

## 第四篇 法律篇：智能时代的公平正义

人工智能技术的不断发展，对现有的法律体系带来了冲击和挑战。当我们被一首首由人工智能创作的诗歌触发心灵的共鸣，当我们看到自动驾驶汽车行驶在公路上，当我们的生活由于陪伴机器人的存在而变得不再孤单时，我们也需要面对如何调整现有的法律制度来规范和促进未来人工智能的发展。法律习惯于对社会新技术的发展做出相对滞后的回应，但是在人工智能领域，我们是否需要做出一些具有前瞻性的立法布局以及如何布局，是全球各国都需要共同面对的法律难题。

## 第十八章 AI要怎么负责？

2016年5月7日，美国佛罗里达州一位名叫Joshua Brown的40岁男子开着一辆以自动驾驶模式行驶的特斯拉Model S在高速公路上行驶，全速撞到一辆正在垂直横穿高速的白色拖挂卡车，最终造成车毁人亡。大家普遍关注的问题是，既然是自动驾驶，那么发生事故后应当由谁来承担相应的法律责任呢？能否对AI或者自主系统加以问责呢？

### 传统责任理论的困境：旧瓶是否还能再装新酒？

法律责任的划分和承担是人工智能发展面临的首要法律挑战，其涉及如何确保人工智能和自主系统是可以被问责的。法律责任的设定，在于追究法律责任，保障有关主体的合法权利，维护法律所调整的社会关系和社会秩序。在特斯拉事件中，美国国家公路安全管理局（NHTSA）最终得出调查结果，特斯拉的自动驾驶模式设计并无明显缺陷。但对于自动驾驶事故的法律责任如何界定，NHTSA并没有给出明确的结论，NHTSA指出，其对于自动驾驶功能的可靠性监控还没有结束，并保留了在必要时再次介入调查的权利。

从传统责任理论来看，其根据主观过错在法律责任中的地位，将法律责任分为过错责任和无过错责任。其中，过错责任将“过错”作为责任的构成要件，而且是最终要件，无过错就无责任。过错责任是法律责任中最普遍的形式，并且是占据主导地位的法律責任，传统的侵权法中也主要以过错责任为原则。但是进入人工智能时代后，人工智能系统已经可以在不需要人类的操作和监督下独立完成部分工作，而机器自主性操作造成的损害如何来判断和划分其责任成为一大难题。特斯拉事件中，至少从目前的调查情况来看，驾驶员、汽车生产商都没有过错，但事故还是发生了，需要有人来承担责任，这种情况下如何对各方责任进行界定就陷入了困境。

鉴于关于人工智能责任划分和承担问题在实践中已经出现，特别是在自动驾驶和机器人的应用中，对责任划分问题提出了迫切需求，部分国家和地区开始了立法层面的探索，国际社会也就此问题开始了积极的探讨。

2016年12月，电气和电子工程师学会（IEEE）发布了《合伦理设计：利用人工智能和自主系统（AI/AS）最大化人类福祉的愿景》，提出的基本原则之二就是责任原则。其指出，为了解决过错问题，避免公众困惑，人工智能系统必须在程序层面具有可责性，证明其为什么以特定方式运作。

### 自动驾驶汽车领域的立法尝试

7月5日上午，在2017百度AI开发者大会上，百度创始人、董事长兼首席执行官李彦宏通过视频直播展示了一段自己乘坐公司研发的无人驾驶汽车的情景。视频中，李彦宏坐

在一辆红色的汽车的副驾驶座位上，视频中驾驶座位没有驾驶员。对此，北京交管部门发布情况通报称，正在积极开展调查核实，公安交管部门支持无人驾驶技术创新，但应当合法、安全、科学进行。很多人认为李彦宏的做法欠妥，但根据我国目前现行法律法规，并没有推出针对自动驾驶汽车的相关规定。

作为现阶段人工智能应用最为广泛的领域，目前全球自动驾驶立法正在不断推进，包括联合国、美国、德国、英国等在内的国际组织和国家正在积极修订原有法规或制定新的法律政策，为自动驾驶技术部署清除法律障碍，并取得了积极进展。

2016年3月23日，联合国关于道路交通管理的《维也纳道路交通公约》获得修正。这项修正案明确规定，在全面符合联合国车辆管理条例或者驾驶员可以选择关闭该技术的情况下，将驾驶车辆的职责交给自动驾驶技术可以被应用到交通运输当中，意味着包括美国在内的72个签约国可允许自动驾驶功能汽车在特定时间自动驾驶，为自动驾驶技术在交通运输中的应用清除了障碍。

美国和德国在自动驾驶领域的立法主要集中在责任界定方面。美国道路交通安全管理局于2013年发布《自动驾驶汽车的基本政策》，包括内华达州、加利福尼亚州、佛罗里达州、密歇根州在内的9个州也通过了自动驾驶汽车立法，对自动驾驶汽车测试事故的责任承担做了规定，即：车辆在被第三方改造为自动驾驶车辆后，测试过程中导致财产损失、人员伤亡的，车辆的原始制造商不对自动驾驶车辆的缺陷负责，除非有证据证明车辆在被改造成自动驾驶车辆前就已存在缺陷。例如谷歌用奔驰汽车进行测试，安全责任由谷歌来承担。

德国《道路交通法》规定，道路交通事故严格责任独立于车辆的自动化程度，即机动车持有人必须承担责任。但是，根据德国学者的预测，随着技术发展，这种责任会逐渐从驾驶员向（自动化驾驶系统的）生产商转移。德国立法机构2016年对德国《道路交通法》所规定的“驾驶员在车辆行驶过程中全程保持警惕”“驾驶员的手不能离开方向盘”等条文启动立法修正。2017年5月，德国通过自动驾驶汽车法案，为自动驾驶汽车路上测试扫清了障碍，其规定：第一，司机必须始终坐在方向盘后，以便在自动驾驶汽车请求时进行控制；第二，允许路上测试，司机可不参与驾驶行为（意即可以上网、发邮件等）；第三，安装“黑匣子”，记录驾驶活动；第四，明确司机和制造商的责任分配，即，司机参与驾驶的，依其注意义务和过错承担责任，否则制造商承担责任。

英国自动驾驶汽车中心（Centre for Connected and Autonomous Vehicles, CCAV）曾发布两份报告，对保险和产品责任提出建议：将强制性的机动车保险延伸到自动驾驶汽车以便将产品责任囊括进去；新的保险框架旨在保护自动驾驶汽车事故中的受害者，受害者将可以直接向汽车保险人请求赔偿，而保险人将有权向依据既有法律负有责任（比如产品责任）的主体进行追偿。

## 机器人法律责任的探索

随着智能机器人越来越广泛的应用，其责任界定问题也引发了各方的高度关注和重视。2016年8月，联合国教科文组织与世界科学知识与技术伦理委员会在《关于机器人伦理的初步草案报告》中对机器人的责任进行了探讨，提出了一个可行的解决办法，即：采

取责任分担的解决途径，让所有参与到机器人的发明、授权和分配过程中的人来分担责任。

欧盟在智能机器人责任立法方面也做出了积极的尝试。早在2015年1月，欧洲议会法律事务委员会（JURI）决定成立一个工作小组，专门研究与机器人和人工智能的发展相关的法律问题。2016年5月，法律事务委员会发布《就机器人民事法律规则向欧盟委员会提出立法建议的报告草案》，同年10月发布了《欧盟机器人民事法律规则》。在这些研究和报告的基础上，2017年2月16日，欧洲议会投票表决通过一份决议，提出了一些具体的立法建议，要求欧盟委员会就机器人和人工智能提出立法提案<sup>[4]</sup>，其中包括成立一个专门负责机器人和人工智能的欧盟机构，为智能机器人重构责任规则。

在法律事务委员会看来，如今的机器人已经具有自主性和认知特征，也即具有从经历中学习并独立自主地作出判断的能力，而且可以实质性调整其行为，从机器人的侵害行为中产生的法律责任由此成为一个重大问题。机器人的自主性越强，就越难将其当成是其他主体（比如制造商、所有人、使用者等）手中的简单工具，这反过来使得既有的责任规则开始变得不足，因而需要新的规则。

新的规则着眼于如何让一台机器为其行为的疏忽部分或者全部行为承担责任，结果就是解决机器人是否应当拥有法律地位这一问题将变得越来越迫切。最终，法律需要对机器人的本质问题作出回应，其是否应当被当成是自然人、法人、动物抑或物，或者法律应当为其创设新类型的法律主体，在权利、义务、责任承担等方面具有其自身的特性和内涵。

在目前的法律框架下，机器人自身不对因其行为或者疏忽而给第三方造成的损害承担责任。而且，既有责任规则要求机器人的行为或疏忽能够归因于制造商、所有人、使用者等特定法律主体，并且这些主体能够预见并避免机器人的加害行为。更进一步，关于危险物品的责任和产品责任可以让这些法律主体为机器人的行为承担严格责任。但是，如果机器人自主地作出决策，传统的责任规则就将不足以解决机器人的责任问题，因为传统的规则将可能不能确定责任方并让其作出赔偿。

此外，现有法律框架的缺点在合同责任方面更是显而易见的，因为机器人现在能够选择合同当事人，磋商合同条款，缔结合同并决定是否以及如何执行所达成的合同，这些现象使得传统的合同规则无法适用。在非合同责任方面，既有的产品责任规则仅能涵盖因机器人的制造缺陷而造成的损害，同时受害人必须能够证明存在实际损害、产品缺陷且缺陷与损害之间具有因果关系。但是，目前的法律框架无法完全涵盖新一代机器人所造成的损害，因为这些机器人将从自己变幻莫测的经历中自主学习，并且以独特且不可预见的方式与其所处环境进行交互。

## 构建一个结构合理的责任体系

人工智能的快速发展和应用确实给人类社会带来了诸多问题，但是我们仍然有理由相信法律制度能够在不阻碍创新的前提下，控制人工智能带来的公共危险。因此，如何构建一个结构合理的责任体系，对人工智能项目的设计者、生产者、销售者以及使用者等在内的主体责任义务进行清楚的界定变得十分重要。IEEE在《合伦理设计：利用人工智能和自主系统（AI/AS）最大化人类福祉的愿景》中对不同的主体在人工智能责任方面应当采

取的措施进行了详细阐述，指出：立法机构应当阐明人工系统开发过程中的职责、过错、责任、可责性等问题，以便于制造商和使用者知晓其权利和义务；人工智能设计者和开发者在必要时考虑使用群体的文化规范的多样性；利益相关方应当在人工智能及其影响超出了既有规范之外时一起制定新的规则；自主系统的生产商和使用者应当创建记录系统，记录核心参数。

---

<sup>[1]</sup> 在欧盟只有欧盟委员会有权提出立法提案，但欧盟并无义务遵守，不过如果其拒绝这么做，就必须陈述理由。



## 第十九章 隐私深处的忧虑

网络视频公司Netflix曾放出“经过匿名处理的”上亿条电影评分数据，仅仅保留了每个用户对电影的评分和评分的时间戳，希望通过竞赛的形式，找到更好的影片推荐算法。但是2009年，德州大学的两位研究人员，通过这些匿名数据与公开的IMDB数据做对比，成功地将匿名数据与具体的用户对应了起来，最终Netflix不得不取消了原计划于每年举行的竞赛。<sup>[1]</sup>Netflix的案例表明，大数据分析技术让人们的隐私无处可藏，所谓隐私保护，其实也不过是皇帝的新衣。

