独角兽
法学精品
人工智能

机器人的话语权

Robotica:
Speech Rights and
Artificial Intelligence

〔美〕罗纳德·K

上海人民出版社

上海人民出版社

机器人的话语权

ROBOTICA:
SPEECH RIGHTS AND
ARTIFICIAL INTELLIGENCE

〔美〕罗纳德·K.L.柯林斯(Ronald K. L. Collins) 大卫·M.斯科弗(David M. Skover)/编
王黎黎 王琳琳/译

上海人民出版社

图书在版编目(CIP)数据

机器人的话语权/彭诚信主编;(美)罗纳德·K.L.柯林斯(Ronald K.L.Collins),(美)大卫·M.斯科弗(David M.Skover)编;王黎黎,王琳琳译.—上海:上海人民出版社,2019
书名原文:Robotica: Speech Rights and Artificial Intelligence
ISBN 978-7-208-15989-1

Ⅰ.①机… Ⅱ.①彭… ②罗… ③大… ④王… ⑤王… Ⅲ.①机器人-研究 Ⅳ.①TP242

中国版本图书馆 CIP 数据核字(2019)第 143246 号

策　　划 曹培雷　苏贻鸣
责任编辑 夏红梅
封面设计 孙　康

机器人的话语权
彭诚信 主编
[美]罗纳德·K.L.柯林斯　[美]大卫·M.斯科弗 编
王黎黎　王琳琳 译

出　　版 上海人民出版社
(200001　上海福建中路 193 号)
发　　行 上海人民出版社发行中心
印　　刷 常熟市新骅印刷有限公司
开　　本 635×965　1/16
印　　张 14.5
插　　页 4
字　　数 255,000
版　　次 2019 年 8 月第 1 版
印　　次 2019 年 8 月第 1 次印刷
ISBN 978-7-208-15989-1/D·3457
定　　价 58.00 元

在每一个通信技术的时代——无论是印刷、广播、电视还是互联网——都会有某种形式的政府审查，以规范媒体及其信息。今天，我们看到“机器语言”的现象被复杂的人工智能的发展所加强。罗纳德·K.L.柯林斯(Ronald K.L.Collins)和大卫·M.斯科弗(David M.Skover)认为，《第一修正案》必须为包括和保护机器人表达提供辩护和理由。机器人不是人类，也没有意图，这无关紧要；重要的是，人类认同机器人言论是有意义的。这是对发生在机器人和接收者的交流中的“无意图言论自由”的宪法认可。《机器人的话语权》是第一本针对这些目的展开法律论证的书。本书以法学和传播学学者、律师和言论自由活动人士为阅读对象，在法律与技术的结合点上探索新的重要问题和解决方案。

关 于 作 者

罗纳德 · K.L.柯林斯(Ronald K.L.Collins)是华盛顿大学法学院的哈罗德 · S.谢菲尔曼(Harold S.Shefelman)学者。在进入法学院之前，柯林斯在俄勒冈州最高法院担任汉斯 · A.林德(Hans A.Linde)法官的法律助理、是首席大法官沃伦 · 伯格(Warren Burger)手下的最高法院法官、华盛顿特区新闻博物馆第一修正案研究中心的学者。

柯林斯撰写了宪法摘要，提交给最高法院和其他联邦和州高级法院。除了与大卫 · 斯科弗合著本书之外，他还是《奥利弗 · 温德尔 · 霍姆斯：一个言论自由的读者》(2010)一书的编辑，以及《我们一定不要害怕自由》(2011)一书的合著者。他的最新的独著是《微妙的专制主义：弗洛伊德 · 艾布拉姆斯和第一修正案》(2013)。柯林斯是美国最高法院博客(SCOTUSblog)的图书编辑，每周写一篇博客(第一修正案新闻)，发表在同意意见(Concurring Opinion)网站上。

大卫 · M.斯科弗(David M.Skover)是西雅图大学法学院的弗雷德里克 · C.陶森德(Fredric C.Tausend)宪法学教授。他在联邦宪法、联邦管辖、大众传播理论和第一修正案等领域教授课程、写作和演讲。

