Open Journal of Legal Science 法学, 2023, 11(3), 1232-1240
Published Online May 2023 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2023.113176

Hans汉斯

风险等级下的人工智能法律监管研究

吴玉婷

宁波大学法学院，浙江 宁波

收稿日期：2023年2月28日；录用日期：2023年3月27日；发布日期：2023年5月23日

摘 要

近十年里，人工智能已成为世界各国科技发展的战略重点，同时监管风险和人工智能技术相伴相生，成为人工智能立法监管的导向。人工智能风险的产生主要是由黑箱模式以及运行中产生的风险共同构成。法律是进行风险防控强有力的工具，针对目前我国人工智能应用广泛、产业规模庞大、监管力量薄弱等现实困境，可以根据风险等级划分不同类型的监管主体，并在明确具体主体的基础上，以政府监管为主体，构建风险评估监管机制。落实人工智能风险等级划分，以期政府、社会和企业在各自的范围内承担起相应的法律责任，共同促进人工智能产业的蓬勃发展。

关键词

人工智能，风险等级划分，法律监管

Research on Artificial Intelligence Legal Supervision under Risk Level

Yuting Wu

Law School, Ningbo University, Ningbo Zhejiang

Received: Feb. 28th, 2023; accepted: Mar. 27th, 2023; published: May 23rd, 2023

Abstract

In the past ten years, artificial intelligence has become the strategic focus of science and technology development in countries around the world. At the same, regulatory risk and artificial intelligence technology are accompanied by each other, which has become the guidance of artificial intelligence legislation and supervision. The risk of artificial intelligence is mainly composed of the black-box and generated in operation. The law is a powerful tool for risk prevention and control. In view of the realistic dilemmas of China's extensive application of artificial intelligence, huge industrial scale, and weak regulatory force, we can divide different types of regulatory bodies ac-

文章引用: 吴玉婷. 风险等级下的人工智能法律监管研究[J]. 法学, 2023, 11(3): 1232-1240.
DOI: 10.12677/ojls.2023.113176

cording to the risk level, and on the basis of clarifying the specific subjects, take government supervision as the regulatory body, and build a risk assessment and supervision mechanism. With the implement of the risk classification of artificial intelligence, the government, society and enterprises should assume corresponding legal liabilities within their respective scopes and collectively promote the prosperous development of artificial intelligence.

Keywords

Artificial Intelligence, Risk Classification, Legal Supervision

1. 人工智能运用所带的潜在风险

当今世界，人类社会已经从互联网时代跨入了人工智能时代，我们曾经梦想的科技化时代已经来临。然而，人们在享受人工智能技术带来的精准化服务的同时，也面临着它所带来的社会各领域中前所未有的风险。法律是预防和控制风险的有力工具，人工智能的发展对于目前的法律监管来说，无疑是一个全新的挑战。尽管目前各国都意识到加强人工智能风险法律监管的重要性，关于人工智能系统现实应用中所产生的各种风险，以及监管治理体系和治理措施，学术界也有过讨论，要想消除人工智能系统应用可能产生的各种潜在风险，采用人工智能法律监管措施成为必选项，所以，根据人工智能系统应用的不同场景划分风险等级是监管治理系统建设的前提性条件。对此，有必要对人工智能带来的潜在风险和成因进行深入剖析，从而切实提升对人工智能的法律监管。

1.1. 人工智能在经济方面的损害

首先，人工智能可能会使消费者的决策出现扭曲。第一，使用人工智能算法来操纵信息、误导消费者决策对于商家而言是完全可以做到的。例如，用户评价和排名在算法上的偏差性实施；发布或广告宣传误导信息、筛选信息推送、机器合成虚假信息等，造成信息误导。第二，商家有可能利用人工智能算法为消费者剥削性定价的可能性很大。数字平台以消费者的个人数据为基础进行个性化定价，为完全占有消费者的余量，甚至可以制定出接近完全价格歧视的价格。简而言之，平台企业可以借助人工智能算法，执行更为普遍的个性化营销和定价策略，对消费者的选择进行更广泛的操纵，从而实现将消费者剩余部分更多地占有或全部占有。其次，人工智能应用可能会使市场出现扭曲。第一，人工智能算法会促使商家之间产生价格合谋。算法能够实现实事跟踪竞争对手的价格变动，并快速采取相应后续策略，价格协调所需费用减少，增加惩罚个别商家出卖的有效性，由此促进了价格合谋[1]。第二，主导平台滥用的方式是通过对商户或者第三方合作者进行封锁。第三，人工智能算法提供了更有效的机制，让各种垄断滥用行为在支配平台上得以实施。主导权平台通过算法封锁等手段排挤竞争者。最后，人工智能的运用会加剧就业和收入公平状况。它不同于工业经济时代，机器取代了人类的工作，人工智能系统的应用对于就业所产生的影响，并不是部分取代了人们繁重的体力劳动，而是全面取代了人。