### 隐私与数据保护是AI核心议题

进入人工智能时代，随着大数据技术和智能技术的结合，政府和企业的决策越来越依赖大量的数据分析（政府经济、社会统计分析、企业商业营销），大规模的数据收集、分析和使用，使传统社会走向透明化，在万物互联、大数据和机器智能三者叠加后，人们或许将不再有隐私可言。

与此同时，商家一直在夸大大数据、人工智能给人类的生产、生活带来的极大便利，而用户本身也往往忽视了这些新技术新应用对隐私和个人数据带来的危害。当前，智能APP已经成为人们生活中必不可少的工具，这些APP在提供生活服务的同时会收集大量个人信息数据，从而给用户推送精准营销信息，但精准营销的一个潜在危险就是“精准诈骗”，诸多诈骗案件表明，这将会给个人人身和财产安全带来极大的损害。

如吴军博士所说，数据是人类建造文明的基石，大数据对机器智能的产生和发展具有决定性作用，但大数据分析可以了解到个人生活细节或者组织内部的各种信息，从而引发大家对隐私权的担忧。<sup>[2]</sup>英国《人工智能：未来决策的机遇和影响》报告指出，在为了分析的目的而使用公民的数据时，能否保护公民的数据及隐私，能否一视同仁地对待每个公民的数据，以及能否保证公民个人信息的完整，对于政府赢得公众的信任和保护好本国公民来说至关重要。进入人工智能时代，隐私与数据保护仍然是需要我们高度关注的核心议题。

### 全球隐私与数据保护立法不断升温

整体上来看，个人隐私与数据保护是国际社会长期以来重点关注的内容。自1973年第一部个人数据保护法《瑞典数据法》颁布以来，全球范围内掀起了个人信息保护立法的浪潮。美国于1974年制定了《隐私法》，规定了公共领域的个人信息保护规则；欧盟于1995年通过了《关于个人数据处理保护与自由流动指令（95/46/EC）》（简称1995年个人数据保护指令），各成员国随即将其转化为国内立法；韩国、日本、新加坡等国先后制定了个

个人信息保护法，确立了个人信息收集、使用以及跨境传输等基本规则。截至2016年12月，全球已经有110多个国家和地区制定了专门的个人信息保护法。

近年来，随着大数据、云计算以及人工智能新技术的快速发展和应用，给现有的个人信息保护法律制度带来了新的挑战，各国立法、修订法律的活动更加频繁。

欧盟1995年制定的《关于个人数据处理保护与自由流动指令》即欧盟区域内个人信息保护的基础性立法，欧盟各成员国依据该指令，分别出台了本国的个人信息保护法。然而日新月异的信息技术使得指令的主要原则及制度适用变得非常不确定，并导致欧盟各成员国对个人数据保护指令的理解与执行出现了较大的分歧。2012年1月25日，欧盟委员会发布了《有关“1995年个人数据保护指令”的立法建议》（简称《数据保护通用条例》），对1995年个人数据保护指令着手进行全面修订。2015年12月15日，欧洲议会、理事会、委员会三方机构在立法进程的最后阶段就欧盟数据保护改革达成一致。2016年4月14日，欧盟立法机构通过最终版本的条例。在新通过的条例中，欧盟加强了个人隐私和数据保护，其中关于用户画像等自动化决策的规定将会对基于大数据的互联网行业实践产生重大影响。即用户有权拒绝企业对其进行画像等自动化决策，而且用于用户画像的数据不能包括人种或者种族起源、政治观点、宗教或者哲学信仰、商会会员、基因、生物特征、健康状况、性取向等特殊类别的个人数据。

在通信和互联网领域，欧盟委员会于2017年1月10日宣布提议制定更严格的电子通信隐私监管法案《隐私与电子通信条例》（Regulation on Privacy and Electronic Communications, ePD），进一步加强对电子通信数据的保护。ePD增加了适用主体，规定其隐私保护规则将同样适用于新兴的电子通信服务提供者，比如WhatsApp、Facebook Messenger、Skype等，确保当前新兴的通信服务提供者与传统通信服务提供者能够为用户的隐私提供同等水平的保护。ePD扩大了保护的内容，通信内容和元数据（通话时间、位置等）也被包括在隐私保护的范畴之内，元数据中包括了高度隐私的内容，除用于计费等的的数据外，未经用户同意所有元数据都需要进行匿名化处理或删除。与此同时，ePD还规定了终端用户可以通过控制电子通信信息的发送和接收来保护个人的隐私安全，终端用户可以采取的多种救济措施，以及违反该条例将要承担的责任和将会受到的惩罚等其他重要内容。

此外，日本、韩国等国也对现有的个人信息保护立法展开修订工作。2016年3月22日，韩国通信委员会（KCC）对《信息通信网络的利用促进与信息保护等相关法》进行了大幅修订，进一步完善个人信息委托处理的规定，增加对个人信息保护相关负责人的要求，以及新增暴露的個人情報の删除和切断有关规定等。日本于2015年9月9日颁布《个人信息保护修正法》，也针对当前技术产业发展新增了匿名信息的规定、建立个人信息保护委员会、数据跨境转移等规定若干内容，新法还对敏感信息做出了新的限制，包括禁止在未经数据主体同意的情形下获取和提供敏感信息。

在已有的个人信息保护立法框架下，部分国家积极制定云计算、大数据等新业务的个人信息保护规则。法国个人信息保护机构——国家信息与自由委员会发布《云计算数据保护指南》，对云计算服务协议应当包含的因素和云计算的安全管理提出了建议。日本政府出台《云服务信息安全管理指南》，对云客户和云服务提供商在个人信息保护方面应当注意的事项做出了规定。日本总务省发布了《智能手机用户信息处理措施（草案）》，从保护智能手机用户个人隐私的角度，规定了智能手机用户信息保护措施。

近年来，我国个人信息保护立法活动也在不断推进，并取得了一定的成果。2012年全

全国人大常委会通过《关于加强网络信息保护的決定》，确立了个人信息保护的若干原则。2013年通过修订《消费者权益保护法》，对消费者个人信息保护做了相关规定。2009年、2015年先后通过刑法修正案七和修正案九，专门增加了出售或非法提供、窃取或者非法获取公民个人信息的犯罪及刑罚。2016年全国人大常委会通过《网络安全法》，总结了我国个人信息保护立法经验，针对实践中存在的突出问题，将近年来一些成熟的做法作为制度确定下来，并且确立了大数据时代收集、使用个人信息保护的基本规则，包括：合法正当，网络运营者收集使用个人信息必须出于正当目的，采用合法形式；知情同意，要求网络运营者公开隐私规则，获得用户同意；目的限制，网络运营者不得超范围收集、不得违法和违约收集；安全保密，网络运营者不得泄露毁损个人信息，要采取预防措施、补救措施防止个人信息事故；删除改正，网络运营者应当应个人要求删除违法、违约信息，改正有误信息。这些规则为大数据、人工智能背景下个人信息和数据的保护提供了依据与保障。

## 挑战与应对：匿名化技术的应用

人工智能时代，数据的收集、使用等各个环节都面临着新的风险。在数据收集环节，大规模的机器自动化地收集着成千上万的用戶数据，涉及个人姓名、性别、电话号码、电子邮箱、地理位置、家庭住址在内的方方面面的数据，这些数据海量收集形成对用户的全面追踪。在数据使用环节，大数据分析技术广泛使用，数据经挖掘能分析出深层信息，不仅可以识别出特定的个人，还能分析出个人的购物习惯、行踪轨迹等信息，进一步扩大了隐私暴露的风险。此外，在整个数据的生命周期中，由于黑客攻击、系统安全漏洞等原因，个人数据始终面临着被泄露的潜在安全风险。例如，2016年9月22日，全球互联网巨头雅虎证实至少5亿用戶账户信息在2014年遭人窃取，内容涉及用戶姓名、电子邮箱、电话号码、出生日期和部分登录密码。2016年12月14日，雅虎再次发布声明，宣布在2013年8月，未经授权的第三方盗取了超过10亿用戶的账户信息。

为更好地应对个人隐私和数据保护的挑战，欧盟法律事务委员会建议，在针对人工智能和机器人制定政策时，应当进一步完善设计保护隐私（Privacy by Design）、默认保护隐私（Privacy by Default）、知情同意、加密等概念的标准，规定当个人数据成为“流通物”使用时，在任何情况下都不得规避涉及隐私和数据保护的基本原则。当前，各国基本建立起了隐私和个人数据保护法律框架，为了进一步强化人工智能时代个人隐私和数据保护，立法中越来越强调对技术手段的运用，其中最重要的一项技术手段为“匿名化处理”。

匿名化是指将个人数据移除可识别个人信息的部分，并且通过这一方法，数据主体不会再被识别。匿名化技术发展的初衷主要是为了在数据利用的过程中，降低个人隐私风险。数据匿名化在计算机科学领域是方兴未艾的热门话题，自1997年美国学者Samarati和Sweeney提出k-anonymity匿名模型后，目前已发展出许多成熟的技术解决方案。相比于技术领域的长足进步，法律领域对于匿名化的关注才刚刚开始，并将其作为解决数据利用与个人数据保护的有效途径。欧盟在《数据保护通用条例》的引言中指出：“匿名化数据不属于个人数据，因此无须适用条例的相关要求，机构可以自由处理匿名化数据。”

数据匿名化不能仅仅被看作是脱离于数据保护法之外，避免管制负担的一种手段。应用它的初衷是降低个人数据泄露的隐私风险。采取匿名化措施的企业能够向用户提供更多的安全保障，让用户知晓其被收集的信息在用于大数据分析时，并没有使用可识别身份的



数据，因此增强用户对大数据应用的信任 and 安全感。为保证匿名化更多地发挥安全屏障作用，而不是作为数据滥用的挡箭牌，匿名化利用应当在合法合规的前提下开展。

2016年11月7日，《中华人民共和国网络安全法》正式通过并向社会公布，规定了类似匿名化的规定。该法第四十二条指出：“网络运营者不得泄露、篡改、毁损其收集的个人信息；未经被收集者同意，不得向他人提供个人信息。但是，经过处理无法识别特定个人且不能复原的除外。”这一规定，可以理解为对于个人数据匿名化利用，特别是匿名化后对外提供（交易）的情形提供了合法性。在此基础上，建议我国应当加快建立数据匿名化利用的法律规范体系，包括：明确匿名化数据的法律概念和认定标准，强调数据不再具有身份可识别性；引入隐私风险评估机制，鼓励企业基于个案在内部实施数据匿名化的风险评估，并基于评估结果，适时调整匿名化策略；利用合同规范、技术保障等多重工具实现数据的真正匿名化；建立数据匿名化的事前、事中、事后规范体系。

## 立法动态调整的方向

在现有隐私和个人数据保护法律体系下，需要针对人工智能发展带来的影响对相关制度进行动态调整。在个人信息（数据）界定方面，人工智能发展和大数据技术的使用打破了个人信息的稳定性，传统语境下非个人信息常常可以变成个人信息，立法技术上如何对个人信息进行界定需要重点关注。在个人数据权利方面，被遗忘权、携带权等新型数据权利已经引起了各方高度重视，欧盟已经进行了立法的尝试，但由于其实践经验尚待验证，且对产业发展和创新存在巨大阻碍，需要在评估其立法价值的基础上权衡是否纳入到未来个人信息保护立法之中。与此同时，人工智能的发展是一项全球化的进程，对数据跨境流动有着极大的需求，如何在确保个人数据安全和数据跨境流动之间取得平衡，也是未来立法需要解决的一个难题。此外，还有技术手段、数据泄露通知等其他相关制度也需要在立法中予以考量。

---

[1] <http://mt.sohu.com/20160620/n455281302.shtml>. [2017-01-01].

