

# 人工智能董事法律地位及其信义义务构想

——基于强人工智能视角

刘成杰

**摘要:** 人工智能技术的革新速度与应用场景改变了我们的传统认知,人工智能董事已从理论迈入现实。我国的相关发展规划已提出“明确人工智能法律主体以及相关权利、义务和责任”的人工智能发展保障措施。在强人工智能语境下,应摒弃客体说观点,通过法律拟制赋予人工智能董事法律主体地位。基于人工智能董事的特殊性,传统董事忠实义务趋于消解,注意义务“理性人”标准适用性也将存疑。以前瞻性视角审视和革新人工智能董事的信义义务,忠实义务需因应特殊情形,注意义务的范围则呈窄域化倾向,合规监督义务的内涵和边界将实质扩展。

**关键词:** 强人工智能;人工智能董事;法律主体;信义义务

**中图分类号:** D922.29 **文献标识码:** A **文章编号:** 1003-0751(2024)02-0079-08

人工智能通常是指由机器或其他智能系统执行本由自然人方可完成的智能行为,本质在于“让计算机系统完成人类智能完成的各类事情”<sup>[1]</sup>。一般认为其有弱人工智能、强人工智能和超人工智能之分,强人工智能已具备自主认知、决策和行为的能力,不仅能产生相应的数据以支持其持续的行为,还拥有自我辨认和控制能力<sup>[2]</sup>。20世纪70年代未来学家托夫勒曾预言,人工智能时代总时间应不足百年,而今已行将过半;ChatGPT(一款聊天机器人程序)的出现及快速迭代已成为“强人工智能时代的里程碑”<sup>[3]</sup>,人类即将进入强人工智能时代<sup>[4]</sup>。

基于人工智能强大的数据处理和分析能力,特别是在战略规划与市场预测方面的优异表现,一些公司开始尝试将人工智能引入决策层,用于辅助甚至替代传统董事会的某些职能。在这一背景下,“人工智能董事”的概念应运而生。其实,早在2014年香港一家名为“深度知识创业”的风险投资基金,曾任命一唤作“VITAL”的人工智能体参与公司投资决策,赋予其“投票决定公司是否投资特定公司”等

决策权利,被称为“世界首个机器人公司董事”<sup>[5]</sup>。人工智能董事是人工智能技术在公司治理场景的具体呈现,这种新兴的治理模式不仅仅是技术和模式的革新,同时也在法律和伦理上带来诸多问题和挑战。如人工智能董事是否应承担与自然人董事相同的信义义务?如是,其义务内涵和构成内容是什么?若否,又该如何防止其可能带来的潜在风险或可能的滥用?这些问题不仅关系到公司的决策效力和责任归属,更影响到其自身在公司治理场景中的良性发展,乃至整个社会对人工智能技术的信任度。

有研究预测,再经过25年左右的发展,非生物“智能”将会10亿倍于人类的“智慧”<sup>[6]</sup>。2017年,国务院印发《关于新一代人工智能发展规划的通知》,提出把人工智能发展置于国家战略层面系统布局,牢牢把握战略主动,并要求明确人工智能法律主体以及相关权利、义务和责任等,加快研究制定相关法规,为新技术的快速应用奠定法律基础<sup>[7]</sup>。2019年《新一代人工智能治理原则——发展负责任的人工智能》进一步提出“敏捷治理”等8项基本原

收稿日期:2023-11-10

作者简介:刘成杰,男,北京科技大学法律系副教授(北京 100083)。

则。2023年7月,我国颁行《生成式人工智能服务管理暂行办法》,成为全球首部该领域的法规。其实,世界范围内各方对人工智能治理的研究和认识均在不断深化,立法行动也在提速,号称全球首部人工智能综合治理立法的欧盟《人工智能法案》将于2024年初生效并于2026年正式实施<sup>[8]</sup>。

而今,人工智能体,将越来越像一个集合了“工具”和“心智”的“生灵”<sup>[9]</sup>。在人工智能改变传统边界的强大创新趋势之下,人工智能董事作为一种具有未来意义的新生事物,基于强人工智能的视角对其法律层面合理定位及其信义义务的前瞻性探讨,将有助于我们更好地理解人工智能技术在公司治理中的作用,同时也为未来的立法和司法实践提供有价值的参考。

## 一、人工智能董事的法律地位界定

对于弱人工智能体的“工具”定性与法律客体定位,学界已达成基本共识。当前,“有意义的是去思考已经露出端倪的强人工智能体的法律地位问题”<sup>[10]</sup>。从本质而言,人工智能董事属于人工智能体之一<sup>[11]</sup>,梳理分析关于强人工智能体的学说观点,有助于我们更为准确地把握和界定人工智能董事的法律地位。