Acemoglu 和 Restrepo 指出，人工智能的运用，会使资本取代劳动的趋势越来越突出，企业内部出现了过度自动化的动机，由此造成了严重的失业问题。人工智能的运用，不仅替代了低技术工人的传统模式，并替代了高技术工人，使得高技术工人的职业生命周期变短[2]。更严重的是人工智能的运用加剧了

收入分配不公，Hémous 和 Olsen、Aghion 等的分析均指明，人工智能的有效运用将提高国家总收入中资本收入的比例，降低劳动所得的比例，高技能人群与低技能人群的收入差距将显著拉大[3] [4]。

1.2. 人工智能社会方面的损害

第一，个人隐私被侵犯。人工智能算法通常会对个人数据进行过度采集、识别个体和行为的轨迹跟踪，在对个人数据进行深度分析的基础上，实现更加准确的个人画像，能够更加全面和深刻地理解个人隐私信息，预测个人行为或能够将个体分类分组，尤其是非授权同意使用人脸识别、语音识别等生物识别系统会对个人隐私造成严重的侵害，因此，人工智能算法对个人隐私进行了深度分析，对会影响或操控个体精神状态的个人情绪识别的人工智能系统的运用。第二，危害人类生命健康。例如智能玩具、电子游戏等包含暴力色情的成分，会扭曲特定弱势群体的行为(老人、儿童和残疾人)。此外，人工智能算法还会对特定的个体或群体进行控制和操纵，诱导他们进行超负荷劳动，从而使劳动者的权益受到损害。过多的算法诱导使得行为人或者劳动者为加强监督和考核，增加劳动强度而伤害行为人生理和精神健康，或为了加强企业压榨普通劳动者之目的而使劳动者的工作状态变化。第三，造成社会性歧视。人工智能很有可能会采取歧视性对待特定群体的方式，使特定公民或群体失去参与重大经济社会活动的权利，从而影响社会公平正义。例如人工智能算法设计中存在的偏偏差、低质量数据造成的结果偏差等等，这些都使得人工智能算法被应用于公共服务、教育录用和司法裁定等以及其他针对特定公民个人或者团体的歧视性待遇，导致特定公民个体或团体无法平等地参与或不公平第参与社会活动。另外，人工智能应用还会由人群体的认知、知识层次或者生理原因不同而出现技术鸿沟，进而导致人工智能算法决定下的社会经济活动中，使特定群体如老人、小孩等处于弱势地位，从而被边缘化。第四，对公共安全造成威胁。人工智能在无人驾驶汽车、智能城市和基础设施等方面的广泛应用以及人们对人工智能的过度依赖将显著增加公共安全风险并在特殊情况下会危害公民生命健康，因为人工智能系统存在能力不足或人工智能应用主体人为忽视系统缺陷和风险。甚至会影响到全社会运转的有条不絮。此外，通过对倾向性信息内容的过滤、筛选、屏蔽或定向推送，人工智能算法也得以实现。个别信息平台推送低俗新闻，误导公众；违规发布不实信息和谣言；传播淫秽视频用于追求流量，强化损害社会伦理道德的偏见或偏激意识。人工智能技术深度应用于医疗健康将挑战固有的生命体伦理，人工智能性爱机器人的出现将冲击传统婚姻家庭关系。更严重的可能会对国家政治安全造成影响。算法正成为可能威胁国家政治安全和左右政治意见的工具。例如，在美国政治选举中，选民的政治判断在很大程度上被人工智能算法所推送的信息所左右。整体而言，人工智能越来越成为影响经济社会运行的重要力量，因为人工智能系统在各个社会经济组织中的广泛深入应用，人们对人工智能技术的依赖程度前所未来[5]。但是，经济和社会风险潜藏在人工智能日益广泛的应用中。人工智能算法黑箱、算法复杂度越来越高，算法透明度和设计不够、使用者带有恶意使用人工智能系统等都会造成对个体、组织和社会造成严重损害的结果偏差或错误。因此，位加强人工智能法律监管而划分人工智能风险等级，成为国家治理社会的一项重要任务，也是确保人工智能技术发展对企业、国家和人类社会都有促进作用的必要措施。