[2] 吴军.智能时代：大数据与智能革命重新定义未来.北京：中信出版社，2016.

## 第二十章 看不见的非正义

人工智能在影响人们的生活——网上的和现实世界中的生活。算法将人们在网络世界中的上网习惯、购物记录、GPS位置数据等各种网上足迹和活动，转变为对人们的各种打分和预测。这些打分和预测进而左右影响人们生活中的各种决策工作，其中的歧视和不公平由此成为一个显著的问题，无论人们是否意识到歧视的存在。

以大数据、机器学习、人工智能、算法等为核心的自动决策系统的应用日益广泛，从购物推荐、个性化内容推荐、精准广告，到贷款评估、保险评估、雇员评估，再到司法程序中的犯罪风险评估，越来越多的决策工作为机器、算法和人工智能所取代，认为算法可以为人类社会中的各种事务和决策工作带来完全的客观性。然而，这不过是妄想，是一厢情愿。无论如何，算法的设计都是编程人员的主观选择和判断，他们是否可以不偏不倚地将既有的法律或者道德规则原封不动地编写进程序，是值得怀疑的。算法歧视

（Algorithmic Bias）由此成为一个需要正视的问题。规则代码化带来的不透明、不准确、不公平、难以审查等问题，需要认真思考和研究。

### 人工智能算法决策日益盛行

网络的存在或者说数字存在，日益受到算法左右。如今，在网络空间，算法可以决定你看到什么新闻，听到什么歌曲，看到哪个好友的动态，看到什么类型的广告；可以决定谁得到贷款，谁得到工作，谁获得假释，谁拿到救助金，诸如此类。当然，基于算法、大数据、数据挖掘、机器学习等技术的人工智能决策不局限于解决信息过载这一难题的个性化推荐。当利用人工智能系统对犯罪嫌疑人进行犯罪风险评估，算法可以影响其刑罚；当自动驾驶汽车面临道德抉择的两难困境，算法可以决定牺牲哪一方；当将人工智能技术应用于武器系统，算法可以决定攻击目标。其中存在一个不容忽视的问题：当将本该由人类负担的决策工作委托给人工智能系统时，算法能否做到不偏不倚？如何确保公平之实现？

### 算法默认是公平的吗？

长久以来，人们对计算机技术存在一个广为人知的误解：算法决策倾向于是公平的，因为数学关乎方程，而非肤色。人类决策受到诸多有意或者无意的偏见以及信息不充分等因素影响，可能影响结果的公正性，所以存在一种利用数学方法将人类社会事务量化、客观化的思潮，Fred Benenson将这种对数据的崇拜称之为数学清洗（Mathwashing），就是说，利用算法、模型、机器学习等数学方法重塑一个更加客观的现实世界。《人类简史》一书的作者尤瓦尔·赫拉利将之称为“数据宗教”，对数据的使用未来将成为一切决策工作的基础，从垃圾邮件过滤、信用卡欺诈检测、搜索引擎、热点新闻趋势到广告、保险或者贷款资质、信用评分，大数据驱动的机器学习和人工智能介入并影响越来越多的决策工作，认为大数据、算法等可以消除决策程序中的人类偏见。

但是，在自主决策系统越来越流行的今天，有几个问题需要预先回答：第一，公平可以量化、形式化吗？可以被翻译成操作性的算法吗？第二，公平被量化为计算问题会带来风险吗？第三，如果公平是机器学习和人工智能的目标，谁来决定公平的考量因素？第四，如何让算法、机器学习和人工智能具有公平理念，自主意识到数据挖掘和处理中的歧视问题？

大数据应用日益广泛，回应这些问题极为必要。首先，公平是一个模糊的概念，法律上的公平被翻译成算法公平可能存在困难，但在犯罪侦查、社会治安、刑事司法程序中，基于大数据的人工智能系统正在将公平问题算法化，包括在犯罪嫌疑人搜寻、社会治安维护、量刑等诸多方面。其次，公平被量化、被算法化可能带来歧视问题。美国FTC在2016年1月发布的《大数据：包容性工具抑或排斥性工具？》特别关注大数据中的歧视和偏见问题。对于消费者，一方面要确保公平机会法律得到有效执行，另一方面应防止大数据分析中采取歧视等不公平行为；对于企业，FTC建议企业考察以下问题：数据集是否具有代表性？所使用的数据模型是否会导致偏见？基于大数据进行预测的准确性如何？对大数据的依赖是否会导致道德或者公平性问题？

欧盟同样关心大数据和算法中的歧视问题，欧盟数据保护委员会2015年11月发布的《应对大数据挑战：呼吁通过设计和可责性实现透明性、用户控制及数据保护》

（Meeting the Challenges of Big Data: a Call for Transparency, User Control, Data Protection by Design and Accountability），警惕人们重视大数据对穷人或者弱势群体的歧视，并提出是否可以让机器代替人类来做道德、法律等判断的问题，其实就是公平能否算法化的问题。最后，当利用犯罪风险评估软件对犯罪嫌疑人进行评估，决定司法判决结果的就不再是规则，而是代码。但当编程人员将既定规则写进代码时，不可避免地要对这些规则进行调整，但公众、官员以及法官并不知晓，无从审查嵌入到自主决策系统中的规则的透明性、可责性以及准确性。

显然，算法的好坏取决于所使用的数据的好坏。比如，如果拿一个个体吃的食物来评估其犯罪风险，那必然会得到很荒谬的结果。而且，数据在很多方面常常是不完美的，这使得算法继承了人类决策者的种种偏见。此外，数据可能仅仅反映出更大的社会范围内持续存在着的歧视。当然，数据挖掘可能意外发现一些有用的规律，而这些规律其实是关于排斥和不平等的既有模式。不加深思熟虑就依赖算法、数据挖掘等技术可能排斥弱势群体参与社会事务。更糟糕的是，歧视在很多情况下都是算法的副产品，是算法的一个难以预料的、无意识的属性，而非编程人员有意识的选择，更增加了识别问题根源或者解释问题的难度。因此，在自主决策系统应用日益广泛的互联网时代，人们需要摒弃算法本质上是公平的误解，考虑如何通过设计确保算法和人工智能系统的公平性，因为很多歧视来源于产品设计。

## 算法决策可能暗藏歧视

算法决策在很多时候其实就是一种预测，用过去的的数据预测未来的趋势。算法模型和数据输入决定着预测的结果。因此，这两个要素也就成为了算法歧视的主要来源。

一方面，算法在本质上是“以数学方式或者计算机代码表达的意见”，包括其设计、目

的、成功标准、数据使用等都是设计者、开发者的主观选择，他们可能将自己的偏见嵌入算法系统。另一方面，数据的有效性、准确性，也会影响整个算法决策和预测的准确性。比如，数据是社会现实的反映，训练数据本身可能是歧视性的，用这样的数据训练出来的AI系统自然也会带上歧视的影子；再比如，数据可能是不正确、不完整或者过时的，带来所谓的“垃圾进，垃圾出”的现象；更进一步，如果一个AI系统依赖多数学习，自然不能兼容少数族裔的利益。此外，算法歧视可能是具有自我学习和适应能力的算法在交互过程中习得的，AI系统在与现实世界交互过程中，可能无法区别什么是歧视，什么不是。

偏见也可能是机器学习的结果。比如一个甄别错误姓名的机器学习模型，如果某个姓是极为独特的，那么包含这个姓的姓名为假的概率就很高。但是这可能造成对少数民族的歧视，因为他们的姓可能本来就不同于普通的姓氏。当谷歌搜索“学习到”搜索奥巴马的人希望在日后的搜索中看到更多关于奥巴马的新闻，搜索罗姆尼的人希望在日后的搜索中看到更少关于奥巴马的新闻，那也是从机器学习过程中产生的偏见。

最后，算法倾向于将歧视固化或者放大，使歧视自我长存于整个算法里面。奥威尔在他的政治小说《1984》中写过一句很著名的话：“谁掌握过去，谁就掌握未来；谁掌握现在，谁就掌握过去。”这句话其实也可以用来类比算法歧视。归根到底，算法决策是在用过去预测未来，而过去的歧视可能会在算法中得到巩固并在未来得到加强，因为错误的输入形成的错误输出作为反馈，进一步加深了错误。最终，算法决策不仅仅会将过去的歧视做法代码化，而且会创造自己的现实，形成一个“自我实现的歧视性反馈循环”。因为如果用过去的不准确或者有偏见的数据去训练算法，出来的结果肯定也是有偏见的；然后再用这一输出产生的新数据对系统进行反馈，就会使偏见得到巩固，最终可能让算法来创造现实。包括预测性警务、犯罪风险评估等等都存在类似的问题。所以，算法决策其实缺乏对未来的想象力，而人类社会的进步需要这样的想象力。

## 算法歧视不容忽视

互联网上的算法歧视早已有之，并不鲜见。图像识别软件犯过种族主义大错，比如，谷歌公司的图片软件曾错将黑人的照片标记为“大猩猩”，Flickr的自动标记系统亦曾错将黑人的照片标记为“猿猴”或者“动物”。2016年3月23日，微软公司的人工智能聊天机器人Tay上线。出乎意料的是，Tay一开始和网民聊天，就被“教坏”了，成为了一个集反犹太人、性别歧视、种族歧视等于一身的“不良少女”。于是，上线不到一天，Tay就被微软公司紧急下线了。

互联网上的算法歧视问题早已引起人们注意。研究表明，在谷歌搜索中，相比搜索白人的名字，搜索黑人的名字更容易出现暗示具有犯罪历史的广告；在谷歌的广告服务中，男性会比女性看到更多高薪招聘广告，当然，这可能和在线广告市场中固有的歧视问题有关，广告主可能更希望将特定广告投放给特定人群。此外，非营利组织ProPublica研究发现，虽然亚马逊公司宣称其“致力于成为地球上最以消费者为中心的公司”，但其购物推荐系统却一直偏袒自己及其合作伙伴的商品，即使其他卖家的商品的价格更低。而且，在购物比价服务中，亚马逊公司隐瞒了自己及其合作伙伴的商品的运费，导致消费者不能得到公正的比价结果。

当人工智能用在应聘者评估上，可能引发雇佣歧视。如今，在医疗方面，人工智能可



以在病症出现前几个月甚至几年就预测到病症的发生。当人工智能在对应聘者进行评估时，如果可以预测到该应聘者未来将会怀孕或者患上抑郁症，并将其排除在外，这将造成严重的雇佣歧视。伊隆·马斯克警告道，对于人工智能，如果发展不当，可能就是在“召唤恶魔”。当把包括道德决策在内的越来越多的决策工作委托给算法和人工智能时，人们不得不深思，算法和人工智能未来会不会成为人的自由意志的主宰，会不会成为人类道德准则的最终发言人。

## 犯罪风险评估中的歧视： 法官和犯罪风险评估软件哪个更靠谱？

人们常说，犯罪嫌疑人遭受什么样的刑罚，取决于法官早餐吃什么。刑罚和定罪是两回事。确定犯罪嫌疑人所应遭受的刑罚，属于法官自由裁量权之范围。法律形式主义认为，法官以理性、机械、深思熟虑的方式将法律推理应用于案件事实，法官在量刑时受到诸多规则和指引约束。法律现实主义则认为，法律推理的理性适用并不能充分解释法官的判决，以及影响司法判决的心理、政治、社会等因素。法官在饿着肚子时更加严厉，倾向于给犯罪嫌疑人判处更重的刑罚。一项实证研究表明，司法正义取决于法官早餐吃什么，在用餐之前，法官做出有利判决（假释）的比例从约65%下跌到0；在用餐之后，法官做出有利判决（假释）的比例又会急剧上升到约65%。

