

# 论通用人工智能时代的企业合规<sup>\*</sup>

印 波 张笑宇

**内容提要** 通用人工智能是新型的社会治理主体,其凭借对数据更强力的汲取与计算能力,建构与真实世界平行的“数据世界”,令企业实现从“粗放式”合规到“精细化”合规、从“框架型”合规到“精准型”合规的转型。在通用人工智能时代,人工智能代替了传统合规人工操作方式,通过智能处理能够直接“穿透”企业的违规风险,为企业生成具有差异化导向的合规计划,使企业实现了有效合规。企业合规发生了从义务识别到风险导向、从行为合规到全生态合规、从人工治理到技术治理的转向,事前合规的形式成为企业合规主导。然而,通用人工智能也带来了合规技术标准、数据安全、人机关系的新问题。在“智能合规”的中国图景中,应当对通用人工智能算法可解释性、新型涉及国家安全数据进行深入研究,并始终让人类对决策负有最终责任,以维护人机关系和谐。

**关键词** 通用人工智能 企业合规 差异化合规 技术治理

DOI:10.16091/j.cnki.cn32-1308/c.2024.04.005

## 引 言

2023年,OpenAI推出大型多模态模型ChatGPT-4,其能够阅读文字、识别图像并自动搜索网络资料,也可以处理多种媒体数据并将其整合到统一的语义空间之中。《通用人工智能的火花:GPT-4的早期研究》<sup>①</sup>一文指出,GPT-4能力的通用性可以与跨越诸多领域的众多能力相结合,它在广泛的范围上有着达到或超越人类水平的表现。GPT-4是迈向通用人工智能(artificial general intelligence,简称AGI)的重要一步。以GPT-4为代表的生成式人工智能标志着人工智能由专用智能迈向通用智能,加速了AGI走进现实的进程。

---

<sup>\*</sup> 本文系中央高校基本科研业务费专项资金资助项目“中国式犯罪治理现代化视野下的新犯罪学理论体系研究”(项目编号:24CXTD02)的阶段性成果。

<sup>①</sup> S. Bubeck, V. Chandrasekaran, R. Eldan, et al., “Sparks of Artificial General Intelligence: Early Experiments with GPT-4”, <http://arxiv.org/abs/2303.12712>.

运用 ChatGPT 是目前最可能实现 AGI 的路径,因此,对早期阶段 AGI 的理解必须立足于 ChatGPT 的训练模式。从 ChatGPT 的本质看,“GPT 模型是通用一个大一统模型认识世界,再用训练出来的认知对具体领域进行降维打击”。<sup>①</sup> 在这个前提下,AGI 不会对依靠人类身体劳动的产业产生根本性冲击,<sup>②</sup>冲击会从依靠系统知识培训的脑力劳动的产业开始,尤其会冲击以具有深度复杂性、工作人员高智商为典型特征的非决策性工作。换言之,尽管早期阶段的 AGI 所带来的深刻变化会体现在产业和技术领域中,但更会体现在治理与政策选择上,<sup>③</sup>因此我们更需关注治理与政策风险。企业是创新的前沿,科技发展所带来的变革会首先反映在企业的运营中,带动企业在产业与管理上的创新。企业合规是国家治理方式、治理理念完成重大转变的重要环节,“国家-企业”共治是国家治理、企业治理与产业发展的交融点。AGI 会引发企业治理转型,会促使国家治理方式、治理理念进一步转变,推动企业合规的范式革命,但其也会给国家、企业带来新一轮的治理与政策风险。因此,必须面向未来,深入认识和理解 AGI 构成的未来企业合规的生态图景,把握未来企业合规的换轨升级,以服务于企业合规现代化转型。

对 AGI 时代的展望也是饱受质疑的。基于 AGI 风险的不可控性,人类社会对 AGI 的态度必然是“控制发展”,而非“放任自流”。这决定了 AGI 的发展必然是阶段性的,只有当人类理性能够认识特定阶段的 AGI 并确保其发展的可控性时,AGI 的技术发展才能够进入下一阶段。在 AGI 早期阶段,基于数据安全和 AGI 决策可控、可信的要求,人类社会必然呈现“智能+人工”的人机协作模式。当前,数据仅在必要的使用范围内流通,如此才能保证数据安全,并使人类对最终决策负责。笔者仅对这一 AGI 早期阶段进行展望,具体讨论 AGI 技术对企业合规的推进可能及其带来的治理与政策风险。

## AGI:迈向差异化合规的技术图景

企业合规是指“企业为有效防范、识别、应对可能发生的合规风险所建立的一整套公司治理体系”。<sup>④</sup> 企业合规的目的在于发现和预防企业内部犯罪。以此视角复盘企业合规的演进可以发现,国家主导是企业合规的根本内容,通过差异化合规实现对企业气质的关照是有效合规的基本逻辑。但是,国家主导天然决定了以差异化合规实现有效合规具有内在困境,AGI 则提供了以差异化合规实现有效合规的可能性。

