

人工智能产业的包容审慎监管： 理论内涵与实现路径

戚聿东 刘 健

【摘要】人工智能产业是具有“新技术、新产品、新经济、新风险”等新特征的新兴产业,对人工智能产业监管模式与监管路径的研究是实现创新与风险相平衡的核心与关键。包容审慎监管是集包容创新与风险防范于一体的监管模式,具有有效市场与有为政府相融合的理论内涵,是市场与政府关系理论在监管领域的实践表出。按照分层分类的研究思路与方法,人工智能产业基础层、技术层与应用层具有不同的技术、经济、风险特征,包容审慎监管契合人工智能产业跨领域、跨主体、跨行业的技术经济特征,能有效防范人工智能产业多层次、多元化、复合型的风险属性。是适合于人工智能产业等新经济的有效监管模式。应以底线思维和以人为本为监管原则,以监管沙盒为主要运行机制,以分级分类、全过程动态和多主体协同为主要手段,有效实施对人工智能产业的包容审慎监管。

【关键词】人工智能产业;包容审慎监管;理论内涵;适应性;实现路径

【作者简介】戚聿东(1966-),男,吉林东丰人,博士,北京师范大学经济与工商管理学院教授,博士生导师,从事数字经济研究(北京 100875);刘健(通讯作者),北京物资学院经济学院(北京 101149)。

【原文出处】《兰州大学学报》:社会科学版,2024.4. 154~165

【基金项目】国家社会科学基金重大项目“技术标准与知识产权协同推进数字产业创新的机理与路径研究”(19ZDA077)。

2024年6月20日,国家主席习近平向2024世界智能产业博览会致贺信指出,人工智能是新一轮科技革命和产业变革的重要驱动力量,将对全球经济社会发展和人类文明进步产生深远影响^①。目前,人工智能已广泛应用并赋能各行各业,以其强大的通用技术外溢性体现在数字经济发展中。然而,创新与风险往往是一对相伴而生的孪生体,人工智能产业在快速发展的过程中,不可避免地会带来一系列的市场风险与社会问题,如何在有效防范风险的前提下促进人工智能产业的健康发展是摆在我们面前的一道现实难题。在这一背景下,2017年1月,国务院办公厅发布《关于创新管理优化服务培育壮大经济发展新动能加快新旧动能接续转换的意见》,首次提出“探索动态包容审慎监管制度”,同年的《政府工作报告》进一步明确:应“本着鼓励创新、包容审慎的原则,制定新兴产业监管规则”。自此,包容审慎监管成为政府监管新技术、新产业、新业态、新模式等新经济的主导模式。

目前国内学术界对人工智能的研究往往从技术上谈得多,从产业上谈得少;从治理上谈得多,从监管上谈得少。即便在产业监管上,也多主张加强监

管,对“包容审慎监管”重视不够,对人工智能产业的“包容审慎监管”研究更是凤毛麟角。随着人工智能技术的飞速发展和人工智能产业的蓬勃兴起,如何有效实施对人工智能产业的包容审慎监管,无论是对企业、政府还是社会公众,都极为重要也十分迫切。本文从包容审慎监管的源起入手,阐释包容审慎监管的理论内涵,结合人工智能产业的“技术—经济—风险”特性,分析包容审慎监管适应于人工智能产业的必然性和可行性,最后从监管原则、监管措施与监管机制三个层面提出人工智能产业包容审慎监管的实现路径。期待能在理论上填补有关人工智能产业包容审慎监管的研究空白,并在实践上推动对人工智能产业包容审慎监管的有效实施。

一、包容审慎监管的理论内涵

(一)包容审慎监管的源起

监管理论起源于以庇古为代表的公共利益理论,认为政府干预能纠正市场失灵,从而使经济的资源配置达到最优,社会整体福利达到最大。该理论在20世纪初期到20世纪90年代基本占据主流,以美国为代表的管制型国家是该理论的坚定践行者,加强监管成为各个国家对付市场失灵的主要手段,

政府监管的主要模式也以审慎监管为主。20 世纪 90 年代末期至 21 世纪初期,在监管实践中逐渐出现了“规制失灵”与“规制俘获”的现象,学术界对加强监管的批评增多,以斯蒂格勒为代表的部门利益理论或规制俘获理论开始对传统的监管理论提出挑战,该理论认为监管部门也具有经济人属性,政府管制往往来自产业游说,是设租与寻租的结果,规制俘获导致了规制失灵。这一理论揭示了政府监管失灵的可能性和破坏性,主张放松监管,倡导包容性的政府监管模式。由此可见,强化监管的审慎与放松监管的包容更多的是一种政府与市场关系的博弈,而这一博弈在金融业中体现得最为充分和明显。在对以银行为代表的金融企业的监管上,审慎监管原则长期以来得到了世界各国的普遍认可和执行,巴塞尔银行监管委员会于 1997 年 9 月发布的《银行业有效监管核心原则》从资本充足率、风险管理、内部控制、跨国银行监管和纠正措施等五个方面制定了审慎监管的基本原则,成为金融业实施审慎监管的纲领性文件。按照迈克尔·泰勒“双峰监管”理论的观点,维护金融机构的稳健经营属于微观审慎,而维护金融体系的稳定则属于宏观审慎^[1]。无论是微观审慎还是宏观审慎,审慎的含义都意味着稳健,意味着对风险的严格防范,意味着政府对市场的干预和管制^[2]。我国政府对金融业的监管也遵从传统的审慎监管基本原则,在微观审慎层面,对金融行业的市场准入采用高门槛的行政审批制度,并对金融企业的日常运营实施严格的审慎监管,确保金融企业微观个体的稳健经营;在宏观审慎层面,从中央到地方的各级政府都严守“不发生系统性金融风险”的底线,保障金融行业整体的安全稳定。这一传统的审慎监管原则和模式,对风险属性和信息不对称程度均较高的传统金融行业较为适合,能够在一定程度上有效防范金融行业的各种风险和不确定性。因而,直到 20 世纪末 21 世纪初,审慎监管在金融业监管中都占据主流,通过政府干预来弥补市场失灵在金融行业发展中发挥较为重要的作用。