2. 人工智能风险产生的成因

2.1. 人工智能的黑箱模式

所谓人工智能系统，就是指采用一种或多种有关技术与方法开发出来的软件，针对所给预设目标系列，能生成对它们交互环境有影响的内容、预测、建议或决策等的输出。实现人工智能的基本途径在于算法。人类在计算机的帮助下，用程序语言录入经过证实的人类经验，形成了模仿人类思维的模型。但是，这一技术的算法运行逻辑有一个隐秘层面，很难被探索，也就是我们常说的“黑箱”。它指的是人

们不知道的既无法开启、也不可能直接从外到内观察系统内部状态。而所谓的“算法黑箱”，就是指在机器深度学习过程中，它那错综复杂的神经元网络，其中有一些隐层并没有被人们直观地捕获。人工智能的迅猛发展以及人类认知能力的限制，这就造成了人工智能技术能力与人的认知能力之间的鸿沟越来越大。基于机器学习的人工智能多使用黑箱模式，人们无法准确地了解它的运作，解释不了算法自主决策结果的形成过程，存在着决策模糊性明显以及结果的不可解释性等。人工智能在其自身的发展历程中，一种公认的看法认为，最精准的模型固有其不可解释性与复杂性，也就是说，不可解释性对于达到决策精确度至关重要。针于使用黑箱模式进行机器学习的人工智能，以其为基础的设计者一般也不能清楚地解释各变量之间的结合及相互影响，由此得到最后决策结果。

人工智能的黑箱模式有以下几方面：第一，人工智能的技术与其应用场景中产生的某些风险具有不可解释性。由于人类思维方式和认知能力存在差异，使得人工智能无法对复杂问题进行准确地推理。人工智能主要通过对大数据的分析，发现数据变量之间的关联性，但它自身并不能很好地解释数据变量间的因果关系，这多少显示了人工智能技术不可解释的特点。同时，由于人工智能系统自身存在缺陷，导致它不能正确地理解和处理复杂问题，因而也会使其出现难以预测和控制的情况。二是人工智能系统结果的不可解释性。由于人工智能系统自身存在着一些缺陷，比如计算机复杂性较高、信息隐藏性差等因素影响了人工智能系统对知识的获取以及推理过错的实现。人工智能决策由机器独立进行，人工智能系统的运作与人类并无相同且相似的推理逻辑，人类对人工智能的工作过程及结果并不能完全了解。同时，由于计算机自身的局限性以及人工智能研究人员缺乏必要的专业知识和经验等原因导致人工智能系统的运行出现偏差甚至失效。三是人工智能系统工作的动态性与结果的不确定性。人工智能的发展与社会经济环境紧密相关，随着人工智能技术的不断成熟，人工智能系统的应用范围也在不断扩大。人工智能多采用取机器自主学习模式，该算法依据环境进行自主学习和动态优化，这样就使人工智能系统处在不断地进行着操作和优化，人们由此不能准确地预测自己的运行结果。最后是人工智能用户故意没有披露算法的操作过错，从而产生模糊性。由于技术发展水平的限制和人工智能本身的复杂性，导致人工智能系统存在着明显的“黑箱”模式。外部无法了解人工智能算法软件的源代码和运行逻辑，原因是知识产权法律保护和人工智能系统使用者出于商业秘密保护，成为人工智能系统风险的来源。

2.2. 人工智能系统运作造成的风险

人工智能系统将输入的数据转化为有用的输出，从而辅助人们做出更好的决策，这是基于复杂的计算和程序运行。人工智能系统本质上是一个投入与产出的过程，包括系统开发、数据投入、算法运算和决策结果的产出与运用等，其中任何一个环节的不足都可能导致系统风险，会对社会造成各方面的危害。例如人工智能系统设计缺陷、输入数据质量缺陷、算法操作错误和解读或使用算法结果错误[6]。