### (一) AGI 的实质与功能

目前,对 AGI 的概念还没有统一的界定。不过,对于 AGI 的基本特征已有如下共识:其一,其无须人类特别编码即可实现人类可以实现的各种目标;其二,其在与环境的相互作用中学习适应情境变化,具有自主学习以适应环境的能力;其三,其能够自主作出决策,而非仅是执行任务;其四,其能进行抽象思考和推理。AGI 的主要功能是对自然意义上的人类智能进行模

<sup>①</sup> 郭全中、袁柏林:《从 GPT 看 AGI 的本质突破:传媒业新挑战与未来》,《新闻爱好者》2023 年第 4 期。

<sup>②</sup> 孔微巍、谭婷婷:《人工智能对我国就业的影响及对策研究》,《理论探讨》2022 年第 3 期。

<sup>③</sup> 尹栾玉、隋音:《技术创新对治理模式的挑战及风险防范》,《中国特色社会主义研究》2020 年第 21 期。

<sup>④</sup> 陈瑞华:《企业合规的基本问题》,《中国法律评论》2020 年第 1 期。

拟,理解和学习完成人类可以执行的任何治理任务。<sup>①</sup>可以说,AGI就是数字法学者们理想中具有法律主体地位的超级人工智能,是可信、可控的社会主体。

AGI本质上能够实现互联网、5G、虚拟现实、沉浸式体验、大数据、区块链、产业互联网、云计算及数字孪生等互联网时代诸种技术的生态级融合。<sup>②</sup>借助于各种场景化的应用,AGI迅速下沉成为整个社会的“操作系统”,汲取了大量真实世界的数据,构建起一个与真实世界孪生的“平行世界”。AGI凭借其高度智能性成为智能时代社会管理主体,在深度卷入社会各项信息的基础上,精准预测、识别每个人的需求,<sup>③</sup>构建“智能化的社会运转模式,精细化的社会交易结构,定制化的需求供应体系”。<sup>④</sup>其可提升每一个社会个体的“声量”,关照每一个社会个体的需求。人工智能作品与人类作品在创作主体蕴含着物质与经济的价值。<sup>⑤</sup>

## (二)有效合规的差异化导向

合规最初是企业的自我监管。企业为在社会中树立守法的良好形象,获取政府、顾客的好感,获取更丰厚的商业利润,会通过规范员工行为来强化自我控制能力。行业协会也会制定合规指南为企业合规提供合规指引。<sup>⑥</sup>内生性的合规缺乏强制性惩罚代价,企业违规获益将远超违规惩罚后果,无法有效预防企业犯罪行为。为保障公共利益,国家力量介入企业犯罪预防,<sup>⑦</sup>形成了以行政合规和刑事合规为主导的现代合规机制。现代合规机制具有典型的科层制治理特征:其一,其合规动力来源于外部力量。国家通过制定合规标准、文件,设定一系列奖励与惩治制度,构建行政合规监督、刑事合规监督机制,强制性地介入公司治理,促使企业合规。其二,由国家干预企业合规整改,通过监督机关确保企业合规标准的有效性,将企业违规结果作为治理基准,并将企业治理体系整改为合规结果。这既是一种责任追究机制,也是一种合作型的治理模式。其三,国家仅对具体企业的“健康状况”进行干预,以“单体治疗”为治理策略。在国家干预的前提下,将企业内部治理体系作为合规产品的原因在于:首先,企业的独立性是企业积极能动性的来源,将合规内化为企业治理体系内容有利于保证企业意志的独立性,激发市场经济活力。其次,企业经营会产生海量数据,主要依靠人工操作的传统监管模式无法有效发现和预警犯罪,会导致“外部监管失灵”。<sup>⑧</sup>因此,只能令企业自我监管,防患未然。最后,合规过程需要大量的治理资源,企业承担主要的合规成本有利于降低国家治理压力。这决定了企业合规能够使企业自主承担犯罪预防职责,保证企业高度自治。

以国家干预为主导的企业合规的问题在于,国家机器的运行过程主要依靠人工操作,即通过科层制在权力顶端制定框架式的规范及示范案例,为下级机关提供治理依据。以国家为主导的企业合规决定了合规有效性评估标准是一种宏观标准,其是通过发掘企业违规行为的共