### (一) 强人工智能定性的主体化趋势

“就发展趋势而言,对于精巧的自主性人工智能体,应赋予特殊的法律地位”的观点,其实已不再是纯粹的学术上讨论,已经体现于域外法案<sup>①</sup>。“强人工智能”概念,来源于数学家 Irving John Good 在1965年提出的“智能爆炸”概念,其中提到“超智能机器”<sup>[12]</sup>。经过数十年的发展,在比较法视野下,域外除智能爆炸说,围绕强人工智能还存在智能融合说、加速进化说等学说。各学说虽然观点、主张各异,但强人工智能体不应再被视为客体之“物”,而应确立其主体地位,并赋予其权利、义务等社会属性,成为共识性的意见<sup>[13]</sup>。

当前,国内学术界则有不少学者持客体说,认为人工智能体虽具有高度智性,却不具“心性”和“灵性”,其“智性”仅为模拟和扩展人类智能,尚不足以赋予其独立的法律主体地位<sup>[14]</sup>。部分持此学说的学者虽认可“强人工智能乃人类级”,但认为其终究受人类“终极管控”,因而仍应被视为客体<sup>[15]</sup>。还有持此观点的学者以新的研究范式论证,认为强人工智能体或与“理性”法律主体的预设相契合,但基

于拉康的“欲望主体理论”,人工智能仅是人类技术理性的延伸,并不具备欲望的机制,因而很难成为适格的法律主体<sup>[16]</sup>。与客体说观点相对,主体说的支持者认为,“物”具有“被支配性”特征是共识性观点<sup>[17]</sup>,但人工智能是基于复杂神经网络进化路径而实现的,这意味着开发者虽然开发和设计了人工智能体,但随着不断的“数据训练”和“深度学习”,其优化过程其实更多是“进化”而不是“设计”,开发者本人可能也无法清楚解释人工智能的完整决策思路<sup>[18]</sup>。当初,“法人人格制度”的出现源于经济发展需要,促进了立法技术的进步。那么,对于强人工智能体,是否也可以通过“法律拟制”的方式,赋予其一种类似于人类或法人的法律主体地位?支持者还指出,随着技术的发展及元宇宙概念的出现,强人工智能将越来越具备“自主性”的行为能力,因而应将其纳入法律范畴来规范其行为并保护其权益,超越人类中心主义的藩篱,借鉴“法人人格制度”解决当前人工智能相关法理与法律适用的困境<sup>[19]</sup>。

关于对客体说的回应,尤其是基于“心性”“灵性”以及“终极管控”等否定强人工智能体法律主体地位的观点,我们还可以从民法对于胎儿和无民事行为能力人的“权利主体”确认中获得启示。虽然强人工智能体可能并不具备与人类完全相同的“心性”和“灵性”,但其作为高度智能化和自主性的存在,应借鉴民法中对于特殊实体的处理方式,对强人工智能体的法律地位进行更为理性的审视。其实,强人工智能体与公司法人“是一种经济生活的客观现实与法律技术运用相结合的产物”<sup>[20]</sup>的定位并无本质不同。有研究提出,强人工智能语境下,无论从主体性层面还是法理层面,赋予其法律主体地位的条件均已成熟<sup>[21]</sup>。因而,结合上述主体说的说理,我们应该持更为开放和包容的态度,基于“法律拟制”赋予其法律主体地位和相应行为能力。这样既可以为其自身和第三人提供必要的权益保护,在宏观上也有利于促进人工智能的合理应用和良性发展<sup>[22]</sup>,在很大程度上也体现出对人类未来社会发展的负责与担当。

### (二) 人工智能董事法律主体地位辩明

人工智能在智能家居、自动驾驶,以及医疗和金融领域的应用,已逐渐为人所熟知和接受。但人工智能进入企业并参与决策,尤其实质替代董事的新兴现象,挑战了我们对传统公司治理的认知。其实,具有标志性意义的人工智能董事“VITAL”的出现,距今已近10年。在这期间,人工智能技术已取得长

足发展和跃升,尤其生成式人工智能的发展现状,已远非10年前可以想象。世界经济组织发布的《深度转变:技术转折点与社会影响》报告显示,近一半受访者预计第一台真正能胜任董事职能的人工智能体将在2025年之前加入公司董事会<sup>[23]</sup>。的确,即便弱人工智能技术,也能够有效促进公司治理,特别是在复杂数据分析以及财务重述治理等方面展现强大能力<sup>[24]</sup>。鉴于此,基于上述强人工智能主体地位的探讨,我们或许应展现出更大的灵活性和适应性,界定其法律主体定位,以促进人工智能技术在公司治理中的价值发挥。

### 1. 人工智能董事法律主体定位的可行性

探讨人工智能董事的法律主体定位,无法回避两个根本性问题<sup>②</sup>:一是人工智能董事是否具有基于理性的“自主性”,二是实践中人工智能董事违反义务时如何“担责”。对于这两个问题,我们可以从技术可行性、实践可行性以及人工智能担责路径三个方面来具体分析。