21 世纪以来,随着互联网、大数据、区块链、人工智能等新技术在金融行业的应用,金融科技的创新特性使得传统的审慎监管模式的弊端暴露无遗。偏重于风险防范的审慎监管往往对新技术、新业态、新模式等创新事物予以抑制甚至扼杀,不利于金融科技等新经济的发展,加上监管实践中的规制俘获与规制失灵事件屡见不鲜,学术界和实务界对放松监管的呼声加大,强调弱化政府干预,主张进一步发挥市场自身的调节作用,这对管制型政府监管产生了

一定的冲击和压力。于是,如何在鼓励创新的同时防范风险,如何平衡政府与市场的关系,成为新时期摆在政界、学术界和实务界的一道难题。在这一背景下,我国政府发挥政治智慧,运用我国传统文化中“草木有情皆长养,乾坤无地不包容”的包容理念,创新性地提出了对新技术、新产业、新业态和新模式等“四新”经济采用包容审慎监管的思路和举措。2017 年 3 月的《政府工作报告》中首次提出“包容审慎”的监管原则,对新经济的包容审慎监管正式登上历史舞台。

(二) 包容审慎监管的概念界定

自我国政府正式提出包容审慎监管的理念与原则之后,国内学术界对包容审慎监管的概念与内涵进行了相应研究。刘太刚^[2]从需求溢出理论的角度,对包容审慎监管进行了学理阐释,认为监管方式与原则的选择取决于对公共风险治理深度和强度的设定,而包容审慎监管在公共风险治理深度和强度上,是一种介于“审慎”与“包容”两者之间的一种折中的监管方式。廖凡^[3]认为金融科技的创新特性是包容监管的立足点,而金融科技的风险特性则是审慎监管的着眼点,包容审慎监管需兼顾金融、科技与创新,实现创新与安全之间的平衡。张友连等^[4]从创新与安全平衡的角度分析了包容审慎监管的内涵,认为审慎中需有包容,包容中亦需有审慎,二者是辩证统一的关系。钱贵民等^[5]分析了包容审慎监管的基本理念,认为包容审慎监管应“对产业包容、对企业审慎”“对创新包容、对技术审慎”,针对不同平台的不同特性,制订有针对性的监管举措。刘权^[6]认为包容审慎监管是政府在新经济发展过程中的必然选择,也是实现创新与安全之平衡的必然要求,包容审慎监管赋予了行政监管部门更大的自主空间和自由裁量权,从而对监管机构的监管水平和能力提出了更高的要求和挑战。田晓祥^[7]从内涵、时间和措施三个维度阐释包容审慎监管,认为包容审慎的内涵是“放”与“管”的辩证统一,既包容创新,又严守底线;时间维度上,包容审慎监管更倾向于事中和事后监管;监管措施上,须分类监管,即根据不同经营主体的特性,采取不同形式的监管措施。总的来说,以往学者大多从监管的目标角度阐述包容审慎监管的内涵,且将包容审慎监管侧重为对新技术、新产业、新模式、新业态等新经济的包容,较少论及包容监管与审慎监管二者之间的关系,也忽略了对创新行为与创新风险同步共生关系的探讨。

我们认为,创新与风险是相伴而生的一体两面关系。创新越强,风险越大;创新越密集,风险越集

中。在某种意义上,市场风险是市场为鼓励创新所必须承受的代价,而市场垄断则是社会为阻止创新所必须付出的成本。对创新风险、市场垄断以及社会风险的包容,有利于对市场主体创新的保护和鼓励,有利于对经济发展新动能的激发和促进,有利于市场效率与活力的提升。与此同时,包容并不意味着放任不管,基于市场规制与秩序之上的公平竞争是创新生发所必需的环境和土壤,政府应通过审慎监管维护公开公平公正的竞争环境与秩序,严厉查处和惩戒利用人工智能新技术侵犯他人权益的不法行为,同时对系统安全风险、市场应用风险等进行动态监测和适用性监管,守住不发生系统性风险的底线,并对社会群体性事件进行妥善处理和应对,最小化人工智能产业应用所引致的社会调整成本。因此,维护市场秩序、防范市场风险和保障社会安全也是包容审慎监管的应有之义。综上所述,人工智能产业的包容审慎监管可以定义为以平衡创新与风险、统筹效率与安全为基本目标,以维护公平的市场秩序与竞争环境为基本前提,通过包容人工智能技术创新、产品开发以及产业应用过程中的市场风险、市场垄断以及必要的社会成本,激发和鼓励市场主体的创新行为,并审慎对待上述市场风险与社会问题的政府监管体系。

(三) 包容审慎监管的理论内涵

党的二十届三中全会审议通过的《中共中央关于进一步全面深化改革、推进中国式现代化的决定》指出:“必须更好发挥市场机制作用,创造更加公平、更有活力的市场环境,实现资源配置效率最优化和效益最大化,既‘放得活’又‘管得住’,更好维护市场秩序、弥补市场失灵,畅通国民经济循环,激发全社会内生动力和创新活力”。“放得活”体现有效市场的逻辑,核心在于如何“放”,如何“更好发挥市场机制作用,创造更加公平、更有活力的市场环境”来“实现资源配置效率最优化和效益最大化”;“管得住”体现有为政府的逻辑,核心在于如何“管”,如何“维护市场秩序、弥补市场失灵”来“畅通国民经济循环,激发全社会内生动力和创新活力”^[8]。包容审慎监管具有有效市场与有为政府融合的理论内涵,是“放得活”与“管得住”二者的辩证统一在监管领域的反映和表现。

首先,包容监管内涵了“放得活”的市场逻辑。人工智能产业的创新行为是一种自主自发的市场行为,只要市场主体的创新项目、创新行为、创新产品没有违背市场规则、没有损害他人正当权益,政府就应予以包容,以“放得活”为基本准则,从广度和深度

