纳不同相关领域的专家、学者、实务人才共话新时代法律与技术问题，我们创办了《人工智能法学研究》，旨在打通人工智能学界与法学界的对话通道，为人工智能

与法律交叉融合研究成果搭建专门的发表平台，希冀助推法学界、人工智能学界展开理论联系实际的合作研究，培养从事人工智能与法律研究的青年学者，培育人工智能与法律研究的重大研究成果，为国家人工智能发展和法治建设提供理论参考。

21 世纪往后，我们无限展望人工智能的纵深发展，相信通过法律人自觉自省的持续刷新研究，技术可以更好地服务于经济社会，更规范地为人类所用。在未来世界，人类终将是传奇。

A handwritten signature in black ink, reading 'He Dongde' (侯东德) in a cursive style.

2018 年 9 月

001 理论争鸣

003 阿西莫夫的教诲：机器人学三法则的贡献与局限

——以阿西莫夫短篇小说《汝竟顾念他》为基础/张建文

021 论人工智能的法律主体地位/赵 磊 赵 宇

038 智能机器人法律地位的审视/付子堂 赵译超

053 部门法视野

055 区块链合同的法律属性

——概念、风险与监管建议/蒋大兴 王彦光

071 近代人格理论建构与人工智能民法定位/刘云生

080 人工智能投资顾问在我国的法律界定

——从“智能投顾”到“智能财顾”再到“智能投顾”/邢会强

096 智能投顾的信义义务/李文莉 杨玥捷

117 跨界对话

119 规制人工智能：一个原则性法律框架研究/涂永前

133 人工智能开发的法律规制对象、目标及措施/岳彩申 耿志强

- 004 145 域外观察
- 147 首次代币发行淘金热：计一场，梦一场，祸福一场  
——兼论其法律监管/德克·A. 思德 罗斯·P. 巴克利  
道格拉斯·W. 华纳 莱纳斯·弗尔 林少伟 译
- 185 澳大利亚智能投顾监管及借鉴/李 晴
- 227 **Abstract**
- 234 约稿函

理论争鸣



# 阿西莫夫的教诲：机器人学三法则的贡献与局限\*

——以阿西莫夫短篇小说《汝竟顾念他》为基础

张建文\*\*

**摘要：**阿西莫夫的机器人学三法则并非法律，更非法律原则或者法律规范，而是建立人类与机器人关系的伦理原则，体现了在机器人发展给人类带来难以清楚预估的影响面前突出和保障人类个体乃至人类整体的安全的优先价值。阿西莫夫透过机器人学三法则的警告和教诲在今天并未过时，而是应当成为在人工智能发展面临革命性突破的今天创制规范人工智能研发、资助、产品制造、商业化与应用等方面的人类与机器人关系中的伦理规范和法律规范的基础。没有安全保障的发展，难谓真正的有意义的发展。重申人类尊严和安全的优先价值，重新确立人类中心主义的伦理与立法价值，并不是不合时宜的。

**关键词：**阿西莫夫的教诲 机器人学三法则 人类与机器人的关系 人类中心主义

被读者誉为“神一样的人”，被美国政府授予“国家的资源与自然的奇迹”这个独一无二的称号，以表彰他在“拓展人类想象力”上所做出的杰出贡献的俄裔美籍科幻作家阿西莫夫<sup>①</sup>，1942年在自己的短篇小说《圆舞》中提出的机器人学三法则在科幻作品中是作为机器人的强制行为

---

\* 本文为作者主持的司法部2016年度国家法治与法学理论研究项目“新生权利的理论与实践问题研究”（项目编号：16SFB2001）、教育部人文社会科学重点研究基地重大项目“新兴权利的基本问题研究”（项目编号：16JJD820031）、西南政法大学人工智能法律研究院教师研究创新项目“人工智能机器人的法律调整：阿西莫夫法则的贡献与局限”（项目编号：2018-RGZN-JS-ZD-10）的阶段性成果。

\*\* 张建文，西南政法大学教授，博士生导师，法学博士，西南政法大学人工智能法律研究院副院长。研究方向：人格权法、信息法学。

① [美]阿西莫夫：《阿西莫夫：机器人短篇全集》，叶李华译，南京：江苏文艺出版社，2014，扉页。

规则<sup>①</sup>，也被称为“当代机器人学的基本法则”<sup>②</sup>“现代机器人学的基础”<sup>③</sup>，堪称“科幻小说中关于人工智能最具启发性的构想”<sup>④</sup>。