第一,技术可行性。基于现行法律体系及法学理论,一般认为,无论是自然人还是法律拟制的法人,成为法律主体的基础和前提,是具有“人性”特征的自主理性。其实,人工智能“弱-强-超”的发展路径,已经清晰预示了未来人工智能“主体地位”的总体发展趋势。就本质而言,人工智能董事是在强人工智能与商业智能决策系统(BI)结合的基础上,实现决策自主性且具备智能交互功能的智能体,其“自主”实现的关键在于以算法为中心的智能技术的提升。中国科学院张钹院士的研究成果显示,“安全、可信、可靠与可扩展”的第三代人工智能,将第一代与第二代人工智能的“知识驱动”和“数据驱动”实现优势互补和结合,更进一步同时运用“知识、数据、算法和算力”4个要素,具有远超前两代人工智能的强大性能,第三代人工智能在决策中会“选择最优的策略,使得预期的累计奖励最大,即值函数取得最优值”<sup>[25]</sup>。可见,相关技术当前已实现一定程度的突破。

第二,实践可行性。在现代企业治理结构中,董事会是公司的决策核心,其决策往往代表着公司的战略方向 and 未来发展。但这些决策并不是由某一位董事单独作出,而是基于董事会成员集体的智慧、经验和判断。每位董事,根据其独特的知识背景、专业领域、与各种资源的联系以及个人的逻辑推理能力,均为决策过程带来了独特的视角和思考。客观而言,也正是基于这样的多样性,不同董事之间可能存

在认知上的偏见或差异。这些认知上的偏差,无论是个人的经验、文化背景还是心理因素所造成,都有可能影响到最终决策的公正性和客观性。基于董事会成员的多元化和决策的多维性,即便人工智能董事偶尔存在疏漏和出现偏差,也不会在本质上或持续地影响公司的整体决策或长远利益,人工智能董事的引入,或许能为董事会决策提供一个相对客观且实用的平衡机制。人工智能董事凭借其强大的数据分析能力、模式识别技能,以及对庞杂、多元信息的快速处理能力,更易于提高董事会决策的效率,拓展决策思维的深度与广度。因而,这种结合人工智能与人类智慧的决策,更加有助于促进董事会决策的全面性、客观性和稳健性。

第三,责任承担问题。人工智能董事具有特定法律人格时,确实还存在一个核心问题:如其决策失误或违法行为导致公司或第三方受损,如何担责。其实,在公司治理之外的领域,这可能难以解决,但在公司治理领域,这似乎并不是一个不可逾越的障碍。关于董事责任问题,国内外公司治理实践已经提出多种解决方案。首先,董事责任保险制度,作为一种特殊的职业责任保险,旨在为公司的董事、监事、高级管理人员等相关决策人员在履行职务过程中,因实际或被指控的疏忽、错误、误导性陈述、遗漏、违反职责或其他不当行为导致的个人及公司损失提供经济保障。例如,“康美药业案”中的独立董事追责判决,促使董事责任保险制度在我国逐渐得到更为广泛的认可和接受,某公司年投保保额已达7.5亿元<sup>[26]</sup>。可以预见,相关制度嵌入公司治理会逐渐趋于成熟<sup>[27]</sup>。此制度的完善,无疑为人工智能董事的法律主体地位确立奠定了较好的现实基础。此外,欧洲人工智能领域通过构建“一项基金来补充保险机制,以确保在没有保险覆盖的情况下可以对损害进行赔偿”<sup>③</sup>的制度理念,也为我们提供了人工智能董事责任问题解决的视角及可行路径。例如,可以针对用于人工智能董事的智能体设计专门机制,要求人工智能体生产商或相关主体在将其投放市场时一次性支付一笔费用,以创建一种此类智能体的专用补偿基金。

总体而言,基于人工智能技术的发展趋势,在可预期的一段时期,强人工智能体能够在复杂多变的市场环境中迅速响应并作出决策;而且,已有的实践应用,也显示出其在公司治理中的巨大潜力。同时,如能合理设计和架构董事责任险机制及补偿基金制度等机制,人工智能董事的责任归属难题和法律执

行困境也将得到化解。最为重要的是,当前人工智能董事即将甚至已经走上台前,顺应人工智能快速发展迭代的趋势,赋予强人工智能之人工智能董事法律主体地位,似更为务实和开明。

## 2.人工智能董事法律主体地位的逻辑自洽性

赋予强人工智能体属性的人工智能董事法律主体地位,还需充分考虑公司场景的特殊性,以及人工智能董事这一新兴事物可能带来法学理论与常识方面的挑战。强人工智能体在公司的“任命-解聘”流程中,可能会经历从客体到主体再到客体的身份转变,且在任职期间还可能存主客体身份的双维度并存现象。为确保人工智能董事法律主体地位的逻辑自洽性,需要借助相关法学理念或理论进行合理阐释。

第一,关于客体到主体再到客体的身份转变问题。罗马法上存在“位格减等”制度,或可以对问题解决提供启示与合理解释路径。罗马法上“位格减等”通常译为“人格减等”,即具有完整法律人格的罗马公民,若自由权、市民权和家族权三类身份权其一或其二丧失,则其会成为法律位格不完整者;当三项皆失,则“人格大减等”而沦为奴隶,法律意义上等同死亡。该“生物机体与其法律位格相分离”理论,为现代法人的法律拟制逻辑提供了关键技术支持,现代法人制度恰似“位格减等”的逆向机制,运用“位格加等”逻辑赋予了公司法律主体地位。可见,当法律位格的赋予仅为法律技术运用的产物,与主体是否为生命体无必然联系时,无生命体获得法律位格的可能性便得以出现<sup>[28]</sup>。