上推进市场化改革,大幅度减少政府对资源的直接配置,推动资源配置依据市场规则、市场价格、市场竞争实现效益最大化和效率最优化。在这个过程中,应给予企业充分的自主发展的时间和空间,让其按照市场的逻辑“自主经营、自主发展”。应该强调的是,政府不应过分担心由于创新所带来的市场风险,更不能以创新会带来市场风险为由干预市场主体的创新行为。因为市场风险衍生至创新行为,风险的防范和管理都必须从源头的风险认知、风险识别和风险判定等入手,而政府相对于企业来说,获取信息能力更弱,认知局限更大,信息不对称程度更强,因而以市场风险为由干预市场创新行为是非理性的,不但可能造成监管失灵,而且还会有损市场经济的健康发展。并且市场本身具有强大的内循环机制和功能,在短期内,市场风险会带来一定的不稳定因素,但是只要在可承受的范围之内,都是市场为创新所必须付出的代价;在长期内,给予市场自主调节的时间和空间,市场风险会被市场自身消化和处理。人工智能技术在过去十余年的革命性突破就已证明,技术领先性、产品市场力与风险投资回报之间形成有力的正向循环,使得市场自身能够有效应对技术革命带来的巨大风险^[9]。

其次,审慎监管内涵了“管得住”的政府逻辑。人工智能产业的创新行为必须在遵循市场规则、遵守市场秩序、不损害公共利益、不侵犯他人权益的前提下进行。政府在人工智能创新过程中,必须有效履行政府职责,对市场主体的违法违规行为零容忍,守住市场经济的安全底线。也就是说,在监管实践中,政府一定要管住干预之手,该包容时一定要“放得活”;同时用好惩戒之手,该作为时一定要“管得住”。如果对市场主体主观上的道德风险防范和约束不力,对违法违规行为查处不严,那么政府的市场监管体系将形同虚设,市场监管的职能就会失效,市场将陷入“一放就乱”的局面。因此,审慎监管要求政府切实全面履行“市场守夜人”的政府职责,转变政府职能,优化政府组织结构,健全监管体系,对人工智能技术开发与产业应用中的各类风险保持全过程的动态监测和防范,并制定有针对性的科学有效的监管措施,以防范系统性安全风险并降低社会整体的变革成本。

最后,包容审慎监管内涵了“放得活”的有效市场与“管得住”的有为政府的融合与统一。包容是对市场行为的尊重,既对创新主体的创新行为也对创新过程中的创新风险予以包容;审慎是对政府职责的履行,既对源自道德风险的违法违规行为予以严

厉打击,也对源自创新的市场风险科学研判、积极应对。包容是以审慎为前提的包容,只要市场主体的创新行为是基于规则与秩序之上的,就应该予以包容和接纳。而审慎是以包容为基础的审慎,只有尊重、承认和确立市场主体的自由选择和自主行为,制定和维护相应的规则与秩序才有存在的意义和价值。包容与审慎是既对立又统一的融合体,是“放得活”的有效市场与“管得住”的有为政府在监管实践的表出。正如习近平总书记在十八届中央政治局第二十八次集体学习时所指出的:“我们要坚持辩证法、两点论……既要‘有效的市场’,也要‘有为的政府’”^[10]。要“放得活”,就必须有“有效的市场”,各市场主体才能充分发挥活力和创造力;要“管得住”,就必须有“有为的政府”,才能充分保障市场秩序和各市场主体的利益。

二、人工智能产业的“技术—经济—风险”特征

对人工智能产业采取何种监管模式以及如何进行有效监管,都应以人工智能产业自身的特征为依据。本文在人工智能产业概念界定的基础上,分部门、分类别地分析和阐述人工智能产业的“技术—经济—风险”特征。

(一)人工智能产业的概念界定

人工智能在产业链上可划分为上游基础层、中游技术层和下游应用层三个层级。基础层包含算法、算力和数据等人工智能软硬件基础设施;技术层由计算机视觉、语音识别、语言技术处理、深度学习等涉及数据认知、学习、分析和处理等核心技术构成;应用层是对人工智能技术和产品在某一领域或场景的具体应用。基础层与技术层的结合能开发具有专用性或通用性的人工智能创新产品,如具有专有性特征的人工智能自动驾驶就是以上游基础层的数据、算法和算力为支撑,融合中游的传感器、计算机视觉、路径规划与控制,以及机器学习等技术所研发的人工智能产品,将该产品应用于交通与驾驶领域,就构成了下游的产业应用层级。而具有通用性特征的新一代生成式人工智能,如 ChatGPT 等,可以“类人”似地处理各个领域、场景复杂的智能任务,从而使得下游的应用层可以拓展至各行各业。

国家统计局 2021 年颁布的《数字经济及其核心产业统计分类(2021)》将数字经济产业划分为数字产业化部分和产业数字化部分^②。数字产业化部分是数字经济核心产业,是指为产业数字化发展提供数字技术、产品、服务、基础设施和解决方案,以及完全依赖于数字技术、数据要素的各类经济活动。产业数字化部分是指应用数字技术和数据资源为传统

产业带来的产出增加和效率提升,是数字技术与实体经济的融合。按照国家统计局关于数字经济产业的定义及分类,结合人工智能产业链的层级,本文将人工智能产业划分为人工智能产业化部门与产业智能化部门。人工智能产业化部门主要包括数据、算法、算力和机器学习等与人工智能技术研发与产品开发相关的计算机软硬件行业,主要涉及人工智能产业基础层与技术层;产业智能化部门则指运用人工智能技术改造传统产业过程中创造的新产品、新技术、新工艺、新模式和新业态,主要涉及人工智能产业应用层。人工智能产业化部门是关键要素的产出部门,产业智能化部门则是关键要素的投入部门,人工智能产业发展是两个部门创新循环形成和演进的结果^[9]。

(二)人工智能产业的“技术—经济—风险”特征分析

人工智能产业化部门与产业智能化部门技术经济特性不同,所带来的风险迥异。本文秉持分类分层的思路,对人工智能产业的“技术—经济—风险”从两个部门三个层级分别进行论述。