一是人工智能系统开发设计的风险主要表现在：1)、算法结构的缺陷，如错误的假设、算法逻辑结构的缺陷等，由于人的认知能力受到限制而导致算法设计缺陷；2)、算法技术代码缺陷，导致模型技术不当、代码出错、算法程序设计出错等算法运行中产生偏差的结果；3)、算法技术的发展单纯以技术或经济目的为中心，社会规则和伦理规范缺失，造成人们的偏见很容易被带入到算法开发中的人工智能算法程序中。二是人工智能系统数据投入的风险主要取决于数据质量，而算法训练和决策结果在很大程度上取决于数据质量。在算法运行时必须保证数据具有足够数量的一致性，否则将会造成错误的结论甚至完全相反的判断。算法中多用到的数据被划分为投入数据、训练数据和反馈数据三大类。投入数据指数据集中所有被标注为正确或错误的样本。如数据不够全面、不够及时、缺乏关联，或者由于偏见存在已久，造成了数据自身含有历史性偏差或者行为性偏差，产生选择性偏差、数据质量不到位和数据采集技术运用不恰当等等，均可能导致数据集存在质量缺陷。在进行决策分析时，必须对数据集中所有可能存

在的问题进行检测并剔除其中一部分。数据集上的微小瑕疵就可能造成系统性偏差结果，从而在人工智能依赖数据训练算法和决策的过程中产生算法决策偏差的风险。三是人工智能操作风险主要体现在：1)、任务分配出现偏差。这主要是用户对算法程序使用目的的理解有误，或对任务分配有误，造成人工智能系统在运行时与初衷产生偏差；2)、虚假模式等情况出现在人工智能系统的训练和验证中。算法在技术上欠缺严密性，在概念上欠缺合理性，致使人工智能算法正在被训练中、在试验或检测时输出是错误的，因而出现了数据使用方面的偏差；3)、人工智能算法系统存在安全漏洞。人工智能算法设计中的安全漏洞，使得算法系统在运行过程中受到问题或不正确的外部攻击造成黑客攻击得以实现。四是人工智能算法产出的应用风险，主要表现为算法结果的解读不当，算法使用者由于对算法的误用或过度依赖，对算法结果的错误场景应用的前提假设，以及对算法结果的应用场景需求被忽略等等。例如，无人驾驶汽车主要是汽车驾驶员对自动驾驶系统的过分信任而导致交通事故的发生。此外，把人类偏见引入人机交互的人工智能系统中、人工智能系统用户的恶意使用、误有用或选择地使用等等，均有可能带来风险。

3. 人工智能的法律监管治理挑战

人工智能时代的崛起作为一种普遍现象，它所带来的的治理挑战，是各个国家共同面临的难题。为规范人工智能的快速发展更好的对其进行法律监管，国外陆续有一些人工智能监管相关的法律文件发布，为的是积极应对人工智能的监管。美国自 2019 年起陆续发布了《2019 年算法问责法案》[1]和《人工智能应用监管指南备忘录(草案)》[2]，《人工智能与数据保护指南》[3]于 2020 年由英国制定并颁布。《欧盟数据战略》[4]和《人工智能法》[5]等相关法律法规与政策文件于 2020 年先后在欧盟发布，主要是为了能够加强人工智能法律监管方面的治理，从而构建稳定的和有利于人工智能创新发展的法律监管环境，进而为划分人工智能风险等级制度提供制度性的保障。

目前，我国已经在探索与改进人工智能监管方面的法律法规了，在《中华人民共和国网络安全法(含草案说明)》和《中华人民共和国数据安全法(含草案说明)》等法律法规之中发布了与人工智能法律监管相关的条款，以此来加强网络和数据安全监管，但是相对来说，人工智能监管的法律制度的提供仍然是比较欠缺的。中华人民共和国科学技术部于 2019 年和 2021 年发布了《新一代人工智能治理原则——发展负责任的人工智能》[6]和《新一代人工智能伦理规范》[7]两部规范人工智能法律监管的指引性文件，2021 年国家市场监督管理总局发布的《互联网平台落实主体责任指南(征求意见稿)》[8]第 19 条专门规定了“算法规制”问题。总体而言，我国现阶段还没有起草制定《人工智能法》等专门法律来监管人工智能，对人工智能进行系统监管的法律还比较欠缺。对于人工智能本身的风险等级也未明确规定，这就使得监管人工智能遭遇多重障碍。在没有对人工智能进行分类分级的情况下，人工智能领域法律监管主体不明确、