循此法理,强人工智能体在公司的“任命-解聘”流程中,并无实质障碍。人工智能董事的法律主体地位,是从其获得公司任命之时开始成立,获得任命之前,其本质上仅属于特定法律关系的客体。自公司正式任命时起,类似“位格加等”,人工智能董事始成为公司治理结构中的一部分,成为具有法律主体地位的“董事”,享有法律法规和公司章程规定的相应权利和承担特定的义务。但当公司决定解聘人工智能董事时,“位格减等”,其法律主体地位也随之结束回复客体身份。

第二,关于主客体身份的双维度并存现象的理论纾解。所谓“双维度并存”,是指强人工智能体获公司任命至解聘之间,其存在“公司董事”维度法律主体地位的同时,作为“实在物所有权”维度还属于客体;作为人工智能体之“实在物”,所有权归属而言,或属于公司所有或是自第三方专业主体租赁而

来。这种看似主客体兼而有之的并存现象,显然有违理论常识。但实质上,或许如公司股权,到底属于所有权、社员权之争一样,“可能是一个‘伪命题’”<sup>[29]</sup>。自通过法律拟制赋予公司法律主体地位以来,关于股权之法律性质的定性,各种学说纷呈,虽然公司法人制度的巨大社会价值早已毋庸置疑,但定性争议至今未熄<sup>[30]</sup>。

其实,人工智能董事的法律主体地位问题,与公司通过法律拟制获得法人资格,本质上并无二致,都是社会发展至一定阶段的客观现实对法律技术运用的诉求和督促。例如索菲亚机器人的出现,有研究成果即提出智能体的交互活动越来越频繁涉及权利和义务分配以及责任承担,人工智能的发展趋势“迫使”人们必须着眼考量赋予其合适的法律身份,以便厘清法律和伦理上的权益关系<sup>[31]</sup>。就公司应用场景而言,强人工智能体作为“具有法律主体地位的公司董事”维度,用于解决公司治理层面的权利义务问题;而其作为“客体”的维度,则在于表征权利主体的所有权;赋予人工智能董事法律主体地位,并不影响其所有权人基于特定规则行使转让等权利,犹如公司股东基于公司章程和公司法规定进行股权转让的权利行使行为,与公司本身作为法律主体并不存在逻辑矛盾。

## 二、人工智能董事对传统信义义务的冲击

董事的信义义务是公司董事制度的支撑性制度,不仅确立了司法对公司经营行为的干预边界,还在平衡董事权利与义务关系的过程中扮演着至关重要的角色<sup>[32]</sup>。从理论上讲,以忠实义务和注意义务为核心的信义义务体系规制董事,要求董事最大程度维护公司利益、促进公司持续稳定发展,并为股东和其他利益相关者最佳利益服务。但是,人工智能董事自身的特殊性将会对董事传统信义义务带来一系列冲击。

### (一)传统忠实义务的消解

从本原意义上讲,科学技术性与客观中立性应是人工智能的天然特质。然而,当前弱人工智能的实践证明,由于人工智能的复杂性和不透明性等原因,算法歧视及大数据偏见等有违其客观中立性的现象普遍存在。但是,在强人工智能阶段,我们有理由相信客观中立的问题将得到更好解决。一方面,随着技术的发展,我们将能够开发出更加先进和可

靠的算法,这些算法将能够更好地处理不完美的数据并减少偏误的影响。另一方面,随着人们对人工智能伦理问题的认识不断加深,算法设计的公正性和公平性将得到更多关注,人工智能之思考与行为的理性、客观性和中立性将会得到更为完备的保障。而且,随着人工智能的快速迭代,对具有“自主性”人工智能体的监管会趋向严格,以确保其客观中立性。2023年12月8日发布的欧盟《人工智能法案》“基于风险”强化对人工智能领域的全面监管<sup>[8]</sup>,即是明确且具有趋向性的信号。因而,可以断定,正在形成中的强人工智能,其算法设计和实施必定受到超乎当前水平的更为严格的监管与审查,此趋势下,设计者必须更加注重算法的可解释性以及中立性。然而,此中立性似乎否定了人工智能董事与公司利益冲突的可能<sup>[33]</sup>。按照传统理论,公司董事的忠实义务主要源于董事与公司之间的代理关系,董事忠实义务要求董事在处理公司事务时必须保持诚实和忠诚,不得利用自身地位谋取私利或损害公司利益。但是,对于人工智能董事而言,传统自然人董事所伴随的“自身利益”已丧失基础,至少,从实质上淡化了其履行忠实义务的必要性<sup>④</sup>。某种意义上,强人工智能董事或将成为最忠于“职守”的存在,基于代理成本理论而构建的现代公司治理机制或将发生巨大改变。