① 江怡、董化文:《论人工智能与人类智能的双向互动》,《自然辩证法通讯》2023年第11期。

② 喻国明:《AGI崛起下社会生态的重构》,《人民论坛·学术前沿》2023年第14期。

③ 孟天广:《智能治理:通用人工智能时代的治理命题》,《学海》2023年第2期。

④ 周佑勇:《论智能时代的技术逻辑与法律变革》,《东南大学学报(哲学社会科学版)》2019年第5期。

⑤ 韩雨潇:《生成式人工智能作品版权归属问题研究》,《扬州大学学报(人文社会科学版)》2023年第4期。

⑥ 陈瑞华:《论企业合规的性质》,《浙江工商大学学报》2021年第1期。

⑦ 丁胜明、张阳:《合作治理视域下企业合规的刑法立法研究》,《学海》2022年第6期。

⑧ 陈瑞华:《论企业合规的基本价值》,《法学论坛》2021年第6期。

性原因,形成企业共同合规义务,即合规要素。合规要素过于宏观化、形式化,是合规不能深入具体企业气质,“穿透”企业违规行为的深层次原因。国家缺乏智能治理的技术手段,无法关照企业具体的合规计划,企业仅需达成“书面合规计划有效性”,对不同的合规要素作出不同的排列组合,<sup>①</sup>对合规计划进行形式模仿,即可完成合规整改。同时,合规有效性标准的宏观性也难以指导监管机关规范评估,难以确认合规结果的有效性,企业仅通过“纸面合规”即可规避违规责任。

晚近以来,一些国家越来越强调“合规文化”在有效合规中的引领作用。例如,美国就将企业文化作为合规审查标准,以求通过“常识性和务实的方法”评估合规计划的有效性,解决“纸面合规”问题。<sup>②</sup>我国也采用了相同的措施。合规文化作为独立的治理标准,平行于企业内部治理体系,甚至有时被作为企业内部治理体系的上位标准。<sup>③</sup>因此,合规文化也被认为是有效合规之实质标准。<sup>④</sup>但是,合规文化并非一种审查标准,合规文化的真实功能在于赋予监管机关对合规计划、合规结果有效性进行差异化审查的权力,以解决合规有效性标准宏观性的问题。公司文化是公司治理结构的衍生品。<sup>⑤</sup>合规文化的建设依赖于技术指标,而没有独立的实践路径。因此,合规文化并不具有审查功能。英国金融行为监督局(Financial Conduct Authority)认为,“文化可能无法衡量,但它是可以被管理的”。显然,之前试图衡量合规文化的尝试都失败了。<sup>⑥</sup>既然不能确认合规文化是否建立,或者合规文化建设到了何种程度,便只能寄希望于监管机关在合规实践中凭借经验、能力与“良心”实现合规差异化审查,探寻符合企业气质的差异化合规计划就成为有效合规演进的基本要求。

### (三)通用人工智能与有效合规的逻辑关系

卢曼认为,“法律系统一方面在运作上保持封闭,仅仅依靠自己的合法/非法符码进行运作;另一方面又在认知上向环境中的各种讯息保持开放”。<sup>⑦</sup>合规文化加强了这种法律自治,使监管机关能够在合规要素之外广泛地探寻企业的现实状态与外在环境,建立了合规计划与企业的适配关系。合规要素提供了企业内部治理体系有效/无效的二维代码,其宏观化打开了企业对现实状态与外在环境的认知。考察企业并细致化合规要素以整改企业内部治理体系,是合规要素在认识上开放的方式。然而,合规文化的出现实质上将有有效/无效的二维代码判定完全交送于环境,使合规标准完全开放。那么,是否可以将合规有效性标准的制定完全交由监管机关?答案显然是否定的。其一,合规文化本身并没有为合规提供有效标准,如何针对企业涉罪性质及企业行业特点、业务范围等达成有效标准缺乏规范内容。企业文化既缺乏实践供

① 陈瑞华:《有效合规的中国经验》,北京大学出版社,2023年,第2页。

② 李勇:《涉罪企业合规有效性标准研究——以A公司串通投标案为例》,《政法论坛》2022年第1期。

③ 刘艳红:《涉案企业合规建设的有效性标准研究——以刑事涉案企业合规的犯罪预防为视角》,《东方法学》2022年第4期。

④ 张远煌、秦开炎:《合规文化:企业有效合规之实质标准》,《江西社会科学》2022年第5期。

⑤ 时延安:《合规计划实施与单位的刑事归责》,《法学杂志》2019年第9期。

⑥ W. M. Burdon, M. K. Sorour, “Institutional Theory and Evolution of ‘A Legitimate’ Compliance Culture: The Case of the UK Financial Service Sector”, *Journal of Business Ethics*, Vol. 162, No. 1 (2020), p. 49.