斯科弗毕业于普林斯顿大学伍德罗威尔逊国际和国内事务学院。他在耶鲁法学院获得了法律学位，当时他是《耶鲁法杂志》的编辑。此后，他在康涅狄格州联邦地方法院和美国第二巡回上诉法院担任乔恩 · 奥纽曼(Jon O.Newman)法官的法律助理。除了与罗纳德 · 柯林斯合著本书外，他还与皮埃尔 · 施拉格(Pierre Schlag)合著了《法律推理策略》

(1986)。

柯林斯和斯科弗共同创作了《话语之死》(1996、2005)、《莱尼·布鲁斯的审判：美国偶像的衰落与崛起》(2002、2012)、《狂躁：愤怒和暴行的生活》(2013)、《异见：它在美国的意义》(2013)、《金钱至上：麦卡锡的决定》、《竞选财务法》和《第一修正案》(2014)、《法官：26条权谋的教训》(2017)。他们合著的许多学术文章在多种期刊上发表，包括《哈佛法评论》、《斯坦福法律评论》、《密歇根法律评论》和《最高法院评论》等。《莱尼·布鲁斯的审判：美国偶像的衰落与崛起》(修订版和扩充版)、《狂躁：愤怒和暴行的生活》、《异见：它在美国的意义》、《金钱至上：麦卡锡的决定》和《法官：26条权谋的教训》已经有了电子书形式。

主编序

彭诚信

前段时间，一篇名为《谨防法学研究的人工智能泡沫》的文章广泛流传并引发热议。其实，学术同行间或也会讨论我国人工智能法学研究是否真的存在学术泡沫。某种质疑或反思声音在社会上的出现，往往事出有因。2016年被认为是我国人工智能研究的元年，相关数据显示，仅仅一年的时间，2017年人工智能法学研究的论文就已井喷。但不可否认，人工智能法学研究在我国毕竟是一个崭新的学术热点，多数学者也是初涉该领域；更为关键的是，我国乃至世界范围内实际发生的人工智能案例又着实太少，这些原因导致当下人工智能学术研究缺乏针对性且不够深入。然而，我们也不应扩大这些不足，认为整个有关人工智能的学术领域都存在泡沫。

相反，只要我们承认互联网时代已经到来，人工智能时代的到来就不可避免。因为在互联网世界中产生的数据、信息，犹如虚拟世界中的阳光、空气与土壤，孕育出了人工智能的果实。当下人工智能已经渗入工业、商业、医疗、金融、交通、法律、军事等各个领域，型构了人类生活的人工智能图景，这些应用具体包括人脸、语音识别在内的智能防控系统，外科机器人在内的人工智能辅助医疗系统，无人机、无人车、无人艇在内的自动驾驶系统，各种智能生活工具、智慧城市系统的设立等。在这样的背景下，法律人如果回避人工智能的学术研究，便等于是对现实生活的无视。实际上，目前诸多企业对人工智能的研究与开发已远远走在理论研究者前头，这更应引发学界正视人工智能法学研究。

正是基于这一紧迫的时代情势，“独角兽法学精品 · 人工智能”第

二辑继续推介国外法学学术精品，期待通过持续努力，能够为人工智能的法学学术研究提供有益素材，为“谨防法学研究的人工智能泡沫”作出实际贡献。

一

本辑是“独角兽法学精品 · 人工智能”的第二辑，在第一辑出版三部译著的基础上，本辑又精选了三部著作进行翻译，分别是美国学者柯林斯(Ronald K.L.Collins)和斯科弗(David M.Skover)教授的《机器人的话语权》(王黎黎、王琳琳译)、以色列学者哈列维(Gabriel Hallevy)教授的《审判机器人》(陈萍译)以及英国学者赫里安(Robert Herian)教授的《批判区块链》(王延川、郭明龙译)。