### (二) 注意义务“理性人”标准适用性存疑

董事承担注意义务,是公司治理的核心原则之一。这要求董事在履行职责时必须秉持合理注意和谨慎,以公司的最佳利益为依归,合理管理风险,并遵守相关的法律、伦理标准<sup>[34]</sup>。董事注意义务,是确保公司稳健运营和持续发展的关键因素,我国公司法上称之为“勤勉义务”,体现于新《公司法》第180条。在公司治理领域,英美法上“理性人”标准,大陆法上的“善良管理人”标准,通常被用作评估董事行为是否符合注意义务的规则<sup>[35]</sup>。其本质在于衡量董事在行使职责时应展示出一般理性人在相似情况下所应达到的注意和谨慎程度,以确保公司的决策和行动是基于充分的信息和理性的分析。然而,对于人工智能董事而言,由于人工智能技术的复杂性和不可解释性,如何判断人工智能董事是否尽到了合理的注意义务,评估董事行为的“理性人”标准可否适用存在一定困难。“理性人”标准是基于人类长期经验积累的判断能力和情感的伦理观念而设计的,但人工智能基于算法“黑箱”,其数百万个参数<sup>[36]</sup>如何发生作用和相互协作,客观上难以明

确<sup>[37]</sup>,因而带来所谓决策逻辑的不透明。人工智能董事的决策或许是正确而明智的,但其很难向自然人董事和利益相关者清晰解释其决策过程或决策逻辑<sup>⑤</sup>,使得股东和其他利益相关者难以理解其如何平衡不同利益并作出决策。因此,对于人工智能董事作出的决策,界定其是否违反注意义务时,还能否适用“理性人”标准,目前存在疑问。

## 三、人工智能董事信义义务的意蕴革新

不可否认,人工智能董事的出现,对董事传统信义义务体系带来了新挑战。那么,人工智能董事是否还是信义义务的适格主体?这一问题的根本在于:公司法为自然人董事制定的代理规则是否也适用于人工智能代理。答案显然是肯定的,毕竟在特定的决策中,人工智能董事与自然人董事的地位和代理关系并无本质不同。因此,人工智能董事也应承担信义义务。但是,基于人工智能董事的特殊性,其信义义务内容必定呈现新的变化。

董事信义义务,本质上是董事担负的一种基于最大善意、忠实、信用及公平的义务,要求其以最高程度的忠诚和正直为公司和股东最佳利益而行为。不过,虽然在具体案情或特定关系中,法庭不难把握是否对当事人科以“信义义务”,但判定“信义义务”的具体内容却并不确定和清晰;以至于英美法上“Fiduciary”一词,已经被界定为“不是英美法上最易误解的词汇,也至少是最难界定的词汇之一”<sup>[38]</sup>。包括我国在内的不少国家,信义义务的适用尚停留在松散的“义务束”状态,且缺乏逻辑主线,其具体内涵也并不清晰<sup>[39]</sup>。因此,信义义务及其所涵盖的忠实义务、注意义务与合规监督义务等,本身就具有较高的可塑性,使得人工智能董事基于自身特性带来义务内涵的个性化成为可能。

### (一) 忠实义务之特例因应

忠实义务的本质在于规制“不得为而为之”的利益冲突行为。但基于强人工智能的中立性特质,人工智能董事本身并无自利的主观意图,客观上,正常运行的智能体也不可能利用职务便利,作出为自身或第三人利益而牺牲公司利益的行为。因此,传统董事忠实义务趋于消解,包括“不得利用职务之便挪用、侵占公司资产”“不得利用公司机会为个人谋利”“进行关联交易”等传统忠实义务内容已不复存在。但我们还需要考虑到,随着强人工智能时代即将到来,未来人工智能董事逐步获得认可,在普及

化使用的过程中,仍有可能出现违反忠实义务的特别情形。

现代商业环境中,董事同时服务于多家公司的情况并不罕见,人工智能普及化过程中,使用效果较好、信誉较高的强人工智能体,也可能出现服务于多家公司的形态。然而,人工智能董事在一家公司任职期间接触到的数据,尤其是商业秘密,一旦泄露或为其他公司所用,可能触发对忠实义务的违反。此时,对前公司而言,较为彻底的解决方案,是对人工智能董事进行全面重置和重新训练。但这显然不利于人工智能技术的发展和运用。人工智能董事在履职过程中累积的经验和技能属于社会的宝贵财富,全面重置将导致这些经验和技能丧失,这无疑是对资源的浪费和对技术进步的阻碍。但如果仅删除相关数据和商业秘密,新公司利用人工智能董事在前公司获得的经验和技能时,尤其是当这些经验和技能涉及前公司的经营策略和经营方法等个性化内容时,就可能对前公司业务造成直接或间接的伤害。人工智能董事可能会将在前公司积累的经营策略或方法等信息披露或应用于新公司,从而与前公司形成利益冲突,此时,即便是关联公司而不是竞争对手,也可能导致违反忠实义务的问题。