[1]Congress of the United States, Algorithmic Accountability Act of 2019, 2019-10-17, https://www.congress.gov/bill/116th-congress/house-bill/2231/text?__cf_chl_jschl_tk__=ZlDyqi04ow0qGZPxgOSzCiMcY1JCNF.jjXt1WsrIKGM-%20164023%206657-0-gaNycGzNCf0, 2022-8-30.
[2]The White House, Guidance for Regulation of Artificial Intelligence Applications, 2020-03, https://ieeeusa.org/?s=The+White+House%2C+Guidance+for+Regulation+of+Artificial+Intelligence+Applications, 2022-8-30.
[3]ICO, Guidance on AI and Data Protection, 2020-07, https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/, 2022-8-31.
[4]European Commission, European Data Strategy, 2020-02, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en, 2022-8-31.
[5]European Commission, Artificial Intelligence Act, 2021-05, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, 2022-8-31.
[6]国家新一代人工智能治理专业委员会：《新一代人工智能治理原则——发展负责任的人工智能》，2019 年 6 月 17 日，http://www.most.gov.cn/kjbgz/201906/t20190617_147107.htm，2022 年 8 月 31 日。
[7]国家新一代人工智能治理专业委员会：《新一代人工智能伦理规范》，2021 年 9 月 26 日，https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html，2022 年 8 月 31 日。
[8]国家市场监督管理总局：《互联网平台落实主体责任指南(征求意见稿)》，2021 年 10 月 29 日，https://www.samr.gov.cn/hd/zjdc/202110/t20211027_336137.html，2022 年 8 月 31 日。

模式不明确、内容不明确和监管责任不明确等一系列问题。未来，随着国家对人工智能产业的发展重视程度的不断提升以及技术进步推动行业快速发展，并逐步完善人工智能相关法律法规和标准体系，有利于促进智能社会建设，实现对人工智能法律监管的有效规制。对于人工智能的监管治理的研究也成为了国内外关注的热点，许多学者从社会、经济以及伦理等多个视角出发来研究人工智能所带来的各种风险。目前已有研究关注三个方面：一是人工智能的不透明性、不可解释性以及由此带来的风险。“Burrell 认为人工智能算法模糊、透明度不足是算法治理中需要解决的关键问题”[7]。“Zarsky 认为人工智能算法在决策过程中存在模糊性与自动性，这将诱发降低经济效率与影响公平两方面的风险。”二是构建整体框架，对人工智能算法进行调控和治理。“贾开和蒋余浩提出了以算法治理为核心的议题，即算法制定权和相应的监管程序，通过分析人工智能的技术性特征治理所面临的挑战”[8]。“吴汉东建议应建立以伦理为首要，以安全为主线，以技术和法律为主导的风险防控机制”[9]。“张凌寒认为应当构建限权与赋权相统一的算法治理制度，并明确算法适用范围、赋予公民个人数据权利和救济权、加强行业自律、构建合作治理体制和算法责任追究体系等建议”[10]。三是聚焦人工智能风险等级监管治理政策。“沈伟伟认为应构建以算法透明所代表的算法治理体系，把事前监管和以算法问责为表达的事后监管结合起来”[11]。“刘友华认为应强化算法使用者和设计者之间的责任，遵循公平、透明和可问责的原则”[12]。

总之，关于人工智能风险的法律监管已经成为研究的理论热点。为了应对人工智能带来的如技术滥用、违反伦理道德或歧视等各种风险，需要加强人工智能发展过程中的风险防控的力度以应对人工智能时代带来的各种挑战。

4. 人工智能法律监管体系构建

人工智能是把双刃剑，它为社会提供了诸多利益，但也造成了巨大的危害。它既可以帮助人类提高工作效率，同时又可能产生一些负面后果。所造成的危害，如不加以制约，会给社会造成巨大的破坏。因此，我们应该积极采取对策来防止人工智能发展可能引发的各种风险。政府在人工智能治理中必须扮演好主体角色，人工智能系统法律监管有效，防范多种潜在风险，使科技更能推动社会的进步，由于人工智能短发具有模糊性特点、系统性与使用者谋求私利，单纯靠企业或者用户的努力，显然无法很好的处理人工智能系统应用中存在的各种风险。因此，政府应该积极介入带人工智能的法律监管中来，加强自身建设，提高社会公信力。人工智能监管治理，就是要通过出台一系列法律、规则和政策来实现，建构多元主体共同参与治理体系才能确保安全、信任与有利于社会的途径，在最大程度上杜绝了人工智能系统在应用于发展过程中产生的法律风险。