机器人学三法则是阿西莫夫所有机器人科幻作品的统一的主题。从法学的角度，认真细致地研究阿西莫夫在自己的小说文本中提出的机器人学三法则是一探阿西莫夫关于机器人与人类关系的卓越思考的桥梁和钥匙，也只有依靠和通过阿西莫夫的相关小说文本才有可能去尝试接近和理解阿西莫夫的教诲。在研究的态度上，不能因为机器人学三法则是在76年前提出的就断然认定阿西莫夫的教诲已经被后人们所超越，当今研究者的智慧一定就比阿西莫夫的智慧更高，而是应当虚心地真诚地倾听“最伟大心智之间的交谈，因而也在于研习那些伟大著作”<sup>⑤</sup>。

现代科学技术，更甚于作为技术根源的希腊哲学，令“人的人性有毁灭之虞”<sup>⑥</sup>，给使人们尤其是学者们带来的冲击，乃至恐慌，产生了某种焦虑，无论是国内还是国外，都有学者发出“面对日新月异的人工智能技术，未来留给我们用于思考的时间已经不多了”<sup>⑦</sup>或者“距离机器开始思考并有意识地完成某些行为，也许所剩时间无多”<sup>⑧</sup>的焦虑。到今天为止，人类对人工智能特别是人工智能机器人的态度尚未走出阿西莫夫所展示的思考与焦虑，尤其是体现在其如同哲学对话录般的短篇小说《……汝竟顾念他》（以下简称《顾念他》）之中。正如2015年Edge的问题：“你怎么看待那些会思考的机器？”“这些机器会想什么？它们会想要拥有民事权利吗？它们会拥有意识吗？AI会为我们选择怎样的政府？它们会形成什么样的自己的社会？或者说它们自己的社会将会和我们的社会一样吗？”

① Issac Asimov, *Runaround: I Robot* (New York: Doubleday, 1950), p. 40.

② [美]阿西莫夫：《阿西莫夫：机器人短篇全集》，叶李华译，南京：江苏文艺出版社，2014，扉页。

③ [美] John Frank Weaver:《机器人也是人：人工智能时代的法律》，郑志峰译，台北：元照出版有限公司，2018年，第3页注释1。

④ [美] John Frank Weaver:《机器人也是人：人工智能时代的法律》，郑志峰译，台北：元照出版有限公司，2018年，第5页。

⑤ [美]列奥·施特劳斯：《古典政治理性主义的重生》，郭振华等译，北京：华夏出版社，2017，中译本前言第1页。

⑥ [美]列奥·施特劳斯：《古典政治理性主义的重生》，郭振华等译，北京：华夏出版社，2017，中译本前言第2页。

⑦ 赵万一：《法律应当如何重新定位人》，载[美] John Frank Weaver《机器人也是人：人工智能时代的法律》，郑志峰译，台北：元照出版有限公司，2018年，推荐序第22页。

⑧ Дороганов Виталий Сергеевич, Баумгартен Михаил Ицкович Возможные проблемы, возникающие при создании искусственного интеллекта//Вестник КузГТУ. 2013. №4 (98). С. 133.

我们和会思考的机器会相互进入同情的圈子吗？”<sup>①</sup>

## 一 阿西莫夫的创造：机器人学三法则过时了吗？

尽管阿西莫夫在提出机器人学三法则的时候，尚不存在像我们今天这样的人工智能发展成就和应用环境，但是，值得思考的是，机器人学三法则是否已经过时了？<sup>②</sup>从法学的角度，法学家是否或者是否应当关注机器人学三法则？除了之前提到的欧洲议会动议中和俄罗斯科学院国家与法研究所的关注之外，其实美国的法学家也对阿西莫夫在三法则提出了第四和第五法则，可以说，法学家更加关注机器人学三法则，而且接过科幻作家的接力棒，用法学的视角、技术和经验推动机器人学三法则继续发展。不过，法学家的关注和发展不如阿西莫夫的三法则名气大，至于其质量优劣尚待进一步分析。

由于在我国对机器人学三法则的翻译存在不准确的地方，笔者根据对该三法则的英文文本<sup>③</sup>和俄文文本<sup>④</sup>的对比，提出自己的译本。第一法则：“1. 机器人不得伤害人类个体，也不得以其不作为致使人类个体受到伤害。”第二法则：“2. 机器人应当服从人类个体给予的所有命令，除非该命令违反第一法则。”第三法则：“3. 机器人应当在在不违背第一法则或第二法则的范围内关注自己的安全。”

值得注意的是，阿西莫夫机器人学三法则本身也存在一个自身的发展，在提出机器人学三法则 44 年之后，阿西莫夫 1986 年在其长篇小说《机器人与帝国》中提出了所谓的零号法则（Нулевой закон），即“0. 机

---

① Что мы думаем о машинах, которые думают: Ведущие мировые ученые об искусственном интеллекте//Джон Брокман; Пер. с англ. -М.; Альпина нон-фикшн. 2017. С. 22.