《机器人的话语权》主要是围绕美国宪法《第一修正案》为什么必须包含并保护机器人表达提供辩护和理由。作者对通信技术及其引发的审查制度进行了历史性回顾，提出机器人表达所传输的是“实质性信息”，即使是机器人发送或者接收的信息，但只要这些信息对于被接收方而言是可识别的，那么这些信息就是交际性言论，而应被视为“言论”。在此基础上，通过“无意图言论自由”规则界定机器人表达具有“效用”价值，从而提出《第一修正案》能够包含并保护机器人的表达。较为难得的是，本书作者特意邀请的几位评论教授也作出了针对性评论，甚至是争论，如格林梅尔曼(James Grimmelmann)教授明确指出将机器人传输视为言论的观点有待商榷，甚至并不正确；诺顿(Helen Norton)教授认为作者提出的“效用”准则也值得质疑。作者对这些评论和质疑作出积极回应：本书关注的是机器人言论表达的潜力；要正确区分《第一修正案》可能包括的活动和受其保护的言论之间的区别；强调“效用”是一种保护机器人语言的概念框架；针对潜在危险，可以通过技术和法律进行功能性解决，同时不影响保护知识的生产者。因此，机

器人表达在许多情况下不仅需要被《第一修正案》所囊括，而且也需要被宪法所保护。相信作者与评论者的评议以及针对评议的回应，将有助于读者更为深入且有趣味地了解本书主题。

《审判机器人》是以色列奥诺学院法学院哈列维教授探索人工智能刑事责任的最新力作，并且是用中文在全球首发。哈列维教授是国际社会中较早关注人工智能刑法问题的法学专家，他的系列文章和相关著作在全球学术界已经产生非常广泛的影响。本书试图解决的问题是，随着人工智能在商业、工业、军事、医疗和个人领域的使用日益增多，如果人工智能系统对人类社会造成损害，现有的刑法制度该如何应对？ 哈列维教授的答案十分明确：在世界各国现有的刑法体系中，追究刑事责任都要求事实要素和心理要素；对于这两种要素的要求，人工智能都能够符合，因此其可以承担刑事责任。对人工智能的刑事处罚，也与自然人一样，涵盖死刑、自由刑、财产刑、社区服务、缓刑。他同时强调，人工智能实体承担刑事责任，并不减少涉案自然人或法人的刑事责任，根据不同情况，可以通过间接正犯、可能的后果责任机制等对其予以追责。据此，哈列维教授阐述了一个关于人工智能刑事责任的综合性法学成熟理论，从现有刑法中识别并选择出类似原则，提出针对多元情形下各种自主技术的刑事责任的具体思考模式，并通过列举人工智能现实应用场景中可能发生的犯罪案例，据其理论作出了相应解答。

《批判区块链》一书在肯定区块链颠覆性的基础上，对目前的区块链“生态系统”进行了多层次批判。作者认为区块链的应用并未惠及普通民众，而只是大企业赚取钱财的工具。从这个角度上而言，区块链偏离了设计者的初衷。国家对区块链这个新业态的发展，基本上持观望态度，目前针对区块链的规制模式亦因此主要表现为民间模式。由于缺乏国家力量的推动，区块链为全民服务这个目标难以实现。作者认为，为了实现“区块链向善”的目标，应该杜绝“区块链技术万能论”这种错误观点的炒作，加强对普通民众的区块链教育，同时建议政府提前介入，引导区块链走向促进社会福祉的道路上来。

整体来看，本辑既有从宪法话语权视角对人工智能体(机器人)言论问题的讨论，也有从刑法视角对人工智能体犯罪问题的研讨，更有从政府管制或政治经济学视角对区块链规制问题的审视。与第一辑集中于人工智能私法具体问题的讨论相较，本辑三部译著的立意更为宏大与高远，两辑相应和共同构成了人工智能法学研究的多维视角。这也从另一个侧面说明有关人工智能的研究已经渗透到更为广泛的法学领域，如果不是全部的话。