1. 人工智能产业化部门的“技术—经济—风险”特征分析。

(1)基础层的算法具有“算法非中性”和“算法黑箱”的技术特性。尽管技术是中性的,但技术开发与设计背后的人具有价值非中性特点,开发者往往不自觉地将自身偏好的价值观念嵌入编码设计中,使人工智能产品在信息预测、行为决策和任务生成时无法做到客观公正,给人工智能带来伦理上的算法歧视和决策偏差风险^[11]。“算法黑箱”技术上的不透明和不可解释性又给人工智能系统带来结构性漏洞,可能面临不可预知的风险,通过在输入数据中添加扰动,使系统做出错误判断,从而带来系统安全风险。基础层的数据和算力作为人工智能系统的“燃料”和“引擎”,使得人工智能系统具有很强的“数据依赖”和“算力支撑”特性,算力的飞速发展进一步加大对大数据的需求,而数据在采集和传输的过程中不可避免会带来虚假信息、隐私泄露、知识产权侵权等数据安全风险^[12],如果涉及商业机密或国家机密,还有可能产生政治安全风险。

(2)技术层具有模块化的技术特性。人工智能产品可以看作是一个由深度学习、自然语言处理、计算机视觉和语音识别等众多功能子模块,以及组合子模块之间的链接所构成的模块化系统。模块化系统具有局部创新、“背对背”竞争和“赢家通吃”的经济特性。局部创新意味着创新的客体由整个系统转

向局部子模块,创新模式由“破坏性系统创新”转向“颠覆性局部创新”,创新主体由纵向一体化大企业转向专精特新中小企业,创新主体之间通过竞争与合作实现跨学科、跨领域和跨组织的技术重组和互补创新,从而使得市场的创新竞争更为激烈,市场创新能力也更强,并且创新主体之间“背对背”竞争的结果往往是“赢者通吃”。以号称“AI时代大赢家”的英伟达为例,2023年,在包含GPU等公开销售产品以及谷歌、亚马逊等云服务商自研加速器在内的整体AI芯片市场,英伟达占有85%的市场份额,其公开市场上最大竞争对手英特尔、AMD占有率则分别仅有2%和1%^③。这些技术经济特性不可避免地会带来收入分配差距与数字鸿沟等问题,从而加大社会分化与收入不平等。与此同时,在子模块研发、模块链接以及模块化系统运行的过程中,如果编码化技术被不法黑客所利用,植入不利于子模块或整体模块系统正常运行的编写代码,将对人工智能技术系统的局部或总体产生系统安全冲击,带来技术风险与系统安全风险。除此之外,当人工智能产品在实际应用过程中出现问题或产生风险时,由于人工智能任务生成和自动化决策受众多子模块以及子模块之间的互动融合的影响,加上各功能子模块在技术上的不透明和不可解释性,使得问题与风险的责任主体难以界定,从而存在一定的责任主体界定风险。

2. 产业智能化部门的“技术—经济—风险”特征分析。

产业智能化部门是将人工智能技术应用于生产与生活的产业部门,其“技术—经济—风险”特征主要体现为经济特性与风险特性。

(1) 渗透性。在2018年10月31日中共中央政治局第九次集体学习上,习近平总书记指出:“人工智能是引领这一轮科技革命和产业变革的战略性技术,具有溢出带动性很强的‘头雁’效应”^④。作为新一轮科技革命和产业变革的技术引擎,人工智能的研发与应用可以极大地提升传统产业的生产效率,人工智能可以溢出至各行各业,全方位全领域地渗透和影响整个经济社会^[13]。在人工智能产业渗透的过程中,由于任何一项技术研发产品都不可能做到百分百完美,不可避免地会存在内生性的技术缺陷,从而带来一定的市场应用风险。如火爆出圈的萝卜快跑在武汉市投入使用后,武汉市民对萝卜快跑无故停车、违规停车所导致的交通拥堵和安全隐患等出行问题进行了多项投诉,作为新生事物,无人驾驶技术尚不成熟,仍需完善改进^⑤。

(2) 协同性。人工智能可以与人类智能协同工作,能够有效地处理复杂的数据和资料,按照算法模型进行信息预测,从而协同人类做出行为决策。如人工智能可以协助金融企业进行信息和数据处理,增强金融企业风险管理的能力;基于人工智能技术的金融企业虚拟客户经理,可以协同企业为客户提供更优化更高效的服务,同时有效降低金融机构服务成本。类似地,人工智能在医疗、教育、科研、文旅等各个行业和领域目前都展现十分强大的协同性。但人工智能在赋能生产生活的过程中,也有可能“赋能”违法犯罪活动,如不法人员可以利用人工智能技术非法获取数据信息,主观恶意生成违法犯罪信息,借助虚假信息实施诈骗等违法犯罪活动,从而产生由主观道德风险所引起的各种侵犯他人权益的不法行为,引发新型的社会治安和刑事风险。

(3) 替代性。随着人工智能在经济领域的全方位渗透,数据与技术逐步成为主要的生产要素,对传统的劳动和资本等生产要素予以替代。例如在生产制造领域,人工智能可以带来生产流程的自动化和加速化,从而替代许多繁琐重复的工作,极大地节省人力资源和时间成本^[14]。随着ChatGPT等生成式人工智能产品的开发和应用,人工智能对技能型蓝领和智力型白领等工作的替代性进一步增强,这一方面能有效节约企业成本并提升生产效率,但同时不可避免地将对现有社会就业体系产生冲击。“职业替代风险”^[15]很有可能引致群体性结构性失业问题,从而加重社会负担,带来巨大的社会调整成本。

(4) 创造性。人工智能应用过程中必然创造并催生新产品、新技术、新工艺、新模式、新业态乃至新产业,从而使整个经济体呈现出全新创新性的新经济特征。以汽车制造产业为例,人工智能已经渗透研发设计、生产制造和自动驾驶等全产业链,创新出智能网联汽车新产品,催生智能汽车新产业,在公共交通领域又创造出自动驾驶出行服务的新模式和新业态。由此可见,人工智能在引领和创造新经济的同时,必然会创造新的市场机会、新的市场部门以及新职业和新岗位。在这个过程中,新的职业、新的部门、新的市场机会将为人工智能产业从业人员创造巨大的新兴市场红利,但同时也会造成数字鸿沟与收入鸿沟的“双重鸿沟效应”^[16]。