② 腾讯研究院等：《人工智能》，北京：中国人民大学出版社，2017，第300页。

③ 0. A robot may not harm humanity, or, by inaction, allow humanity to come to harm. 1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

④ 0. Робот не может нанести вред человечеству или своим бездействием допустить, чтобы человечеству был нанесён вред. 1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред. 2. Робот должен повиноваться всем приказам, которые дает человек, кроме тех случаев, когда эти приказы противоречат Первому закону. 3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму законам.

机器人不得伤害人类，或以自己的不作为致使人类受到伤害”。<sup>①</sup>从这个法则的编号上看，意思是这个法则应当居于三个法则之前，其中所蕴含的地位之重要性自不待言。只不过是出于零号法则提出较晚，不如机器人学三法则名气大，就阿西莫夫的天才思考而言，其实是机器人学四法则，不过，在科幻小说和法学界习惯上仍称之为机器人学三法则（три закона роботехники）。

在此之前，英国专门负责国内科技领域调整的国家机构“工程与科学研究委员会（EPSRC，即 Engineering and Physical Sciences Research Council）”于2011年在阿西莫夫机器人学三法则的思想基础上提出了自己的五条方案。“1. 不得专为毁灭或者伤害人们的目的而研发机器人。2. 承担责任的人是人，而不是机器人。机器人是达成人类目的的工具。3. 应当根据其使用的安全性研发机器人。4. 机器人是人工造物，它们不应当在有感觉的人类情感中发挥作用。机器人必须与人类相区别。5. 任何时候都应当可以找出在法律上为该机器人承担责任的人。”<sup>②</sup>

经过长期的发展，现在已经成为法学界关注的对象。2016年，美国法学家马克·罗腾堡教授直接在阿西莫夫的机器人学三法则的基础上提出了第四法则“机器人应当公开识别自己”和第五法则“机器人应当能够公开证明自己的行为”<sup>③</sup>。

2016年微软公司负责人Наделл提出了六项发展人工智能的规则。“1. 人工智能应当是为了帮助人类而建造。2. 人工智能应当是透明的：任何时候都应当有机会查明它是如何工作的。3. 人工智能应当在不侵犯人们尊严的前提下促进解决问题效率的提升。4. 人工智能应当保持合理的机密性并通过保护所交付的信息来赢得信任。5. 人工智能在算法上应当受到监督：人类个体有机会消除非故意给其造成的损害。6. 人工智能应当被保护免受干扰，并平等对待所有人。”<sup>④</sup>

2017年2月16日，欧洲议会《关于机器人民法规范的动议》明确提出：“阿西莫夫机器人学三法则必须被视为针对机器人的设计者、生产者

① Азимов, Айзек. Роботы и Империя. -Москва : Эксмо, 2006. 5 - 699 - 17608 - X.

② Андрей Незнамов. Новые законы робототехники: как в Европе регулируют права роботов. Популярная механика. 2017. №8.

③ Андрей Незнамов. Новые законы робототехники: как в Европе регулируют права роботов. Популярная механика. 2017. №8.

④ Андрей Незнамов. Новые законы робототехники: как в Европе регулируют права роботов. Популярная механика. 2017. №8.

和操作者，包括被赋予自主能力和自学能力的机器人的法则。”<sup>①</sup> 机器人技术发展的伦理层面令欧洲议会觉得不安，因此在该议会会议的附件中提出了供机器人技术研发者自愿遵守的《伦理规则》(кодекс этики)，其中所提出的得到所有人同意的核心性原则“造福”“不伤害”等均来自阿西莫夫著名的“机器人学法则”，<sup>②</sup> 此外，还有自主性原则（即人类自愿地知情同意与机器人互动）和公平原则（即正确分配在此情况下所产生的利益）。2018年2月15—16日，俄罗斯科学院国家与法研究所举行的法学会议中举行了纪念阿西莫夫的系列讲座。<sup>③</sup>