二

三部译著宏大理论视野中体现出来的观念与理论分歧，仍在提醒着学界切莫忽视有关数据、信息及人工智能等基础理论研究，因为所有争论在某种意义上都可归因于相关基础理论的模糊甚至缺失。

《机器人的话语权》是有关机器人话语权应否适用《第一修正案》的争论。其要点在于，在越来越多具有语音或语言功能的机器人中，其表达是法律上的言论，抑或仍为机器人处理和传输的数据？ 如果构成言论，那是谁的言论，是机器人自身，还是公司、公司员工、数据源、用户或其他对培训数据作出回应的用户的言论。总结说，核心的法律问题是，机器人能否作为独立的法律主体？ 其表达内容是言论，是算法，抑或是数据的表现形式？ 只有清楚了这些基本问题，才能更好地讨论机器人的话语是否适用《第一修正案》。

尽管哈列维教授在《审判机器人》一书中的观点明确而肯定，即人工智能体符合承担刑事责任的事实要素和心理要素，也可跟自然人一样，也可被处以死刑、自由刑、财产刑、缓刑、社区服务等各种刑罚；法律并可根据具体情况，依据间接正犯或其他可能的后果责任机制，令相关的涉案自然人或法人承担相应的刑事责任。但具体到机器人的犯罪构成与刑罚承担，还是有诸多问题需要深入讨论。例如，应如何判断机

器人犯罪事实构成中的行为要件以及故意或过失心理要素？是判断具体机器人的行为或心理，还是判断算法设计者或机器人生产者的行为或心理？当依据间接正犯理论判由机器人或相关自然人或法人分别承担刑事责任时，应如何确定机器人与相关自然人或法人的法律关系？对于诸如正当防卫、紧急避险等免责事由的判断，对人工智能系统应如何认定？所有这些疑问涉及的核心问题依然是，机器人能否以及如何作为法律主体？如何认定机器人的意志与行为？机器人与算法设计者、机器人生产者的关系如何？上述问题不仅涉及法律，而且也涉及伦理学、工程学等多学科知识。

《批判区块链》一书尽管形式上讨论的是政府对区块链的管制，甚至论及了政治与社会经济政策问题，但实质上，该书却触及更为深刻的有关数据与人工智能的核心基础理论。应该说，区块链(更广意义上的智能合约)当中所有的数据都应该是公用的，亦即区块链上的数据皆应为公共数据。在此基础上，区块链事后记录不能被更改或删除，唯此方能确保区块链的永恒不可变性，这被认为是区块链最令人满意的特性。区块链所谓的透明度创造和信任能力的培养都是基于该特性产生。但 2018 年 5 月欧盟《一般数据保护条例》(General Data Protection Regulation, GDPR)的颁布，尤其是其中个人数据删除权(被遗忘权)的赋予，与区块链的设计原则相冲突，并从根本上动摇了这一特性。因为在区块链环境中，其体系结构的设计目的就是要在技术上防止个人数据的删除或擦除。若承认并严格遵守个人数据删除权制度逻辑的话，则要建立实施解锁链(undoing chains)机制，区块链透明度的创造以及培养信任的能力便会从根本上发生动摇，其永恒不可变性自然也会因此消解，甚至可以说区块链的生存环境便不复存在。由此提出的更为核心的法律问题是：在区块链环境中谁来控制个人数据？个人数据本应属于谁、应由谁来控制？这是《一般数据保护条例》与区块链争论的关键问题。要厘清这些问题，就必须回到人工智能法学研究的基础问题上来，即须从数据、信息以及隐私的基础理论说起。

三

数据是一种资产或资源，有观点甚至认为，如果数据(尤其是在网络空间中)不让人(含企业法人、非法人组织等)自由利用的话，便是一种道德上的恶，因为它阻碍了数据以及人工智能产业的发展，也从根本上阻碍了互联网社会的发展。这种观点尽管凸显了网络数据的积极价值，但却忽视了任意使用数据(信息)的潜在危害，尤其是数据中包含个人隐私的话。包含个人隐私的数据若被他人任意使用，不仅会损害人的尊严，而且也会造成人在社会上更大的不自由；这不仅违反了法律，也是更深层道德意义的恶。因此，在法律上厘清数据、信息以及隐私的关系便尤为重要。

(一) 何为隐私?