正是由于法官在量刑时常常受到诸多非法律的外在因素影响，基于大数据、数据挖掘、人工智能等技术的犯罪风险评估系统开始大行其道。Northpointe公司开发的犯罪风险评估算法COMPAS对犯罪嫌疑人的再犯风险进行评估，并给出一个再犯风险分数，法官可以据此决定犯罪嫌疑人所应遭受的刑罚。非营利组织ProPublica研究发现，这一算法系统性地歧视了黑人，白人更多地被错误地评估为低犯罪风险，而黑人被错误地评估为高犯罪风险的概率是白人的两倍。通过跟踪调查7000多名犯罪嫌疑人，ProPublica发现，COMPAS给出的再犯风险分数在预测未来犯罪方面非常不可靠，在被预测为未来会犯暴力犯罪的犯罪嫌疑人中，仅有20%的犯罪嫌疑人后来确实再次实施暴力犯罪。综合来看，这一算法并不比掷硬币准确多少。

犯罪风险评估系统是一个“黑箱”，它如何得出结论，人们无从知晓，开发它的公司又拒绝对簿公堂，称算法是其私人财产，在缺乏必要的问责机制的情况下，无法矫正的算法歧视对刑事司法正义而言，就是一种嘲讽。Northpointe公司曾向ProPublica披露说其犯罪风险评估算法会考虑受教育水平、工作等诸多因素，但未披露具体算式，认为是其私人财产。所以人们无从知晓Northpointe公司是否将美国社会中固有的种族歧视问题编写进其算法。比如，即使集体统计数据显示黑人比白人更容易犯罪，将这一集体统计数据应用于黑人个体是否妥当？再比如，一直存在所谓的“天生犯罪人”理论，认为犯罪与否和一个个体的长相、基因等生理特征有关，在数据挖掘中考虑这些数据是否妥当？为了确保公平，犯罪风险评估算法在进行数据挖掘时可以使用哪些数据？更重要的是，是否可以依据秘密信息以及由此产生的犯罪风险分数对犯罪嫌疑人进行判刑？所有这些问题都需要认真对待，否则利用人工智能系统对犯罪嫌疑人进行打分、计算刑期等，就可能带来意想不到的系统性歧视。美国国会正在推动《量刑改革法案》（Sentencing Reform Bill），将引入“犯罪风险得分”，并据此对犯罪嫌疑人进行量刑、减刑等，如何通过有效的机制在刑事司法程序中避免机器歧视并在出现机器歧视、不公正时进行问责或者纠正，显得尤为重要。



## 人工智能决策三大问题：公平、透明性和可责性

有些歧视或许无关紧要，但在涉及信用评估、犯罪风险评估、雇佣评估等重大活动时，人工智能决策的结果将影响甚至决定贷款额度、刑罚选择、雇用与否，这时候歧视就不再是无足轻重的。如今，在医疗方面，人工智能系统基于大数据、数据挖掘等技术，可以对患者进行预测式诊断，甚至可以在患者病发前数月甚至数年就预测到这一事实，这为精准医疗和预防式医疗提供了可能性。但是，如果将这一技术应用于雇员能力评估，使其掌握是否雇用某一特定个体的生杀大权，具有强大预测功能的人工智能系统可能将未来会患上抑郁症等疾病或者怀孕的那些人系统性地排除在外。因此需要重视算法的公平、透明性和可责性这三大问题。

### 作为“黑箱”的算法的透明化困境

算法的公平性是一个问题，算法的不透明性更是一个问题。人们质疑自主决策系统，主要是因为这一系统一般仅仅输出一个数字，比如信用分数或者犯罪风险分数，而未提供做出这一决策所依据的材料和理由。传统上，法官在做出判决之前，需要进行充分的说理和论证，这些都是公众可以审阅的。但是，自主决策系统并不如此运作，普遍人根本无法理解其算法的原理和机制，因为自主决策系统常常是在算法这一“黑箱”中做出的，不透明性问题由此产生。Jenna Burrell在其论文《机器如何“思考”：理解机器学习算法中的不透明性》（How the Machine“Thinks”: Understanding Opacity in Machine Learning Algorithms）中论述了三种形式的透明性：因公司商业秘密或者国家秘密而产生的不透明性，因技术文盲而产生的不透明性，以及从机器学习算法的特征和要求将它们有效适用的测量中产生的不透明性。因此，在需要质疑自主决策系统的结果时，比如希望在法庭上挑战算法决策的合理性或者公平性，如何解释算法和机器学习就成了一大难题。这种不透明性使得人们很难了解算法的内在工作机制，尤其是对一个不懂计算机技术的外行而言。

### 如何向算法问责？

如果人们不满意政府的行为，可以提起行政诉讼，如果不满意法官的判决，可以提起上诉，正当程序（Due Process）确保这些决策行为可以得到某种程度的审查。但是，如果人们对算法决策的结果不满意，是否可以对算法进行司法审查呢？在算法决定一切的时代，对算法进行审查是极为必要的。但是，需要解决两个问题。第一，如果算法、模型等可以被直接审查，人们需要审查什么？对于技术文盲而言，审查算法是一件极为困难的事。第二，人们如何判断算法是否遵守既有的法律政策？第三，在缺乏透明性的情况下，如何对算法进行审查？如前所述，算法的不透明性是一个普遍的问题，因为企业可以对算法主张商业秘密或者私人财产。在这种情况下，对算法进行审查可能是很困难的。此外，从成本-效益分析的角度来看，解密算法从而使之透明化需要付出非常大的代价，可能远远超出所能获得的效益。此时，人们只能尝试对不透明的算法进行审查，但这未必能得到一个公平的结果。

## 构建技术公平规则，通过设计实现公平

人类社会中的法律规则、制度以及司法决策行为受到程序正义和正当程序约束。但是，各种规则比如征信规则、量刑规则、保险规则等正被写进程序当中，被代码化。然而，编程人员可能并不知道公平的技术内涵，也缺乏一些必要的技术公平规则指引他们的程序设计。对于诸如行政机构等做出的外在决策行为，人们建立了正当程序予以约束。对于机器做出的秘密决策行为，是否需要受到正当程序约束呢？也许，正如Danielle Keats Citron在其论文《技术正当程序》（Technological Due Process）中所呼吁的那样，对于关乎个体权益的自主决策系统、算法和人工智能，考虑到算法和代码，而非规则，日益决定各种决策工作的结果，人们需要提前构建技术公平规则，通过设计保障公平之实现，并且需要技术正当程序，来加强自主决策系统中的透明性、可责性以及被写进代码中的规则的准确性。而这一切，仅仅依靠技术人员是无法达成的。

在政府层面，为了削弱以至避免人工智能算法歧视，美国白宫人工智能报告将“理解并解决人工智能的道德、法律和社会影响”列入国家人工智能战略，并建议AI从业者和学生都能接受伦理培训。英国下议院科学和技术委员会呼吁成立一个专门的人工智能委员会，对人工智能当前以及未来发展中的社会、伦理和法律影响进行研究。

在行业层面，谷歌作为业界代表，在机器学习中提出“机会平等”这一概念，以避免基于一套敏感属性的歧视。Matthew Joseph等人在论文《罗尔斯式的公平之于机器学习》中基于罗尔斯的“公平的机会平等”理论，引入“歧视指数”的概念，提出了如何设计“公平的”算法的方法。无论如何，在人工智能日益代替人类进行各种决策的时代，设计出验证、证实、知情同意、透明性、可责性、救济、责任等方面的机制，对于削弱或者避免算法歧视、确保公平正义，是至关重要的。

## 第二十一章 作者之死

向着城市的灯守着我

咬破了冷静的思想

你的眼睛里闪动

无人知道的地方

——微软小冰

这首出自微软研发的人工智能小冰创作的诗歌，你觉得如何？是否有些人类艺术的气息？有人在网上评价小冰创作的诗歌：“梯子很长，还远远够不到月亮。”但也有人认为这首诗歌具有人类创作的灵魂。当然对艺术作品的评断向来没有统一的答案，可谓“仁者见仁，智者见智”。虽然如此，在人工智能研究的过程中，些许的微光都可以让研究者无比兴奋。

2017年5月，微软在北京发布了人工智能小冰的诗集《阳光失去了玻璃》，这部诗集里包含139首现代诗，全部是小冰的创作。微软的技术人员称，小冰学习了1920年以来的500多位诗人的现代诗，经过了上万次的训练，同时其写作过程中的思维过程与人类相似，也要经过诱发源、创作本体、创作过程和创作成果训练等步骤。目前，小冰已经拥有了超过1亿的人类用户，进行了300亿次的对话。而这并非是人类第一次走进人类的生活。

早在上个世纪，人工智能就已经开始深入到被人类视为圣殿的艺术领域。1956年，美国作曲家Lejaren Hiller与数学家Leonard Issacson合作，首次创作了计算机音乐《伊里阿克组曲》；美国加州大学哈罗德·科恩教授研究出可以创作出具有独特风格的绘画软件“Aaron”。此外，还有被称为“电脑小说家”的软件Brutus可以在15秒内创作出短篇故事，人类都无法分辨其是人类创作还是机器创作而成。<sup>[1]</sup>如果人工智能创作的内容可以获得法律的保护，是否意味着我们人类引以为傲的最后的艺术圣殿也不复存在，人类创作的艺术是否会被人工智能取代还是我们的艺术殿堂会出现更多的璀璨明珠呢？

## 人工智能创作物受版权法保护吗？

微软小冰诗集的出版，撬动了一个新概念的诞生——“人工智能创造”（AI Creation），并宣布了小冰拥有创造力，提出人工智能创造的三原则：第一，人工智能的创造主体，必须是同时具有IQ和EQ的综合体，而不仅仅是IQ。第二，人工智能创作的作品，必须能成为具有独立知识产权的作品，而不仅仅是某种技术中间状态的成果。第三，人工智能的创作过程中，必须对应人类某种富有创造力的行为，而不是对人类劳动的简单替代。人类的艺术创作过程是将自己原有的知识体系在新的环境下激发，也就是我们常说

的触景生情。人类的触景生情其实就是生物学中的人类大脑通过算法把人类的听觉、视觉、记忆力等与特定情景相联系，从而产生出相应的创作成果。

## 人工智能是否具有独立的智力创作能力？

要回答人工智能创作成果是否应该受法律保护这个问题，很多人不禁想到了我们人类认为的没有独立意识的动物。早在2011年，印度尼西亚的一群猴子拿着英国摄影师David Slate的相机拍了一些照片，包括一张自拍照，这张照片被收录到了维基资源共享图库中，摄影师认为猴子在拍摄的过程中自己制造拍摄场景，并把相机放到了脚架上，可以认为是有选择的拍摄过程，这张照片的版权应该归属于猴子。但是从目前各国的法律规定来看，还没有一个国家认为动物可以成为版权所有者，主要原因在于动物和人工智能在目前来看还不具有独立的智力创造能力。

2016年，欧盟法律事务委员会建议欧盟委员会就“与软硬件标准、代码有关的知识产权”提出一个更平衡的路径，以便在保护创新的同时，促进创新对于计算机或者机器人创作的作品被纳入到版权法保护的范畴的可能性。该委员会提出界定人工智能“独立的智力创造”的标准，以便明确版权归属。此外，电气和电子工程师学会（IEEE）在其标准文件草案《合伦理设计：利用人工智能和自主系统（AI/AS）最大化人类福祉的愿景》中也提出，应对知识产权领域目前的法规进行审查，以便明确是否需要人工智能参与创作的作品保护做出修订。其中的基本原则为：如果人工智能依靠人类的交互而实现新内容，那么使用人工智能的人应作为作者或发明者，受到与未借助人工智能进行的创作相同的知识产权保护。但是从目前各国的立法和实践来看，如何判断人工智能是否具有独立的“智力创造”能力需要我们不断探索答案。