总而言之,在人工智能产业化部门的基础层,算法非中性、算法黑箱的技术不透明与不可解释性将衍生出算法歧视、决策偏差风险和系统安全性风险,数据依赖和算力支撑的特性将衍生虚假信息、隐私泄露以及知识产权侵权等数据安全风险,如果数据

涉及国家信息机密,还有可能产生政治安全风险。基础层与技术层融合所创新和开发的人工智能产品具有模块化系统的属性,衍生局部创新、“背对背”竞争和“赢者通吃”的经济特性,在极大地促进新经济颠覆性创新的同时,很有可能带来责任主体界定风险以及社会分化与收入不均衡等社会问题。在产业智能化部门,人工智能的通用目的技术特性使得人工智能表现出强大的渗透性、协同性、替代性与创造性等经济特性,在极大地推动产业智能化部门高质量发展的同时,不可避免地会产生市场应用风险与新型社会治安和刑事风险,并带来结构性失业与社会调整成本以及数字鸿沟与收入鸿沟等社会问题。

三、包容审慎监管的适应性分析

人工智能产业的各级各类风险正全方位地嵌入经济社会当中,给人们的生产生活造成潜在的威胁和损害,同时这些新兴风险给传统政府监管提出新挑战,并倒逼传统监管体系的改革与创新。

(一) 包容审慎监管的必然性

我国目前对人工智能产业的监管主要采取的是纵向监管模式,注重分行业、分领域、分场景的专门性、领域性监管。在监管实践中,采用软法与硬法相结合的监管方式,坚持“技术主义”路径和场景化逻辑。现行监管体系在防范人工智能产业风险、促进人工智能产业健康发展上起到了一定的积极作用,但同时也暴露出诸多弊端和不足。

1. 基于场景的纵向监管模式无法适应人工智能跨领域、跨场景、跨学科的技术特性。

基于场景的纵向监管模式具有分散性、局部性和零星性特征,难以适应人工智能跨领域、跨场景、跨学科的通用性监管要求,无法有效应对人工智能复杂的技术特性与多样化的应用场景,容易陷入“左支右绌”的监管困境。另外,基于场景的纵向监管较为宽松,监管力度比较单薄,这虽然有利于市场主体的创新环境,但不利于对人工智能产品研发与产业应用中各种主观恶意风险的防范和管控,存在“包容有余而审慎不足”的弊端。与此同时,人工智能跨学科、跨领域、跨主体的产品开发特性以及技术上的不透明和不可解释等特性,使得监管者不可避免地存在信息和知识的双重盲区,难以全面、精准地识别和划定人工智能的风险类型和等级,由此可能产生不稳定的监管预期,难以形成有针对性的监管措施。

2. 基于多部门的多头监管体制无法适应人工智能多层次、多属性、多元化的风险特性。

我国《生成式人工智能服务管理暂行办法》第十六条规定:“网信、发展改革、教育、科技、工业和信息

化、公安、广播电视、新闻出版等部门,依据各自职责依法加强对生成式人工智能服务的管理。”。多部门监管虽然有利于国家在不同领域的专业性监管,但是人工智能具有多层次、多属性、多元化的风险特性,因而对人工智能产业的监管往往需要各个机构之间的通力协作。实践中相互独立的监管部门之间往往存在竞争与推诿的现象。对于存在潜在利益的事项,各部门会争相制定相应的监管政策,并积极参与执法工作;而对于情况复杂、涉及面广且潜在利益不足的事项,各个机构则往往会选择逃避推诿。如此多头监管的“九龙治水”不仅无法达成防范风险与促进产业发展的目标,还会造成监管资源的分配不均,进而影响行业竞争格局与公共利益。

3. 政府单一行政监管为主的纵向监管体系无法适应人工智能跨行业、跨主体、跨边界的经济特性。

政府单一主体的监管模式在面对专业性和复杂性都非常强、应用面又特别广的新型人工智能产品时,现行监管的局限性显露无遗。一方面,在面对人工智能应用场景的多元化与应用领域的全局化时,政府单一而分散的行政监管体系无法有效应对人工智能产业跨行业、跨主体、跨边界的经济特性,加上监管主体信息不对称和专业知识的缺乏的双重劣势,使得监管机构在既没有可依据的法规,也没有可依据的原则的情况下,陷入监管空白的境地。另一方面,基于风险预防的监管模式呈现出静态化的特征,而人工智能的产业应用过程中,既存在内生性技术风险,也存在外生性社会风险,并且风险属性呈现出动态、多变和复杂的特征,在这种情况下,政府单一的行政监管往往滞后于人工智能产业发展过程中的风险生发,从而在风险防范和管控上存在不同程度的失效。

(二) 包容审慎监管的可行性

1. 包容审慎监管模式与人工智能产业“技术—经济—风险”特征高度契合。

对人工智能产业的监管应以人工智能产业的技术—经济—风险特征为逻辑起点,探寻与之相契合的监管模式。人工智能产业具有创新性与风险性并存的技术经济特性,包容审慎监管遵循有效市场与有为政府相互融合的理念,对技术创新、经济创新与风险规制之间相互平衡的监管原则与人工智能产业的技术经济风险特性高度适应和契合。一方面,政府注重有效市场作用的发挥,通过放松规制,特别是放松进入规制,取消不合理的审批事项,建立负面清单制度,营造宽松的市场氛围和环境,将市场的自主权完全交给企业,对企业的自主选择和自主经营予

以包容、鼓励和容错,充分发挥市场决定性作用和竞争基础性作用。另一方面,政府同时注重有为政府作用的发挥,通过加强事中事后监管,坚持底线思维,加强社会性规制,对人工智能产业各个部门在技术研发、产品开发与产业应用过程中的技术风险、经济风险、社会风险等进行有效防范和管控,从而在制度上实现创新与风险、效率与安全之间的平衡。