我国《民法总则》同时规定了隐私和个人信息，但并未界分两者的具体内涵，由于这一立法格局在未来《民法典》中也几乎也不会发生变化，因此正视并科学界定隐私与个人信息的区分便成为法释义学上的一项重要任务。互联网时代的到来催生了社会对个人数据与信息流转的迫切需求，也倒逼法律对个人隐私范围作出相对明确的规定，从而为数据、人工智能产业的发展清除模糊区域、扫除相关障碍。

隐私在我国法中的应然内涵应当采取狭义理解，即主要是关涉自然人自然存在与社会存在的在自由与尊严方面不愿为他人所知的信息，如关涉性的取向与选择、基因等信息。既然我国《民法总则》同时规定了隐私与个人信息，那么隐私便不应被个人信息所包括。学界通常使用的敏感信息与非敏感信息等术语，并非严格规范意义上的法律术语。法律上的隐私与敏感信息的关系也远非清楚，因为若敏感信息仅是对《民法总则》中个人信息的分类，那这种分类本不应涵盖应然意义上的隐私内容；反过来，若敏感信息可包含隐私内容的话，那这种分类便混淆了

《民法总则》中隐私与个人信息相互独立的界限。这也是我国学界和实务界混淆隐私与信息内涵的其中一个具体表现。

隐私权的人格权定性无论是在理论界还是实务界，几乎没有任何争议。狭义的隐私内涵与严格的人格定性也决定了隐私不能自由交易和公开，哪怕权利人同意，也不能任意处分，因为它关涉人在社会上自由与尊严的基本存在样态，法律必须要严加保护。无论现实社会对数据的利用有多迫切，数据与人工智能产业的发展也不能突破法律底线：即自然人的隐私权利不容任何方式或理由予以侵害与妨害。

(二) 何为个人信息?

在对隐私采狭义理解的基础上，个人信息应是去除隐私之外作为独立保护客体的信息。对个人信息的这一界定尚有若干要点需要明确。第一，有关个人信息的法律属性是人格利益还是财产利益的争论，尽管有多种观点，但将其界定为人格利益属性还是相对更为合理。因为信息多属于人的社会属性，是人的社会存在形态的体现，而人格便是人的社会存在基础，即自由和尊严。把信息界定为人格利益，体现了对人的社会(包括虚拟世界)存在样态的尊重。个人信息的人格利益定性并不妨碍个人信息中包含财产属性。把隐私跟个人信息严格区分的主要目的之一，便是要让个人信息可以为他人利用。也就是说，信息主体可以将其个人信息行使商品化权或公开化权的方式为他人(含法人等主体)利用，只不过要征得信息主体的同意以及其他法律的外在限制。

第二，至于被界定为人格利益的个人信息如何被利用，现在讨论的核心问题是事前征得信息主体的同意，实践中发展出了所谓的“三重授权许可使用规则”。依据这一规则，开放平台方直接收集、使用用户数据需获得用户授权，第三方开发者通过开放平台 Open API 接口间接获得用户数据，需获得用户授权和平台方授权。然而，这种同意模式一则对信息控制者与使用者来说很不效率，而且信息主体的同意内容也不尽相同，为保证信息主体同意相同内容与提高效率，那也只能通过格式合同的方式。这无疑会大大增加信息控制者与使用者的成本，且在发生争

议时也难以起到应有的风险防范作用。在根本上，这种同意模式也并没有实现对信息主体应有的基本尊重与利益保护。信息主体或是不得不同意；或者即便同意也得不到应有的利益保护，甚至有时交出去的是个人信息，换来的却是伤害，如大数据杀熟、算法歧视、预测性识别等现象的存在。可见，一对一的事前同意规则并不符合网络社会发展的客观要求与应然逻辑。因此，需要予以强调的是，对于个人信息的利用，即便是以个人同意为前提，但也要设计出一种既能符合互联网社会要求，又能体现尊重与保护信息主体利益的现代而科学的同意方式。