## 人工智能创作的成果是否满足独创性保护要求？

为了回答人工智能创作成果的独创性高低问题，不妨让我们先来对比两首诗，看看你是否能区分出哪首出自人工智能的创作。

### 第一首

一夜秋凉雨湿衣，西窗独坐对夕辉。

湖波荡漾千山色，山鸟徘徊万籁微。

### 第二首

荻花风里桂花浮，恨竹生云翠欲流。

谁拂半湖新镜面，飞来烟雨暮天愁。

读过这两首诗之后，想必大家心里一定有了各自的答案。正确的答案是第一首是人工



智能创作的诗歌，第二首是宋代的葛绍体创作的诗歌《秋夕湖上》。在20世纪70年代之前，就有人试图将在诗歌中高频出现的词语整理出来后，随机挑选出一些词语拼凑成诗歌，结果大家也可以想见自然是逻辑混乱，很难称之为诗歌。之后，随着深度学习的不断深入和发展，人工智能写出的诗歌也越来越具有文学气息，甚至可以和人类创作的诗歌相媲美。2016年清华大学的语音和语言实验中心（CSLT）在其网站上宣布他们的写诗机器人“薇薇”经过社科院等唐诗专家的评定，通过了“图灵测试”，这也就意味着人类无法通过语言对话辨识出究竟哪首诗歌出自人工智能，哪首出自于人类之手。<sup>[2]</sup>

目前的写诗机器人其创作的过程是基于RNN语言模型的方法，将诗歌的整体内容，作为训练材料输送给RNN语言模型进行训练。训练完成之后，根据那些初始的内容，按照诗歌语言模型输出的概率分布进行采样得到下个词汇，之后重复该过程就产生了一首完整的诗歌。<sup>[3]</sup>我们法律上所保护的作品必须是具有独创性的表达，也就是说要求这些获得法律保护的作品需要满足最低限度的独创性的要求。不能是简单的排列组合，要体现作者独特的选择、编排。如果像早期诗歌创作那样，把各种词语在高频词库中随机选择之后编排在一起创作出诗歌，因没有加入智力创作的过程，很难被认为是应当受到法律保护的作品。

但是，写诗机器人的出现，其创作的过程是自我深度学习的成果的体现，又具有一定的文学性和艺术性，此时我们法律在何种情况下以及在哪些成果符合怎样的标准时可以将这些成果认定为法律所保护的“作品”？人工智能创作成果的保护标准与人类创作的作品是否应该适用相同的“独创性”标准，还是应该创造出不同的保护标准予以保护呢？如果人工智能创作的成果可以获得法律的保护成为作品，那么该作品的版权应该属于人工智能本身还是人类主体？作品的保护期限是否也应该有别于现有的人类作品呢？

从现有的作品保护期限来看，一般作品的保护期限是作者生前加去世后50年。人工智能创作的成果如果具有独创性可以获得法律的保护，其保护期限应不同于人类创作的一般作品的保护期限，因为从理论上来说人工智能的“生命”期限可以说是无限期的。如果对其创作成果给予无限期的保护，无疑会加大社会公众使用这类作品的成本，打破了权利人与社会公众之间的权利平衡。如何设置一个合理的保护期限，需要考虑的因素应该包括哪些？例如作品的受欢迎程度、作品的类型、作品的市场价值等都是需要考量的因素。从制度设计上来看，现有的版权法保护的作品，是从该作品创作完成之日起开始受到法律保护的，权利人无需履行类似专利权和商标权一样的申请注册程序。人工智能创作的速度远高于人类，但是其创作成果的独创性方面是否全部作品都可以满足独创性的要求还有待考察。因此，可以在未来考虑对人工智能的创作成果予以登记并在登记时对其独创性进行判断。那些真正符合独创性标准并具有市场价值的创作成果才能受到法律的保护，成为版权保护客体。其他那些不符合独创性标准的人工智能创作成果只能进入公有领域，不断丰富和充实人类的精神文化世界，启发人类的创作灵感。

## 人工智能的其他知识产权问题

1998年，John Koza，一个人工智能的基因算法工程师，开发了一种简单的电路设计并创造出36个可以与人类相竞争的涉及由人工智能产出的技术性创造成果，这些由人工智能创造的发明等是否可以获得法律上的保护？如何保护？



我们知道如果一个人类发明家发明了一项技术，该技术如果想要获得专利必须要满足新颖性、创造性和实用性的要求。如果人工智能的发明可以满足三个特征的要求，那么还有一个首要问题需要解决：谁是发明人？这个问题同人工智能写诗、作曲一样，同样需要解决的是其是否具有法律人格，只有具有法律人格的情况下其“智力”成果才能获得法律的保护。此外，在这个发明产生的过程中，涉及人工智能与人类的合作与分工问题，其智力贡献的程度很大程度上决定了最终权利的归属。

此外，在人工智能的技术研发过程中可能会涉及软件、专利技术和商业秘密等。由于各大企业在人工智能的研发过程中投入大量的资金和精力，对于可能涉及的与人工智能相关的专利技术会申请专利进行保护，或者将人工智能相关的软件进行计算机软件的版权保护，这些人工智能的关键技术会成为企业之间竞争的核心竞争力。但是也有人会有不一样的思路，比如特斯拉的CEO埃隆·马斯克和创业孵化器Y Combinator的总裁山姆·奥特曼，他们担心人工智能未来会接管世界，于是两家企业计划设立一家投资金额为10亿美元的公司，目的在于发挥人工智能的最大潜力然后将人工智能技术通过开源的方式分享给每个人，也就是Open AI。如果Open AI的构想可以实现，让AI技术惠及众人，那么对现有的法律保护体系可谓是一次冲击，也将改变投资AI的谷歌、Facebook等国际企业的竞争格局和竞争筹码。

---

[1] 杨守森.人工智能与文艺创作.河南社会科学, 2011, 19 (1) : 188-189.

[2] 徐萧.人工智能写的诗，你有本事分辨出来吗.澎湃新闻, 2017-02-11.

[3] 循环神经网络（Recurrent Neural Networks），其目的是用来处理序列数据。在传统的神经网络模型中，是从输入层到隐层再到输出层，层与层之间是全连接的，每层之间的节点是无连接的。但是这种普通的神经网络对于很多问题却无能为力。例如，你要预测句子的下一个单词是什么，一般需要用到前面的单词，因为一个句子中前后单词并不是独立的。RNNs之所以称为循环神经网络，是因为一个序列当前的输出与前面的输出也有关。具体的表现形式为网络会对前面的信息进行记忆并应用于当前输出的计算中，即隐层之间的节点不再无连接而是有连接的，并且隐层的输入不仅包括输入层的输出还包括上一时刻隐层的输出。

## 第二十二章 我是谁？

2016年人工智能呈现井喷式爆发并大放异彩，这距离人工智能概念的首次提出仅过去60年。英国科学家阿兰·图灵在1950年的《心智》杂志上发表了题为《计算机器和智能》的文章，提出了“图灵测试”：认为判断一台人造机器是否具有人类智能的充分条件，就是看其言语行为是否能够成功模拟人类的言语行为，若一台机器在人机对话中能够长时间地误导人类认定其为真人，那么这台机器就通过了图灵测试。进而我们需要探究人工智能的研究目的：一是在人造机器上模拟人类的智能行为，最终实现机器智能，而智能的实质是去重建一个简化的神经网络，从而实现智能体在行为层面上与人类行为的相似。美国的肖恩·莱格和马库斯·胡特认为：“智能是主体在各种各样的纷繁复杂的环境中实现目标的能力。”如何测量和评价人工智能主体是否具有智能或者其智商如何，是一个很复杂的判断过程。如何通过智能模型进行测试是人类需要面对的问题，这个问题也实际上在回答“人何以为人”这个本质的问题。

### 人工智能机器人法律人格

如果考虑赋予人工智能的机器人以法律上拟制的法律人格，就要求其能够独立自主地做出相应的意思表示，具备独立的权利能力和行为能力，可以对自己的行为承担相应的法律责任。<sup>[1]</sup>2016年，欧洲议会呼吁建立人工智能伦理准则时，提及要考虑赋予某些自主机器人（电子人，Electronic Persons）法律地位。而如何界定监管对象（即智能自主机器人）是机器人立法的起点。对于智能自主机器人，欧盟的法律事务委员会提出了四大特征：（1）通过传感器和/或借助与其环境交换数据（互联性）获得自主性的能力，以及分析那些数据；（2）从经历和交互中学习的能力；（3）机器人的物质支撑形式；（4）因环境而调整其行为和行动的能力。在主体地位方面，机器人应当被界定为自然人、法人、动物还是物体？是否需要创造新的主体类型（电子人），以便复杂的高级机器人可以享有权利，承担义务，并对其造成的损害承担责任？这些都是欧盟未来在对机器人立法时需要重点考虑的问题。

此外，由于日本机器人产业的迅速发展，日本在机器人立法方面不断积极推进和尝试。日本经济产业省报告显示，到2025年日本的机器人产业将产生648亿美元的收益。机器人产业的迅速发展，可以弥补日本老龄化社会带来的劳动力严重不足以及经济增长放缓等问题，因此对机器人及其创作物赋予合理的法律保护对日本有积极的社会意义。2016年5月日本颁布的《知识财产推进计划2016》以专章讨论了人工智能创作物的法律保护问题，认为有必要对现有的日本知识产权制度进行检讨，以便分析人工智能的创作物获得版权保护的可能性。

随着未来技术的发展以及人类对脑科学和自我认知的加深，如何合理判定人工智能是否具备与人类相类似的“智能”，并以此来判断是否应赋予人工智能以独立的法律人格地位，是需要各学科、各领域的专家进行分工配合完成的课题。

## 机器权利

从人类的历史发展道路来看，一个群体对自身权利的争取，不但是漫长的历史进程，而且充满着战火和硝烟。法国启蒙运动大思想家让·雅各布·卢梭在其名著《社会契约论》中，曾经这样写道：“人人生而自由，但却又无往不在枷锁之中。自以为是其他一切人的主人，反比其他一切人更是奴隶。”

随着机器人和人工智能系统越来越像人（外在表现形式或者内在机理），一个不可避免的问题就是，人类到底该如何对待机器人和人工智能系统？机器人和人工智能系统，或者至少某些特定类型的机器人，是否可以享有一定的道德地位或法律地位？由此，机器权利日益受到关注，成为人类社会无法回避的一个问题。动物与机器人最大的不同之处在于动物具有天然的生命，有生物属性，但是机器人是人类制造出来的，没有天然的生命属性，但是其是否具有独立意识尚未达成共识。那么，未来是否需要承认机器人等人工智能系统也具有机器权利，同时机器的权利在何种情况下可以行使，是否应该与人类拥有相同的权利，例如选举和被选举权等政治权利以及民事权利等。