4.1. 实行以人工智能风险等级为基础的监管

由于国内法律对人工智能风险等级并没有划分，我们可以结合我国人工智能系统的应用情况，从国外已有的相关法规中研究人工智能系统监管条例。例如，借鉴欧盟《人工智能法》中划分人工智能系统风险等级的方法，找到相关的理论依据，对我国人工智能系统进行分级监管。首先要明确 AI 系统的定义：是指针对一群特定人群定义的目标，能够产生输出的软件，例如内容、预测、建议、决定影响等，使用一种或多种方法和技术研发出来的程序。其次，将风险等级落实到不同应用场景的人工智能系统上，并提出不同的监管路径。将人工智能系统划分为四个风险等级，分别为“不可接受、高、有限、极小”：第一，将人工智能系统视为“不可接受，这是对人的基本人权和社会公平造成显著威胁划分的”；第二，将人工智能系统视为“高风险”，这是基于对公民人身安全和生活基本权利场构成威胁划分的；第三，将具有透明义务的人工智能系统视为“有限风险”；第四，将人工智能系统视为“极小风险”，该系统提供简单的工具功能。“有限风险”具体包括，在使用人工智能系统时，能够意识到其本身正在与人工

智能系统进行交互，这类系统需要履行用户的透明义务，如用户告知、保障用户的知情权和选择权等，采用事前披露和事后控制的处理方法在风险产生后。“极小风险”如提供电子游戏、短视频、邮箱等简单功能，由于这些人工智能系统对公民权利和财产安全的风险很小，因此这样的人工智能系统是不实施干预的。“不可接受风险”这类系统应绝对禁止在国内市场上的部署和应用，其具体应用场景包括：1) 公共机构开展的社会信用评分；2) 基于执法目的的非特殊情况下的公众场所进行实时远程生物识别；3) 操纵人类意识和行为的场景，例如平台算法推荐。(如大数据杀熟)；4) 算法偏见的应用场景会对儿童及残疾人等弱势群体造成危害。“高风险”这类系统实行的是事前评估和风险预警，其具体应用场景包括关键基础设施领域、就业领域、教育和职业培训等领域。[9]

《关键信息基础设施安全保护条例》[10] 对铁路、民航、邮政、水利、金融等行业进行了规定，而卫生健康、社会保障和国防科技工业等是我国网络安全的重点行业。国家鼓励其他重要行业制定相关标准或指南，加强对重要行业网络与信息安全保障工作的指导。《中华人民共和国数据安全法(含草案说明)》对国家核心数据进行了界定，也就是事关国家安全、国家的经济命脉、重要民生、重大公共利益及其他数据。其中，“国家核心数据包括国家安全领域中与经济发展密切相关的各类数据”。“国家核心数据分为基础型、战略型、创新型三个层次，分别对应不同级别的安全需求”。“关键信息基础设施”和“国家核心数据”主要以国家公共安全为依据，明确监管重点，所以“关键信息基础设施”和“国家核心数据”并非适合人工智能的监管，由于人工智能所带来的风险，既包括私人权益受损，还存在公共利益受损问题，还涉及到社会伦理方面的内容。对于人工智能领域中的可接受风险以及高风险人工智能能当区别对待。对于不可接受风险和高风险人工智能，特别是对不可接受风险的人工智能应用，有必要通过《人工智能法》的出台，对其作出进一步的明确。

4.2. 构建一体化的人工智能政府监管体制

事前以制定安全标准为主的工智能系统应用评估认证制度。以防患于未然，建立人工智能系统研发和应用的基本质量和安全标准，对于高风险人工智能系统，要执行必要认证批准程序和达标认证制度。建立第三方实施人工智能安全达标评估认证体系，高风险人工智能系统须经专业机构人工智能系统的安全评估认证，未经安全评估或者达不到标准的人工智能系统是无法投入使用的。