第三，掣肘个人信息使用的深层矛盾不在于个人信息的定性(无论是否定性为人格利益，都可以让渡)，也不在于同意规则的设计(人类总能设计出更为理想的同意规则)，而在于如何让个人信息所有者在最低程度上不受到侵害，在中级程度上得到信息利用者所得利益的分配与分享，在终极意义上感受到对自由与尊严的尊重。经常有人说，个人信息在单个人手中并无价值，而只有通过收集、加工与整理等过程，其价值才得以彰显。但这并不能证成让信息所有者放弃其个人信息的正当性，一个人不看电视，并不能证成放在其家里的电视可被他人随意取走。上述观点更不能证成信息加工者任意、无偿取得他人信息的正当性，何况这会给信息加工者带来利益。尤其是当信息加工者通过大数据的整理，反过来又侵害无偿信息提供者时(如大数据杀熟现象的产生)，就更具有“恩将仇报”的意味。因此，如何设计出能让信息提供者切身感受到信息利用的制度红利，或许是解决当下个人信息使用现存矛盾的关键，如通过设立特定税收、基金或信托方式，均不失为一种新的探索途径，但关键是这些税收、基金、收益的使用目标与路径，能让信息提供者获得制度红利的反哺。在这个意义上，除了隐私，没有不可让渡不可利用的信息，关键是利益分配方式。这种利益分配(直接表现为个人信息的价格)未必是信息控制者、利用者与信息所有者之间谈定，而是应通过符合互联网社会发展的应然制度设计。我们有必要再次强调，让信息提供者感受到制度红利，未必是利用者与信息提供者之间通过合同买卖来实

现，或者本不应通过此种途径实现，而是可通过宏观的分配制度间接实现信息提供者利益的保护。

第四，对个人信息主体予以尊重的法律途径是赋予其特定权利，比如同意权、知情权、查询权、可携带权、删除权(被遗忘权)以及红利受益权等，其中有些权利名称(如红利受益权)主要是在描述意义而非在法律规范意义上使用。这些权利的具体实现可通过各种更为实效的符合互联网络思维的路径来实现，比如同意权的设计，即需要突破现行的一对一签约形式；知情权、可携带权的范围也要进行准确界定；删除权与区块链的矛盾也要从根本上解决；而红利受益与分配制度，更要通过多元的制度相配合等。

(三) 何为数据?

尽管《民法总则》也同时规定了个人信息与数据，但从个人角度来看，个人数据与个人信息难以实际区分，或者说两者在本质上应该是相同的。如果一定要从信息学技术意义上把两者区分开来，将数据(date)界定为以 0 和 1 二进制单元表示的信息，数据就是以适合通信、解释或处理的形式表现的可复译的信息，国际标准化组织(ISO)即采此定义；而信息即指在特定上下文中具有特定含义的关于特定对象(例如事实，事件，事物，过程或想法，包括概念)的知识。然而，这样界定的个人数据难以具有法律意义，具有法律意义的实为信息。当我们在法律内部谈论数据时，主要不应站在个人的立场。而现在在世界范围内，站在个人立场设计的相关法律文件，无论使用的是个人数据，还是个人信息，在本质上都是个人信息。

其实，数据概念主要是为信息主体之外的数据控制者、加工者与利用者所使用，对于这些主体而言，他们拥有的客体仅为数据。尽管在这些数据中也包含众多主体的个人信息，但由于数据经加工已对个人信息进行脱敏、加密处理，该信息已不再跟具体个人发生关联，所以可把他们整理、交易的客体称为数据。