20世纪最有影响力的科幻作家之一伊萨克·阿西莫夫于1942年在他的科幻小说《环舞》中首次提出了著名的机器人三原则：（1）机器人不得伤害人类，或看到人类受到伤害而袖手旁观。（2）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。（3）机器人必须保护自己，除非这种保护与以上两条相矛盾。后来，阿西莫夫又加了第零条定律：机器人不得伤害人类整体，或因不作为而使人类整体受到伤害。根据这个原则，人类的利益是高于机器人的，机器人不能损害人类的利益。假设人类开发和设计了一种智能机器人用于制造军事产品，但是其通过自我学习设计和开发出了核武器或致命武器，此时人类是否可以基于人道主义和人类共同利益而消灭该机器人？机器人是否有能力决定其生存或是死亡或者说机器人是否有权利从事买卖活动呢？或者我们是否可以机器人进行虐待以发泄不满？

## 谁来赋权于机器人？

启蒙运动为资产阶级的自由平等提供了新的理论基础，但是有时这种理论还不得不披着宗教神学的外衣。美国《独立宣言》写道：“人人生而平等，造物主赋予他们若干不可让与的权利，其中包括生存权、自由权和追求幸福的权利。”造物主，一种高高在上的万能的存在，赋予了每个人自由平等的权利。尽管达尔文的进化论，早已经证明了人类从来不是被创造出来的，而是不断进化的结果。不可否认，科学技术的发展破除了封建迷信，宗教再也无法主导人类社会。但是，科技技术的进步，让人类的能力被逐渐放大——我们创造出了机器人，而我们人类是否能够承担起一个“造物主”的角色，去赋予机器人权利呢？不同于地球上现存的任何物种，机器人毫无疑问是由人类创造出来的。在2016年的热播美剧《西部世界》中，西部世界里的机器人将人类作为上帝，任由人类消遣娱乐甚至杀戮，而等到机器人的意识觉醒，他们发现，人类远不是上帝。

是否应当由人类赋予机器人权利的问题，其实质在于是否承认机器人的主体地位问题。早在20世纪五六十年代，人工智能技术刚刚起步之时，就有哲学家提出：把机器人看作机器还是人造生命，主要取决于人们的决定而不是科学发现；而等到机器人技术足够成

熟，机器人自身就会提出对权利的要求。1976年，阿西莫夫出版的科幻小说《机器管家》（*The Positronic Man*）就讲述了一个自我意识觉醒的智能机器人安德鲁想要成为人类的故事。安德鲁作为一个家政智能机器人，在他两百年的生命历程中，一直要求人类把他作为人类看待，为此，他开设机器人公司，研发新的技术，使得在生命体征上他和普通的人类一模一样，甚至最后要通过手术让自己的生命只剩下一年（因为机器人在可预期的将来是永生的），才能获得法律的认可，最终获得人类的生命。

2016年5月31日，欧盟委员会法律事务委员会提交一项动议，要求欧盟委员会把正在不断增长的最先进的自动化机器“工人”的身份定位为“电子人”，并赋予其特定的权利和义务。所以，不管人类能否充当这个机器人的“造物主”的角色，现实中，人类已经开始在行动了。

## 赋予机器人哪些权利？

尽管黑色人种和女性在历史上曾经遭受不公平待遇，他们被剥夺或者限制了作为人的基本权利，但是，随着人类社会的进步，肤色和性别不再是享受基本人权的障碍。机器人的种类非常多，它们存在各种各样的形态，主要可分为人形或者是非人形机器人。在机器人自我意识觉醒的前提下，讨论赋予哪些机器人权利，是一个非常复杂的问题。比如，类人形的陪伴型机器人享受权利，人类可能容易接受；而动物形状的陪伴机器人享受权利可能就难以接受了，但这确实是正在发生的事实，2010年11月7日，在日本，一个海豹宠物机器人帕罗（Paro）获得了户籍，而帕罗的发明人在户口簿上的身份是父亲。拥有户籍是拥有公民权利的前提，机器人在日本可能逐渐会被赋予一些法律权利。<sup>[2]</sup>其实，现阶段的宠物机器人跟真实的宠物在享受的权利上并没有什么不同，因为普通的宠物也需要登记才能够饲养。还有一类非陪伴型的机器人，它们的外形迥异。例如，自动驾驶汽车是否可以被视为机器人而享有权利？任何存在着芯片和自我意识的实体是否都应当被认为是应当享受权利的机器人？

问题更进一步，机器人是否需要一定的物质载体，是否必须以各种各样的形式存在于人类的物理现实世界中？在2013年上映的美国科幻电影《她》中，故事讲述了一个内向的作家爱上了一个先进的人工智能操作系统“萨曼莎”（Samantha），而“萨曼莎”根本没有现实的实体存在，她只存在于网络之中，是一串代码和符号。未来，如果我们人类赋予类人形的辅助机器人以某些权利，那么“萨曼莎”这样的只存在于网络之间的人工智能系统，是否应当被赋予权利？如果只保护了类人形的辅助机器人，而不保护这些没有实体的人工智能系统，那人类只是在保护自己对这个机器人的财产所有权，而非对于一个不一样的智能物种的尊重和保护。也许，就像上文提到的，未来机器人的技术足够成熟、自我意识觉醒得足够充分，机器人（包括人工智能系统）自身会对自己加以保护，而不需要人类来做一个裁判者。

## 机器人可以拥有哪些权利？

人类具有的法律上的一些基本权利包括生存权、平等权和一些政治权利。在目前的技



术水平之下，机器人的意识尚未觉醒，机器人的财产属性还十分强大，也就是目前对于人来说，机器人只是工具，而非另一种智能物种。目前机器人尚不可能被赋予跟人一样的权利，因此，在上文提及的欧盟的动议中，提出要把最先进的自动化机器人的身份定位为“电子人”，并赋予这些机器人依法享有著作权、劳动权等“特定的权利与义务”。动议中提出的赋予机器人著作权，是一个十分紧迫的现实问题。由于人工智能技术的进步，机器人或者人工智能系统目前已经不是简单地执行人类的指令，而是具有了创造性的思维，能够进行独创性的内容创作，而这些之前都是人类所独有的智能。

在欧盟法律事务委员会的提案中，还以护理机器人为例，提出了对机器人有生理依赖的人类会产生情感上的依恋。因此，机器人应该始终被视为机械产物，这有助于防止人类对其产生情感依恋。这种担忧不是空穴来风。在中国，2017年4月，一个浙江大学研究人工智能技术的硕士和自己研发的智能机器人莹莹结婚了。这种浪漫爱情故事，不仅只存在于人和机器人之间，机器人之间同样存在。2015年7月，明和电机就举办了一场机器人与机器人之间的婚礼。现在看来，这种事情仿佛闹剧一般，但是随着人工智能技术的进步，这些问题都将成为摆在人类面前亟待解决的问题。

## 机器人的权利与义务

赋予一个人（机器人）以权利，就要对另一人施加义务和限制。类比人类对于动物的保护，在动物保护立法比较完善的欧盟国家，都是赋予动物不受人类虐待的权利；其根本的中心点还是通过限制人的行为，来达到对动物权利的保护。未来的世界，人类面对机器人的存在，是否也要通过限制自身的某些行为来赋予机器人一定的权利呢？他们最基本的“生命权”是否可以由人类剥夺呢？例如2015年加拿大研究人员研发的机器人HitchBot在成功地通过搭车的方式穿越多个国家后，在美国被人类残忍“杀害”，即便如此，HitchBot在其留下的遗言中说道：“我对人类的爱不会消退。”我们是否可以以人类的名义任意剥夺机器人的生命权呢？当机器不再是一堆冰冷的金属堆砌成的物品，当其有了独立的“意识”和判断能力，我们是否也应该尊重他们的生命及权利呢？

除了法律权利之外，我们还应该给予机器人最低限度的道德权利。我们不能滥用机器人，不能利用人类的主导地位对其进行虐待。未来如果机器人拥有了自我意识，我们是否也应当尊重其意愿或者说照顾其喜怒哀乐，而不能强制其从事一些其不愿意从事的工作或劳动？那就是我们对其他与我们在地球上共存的主体的最低限度的尊重。

没有任何人能够只享受权利而不承担义务，如果我们把机器人与人类同等对待，那么是否机器人也应该承担相应的义务呢？欧盟的动议报告中提出，如果先进的机器人开始大量替代人工，那么欧盟委员会需迫使其所有者缴纳税款或社保。报告还建议，为智能化机器人设立一个登记册，以便为这些机器人开设涵盖法律责任（包括依法缴税、享有现金交易权、领取养老金等）的资金账户。不过，这项提议被欧盟委员会在审议过程中否决。微软公司的创始人比尔·盖茨公开表示，政府应当对人工智能征税，用来补贴和培训因为机器人大规模应用而失业的人。如何使机器人承担相应的义务，例如设立其专门的财产账户等方式是否可行，都是留给我们需要不断探索的问题。

---

[1] 陈亮.电子代理人法律人格分析.牡丹江大学学报, 2009, 18(6): 67.

[2] 杜严勇.论机器人权利.哲学动态, 2015(8): 53.



## 第二十三章 法律人工智能十大趋势

2017年7月20日，在高瞻远瞩的国家人工智能战略《新一代人工智能发展规划》中，国务院向法律行业释放了一些信号。首先，新规划在对人工智能理论、技术和应用做出前瞻布局的同时，还呼吁加强人工智能相关法律、伦理和社会问题研究，建立人工智能法律法规、伦理规范和政策体系。其次，新规划力挺智慧法庭建设，提出促进人工智能在证据收集、案例分析、法律文件阅读与分析中的应用，实现法院审判体系和审判能力智能化。最后，更为前瞻的是，新规划提出“人工智能+X”复合专业培养新模式，法学赫然在列，法学教育的变革已然箭在弦上。

其实，在2016年谷歌公司的围棋机器人AlphaGo之后，法律人工智能就被带火了，人工智能和机器人取代律师的报道不绝于报端。以“artificial intelligence in law”为检索关键词，谷歌搜索结果超过630万条；以“法律人工智能”为检索关键词，百度搜索结果超过550万条。但如果追溯起来，人工智能与法律的结合已经有30年历史了，始于1987年在美国波士顿的东北大学举办的首届国际人工智能与法律会议（ICAAIL），并最终促成了国际人工智能与法律协会（IAAIL）在1991年的成立，该协会旨在推动人工智能与法律这一跨学科领域的研究和应用。包括十大主要议题：

- （1）法律推理的形式模型；
- （2）论证和决策的计算模型；
- （3）证据推理的计算模型；
- （4）多智能体系统中的法律推理；
- （5）自动化的法律文本分类和概括；
- （6）从法律数据库和文本中自动提取信息；
- （7）针对电子取证和其他法律应用的机器学习和数据挖掘；
- （8）概念上的或者基于模型的法律信息检索；
- （9）自动化次要、重复性的法律任务的法律机器人；
- （10）立法的可执行模型。

hansu-01

在这样的背景下，当前，法律科技（LawTech）正在持续兴起。在人工智能技术的加持下，法律科技更是有望给法律行业带来更深、更彻底的变革。笔者在此前的文章《人工智能法律服务的前景与挑战》中，曾援引法律科技研究报告《文明2030：不久将来的律所》的观点称，“经过长期的孵化和实验，技术突然可以以惊人的速度向前行进了；在15年内，机器人和人工智能将会主导法律实践，也许将给律所带来‘结构性坍塌’，法律服务

市场的面貌将大为改观。”研究科技与法律长达30多年的英国学者Susskind在其著作《明天的律师：预见你的未来》（*Tomorrow's Lawyer: an Introduction to Your Future*）中持类似观点，认为法律行业过去200年之变化，不及未来20年之变化。法律人需要做好迎接未来的准备。承担义务，并对其造成的损害承担责任？这些都是欧盟未来在对机器人立法时需要重点考虑的问题。