在2018年5月出版的一本关于人工智能时代法律的著作《机器人也是人：人工智能时代的法律》中，该书的三个部分的标题分别为“第一部分 机器人会伤害人类吗？”“第二部分 机器人必须服从人类的命令吗？”“第三部分 机器人将会保护自己吗？”，这三个标题与阿西莫夫机器人学三法则之间的相似性和关联性，显示了阿西莫夫提出的不伤害原则、服从原则和自我保护原则对该作者思考和提出人工智能和机器人相关法律问题的直接和重大的影响。作者自己也指出：“如果仔细翻看本书的三个部分，你就会发现它们正是基于这三大法则而展开的。机器人学三大法则力图确保机器人能够增进人类福祉，而不是给人类带来伤害。”但是作者的思考正是针对三大法则的质疑：“正如罗杰·克拉克（Roger Clarke）所说，阿西莫夫的小说本身就提供了充分的证据，根本不可能透过制定一套规则去约束机器人的行为。”<sup>④</sup> 其实从作者对该著作三个部分的标题所使用的问号，就可以直接明了作者对自己的意图和基本立场的暗示。

阿西莫夫关于机器人学三法则的教诲不是教条式的律法主义，不是规范式的法律条文，其关于机器人的系列短篇小说是奉献给三法则以及可能的违反行为的原因及其后果的，相反在另一些作品中，则主要是研究遵守机器人学三法则的不可预见的后果。据说他是非常具有诙谐搞笑的天赋

① 参看欧洲议会于2017年2月16日《机器人民法规范》[P8-TA(2017)0051第T段“Civil law rules on robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics”2015/2013(INL)]。

② Андрей Незнамов. Новые законы робототехники: как в Европе регулируют права роботов. Популярная механика. 2017. №8.

③ Андрей Незнамов. От редактора. Дайджест Робоправа. Выпуск января 2018. с. 1.

④ [美] John Frank Weaver:《机器人也是人：人工智能时代的法律》，郑志峰译，台北：光照出版有限公司，2018年，第6页。

的，甚至还写过一本《笑话集》，到了晚年，开始变得“好色”，还出版了系列两性话题的“黄书”。总体而言，在今天断言阿西莫夫所提出的机器人学三法则已经过时是武断的、轻率的和盲目的，现代学者们特别是法学学者们或者支持或者批判的争论中，就凸显了阿西莫夫教诲的吸引力、穿透力和近乎难以解释的神奇魅力。

## 二 阿西莫夫的玩笑：机器人学 三法则真的是法律吗？

008

值得追问的问题是，机器人学三法则真的是法律吗？如何看待机器人学三法则的价值和地位？“Three laws of Robots” “три закона роботехники”的表述，似乎看起来机器人学三法则就像法律条文或者法律规定那样。然而，这一套法则是由科幻作家提出的而不是由法学家提出并得到司法认可的或者由立法通过的法律文件。实际上，更准确地说，阿西莫夫的机器人学法则，并非机器人技术发展的法律，而是在其基础上建立机器人伦理规则的基石（фундамент）或者构成（конституция）。其具有丰富的伦理价值和道德意蕴，机器人学三法则产生的时代背景，正是在1920—1930年许多作品中都以反叛并毁灭自己的创造者的机器人为主题的氛围中，阿西莫夫极为厌恶在这类作品中所提出的警告<sup>①</sup>。

对于机器人学三法则的由来，有一段有趣的故事，据说在1940年12月23日，阿西莫夫与其朋友、*Astounding*杂志主编坎贝尔（Джон Кэмпбелл）就机器人故事进行一场交谈，坎贝尔表述了后来成为著名的机器人学三法则的东西。后来坎贝尔说，他只是将法则从阿西莫夫已经写好的东西中抽取出来。阿西莫夫自己总是把三法则作者地位的荣誉归于坎贝尔。<sup>②</sup>好几年之后阿西莫夫的另一朋友卡贝特（Рэндал Гаррет）将法则的作者归于两人，阿西莫夫欣然接受了这个表述。

一般来说，在阿西莫夫的作品中，机器人学三法则是逐渐出现的：最初的两篇机器人小说（1940年的《罗比》和1941年的《理性》）中没有明确提到三法则，但是在其中已经蕴含着（подразумевается）机器人具

<sup>①</sup> Азимов, Айзек. Улики/Мечты роботов. —М.: Эксмо, 2004. С. 142 – 169.

<sup>②</sup> David Langford, *Laws of Robotics/SFE: The Encyclopedia of Science Fiction*, online edition, 2011.; Айзек Азимов; Человек, который писал еще быстрее. Русская фантастика (1994). Проверено 14 января 2007. Архивировано 25 января 2012 года.