因此,为了确保人工智能董事在多公司任职场景中能够妥善履行忠实义务,有必要进行前瞻性研究,构建相应的监管机制、行业规范和行为准则。例如,必须从技术和机制上确保所服务公司之间的个性化信息和非公开性内容不会发生调用或披露。若技术和机制上无法确保,从公平角度出发,应限制同一人工智能董事同时在两个或更多不同公司担任董事职位,否则人工智能董事将可能承担违反忠实义务的责任。

## (二) 注意义务之范围窄化

董事注意义务主要涵盖决策注意与监督注意两个方面。近年来,随着董事监督义务日益成为一种相对独立的董事义务类型<sup>[40]</sup>,决策注意成为注意义务中较为核心的内容。通常情况下,董事决策注意的范围主要涵盖信息收集、信息评估以及谨慎决策三个层次<sup>[41]</sup>。具体而言,首先,董事在作出任何决策之前,需要积极获取和更新与公司业务相关的信息,并充分了解相关的情况和背景信息;其次,董事应基于上述信息,参照相关规定与公司章程等规范和要求,并结合商业知识和判断能力,对综合信息进行整体评估;最后,董事应关注和考虑公司最佳利益并谨慎决策。因此,自然人董事履职期间,必须投

入足够的时间、精力熟悉和掌握公司情况,以理性人标准充分利用自身商业经验和技巧评估综合信息,并秉持善意和谨慎原则进行决策<sup>[34]</sup>。

然而,强人工智能体语境下,人工智能董事的注意义务无疑会发生重大变化。强人工智能体在获得任命并成为人工智能董事之前,已接受相对完备的专门“训练”和“培训”,应当具备胜任董事的商业专业知识和经营判断技能。当其履职时间和工作精力不存在限制,并且完全满足胜任性标准,基于“理性智能体”的法律本质定位<sup>[42]</sup>,人工智能董事在信息收集以及信息评估层面的能力,无疑远超自然人董事。如无外在因素影响和干涉,人工智能董事完全可以满足信息收集与信息评估的注意要求,且不会导致注意义务违反问题。因此,人工智能董事的注意义务范围将收窄,且主要集中于“谨慎决策”层面。同时,基于人工智能决策更注重程序化和标准化的特质,以及实践上注意义务“重在关注决策程序及其过程中瑕疵”的判断标准<sup>[43]</sup>,可以预见,人工智能董事虽有违反决策注意的可能<sup>⑥</sup>,但其注意范围相较自然人董事将明显呈现出窄域化倾向。

## (三) 合规监督义务之内涵新解

董事合规监督义务的性质属于忠实义务还是注意义务,学界观点不一<sup>[44]</sup>。但董事合规监督义务日渐成为董事的一种独立义务类型,已成为共识。董事合规监督义务既包括督促公司依法依规完善信息披露,也涵盖董事检查督促公司行为的合规性,还包括及时发现并督促改正公司的信息披露问题以及必要时的监督举报。人工智能董事无法从主观上实施积极合规监督和主动举报行为,似乎其并不存在违反董事合规监督义务的问题。但是,结合违反董事合规监督义务的实践案例,可以发现,当前过错认定基本持“客观化”原则。当人工智能董事参与公司治理后,基于庞大的数据与算力支持,以及“客观中立”的本质特征,人工智能董事理应能够更好地促进董事会及公司行为的合规性,但如果其未能有效履行该职责,则涉嫌违反合规监督义务。

同时,人工智能董事介入董事会决策后,势必导致董事会决策环境的改变,这一方面“可能引发自然人董事义务回避的道德风险”,但另一方面“更会抬升自然人董事的注意义务履行标准”<sup>[45]</sup>。因而,人工智能董事参与董事会决策,不仅引入了新的决策维度,还会对董事会现行决策生态产生类似“鲑鱼效应”的影响。

具体而言,这种影响体现在以下两个方面:一方

面,基于人工智能董事在数据处理和分析方面所具有的明显优势,它能够迅速、全面地掌握相关资料。人工智能董事高效的信息处理方式不仅会提升信息披露的合规性,还会对其他自然人董事的决策过程施加积极影响,进而对公司的整体决策和运营产生合规监督效果。另一方面,基于强人工智能的中立性,人工智能董事的决策意见和报告具有较高的客观性。这不仅有利于提升董事会决策的整体合规质量,也会对其他自然人董事产生正向效应,促使他们在决策时更加客观透明、合法合规,客观上也实现了另一层面的监督。因此,人工智能董事虽无法实现积极的监督与举报功能,但实质上其一定程度扩展了董事合规监督义务的内容和边界。

## 结 语

马克思在《对民主主义者莱茵区域委员会审判》中指出:“社会以法律为基础,只是法学家的幻想。相反,法律应以社会为基础。”面对人工智能时代急速变迁的社会,我们秉持更为包容和开明的姿态,对强人工智能语境下人工智能董事法律主体地位及其信义义务进行探讨,尤其对人工智能董事忠实义务、注意义务与董事监督义务进行构想式理论分析和展望,更易于避免法律的滞后性,促进未来社会发展的多样性。