事中着力强化监管机构人工智能系统应用风险监测与合规审核工作，以强化系统使用者的主体责任为主，需要它在内部风险管理中建立起有效技术，然后再组织保障。由于人工智能系统具有动态性，这表明，一次性的检查和审核可能不久即被淘汰。所以有必要对人工智能系统进行不断的监测，对全生命周期进行动态监测和审核。人工智能系统用户对监管机构负有公开放数据界面、算法源代码、算法学习机制和算法运行结果等的责任，为便于监管机构进行算法监控和审查。由于算法审计属于专业较强的工作，因此政府监管机构和第三方机构之间存在保密义务，为了保障算法知识产权，保护人工智能系统用户的利益，监管机构可以授权交由第三方专业机构实施。

强化事后不法行为的追责力度。加大对人工智能应用违法行为的惩罚力度，本文着重探讨了如下问题：1)、厘清人工智能系统用户的法律主体责任。尽管人工智能算法的运行是机器基于神经网络自主学习来实现的，没有人为干预，而用户保障人工智能安全性和可信性等主要职责的人工智能算法用户，仍然无法以此免责。2)、对不遵从数据治理规则，触犯“红线”要求的违规企业，会受到相应的处罚。增加罚款额度江会减少违法激励，当违法违规行为被查出时，按照法经济学的最优处罚理论来惩罚[13]。例

[9]European Commission, Artificial Intelligence Act, 2021-05, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206&qid=1683794129650, 2022-9-26.
[10]中华人民共和国国务院：《关键信息基础设施安全保护条例》，2021 年 7 月 30 日，http://www.cac.gov.cn/2021-08/17/c_1630785976988160.htm，2022 年 9 月 26 日。

如，欧盟《人工智能法》[11]规定，对于企业违法可处以3000万欧元以上或者违法企业前一年度全球市场营业额6%以下的罚款。约谈、整改、营业资格限制、行政罚款和刑事责任等各项措施，已被确定在我国《中华人民共和国网络安全法(含草案说明)》《中华人民共和国数据安全法(含草案说明)》等法律中，但是需要进一步明确各项措施的具体应用条件，尤其是不利于重大违法行为进行严厉处罚的行政罚款上限，并提出应设立与违法损害相适应的罚款等级，对于违法企业以营业额百分比作为罚款额的计算方法，以更加有力地威慑违法行为。3)、确立民事损害赔偿制度。因在使用人工智能系统过程中而严重损害公民个人权益的违法行为，个人或者组织为维护自身的合法权益，其应当有权享有提起民事诉讼的权利。

4.3. 构建人工智能风险评估监管机制

现有法律并不能解决或者防范所有人工智能领域所带来的潜在风险，监督机构、司法机构以及人工智能参与者必须在某些方面达成一定的监管共识，实现可信赖的人工智能的监管目标，并最终促进人工智能行业的蓬勃发展。针对风险等级视野下的人工智能，监管机构可预先通过风险判断或风险评估，从而确定对具体的人工智能的监管力度和监管标准。

具体步骤如下：首先，通过对人工智能所处的行业领域，分析对人工智能的监管力度和监管标准。一般情况下，与公共政策、公共服务相关的领域，人工智能使用的场景直接或者间接牵涉到大多数人的利益一般被评估为“高风险”而被监管机构给予高度关注。比如我国卫生部针对通过人工智能施行医疗辅助诊断，以及出台了医疗辅助治疗相应的规范标准，这是对“高风险”人工智能法律监管的一个典型示例。此外，对于工业应用程序而言，由于其发生风险安全事件可能给公众造成重大损害，一般也将其评估为“高风险”行业进行监管。其次，通过对人工智能的具体运用及适用用户群体来确定监管力度和监管标准。某些情况下，在“高风险”行业领域的应有并非是产品或者服务是“高风险”的，比如公共交通领域的预约订票系统，虽然人工智能在公共交通领域属于“高风险”，但是在保证数据和信息安全的情况下，此运用不会对公众造成重大风险。同理，一些“非高风险”的行业领域的具体运用，可能存在较高的安全风险。比如针对儿童的智能穿戴设备，可能因系统的缺陷，导致儿童被追踪或者权利受到侵犯，那么此款智能设备对于儿童群体来说就属于高风险。又如，对于需要特殊照顾的老年人或者不能自理者，一些人工智能辅助设备在使用过程可能由于忽略用户的特殊情况而导致较高的风险。以人工智能的运用场景来评估风险等级更有利于监管机构精准对不同行业领域因人工智能导致的损害程度，从而综合确定分析适当的监管规则使风险最小化。最后，对于“非高风险”的人工智能应用，监管力度可适当减小是企业有更大的自我管理和发展。同时可借鉴国外成果经验，比如引进“自愿认证计划”。欧盟委员会高级别组在其公布的《关于人工智能、物联网和机器人对安全和责任影响报告》中，提及对于一部分人工智能项目提倡开展“自愿认证计划”。我国《数据安全管理法(征求意见稿)》规定“国家促进鼓励网络运营者自愿通过数据安全管理认证和应用程序安全认证，鼓励搜素引擎、应用商店等明确标识并优先推荐通过认证的应用程序。”[12]表明在数据安全领域，国家正在开始“自愿认证”的监管尝试。