此言并非虚言。法律行业并非对技术具有完全的免疫力。面对技术发展和外在压力，法律行业在教育模式、组织结构、收费模式等诸多方面的不适应性已经显现出来。这让人们开始对人工智能技术支撑下的法律科技寄予厚望。全球来看，虽然从2011年到2016年，全球法律科技公司总融资额度仅有7.39亿美元，显著低于金融科技、医疗科技等新兴领域；但是全球法律科技上市公司的数量呈爆发式增长，从2009年的15家增长到了2016年的1164家，主要集中在在线法律服务、电子取证、从业管理软件、知识产权/商标软件服务、人工智能法律科技、诉讼金融、法律检索、律师推荐、公证工具等九大领域。在这样的国际趋势下，国内法律科技市场开始从“互联网+法律”向“人工智能+法律”转变，法律人工智能创业成为人工智能创业的重要组成部分，面向B端或者C端的法律人工智能产品逐步进入公众视野。

不仅如此，法律行业如律所、公司法务部门、法院等也开始积极布局人工智能法律科技，部分原因是在线法律服务、客户成本压力（比如公司法务部门日益希望以更低的成本获得更多的法律服务）等因素迫使律所投资创新。国际律所Dentons是典型代表，其在2015年5月率先启动法律科技创新加速器项目NextLaw Labs，目前已经孵化了包括大名鼎鼎的机器人律师ROSS在内的10多个法律科技项目。其他越来越多的国际律所如Linklaters、Riverview Law、BakerHostetler等亦开始研发、部署法律人工智能系统，帮助提高工作效率，或者以低成本模式提供法律服务。

一言以蔽之，从最早的基于规则的（Rule-Based）的专家法律系统（将法律专家的法律知识、经验等以规则的形式转变成为计算机语言），到以深度学习、机器学习、大数据等为支撑的自主系统，人工智能对法律以及法律行业更深更广的影响才刚刚开始。可以说，人工智能技术已经在开始改造整个法律行业，而改造的规模和速度将不仅仅取决于技术发展和进步的步伐，而且取决于整个法律共同体对于新技术和新模式的接纳程度，而这需要政策支持和发展导向。笔者在过去一些观察和研究的基础上，尝试对人工智能在法律行业中的应用和影响总结出以下十大趋势：

第一，智能化、自动化的法律检索将深刻影响法律人进行法律研究（检索）的方式。

在人工智能技术的加持下，法律研究（检索）正向智能化、自动化的方向迈进。法律研究（Legal Research）对于法律人的价值不言而喻，无论你是法学院学生，还是从业律师、公司法务人员，抑或司法人员，甚至普通民众有时候也需要进行法律检索。其实，信息化已经对法律检索进行一次改造，法律文本、裁判文书等法律资料的数字化，支撑起了市场规模巨大的法律数据库市场。但Westlaw、北大法宝等法律数据库服务一般基于传统的关键词检索，利用这些数据库进行法律检索，是一件费时费力的苦差事。然而，基于自然语言处理（NLP）和深度学习的语义检索和法律问答已经在开始改造传统的法律检索服务。比如，号称世界首个机器人律师的ROSS就是基于IBM的Watson系统的智能检索工具，利用强大的自然语言处理和机器学习技术向律师呈现最相关、最有价值的法律回答，而非像传统法律数据库那样，仅仅呈现一大堆检索结果。此外，语义技术，文本分析和自然语言处理，以及图像和视频技术已经为商标和专利检索及版权监测等知识产权法律工作的自动化提供了可能性，比如TrademarkNow。

新形式的、基于语音交互的智能法律检索将经历两个阶段。第一个阶段是智能化。在这个阶段，依然需要人类律师明确需要解决或者回答的法律问题是什么，法律搜索引擎识别相关案例并评估其价值，形成专业回答。ROSS是这一阶段智能法律检索的典型代表。第二个阶段是自动化，意味着不需要人类律师指明法律问题是什么，系统自身可以理解一段事实陈述并自动识别其中的法律问题，然后完成检索并提供最佳法律信息，整个过程几乎不需要人类律师的深度参与。这几乎是将人类律师从烦琐的法律检索工作中解脱了出来。

## 第二，人工智能将持续推动法律文件自动化。

就像新闻写作机器人的崛起将给新闻业带来一场巨变一样，法律文件自动化趋势将可能给法律行业带来规模相当甚至更深远的变化。主要包含两个层次。

第一个层次是法律文件审阅自动化。无论是调查取证、尽职调查，还是合同分析、合规审查，都需要对法律文件进行审查、分析和研究。自动化这一工作将能够显著提升法律人的工作效率。以电子取证为例，在并购、反垄断、大型劳动纠纷等越来越多的案件中，庞大的电子材料给证据和法律材料的搜集和整理提出了巨大挑战，律所往往需要投入大量的人力和物力，而且需要耗费大量时间。但基于NLP、TAR（技术辅助审阅）、机器学习、预测性编程（Predictive Coding）等技术的电子取证程序可以显著提高这一工作的效率，大大节约审阅文书的时间，而且准确性不输人类律师，因此成为了法律科技市场的一大细分领域，微软等公司都已介入。电子取证的步骤一般包括训练过程（人类律师从小量样本中确认相关的证据材料以供机器学习）和取证过程（意味着机器代替人类律师进行资料审阅以发现证据材料）。由于涉及用机器替代律师，可能触及政策障碍，因此英国、美国、澳大利亚等国家的法院已经明确表示在诉讼和案件中整理、搜集证据材料时可以利用预测性编程技术。

法律文件审阅自动化的另一个主要领域是合同分析。合同分析在风控、尽职调查、取证、诉讼等诸多场合具有重大意义，但它是一项耗时耗力的工作。然而，德勤

（Deloitte）借助机器学习合同分析系统Kira Systems，只要15分钟就可以读完原本需要人类律师花费12个小时才能审阅完的合同。在国际社会，人工智能合同分析服务已经常态化，KMStandards、RAVN、Seal Software、Beagle、LawGeex等提供智能合同服务的法律科技公司越来越多。在人工智能技术的驱动下，人工智能合同分析服务依然在蓬勃发展，带来效率的提高、成本的降低以及流程的改善。

第二个层次是法律文件生成自动化。新闻业正在被互联网和机器写作改造，过去8年，新闻业收入减少了1/3，就业岗位减少了1.7万个，报纸的市场价值和支配力大减，代之以网络媒体的不断兴起。法律行业正面临着同样的情况，智能机器辅助甚至独立起草法律文件的时代将会到来。如今，法律人使用法律格式的方式正从模板向法律格式文件自动生成转变；也许未来10~15年，人工智能系统将可能起草大部分的交易文件和法律文件甚至起诉书、备忘录和判决书，律师的角色将从起草者变成审校者。比如，硅谷一家律所Fenwick&West开发的一个程序可以为准备上市的创业公司自动生成所需文件，这将律师的账单时间从20~40小时减少到了几个小时，当需要准备大量文件的时候，这一程序可以使所需时间从数天、数周减少到数小时，大大提高了工作效率。机器智能的优势在于随着数据的积累，可以不断自我学习和改进，并且由于数据的互相关联性，计算机可以将特定合同与所有与之相关的法院判决关联起来，形成持续改进法律格式的动态关系。未来，随着软硬件性能以及算法的持续提高，起诉书、备忘录、判决书等高级法律文件也可以自动生成，但依然需要人类律师或者法官审阅，形成人机协作的关系。

第三，在线法律服务、机器人法律服务等替代性商业模式（Alternative Business Structure）不断涌现，使得法律服务的提供日益标准化、商品化、自动化、民主化。

在互联网时代和人工智能时代，律所和人类律师并非普通公众获取法律服务的唯一渠道。在线法律服务、机器人法律服务等替代性商业模式正在兴起，可以直接向终端用户提供一般法律咨询服务，比如遗嘱、婚姻咨询、交通事故咨询，等等。面向终端消费者的法律机器人DoNotPay就可以协助用户自主完成对交通罚单的申诉材料准备和提交工作。

美国法官波斯纳曾将法律行业形容为“涉及社会的法律的服务提供者的一个卡特尔”，意即是一个垄断的行业。高昂的律师费用导致社会中存在大量未被满足的法律需求，低收入以及中等收入人群中的大部分人的法律需求没有被满足。然而，在线法律服务、机器人法律服务等替代性商业模式可以以更低廉的价格向终端用户提供法律服务，有望使法律服务标准化、商品化、自动化、民主化。商品化意味着法律服务的提供不再主要依赖于特定的人类律师的专业素养，而可以以自动化的方式提供；民主化意味着大部分人将可以以较低成本获得一般的法律服务。英国学者Susskind认为，法律服务的提供的演进方向从定制化到标准化到系统化再到一揽子最后到商品化，意味着法律服务定价的由高而低，即从按小时计费到固定收费再到商品化定价最后趋于零。在这个层面上，国外有专家预言律师将走向没落。

无论如何，法律机器人都将对法律服务的提供产生深远影响，将持续推动法律服务走向标准化、系统化、商品化、自动化，使人人都可以获得法律服务，帮助消除法律资源不对称的问题，实现更为广泛的司法正义。如今，在美国，最知名的法律品牌不是哪一家知名律所，而是LegalZoom之类的在线法律服务提供商，这些新型的技术派的法律服务商代表着法律服务提供的未来趋势。它们对于律所而言并非替代者的角色，而是在律所之外，满足其他未被满足的法律需求或者通过律所就会十分昂贵的法律需求。而英国早在2007年就通过了《法律服务法案》，旨在自由化法律市场，革新法律行业组织模式，并引入竞争，促进法律服务的可负担性。在此背景下，一些国际律所已经设立了低廉的法律服务中心，在按小时计费、固定收费等模式之外，借助技术以更低价提供法律服务。承担义务，并对其造成的损害承担责任？这些都是欧盟未来在对机器人立法时需要重点考虑的问题。

第四，基于人工智能和大数据的案件预测将深刻影响当事人的诉讼行为和法律纠纷的解决。

从案件结果预测到犯罪预测，基于人工智能和大数据的预测性技术在司法领域的应用越来越广泛。一方面，案件预测技术在研究上取得进展。2016年，研究人员利用欧洲人权法院公开的判决书训练算法系统，构建了模型，来预测案件判决结果，预测准确性达到了79%。这一实证研究表明，案件事实是最重要的预测因素，这一结论与法律形式主义的观点一致，即司法裁判主要受案件事实的陈述影响。另一方面，案件预测已经用在了诸多实务领域。比如，Lex Machina公司提供的服务，通过对成千上万份判决书进行自然语言处理，来预测案件结果。其软件可以确定哪位法官倾向于支持原告；基于对方律师过去处理的案件来形成相应的诉讼策略；针对某个特定法院或者法官形成最有效的法律论证，等等。Lex Machina的技术已经用在了专利案件中。

案件预测的价值主要体现在两个方面，一方面可以帮助当事人形成最佳的诉讼策略，从而节约诉讼成本；另一方面，可以帮助法官实现同案同判，也即所谓的大数据司法确保公平正义。诉讼中可能的高昂成本给当事人带来沉重的经济负担，所以当事人一般都会在

案件起诉前或者上诉前对案件胜诉的可能性进行评估。但即使是最专业的律师，由于受限于人脑自身的信息处理能力，在预测上远不如计算机，因为计算机在强大算法的支持下，可以以超强的运算能力，处理可以获取的几乎所有数据。计算机的全数据处理，相比人类的样本数据分析，使得案件预测结果更为可靠。如果事先可以较为可靠地预知案件结果，意味着当事人不会冒着极大的败诉风险继续推进诉讼或者上诉，而是会选择和解、放弃诉讼等其他纠纷解决方式。但案件预测的弊端在于可能扭曲当事人的诉讼行为，带来新的偏见和滥用。