数据主要体现为财产属性，可以为数据主体自由交易，为他人利

用。但数据毕竟在本质上是众多个人信息的整合或集合，只不过经过脱敏等技术使得信息与其主体无法或不能直接发生关联。因此，一定要尽力避免数据交易的法律风险，即一定不能泄露、侵害他人信息与隐私，否则就会有承担民事、行政乃至刑事责任的法律后果。这也决定了数据与个人信息的具体法律关系，必须要由法律专门作出相应规定，比如对信息加工、脱敏的具体要求，如何设计个人可携带权、知情权等权利类型与具体保护方式等。

厘清数据、信息与隐私等概念的基本内涵与法律界限，是讨论虚拟空间中各种法律关系的前提，也是讨论人工智能如何健康发展的前提。正是在此意义上，数据才是人工智能的阳光、氧气、土壤与食粮，是人工智能得以存在的前提。而数据若要成为人工智能的养分，那还必须通过光合作用(即算法)来完成。

四

目前，具有全面思考能力，能够自主从事创造性劳动的通用人工智能体尚未出现，更不要说超级人工智能体了。所以，我们讨论的还仅是单一功能的专用人工智能体，如人脸、语音识别功能、自动驾驶功能、单一医疗诊断功能等。算法也主要是依附于人工智能体所要发挥的具体功能来设计。算法的复杂程度取决于人工智能体所欲实现功能的精细度与准确度等因素。复杂算法的实现则取决于具体算法层次以及各个层次之间的相互关联。算法做到何种程度的层级划分以及实现何种程度的关联，人工智能体才具有深度学习能力，完全是技术问题。法律主要关心的问题是，一旦人工智能体具有深度学习能力，法律则极有可能确认其与自然人相似、相同或甚至超越自然人的意志能力，而将其作为法律主体对待。当人工智能体具有意志能力并有深度学习能力的时候，此时的算法已经摆脱了人的控制。当人工智能体摆脱人为设计的算法控制之

后，才是真正意义上的人工智能体，也将成为法律主体；当下的人工智能体还主要是表现为智能工具、智能产品，在法律上一般还是应作为客体对待。

在算法为王的大数据环境下，基于数据驱动的人工智能自动化决策常常表现出算法“黑箱”，算法解释和监管问题因而浮出水面。否则，内含算法黑箱的人工智能产品由于偏见势必对现实世界造成很大的价值观冲击，极端情况下甚至会引发新的社会治理矛盾与危机。问题是，如何能够实现算法的可解释性？

首要路径是算法透明。然而，对于专用智能体，要求设计者或研发者公布其算法是否具有正当性？ 如果法律强制算法公布，又当如何对算法进行审查？ 实质审查还是形式审查？ 其实，如同自然人的出生，谁能要求公布自然人的产生密码？ 又是如何能够做到必须公布？ 对于人类的产生主要有两大理论：一是上帝造人，另一个是进化论。无论遵循哪一观点，人类一旦产生，我们即不能了解彼此的想法，外人无从知悉我正在思考什么。尤其是当人工智能体有了深度学习能力之后，试图让算法透明更是无从实现。在此意义上，尽管在技术上可以进行各种克服黑箱的尝试，但人工智能体的算法黑箱在本质上已是一种客观存在。

客观存在的算法黑箱是否就无法控制？ 答案显然是否定的！ 就像自然人一样，尽管个个都有自主思考能力，但仍受法律的控制，途径就是课以法律责任。对于人工智能体也同样如此，即便存在算法黑箱，也可以通过课以法律责任予以控制。问题是如何设计出妥适的责任分配机制，是让人工智能体自己承担责任？ 还让其背后的研发者、生产者、销售者、使用者等主体承担责任？ 还是若干主体之间共同承担责任？ 若让人工智能体独立承担责任，此时人工智能体就像公司一样成为独立的法律上的人。当特定股东利用公司为不当行为时，法律利用“刺破公司面纱”制度否认公司人格，从而追究有责股东的责任。同样的，当人工智能体利用算法“黑箱”为侵害行为乃至犯罪行为时，法律也应否定人工智能体的人格而追究那些有意设计或生成恶算法的特定主体的责