随着人工智能技术的快速发展以及应用场景的不断扩大,人工智能参与公司治理将会逐渐普及,人工智能董事也将由理论构想迈入现实。这一变革不仅将重塑董事会的结构和生态,也对现有的法律体系提出新的挑战和要求。前瞻性地对人工智能董事核心制度进行研究,具有实践意义和理论价值。我国的人工智能整体发展水平已居于世界前列,为人工智能董事在我国的应用提供了坚实基础和广阔空间。当然,我们也必须认识到,人工智能技术的复杂性和不确定性也给法律监管带来前所未有的挑战。因此,我们应结合人工智能技术的特点和发展趋势,做好提前因应,及早研究和规划,制定出更加具体和可操作的规范标准,以确保人工智能董事的应用能够在法律框架内有序进行。

### 注释

①参见 European Parliament Resolution with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (European Parliament, 16 February 2017) para.59(f). ②有学者指出,人工智能

担任公司董事还存在非自然人可否成为公司董事的前置性问题。其实,在比较法视野下,已有立法例和公司治理实践证明,该问题不存在实质障碍。我国香港公司条例、澳门商法典以及法国商法典等法律已经明确了非自然人担任董事的合法性。其实,我国新《公司法》仅第178条规定了董事任职的消极条件,有学者指出我国《公司法》也并未明确禁止非自然人担任公司董事。具体论述详见林少伟:《人工智能对公司法的影响:挑战与应对》,《华东政法大学学报》2018年第3期。③参见 European Parliament Resolution with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (European Parliament, 16 February 2017)。④就形式而言,当人工智能董事本身算法或数据存疑或其受到开发者以及利益相关者的偏见或利益影响,从而无法忠实为公司的最佳利益行事,从表象上也构成忠实义务违反。但就实质而言,上述情形显然已形成另一层面法律问题,属于侵权或产品质量问题,而不再属于董事忠实义务违反与否的讨论范畴。⑤也有研究成果认为,为避免出现“算法同质化”和商业秘密泄露,本身即应坚持适度而非过度披露。参见王怀勇、邓若翰:《算法趋同风险:理论证成与治理逻辑——基于金融市场的分析》,《现代经济探讨》2021年第1期。⑥算法发挥影响的过程,可分为算法应用的研发设计、算法部署应用和智能算法自主决策三个阶段,前两个阶段存在人为影响的可能,但第三阶段,算法则表现出与研发设计者行为相分离的特征,参见张凌寒:《算法权力的兴起、异化及法律规制》,《法商研究》2019年第4期。弱人工智能的决策过程,通常非自主性特征较为明显,但强人工智能体已具备高度自主性,当其基于自主性逻辑,未使用经过“训练”和“培训”的信息、知识和技能,或因运用不当作出有违公司最佳利益的商业决策时,会引发决策的适当性问题而构成注意义务违反。

### 参考文献

- [1] 博登.人工智能的本质与未来[M].孙诗惠,译.北京:中国人民大学出版社,2017:3.
- [2] 翟振明,彭晓芸.“强人工智能”将如何改变世界:人工智能的技术飞跃与应用伦理前瞻[J].人民论坛·学术前沿,2016(7):22-33.
- [3] 史爱武.ChatGPT:强人工智能时代的里程碑[N].中华读书报,2023-04-12(18).
- [4] 何立民.人工智能系统智能生成机理探索之六:从弱人工智能、强人工智能到超人工智能[J].单片机与嵌入式系统应用,2020(8):87-89.
- [5] ZOLFAGHARIFARD E. Would You Take Orders from a Robot? An Artificial Intelligence Becomes the World's First Company Director [N]. Daily Mail, 2014-05-19(4).
- [6] 杜严勇.论机器人权利[J].哲学动态,2015(8):83-89.
- [7] 国务院.关于印发新一代人工智能发展规划的通知:国发[2017]35号[A/OL].(2017-07-08)[2023-11-07].[https://www.gov.cn/gongbao/content/2017/content\\_5216427.htm](https://www.gov.cn/gongbao/content/2017/content_5216427.htm)
- [8] 王卫.欧盟《人工智能法案》实行风险分级监管[N].法治日报,2023-12-18(5).
- [9] 基辛格,施密特,胡滕洛赫尔.人工智能时代与人类未来[M].胡和平,凤君,译.北京:中信出版社,2023:266.
- [10] 黎四奇.对人工智能非法律主体地位的解析[J].政法论丛,2023(5):117-127.
- [11] 高诗宇.人工智能体法律主体资格之争的商谈理论解决方案[J].哲学分析,2023(5):121-132.