2. 包容审慎监管有利于企业做出个体理性与集体理性相一致的行为选择。

政府对人工智能产业监管的目的是实现人工智能产业高质量发展的同时有效管控和防范人工智能产业发展过程中的风险,而无论是产业发展还是风险防控,核心落脚点都在于企业自身的行为选择。企业的行为选择是在约束条件下的目标函数最优化,即在给定动机的条件下做出能使自身利益最大化的行为。外因环境决定了行为动机,而在经济、技术与社会文化较为稳定的条件下,政府的监管制度往往对企业的行为动机产生至关重要的作用。假设政府实施严格的审慎监管,对企业来说,技术创新的收益虽然很大,但失败的概率也较大,且创新中的风险极高,加上创新的管制成本也较高,创新的整体成本将高于创新收益,在权衡利弊之后,企业将理性地选择不进行技术创新,如果新技术在传统产业中的应用也受到严格管制,企业也将理性地规避产业应用的行为。因而在严格审慎监管体制下,无论是新技术的研发、新产品的开发抑或新技术的应用,都将受到抑制甚至扼杀。再假设政府实施自由放任的监管体制,在创新收益既定的前提下,企业所面临的创新管制成本下降,从而创新的整体成本下降,加上企业无需承担技术创新与产业应用中的外部风险与社会成本,企业将在自利驱动下,理性地选择进行技术创新以及产业应用,并对可能造成企业损失的内部风险进行有效管控和防范,但对外部风险和社会成本不予理会。企业的这一行为选择虽然将在一定程度上带来人工智能产业的快速发展,但技术创新与产业应用过程中不可避免的外部风险与社会成本则转嫁给其他市场主体和社会大众,造成严重的负外部性,外部风险和社会成本的处理和消化一旦出现问题,有可能转化为全局性的社会危机,给整个经济社会带来系统性的破坏。由此可见,不论是严格的审慎监管模式还是自由放任的监管模式,都将出现个体理性导致集体非理性的结果。只有采用包容审慎的监管体制,才能在降低企业整体创新成本的同时,加强企业对外部风险和社会成本的管控和防范,使得企业在实现自身利益最大化的同时,达到社会

整体福利的最优化,从而使得个体理性与集体理性保持一致。

四、人工智能产业包容审慎监管的实现路径

依据包容审慎的理论内涵,从我国人工智能产业“技术—经济—风险”的实际出发,本文从监管原则、监管措施与监管机制三个层面提出人工智能产业包容审慎监管的实现路径。

(一) 监管原则

在实施包容审慎监管的过程中,我们应该在已有法规体系的基础上,对法规的滞后或真空领域,采取“基于原则的监管模式”,根据监管原则采取相对应的监管手段和措施。

1. 底线监管原则。

市场机制的核心是基于市场秩序之上的公平竞争,政府在人工智能产业监管中务必做好“守夜人”,维护市场的公平秩序和竞争环境。市场主体利用人工智能新技术的恶意欺诈、利用算法黑箱的不透明性的恶意编码植入、商业机密与国家机密信息主观泄露等行为,在本质上是一种基于主观恶意的不法行为,严重损害投资者和消费者的正当权益,扰乱正常的市场秩序,该种行为在人工智能产业的风险等级应划定为不可接受的风险,监管当局应以零容忍的态度对其予以严厉制裁和惩戒。另外,对个人的生命、自由、人格尊严等人身权益有重大影响以及关系到国家安全的关键信息基础设施的人工智能技术与产品,监管机构也必须坚持底线监管原则,对人工智能开发者和提供者组织机构、风险评估、风险披露、安全应急处理等方面提出更高的监管要求。

2. 以人为本原则。

习近平主席在向联合国贸易和发展会议成立60周年庆祝活动开幕式发表视频致辞时提出:“坚持以人为本、智能向善,在联合国框架内加强人工智能规则治理”^⑥。在人工智能产业监管实践中,我们也应秉持这一“以人为本、智能向善”的基本原则。一是基础层和技术层的人工智能技术研发和人工智能产品开发应坚持“科技服务于人”的根本伦理准则,将“以人为本”科技伦理融入人工智能开发和使用的全生命周期。我国在2022年出台了有关人工智能伦理治理的国家政策,提出科技伦理治理的原则和行动方案,贯彻科技进步和创新发展的根本目的是让人民生活更美好这一基本准则,在包容审慎监管的过程中,我们也应始终坚持这一理念。二是在应用层,也应坚持以人为本的理念,将人工智能产业发展对社会的冲击和影响最小化。人工智能产业发展所引起的群体性结构性失业、收入差距两极分化、数字

鸿沟等社会问题,将对传统产业和从业人员产生巨大的冲击,在经济大环境向下、国际形势不稳定等新常态条件下,社会和个人都难以承受社会急剧调整的成本和代价,因而务必综合考虑不同利益主体的诉求,统筹考量社会的稳定和安全。

(二) 监管措施

监管手段是为监管目的服务的,要达到平衡创新与风险、统筹效率与安全的监管目标,就必须采用分级分类监管、全过程动态监管,以及多主体协同监管等监管措施。

1. 分类分级监管。

不同的人工智能产业层级所具有的技术经济特性是不同的,其风险属性也迥异,因此在人工智能产业监管的过程中,应该秉持分类监管的原则,根据人工智能产业“技术—经济—风险”特征,将人工智能产业中的风险划分为归属于基础层和技术层的技术创新风险与产品开发风险,以及归属于应用层的产业应用风险。在监管实践中,应根据不同的风险类型来制定和实施相应的监管对策与措施。对基础层和技术层而言,在监管中应以包容为主、审慎为辅,采取放松准入的规制措施,通过加强事中和事后监管来有效防范创新风险和开发风险。对应用层而言,产业应用风险较为复杂和多变,且涉及更多关系国计民生的社会问题,因而在监管中应以审慎为主、包容为辅,加强社会性规制,强化企业社会责任,防范社会风险,减少社会调整成本。与此同时,同一类别的风险也有不同的风险表现形式,对经济社会的潜在威胁与危害也不相同,因此,在人工智能产业监管过程中,还应秉持分级监管的原则,对不同程度的风险,划分不同的风险等级,根据风险等级来制定和实施相应的监管对策与措施。目前欧盟《人工智能法案》将人工智能风险等级划分为不可接受风险、高风险、有限风险、低或极小风险四个等级,我们可以借鉴这一方法,根据我国人工智能产业发展的实际特点,划定风险等级,有的放矢地制定监管措施。