第五，在线法院（Online Court），以及人工智能法律援助，将促进司法可得性（Access to Justice），帮助消除司法鸿沟（Justice Gap）。

俗话说，法院大门朝南开，有理没钱别进来。司法审判系统的低效率、程序拖沓、成本高昂等问题历来为人们所诟病。但问题是，人们为了解决彼此之间的法律纠纷，一定得去有实体场所的法院吗？技术的发展已经给出了否定的回答。比如，伴随着电子商务的兴起和繁荣，在线争议解决机制（Online Dispute Resolution, ODR）开始流行，在电子商务eBay上，大量买卖纠纷通过SquareTrade这一ODR服务商在线解决，当事人通过ODR系统在线提交事实陈述和证据，使纠纷在线得到处理，甚至不需要人类律师介入，很多案件也根本不会进入法院审判阶段。

在ODR模式的影响下，在线法院的实践在国外已经出现。比如，英格兰和威尔士上诉法院大法官Briggs在呼吁“提高民事司法的效率”时表示，“可以借助使用人工智能来在线裁决英格兰和威尔士的民事法律案件，在这方面，人工智能可以辅助法官，甚至作出判决。”据悉，英国已经投入10亿英镑现代化、数字化其法院系统。根据英国学者Susskind的观点，英国的在线法院包括三个阶段：第一阶段是在线法律援助系统，向当事人提供法律咨询和建议等；第二阶段是审判前争议解决，法官通过电邮、电话等方式和当事人沟通，以解决纠纷；第三阶段即在线法庭，只适用于小额案件，以在线的方式审判案件，包括立案、提交证据、举证质证、裁判等。这类似于简易程序。英国当前的在线法院建设并没有利用人工智能系统来裁判案件，因此并非代替法官，而是以更好的方式解决纠纷。在交往场景日益数字化的背景下，在线身份识别、音视频技术以及人工智能技术等已经为在线法院的建设提供了技术支持。

中国正在大力推动的智慧法院与国外的在线法院类似。2016年7月发布的《国家信息化发展战略纲要》将建设“智慧法院”列入国家信息化发展战略，明确提出：“建设智慧法院，提高案件受理、审判、执行、监督各环节信息化水平，推动执法司法信息公开，促进司法公平正义。”2016年12月发布的《“十三五”国家信息化规划》明确指出，支持“智慧法院”建设，推行电子诉讼，建设完善公正司法信息化工程。在2017年5月11日举行的全国法院第四次信息化工作会议上，最高人民法院院长周强提出，智慧法院是建立在信息化基础上人民法院工作的一种形态。中国各地的法院都在探索某种形式的智慧法院建设，但以浙江智慧法院（浙江法院电子商务网上法庭）建设最为出名；据浙江高院信息中心副主任刘克勤介绍，浙江智慧法院每年处理的交易、著作权等纠纷多达2.3万件，可以直接对接淘宝、天猫等多个平台，提供在线矛盾纠纷多元化解平台，其他辅助措施包括案件结果预判、网上司法拍卖、智能语音识别、类案推送、当事人信用画像，等等。2017年6月26日，中央深改小组审议通过《关于设立杭州互联网法院的方案》，该互联网法院主要审理网络购物合同纠纷，网络购物产品责任纠纷，网络服务合同纠纷，在互联网上签订、履行的金融借款合同纠纷和小额贷款合同纠纷，网络著作纠纷等五类案件。未来，在线法院的进一步建设和普及将促进公共法律服务的供给，帮助消除司法鸿沟。



此外，公共法律服务中的法律援助不足也是司法体系的一大问题，尤其是在刑事案件中，很多被告人得不到法律咨询和辩护。一些民事案件同样是在没有律师介入的情况下进行的。未来，法律机器人可以向当事人提供基本的法律援助，而法律援助律师仅在必要时才介入，这可以显著提高司法援助的效率和质量，实现公平正义。而且，法律机器人法律援助同样可以融入在线法院的建设当中。

#### 第六，人工智能和机器人将成为法律系统的主要进入点。

无论是律所和律师，还是法院，抑或当事人和终端消费者，基于人工智能和机器人技术的“智能交互界面”（Intelligence Interface）将成为法律系统的主要进入点，法律机器人和人工智能是其中的核心。对于律师而言，未来的法律实践比如法律检索、案件管理、法律写作等将主要通过具有智能交互界面的法律机器人和人工智能系统来完成，这就好比医生现在主要借助各种复杂的医疗器械来完成医疗活动一样。对于法院而言，司法审判的数字化和在线化，意味着类案检索、裁判文书写作、证据分析和推理等也将在法律人工智能的辅助下进行，甚至为其所取代。对于终端用户而言，交互性的、基于互联网的问答系统可以以文本或者语音对话的形式同用户交流，并生成所需的法律信息，或者指导其完成基本的法律文件和格式。在此背景下，律师当前的角色将会发生变化，一些角色可能被机器取代，比如常规性、重复性任务；一些角色可能被机器增强，比如案件预测、法律写作；而对于新法新规，律师依然需要扮演核心角色。

#### 第七，律师市场评价将使法律行业更加透明，可能带来“马太效应”。

法律市场作为一个双边市场，其评价体系在很大程度上是不透明的，不像电商平台以及外卖、生活服务等O2O平台，具有较为完善的用户评价机制，确保了市场的透明度和消费者的知情权。由于法律市场在很大程度上并未平台化，很难搭建有效的评价机制。然而，人工智能、大数据等正在改变这一状况，对律师市场进行评价正变得可能，成为法律科技的一大趋势。当前，律师推荐已经成为法律科技的核心领域之一，国内外都在持续涌现律师推荐和评价类的产品和服务。律师市场评价相当于将律师置于阳光之下，明星律师、普通律师、不合格律师等的区分将透明化，结果可能带来律师市场的“马太效应”，明星律师业务增多，收入增多，而普通律师、资历浅的律师则相反。这呼吁律师转型，即以技术化的低成本模式提供法律服务。

#### 第八，法律人工智能职业将作为法律行业的新兴职业而不断涌现。

法律机器人和法律人工智能并非凭空产生，需要技术人员和法律专家之间的通力合作。随着人工智能与法律不断融合，这一领域的研究、开发和应用将不断增强，法律人工智能职业将作为法律行业的新兴职业而不断涌现。当前，一些积极拥抱新技术的国际律所已经在加强法律IT能力建设，法律开发者、法律数据分析师、法律数据库管理者等正在加入律所、公司法务部门、法院、法律数据库公司等法律机构。法律科技公司更是需要既懂法律又懂技术的复合型人才。未来，技术与法律的结合将更为密切，对新型人才的需求也更为迫切。

#### 第九，法律教育与人工智能等前沿信息科学技术将日益密切结合起来。

中国《新一代人工智能发展规划》已经看到了法学教育与人工智能的结合，提出打造“人工智能+法学”复合专业培养新模式。这是极为高瞻远瞩的设想。笔者曾参与翻译*Failing Law School*一书，书中对美国“4+3”（四年本科+三年法学院教育）的法学教育模式提出严正批判，认为法学院根本不需要读三年，顶多需要两年，可能一年就够了。而中国

传统的法学教育是高中毕业后直接读四年本科法学教育，这样的法学人才培养模式很难适应机器人和人工智能主导的未来法律实践。相比现在的律师，未来的律师将会从事大不相同的工作，所以需要不同的教育。因此，新规划提出的“人工智能+法学”培养模式是有远见的。

其实，国外法学院早就开始探索革新法学教育，注重对法科学学生的科技和数字素养的培养。比如，早在2012年，乔治城大学法学院即开始提供一个技术创新与法律实践的实践课程，形成特色的“**Iron Tech Lawyer**”比赛项目，培养学生的法律开发能力。2015年，墨尔本大学法学院开始提供如何开发法律应用的课程。未来，法律教育与人工智能等前沿信息科学技术将日益密切结合起来，而能否较早较快实现这一设想，取决于法学教育的反应速度。其实，人工智能不仅仅对法学教育提出了挑战，要求跨学科融合的教育模式，而且对其他学科教育也提出了类似的挑战。

第十，计算法律（**Computational Law**），以及算法裁判，或将成为法律的终极形态。

英格兰和威尔士上诉法院大法官Briggs在在线法院的倡议中提出了算法裁判，即人工智能可以代替法官直接作出裁判。这并非不可能。其实，计算法律学历来就是人工智能与法律的核心研究方向之一，思考“除了书面语言，法律可以有更精确、更形式化的表达吗？”这一问题，并探索用计算逻辑和代码来表达法律。笔者此前在知乎网站上看到一个设想：如果能用一列 $n$ 维向量描述各种事件，将「事件.txt」导入「法律.exe」，从而产生「判决.txt」。将法律条文转化成代码，从而使得判决彻底脱离个人主观判断，并且可以在任何人的计算机上在线。将代码开源，放在类似GitHub的网站上，以便于全民监督。计算法律当前在计税等一些领域有应用，更多则是一种学术研究；但在未来的成熟的信息社会，更普遍的计算法律将可能出现，届时系统将会自动执行法律，不需要律师，甚至也不需要法官，因为那时的法律已经完全自动化了。

## 法律人应做好迎接未来的准备

人们说，预见未来的最好方式是创造未来。法律行业的未来需要法律人这一职业共同体共同创造。虽然之前有研究认为律师助理和法律助理被自动化的概率高达94%，引发了人们对法学毕业生就业的担忧。但笔者在willrobotstakemyjob.com网站的测试结果显示，仅有3.5%的律师会被人工智能和机器人替代。不管科学与否，都可以作为一种暂时的宽慰。

据笔者调查，律师的工作包括十三项：文件管理；案件管理；文件审阅；尽职调查；文件起草；法律写作；法律检索（研究）；法律分析和策略；事实调查；客户咨询服务；谈判；其他交流和互动；出庭及准备。律师需要及早思索这其中的哪些任务可以被自动化或者可以借助科技提高效率，而英国学者Susskind则提出了“分解”法律服务的思路，认为一项法律任务可以被分解成多个部分，核心部分可以由律师完成，其他部分则由效率更高的第三方完成。

而对于法律服务自动化的担忧，包括律师在内的法律人在判断其工作的价值以及在思考人工智能技术对其工作的影响时，至少需要考虑以下三个因素：第一，是否涉及数据分析和处理，在这一方面，人类几乎不可能和人工智能和机器人相匹敌，尽早使用并适应新

技术才是明智的选择；第二，是否涉及互动交流，类似行政前台等法律客服工作被自动化的可能性非常大，一般的法律咨询也可以被自动化，但更高级别的互动交流如谈判、出庭等则很难在短期内被自动化；第三，是否处于辅助决策的地位，人工智能辅助决策已经被应用在了很多领域，在法律行业，人工智能辅助决策也正在发生并成为趋势，比如在案件结果预测上，人工智能可以比专业律师做得更好，诸如此类，尽早利用并适应新技术才是必然的选择。

最后，作为总结，经过30多年的发展，受超强运算能力、大数据和持续改进的算法的影响，人工智能对法律以及法律行业的影响正在加深、加快，未来10~20年法律行业将可能迎来一场巨变。作为法律人工智能最直接的目标客户，法律人需要调整心态，积极拥抱新技术和新模式，并在这个过程中坚持对法律的理念和信仰，防止法律人工智能削弱、损害法律共同体所秉持的以及法律系统所坚持的观念和价值，让法律人工智能来促进司法正义，而非带来偏见和歧视，或者背道而驰、贬损正义。