5. 结论

站在人工智能时代的风口，全球都在顶层布局，占领人工智能领域制高点，我国是一个优势大国，数据资源丰富，我们要牢牢把握时代发展的机遇，探索建立与人工智能领域相适应的法律监管制度，营造人工智能监管治理的良好环境。目前我国在人工智能监管方面陆续出台了多部政策文件，但其规定过

[11]European Commission, Artificial Intelligence Act, 2021-05, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex, 2022-9-26.
[12]中华人民共和国国家互联网信息办公室：《数据安全管理办法(征求意见稿)》，2019年5月28日，http://www.cac.gov.cn/2019-05/28/c_1124546022.htm，2023年2月28日。

于简单，且没有明确相关条款适用于人工智能的发展阶段。所以我国要不断地创新监管制度，建构相关法律监管制度等，健全相关法律法规，以规范人工智能开发和应用，提升法律监管体系在人工智能领域防范和控制不同风险等级的能力，进而能够有效避免人工智能在开发和应用过程中存在的风险，并对可能产生的冲突和争议进行有效的解决和消解。同时，通过加强立法、健全技术支撑体系、强化执法等方式来提升我国的人工智能监管水平，最终达到促进我国经济发展、社会稳定的目的。综上所述，针对人工智能的不同风险等级进行划分，才能有效监管，从而达到人机和谐共存的目的，促进人类的共同发展。

参考文献

[1] 唐要家, 尹钰峰. 算法合谋的反垄断规制及工具创新研究[J]. 产经评论, 2020(2): 6.

[2] Daron, A. and Pascual, R. (2017) Robots and Jobs: Evidence from US Labor Markets. NBER Working Paper 23285, 35.

[3] Hémous, D. and Olsen, M. (2016) The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality, Barcelona: IESE Business School Working Paper No. WP1110-E, 1.

[4] Aghion, P., Jones, B.F. and Jones, C.I. (2019) Artificial Intelligence and Economic Growth. In: Goldfarb, A.G., Eds., *The Economics of Artificial Intelligence*: *An Agenda*, National Bureau of Economic Research, Inc., Chicago, 237-290. https://doi.org/10.7208/chicago/9780226613475.003.0009

[5] Corinne, C. (2018) Artificial Intelligence and the Good Society. *Science and Engineering Ethics*, **24**, 505-528.

[6] 唐要家, 唐春晖. 基于人工智能监管治理[J]. 社会科学期刊, 2022(1): 116.

[7] Jenna, B. (2016) How the Machine Thinks: Understanding Opacity in Machine Learning Systems. *Big Data and Society*, **3**, 1-12. https://doi.org/10.1177/2053951715622512

[8] 贾开, 蒋余浩. 人工智能治理的三个基本问题: 技术逻辑、风险挑战与公共政策选择[J]. 中国行政管理, 2017(10): 40-44.

[9] 吴汉东. 人工智能时代的制度安排与法律规制[J]. 法律科学(西北政法大学学报), 2017(5): 128-136.

[10] 张凌寒. 算法权力的兴起、异化及法律规制[J]. 法商研究, 2019(4): 63-75.

[11] 沈伟伟. 算法透明原则的迷思——算法规制理论的批判[J]. 环球法律评论, 2019(6): 20-39.

[12] 刘友华. 算法偏见及其规制路径研究[J]. 法学杂志, 2019(6): 55-66.

[13] Becker, G.S. and Stigler, G.J. (1974) Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Economics*, **98**, 371-400.