2. 全过程动态监管。

为规避监管机构信息与知识双重盲区的不足,有效解决监管滞后与监管真空问题,必须对人工智能技术研发、产品开发以及产业应用等全生命周期实施全过程动态监管。在事前,对人工智能项目应进行科学评估,全面了解项目的技术现状、开发趋势以及可能存在的风险因素,在综合评估的基础上给出是否准入的监管意见。在事中,必须对人工智能技术研发、产品开发以及投入应用等全过程跟踪研究,准确把握全过程的“技术—经济—风险”特性,对

风险进行分类分级,确保风险在可控范围,有的放矢地制定和实施相对应的监管措施。在事后,应动态跟踪人工智能产品在产业应用中的后续进展和社会影响,并对事后可能出现的数据滥用、侵犯隐私等不良行为和后果予以关注。最后,还应根据全流程的监管记录,给出涵盖事前事中事后全流程的监管报告,为后续产业监管提供参考和借鉴,也为出台和完善相应的法规体系提供事实依据。

3. 多主体协同监管。

针对人工智能产业跨市场、跨行业、跨领域的特点,对人工智能产业的监管也应采取多主体协同监管的措施,形成政府、行业、企业、公众等多方主体协同的监管格局。政府部门监管应由单一专业化部门监管走向综合跨部门监管,实现政府机构内部的监管协同,并作为人工智能产业监管的第一责任人,发挥市场监管的引领和主导作用,牵头制定人工智能产业的法规制度和监管体系,并主导实施对人工智能的产业监管。行业协会作为介于政府与企业之间的自律性中间组织,具有制定行业规范和标准,引导企业会员遵守行业规则,发挥行业自律性监管的作用。企业既是人工智能产业的市场参与者,同时也是产业监管的自我监督者,承担和履行管控市场风险和加强内部治理的企业社会责任。学界、法律界、媒体和消费者所组成的公众团体则可以在人工智能产业监管中发挥重要的社会监管与外部监管作用,通过媒体监督、消费者投诉、独立第三方机构参与监管调查等确保监管的公平公正和公开。四方主体基于不同的功能作用,在通力协作的基础上可以构建起完整的、多元化的人工智能产业监管体系。

(三) 监管机制

监管沙盒作为创新的监管机制,为践行包容审慎监管理念提供了临时性、局部性的试验场所,既能技术创新留有足够的发展空间,又能推进监管政策的迭代修改,是技术与制度协同创新的实践依托。本文从事前准入、事中监管以及事后退出三个阶段设计人工智能产业监管沙盒的运行机制。

首先,在监管沙盒的项目准入阶段,监管机构应主要采用包容监管的理念和措施,以放宽准入为基本原则,根据我国人工智能产业发展实际和特点,从进入沙盒的项目或产品范围、项目的创新与风险特性、项目的市场价值等方面向市场公开发布监管沙盒的准入标准。在沙盒项目与产品范围上,考虑到人工智能产业基础层和技术层在产业发展中的核心作用,可进一步对基础层的基础技术研发项目与技术层的技术与产品开发项目放宽准入。

在项目的创新与风险特性上,要求拟进入监管沙盒的企业必须对项目或产品的创新与风险特性予以公开说明和披露,以便监管机构判断该项目是否真正具备创新价值且风险是否可控。在项目的市场价值上,要求拟进入沙盒的人工智能项目或产品具有为消费者提供或创造价值的市场前景,以凸显“科技服务于人”的以人为本原则。人工智能企业在具备以上三个方面的准入标准前提下,可以向监管机构提出监管沙盒项目申请,监管机构根据企业提交的项目申请书,对照公开的准入标准进行项目审核,符合准入标准的企业可正式进入监管沙盒开始监管测试。

其次,在事中的监管沙盒测试阶段,监管机构要坚持审慎监管的原则,全流程全方位地对被测试项目进行持续动态的监测和监管。通过对沙盒项目创新和风险的动态监测,监管机构和被测试企业均可以全面学习和了解创新项目或产品的技术与风险特性,企业可以增加学习和技能积累,监管机构则获得有关项目信息与专业知识,降低信息不对称程度,减少知识盲区。监管机构通过实时了解和监测被测试项目的风险属性与风险动态,可以采取相对应的动态监管措施。同时,监管机构要与被测试企业保持密切沟通与合作,共同收集信息和数据,让企业在监管机构的合作和支持下更好地实现创新并管控风险。并且在测试过程中,监管机构还要坚持底线原则,贯彻科技以人为本的理念,如果企业的测试参数超出范围,风险强度超标,监管机构可以发出警告,要求企业立即采取补救措施,如果企业没有做出适当反应,监管机构可以要求终止测试、处以罚款或撤销开展相关测试活动的许可。

最后,在沙盒退出阶段,监管机构应该要求被测试企业妥善处理涉及个人数据、商业机密等隐私数据,并确保其行为符合相关法律法规的要求。监管机构应承担起审慎监管的职责,切实维护消费者的个人隐私权益,让企业在退出沙盒后以安全、可靠的方式存储和使用数据。同时,在退出时,应由独立且公正的第三方机构对沙盒测试项目进行专业评估和安全认证,监管机构依据该评估报告,结合沙盒监管协议和测试结果,对沙盒测试项目或产品做出无条件通过测试(可按测试结果不做更改即向市场提供人工智能产品或服务)、有条件通过测试(在修改测试结果并审核通过后向市场提供人工智能产品或服务)或不予通过测试等决定。此外,在沙盒项目退出时,企业应与监管机构一道共同编制监管沙盒测试退出报告,详细记录并报告沙盒中的人工智能技术