- [12] GOOD I. J. Speculations Concerning the First Ultraintelligent Machine[J]. *Advances in computers*, 1966(6):31-88.
- [13] 王彦雨.“强人工智能”争论过程中的“态度转换”现象研究[J]. *科学技术哲学研究*, 2020(6):26-33.
- [14] 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.
- [15] 冯洁.人工智能体法律主体地位的法理反思[J]. *东方法学*, 2019(4):43-54.
- [16] 龙文懋.人工智能法律主体地位的法哲学思考[J]. *法律科学(西北政法大学学报)*, 2018(5):24-31.
- [17] 杨立新.民法物格制度研究[M].北京:法律出版社,2008:3.
- [18] 刘艳红.人工智能的可解释性与AI的法律责任问题研究[J]. *法治与社会发展*, 2022(1):78-91.
- [19] 范进学.人工智能法律主体论:现在与未来[J]. *政法论丛*, 2022(3):3-17.
- [20] 尹田.论自然人的法律人格与权利能力[J]. *法制与社会发展*, 2002(1):122-126.
- [21] 徐昭曦.反思与证立:强人工智能法律主体性审视[J]. *中共中央党校(国家行政学院)学报*, 2019(3):80-88.
- [22] 朱体正.人工智能时代的法律因应[J]. *大连理工大学学报(社会科学版)*, 2018(2):97-102.
- [23] World Economic Organization. Deep Shift :Technology Tipping Points and Societal Impact[EB/OL].(2015-09-09).[2023-12-16].<https://cn.weforum.org/publications/deep-shift-technology-tipping-points-and-societal-impact>.
- [24] 耀友福,何相亿.人工智能应用的公司治理作用:基于财务重述视角[J]. *工业技术经济*, 2023(3):125-134.
- [25] 张钺,朱军,苏航.迈向第三代人工智能[J]. *中国科学:信息科学*, 2020(9):1281-1302.
- [26] 屈信明.董事责任险为何受到青睐[N]. *人民日报*, 2023-05-15(18).
- [27] 何启豪.董事责任险走俏,如何助力公司治理?[J]. *中国保险*, 2022(4):44-46.
- [28] 张绍欣.法律人格、法律主体与人工智能的法律地位[J]. *现代法学*, 2019(4):53-64.
- [29] 周友苏.新公司法论[M].北京:法律出版社,2006:232.
- [30] 张双根.论股权的法律性质:以成员权之法教义学构造为中心[J]. *中外法学*, 2023(3):687-706.
- [31] 周详.智能机器人“权利主体论”之提倡[J]. *法学*, 2019(10):3-17.
- [32] 梁爽.董事信义义务结构重组及对对中国模式的反思:以美、日商业判断规则的运用为借镜[J]. *中外法学*, 2016(1):198-223.
- [33] 徐凤.人工智能算法黑箱的法律规制:以智能投顾为例展开[J]. *东方法学*, 2019(6):78-86.
- [34] 翁小川.董事注意义务标准之厘定[J]. *财经法学*, 2021(6):48-66.
- [35] 陈本寒,艾围利.董事注意义务与董事过失研究:从英美法与大陆法比较的角度进行考察[J]. *清华法学*, 2011(2):71-78.
- [36] KRIZHEVSKY A, SUTSKEVER I, HINTON G. E. ImageNet classification with deep convolutional neural networks[J]. *Advances in neural information processing systems*, 2012(25):1097-1105.
- [37] CASTELVECCHI D. Can we open the black box of AI? [J]. *Nature*, 2016(7623):20-23.
- [38] FINN P. D. *Fiduciary Obligations* [M]. Sydney: Law Book Co., 1977:26.
- [39] 王莹莹.信义义务的传统逻辑与现代建构[J]. *法学论坛*, 2019(6):27-36.
- [40] 吴越,陈杰.董事监督义务裁量规则研究[J]. *社会科学研究*, 2023(6):117-130.
- [41] 任自力.公司董事的勤勉义务标准研究[J]. *中国法学*, 2008(6):83-91.
- [42] 陈亮,张翔.人工智能立法背景下人工智能的法律定义[J]. *云南社会科学*, 2023(5):162-170.
- [43] 张海棠.公司法适用与审判实务[M].北京:中国法制出版社, 2009:329-330.
- [44] 傅穹.公司利益范式下的董事义务改革[J]. *中国法学*, 2022(6):197-218.
- [45] 程威.人工智能介入董事会的董事义务与责任重释[J]. *东北大学学报(社会科学版)*, 2022(2):100-108.

## Conception of the Legal Status and Fiduciary Duty of Artificial Intelligence Directors — Based on the Perspective of Strong Artificial Intelligence

Liu Chengjie

**Abstract:** The speed of innovation and application scenarios of artificial intelligence technology have changed our traditional understanding, and artificial intelligence directors have moved from theory to reality. Our country's relevant development plan has proposed measures to ensure the development of AI by clarifying the legal subjects, rights, obligations, and responsibilities of AI. In the context of strong AI, the viewpoint of object theory should be abandoned, and AI directors should be granted legal subject status through legal drafting. Based on the particularity of AI directors, the traditional fiduciary duty of directors tends to dissolve, and the applicability of the “rational person” standard of duty of care will also be questioned. Looking at and innovating the fiduciary obligations of AI directors from a forward-looking perspective, the duty of loyalty needs to be tailored to special circumstances, the scope of duty of care tends to be narrow, and the connotation and boundaries of compliance supervision obligations will be substantially expanded.

**Key words:** strong AI; AI directors; legal subjects; fiduciary duty

责任编辑:一鸣