创新过程、衍生风险特性、风险管控机制等。

五、结论与研究展望

(一) 结论

人工智能产业是人工智能技术研发、产品开发以及市场应用的产业部门,可以划分为人工智能产业化部门与产业智能化部门。不同的产业部门具有不同的“技术—经济—风险”特征,在基础层和技术层主要表现为技术特性与技术风险特性,在应用层主要表现为经济特性与市场风险与社会风险特性。人工智能产业不可避免的创新风险、市场风险与社会风险给人工智能产业监管提出了巨大挑战。我国目前对人工智能产业的纵向监管模式无法适应人工智能跨行业、跨主体、跨领域的“技术—经济—风险”特性,只有包容审慎监管体系才能与人工智能产业的实际特征高度契合,并能促使人工智能企业做出个体理性与集体理性相一致的行为选择。在包容审慎监管实践中,应以底线思维和以人为本的监管原则为指导,以监管沙盒为主要运行机制,通过分级分类、全过程动态和多主体协同等监管手段和措施,有效平衡人工智能产业的创新与风险。

(二) 研究展望

本文虽然从监管原则、监管措施以及监管机制三个层面提出了人工智能产业包容审慎监管的实施路径,但仅给出了初步的原则性的监管方案,实施路径的具体步骤与举措以及监管体系的实施细则等还有待后续进一步的细化与深入研究。一是分级分类监管体系的构建,需要进一步对人工智能产业的风险特性进行深入分析和论证,完善和明确人工智能产业的风险分级分类体系,并针对不同类型不同层级的不同风险,制定一一对应的监管措施与方案;二是监管沙盒的具体设计,需要进一步针对人工智能产业不同的部门、不同的层级、不同的场景、不同的项目,制定相对应的沙盒运行机制。并且对人工智能产业包容审慎监管体系的实施与操作还需要在实践中不断地探索和研究,并在实践中不断地修订和完善。

注释:

①参见《国家主席习近平向2024世界智能产业博览会致贺信》(https://www.gov.cn/yaowen/liebiao/202406/content_6958352.htm)。

②参见中国政府网《数字经济及其核心产业统计分类(2021)》(国家统计局令(第33号))(https://www.gov.cn/gongbao/content/2021/content_5625996.htm)。

③参见英伟达:“AI时代的大赢家”是如何炼成的?

(<https://www.sohu.com/a/788143737121124378>)。

④参见《习近平主持中共中央政治局第九次集体学习并讲话》(https://www.gov.cn/xinwen/2018-10-31/content_5336251.htm)。

⑤参见《武汉市民投诉萝卜快跑致交通拥堵,相关部门:无人驾驶技术还不成熟,正改进》(https://news.sohu.com/a/782373242_116237)。

⑥参见《习近平向联合国贸易和发展会议成立60周年庆祝活动开幕式发表视频致辞》(https://www.gov.cn/yaowen/liebiao/202406/content_6956945.htm)。

参考文献:

[1]王兆星.机构监管与功能监管的变革——银行监管改革探索之七.中国金融,2015(3):14-18.

[2]刘太刚.从审慎监管到包容审慎监管的学理探析——基于需求溢出理论视角下的风险治理与监管.理论探索,2019(2):56-62.

[3]廖凡.论金融科技包容审慎监管.中外法学,2019(3):797-816.

[4]张友连,胡洁林.论法治营商环境中的“包容审慎”监管——基于《优化营商环境条例》的分析.浙江工业大学学报(社会科学版),2020(3):267-272.

[5]钱贵民,阳镇,陈劲.平台监管逻辑的反思与重构——兼对包容审慎监管理念的再反思.西安交通大学学报(社会科学版),2022(1):131-140.

[6]刘权.数字经济视域下包容审慎监管的法治逻辑.法学研究,2022(4):37-51.

[7]田晓祥.包容审慎监管的法治逻辑与地方实践.中国市场监管研究,2023(10):44-47.

[8]方福前.既“放得活”又“管得住”.光明日报,2024-08-13(11).

[9]陈龙,刘刚,戚丰东,等.人工智能技术革命:演进、影响和应对.国际经济评论,2024(3):9-51.

[10]习近平.不断开拓当代中国马克思主义政治经济学新境界.求是,2020(16):4-9.

[11]鞠宏磊,申欣雨.生成式人工智能产品的内容安全风险及监管路径.经济问题,2023(12):16-19.

[12]张素华,李凯.生成式人工智能虚假信息风险与治理研究.学术探索,2024(7):129-140.

[13]苏玺鉴,胡安俊.人工智能的产业与区域渗透:态势、动力、模式与挑战.经济学家,2023(2):79-89.

[14]李翔,叶初升,潘丽群.人工智能何以提升中国制造业发展质量——索洛悖论在中国制造业的再检验.兰州大学学报(社会科学版),2023(4):44-58.

[15]王林辉,胡晟明,董直庆.人工智能技术、任务属性与职业可替代风险:来自微观层面的经验证据.管理世界,2022(7):60-79.

[16]解学芳,曲晨.价值对齐:AIGC时代的人工智能文化科技伦理风险与精准共治路径研究.兰州大学学报(社会科学版),2024(3):147-156.

Accommodative and Prudential Regulation of the Artificial Intelligence Industry: Theoretical Connotation and Implementation Path

Qi Yudong Liu Jian

Abstract: The AI industry is an emerging industry with new characteristics such as "new technologies, new products, new economy, and new risks". Research on AI industry regulation modes and paths is the key to balancing innovation and risks. The accommodative and prudential regulation can be regarded as a regulatory model that integrates inclusive innovation and risk prevention. With the theoretical connotation of integrating effective markets with promising governments, the regulation is a practical manifestation of the relational theory of the market and government in supervision areas. According to the research approach of hierarchical classification, the layers of the foundation, technology, and application of the AI industry have different technical, economic, and risk characteristics. Since accommodative and prudential regulation is in accordance with the AI industry's technical and economic characteristics of spanning multiple fields, entities, and industries, and can effectively prevent the AI industry's attributes of multi-level, diversified, and composite risks, it is an effective regulatory model for the new economical forms like the AI industry. In regulatory practice, it is necessary to adopt bottom line thinking and people-oriented philosophy as regulatory principles, regulation sandbox as the primary operating mechanism, and graded classification of risks, dynamic monitoring throughout the entire process, and multi-agent collaborative regulator system as the main methods, to effectively practice accommodative and prudential regulation of the AI industry.

Key words: artificial intelligence industry; accommodative and prudential regulation; theoretical connotation; applicability; implementation path