⑦ 鲁楠、陆宇峰:《卢曼社会系统论视野中的法律自治》,《清华法学》2008年第2期。

能作用,也缺乏对监督机关的限权及行权监督功能。其二,有效监管的缺位是合规整改机制失灵的直接原因,第三方监督评估“形式化”导致合规整改计划流于形式。对此,有学者提出要建立检察机关“全流程再监督”。<sup>①</sup>显然,将监督功能交予第三方的一大原因就是检察机关缺乏充足的治理资源或治理能力,难以进行合规整改。同时,如果由检察机关进行“全流程再监督”,等同于又将合规有效性标准交给了检察机关(检察机关行使无限制的权力,必然使第三方监督以检察机关的意见为标准),那么第三方监督的意义何在?其三,企业在运营过程中会产生海量数据,以人工操作为主的传统合规难以将数据转化为信息,这会限制对企业违规原因的深层次探索。其四,即使对“二次合规”问题证明进行有效合规,由于风险的不完全可控性以及企业运营过程中产生的新风险,其也不能完全防止企业违规问题。因此以违规行为为治理基点的传统合规模式,在时间上体现为“节点式”的合规,不能完全保证治理成效。总体而言,对企业运营全流程监控机制的缺乏与人类行为的不确定性是传统企业合规形式化面临的根本问题。

AGI 节约治理资源的外观体现为对人类劳动的替代,在合规中,AGI 的功能恰恰在于以替代人类劳动的方式实现对传统合规形式化问题的解决。作为各种具体应用的生态级融合工具,AGI 是社会的底层操作系统。随着人工智能系统不断嵌入越来越多的智能设备中,信息时代的一切技术都将在 AGI 中实现交互,这将带动万物互联时代的到来。此时,AGI 不仅是合规的管理者,更是合规的监视者。任何通过 AGI 进行的企业行为均是 AGI 认识企业的行为,这样就能在时间和空间上扩展对企业的监控。这种全流程的监控,使 AGI 对企业的行业特点、业务范围等具体运营细节实现绝对的掌握。相对于传统合规,AGI 能够更有力地“穿透”企业违规的原因。在生成结果上,AGI 更加稳定、权威和规范,不会因技术人员水平参差、监督机关管理能力强弱、人工操作上限等产生不确定的合规结果。换言之,AGI 能够在对企业全流程监控的基础上,对企业合规风险进行预警、发现,并根据企业特点与现实情况生成差异化合规计划,极大程度地超越了传统合规的局限,带来合规范式革命。

### 从 AGI 路向看企业合规的发展趋势

从专用到通用,从工具到主体,AGI 的发展带来的既是人类劳动力的解放,也是人类尊严的提升。传统治理模式中,社会个体被类型化而失去了自身的“面部特征”。AGI 则对社会个体需求作了细致化的考察,还原出了每一个社会主体的具体“面部特征”。AGI 使合规义务的生成由国家主导向企业主导转型,孕育了“再中心化”的合规格局。以企业为合规义务生成主体,通过对企业“健康状况”的监测与对企业外在“适配环境”的探测,以法律法规、商业伦理、行业规范以及实践案例等为映射,AGI 可以实现对合规标准的细致化模拟与反思性完善,对企业在生态运营中所遇到的实践合规风险作出动态的识别、预测与警示。

---

<sup>①</sup> 程雷、曲育铮:《涉案企业合规改革现状及批判性反思》,《上海政法学院学报(法治论丛)》2023 年第 5 期。



### (一)从义务识别到风险导向

合规义务识别是合规计划有效性的基础性建构。已经有企业在事前自主合规体系建设中使用 ChatGPT 实现以智能识别替代人工识别,从立法、司法解释、司法政策变化、专项执法行动和海量判决中识别合规义务。<sup>①</sup> 当下智能实践使用的是以合规要素和类案标准为基准建构的数据,相关数据立足于企业违规共性原因与相似企业违规原因,极大地提升了企业发现与预防犯罪的概率。在信息化时代,信息分享与传播是多节点化的类案合规经验的公开,其也推动了事前合规有效性发展。随着事前合规有效性的提升,合规义务识别的话语权也从国家转移至企业,实现了“去中心化”。多元参与、分散治理的企业事前合规格局初步形成,更孕育了 AGI 时代的合规“再中心化”格局。

“再中心化”在本质上是以合规风险识别为导向的合规机制。AGI 时代合规义务具有两个来源:其一,代码形式的合规要素。<sup>②</sup> AGI 时代法律代码化成为新的趋势,<sup>③</sup> 合规要素以代码的形式存在,在持续发挥框架功能的同时,自动与复杂情景结合,使合规要素向“私人化”方向发展。其二,具体化企业的合规风险。所谓与复杂情景结合,是指企业所面临的现实的或未来的合规风险。在全面构建数据世界的基础上,AGI 能够对真实世界的复杂性进行分析,准确认识影响企业行为的全部因素,发现和预警企业违规行为,在每一个事实因果关系上进行“点对点”的治理,以风险为导向,形成差异化的合规计划。

法律的沟通不再是从合规要素出发认知现实环境,而是将合规要素作为现实环境的一环。这就使得以事后合规为主的合规机制向以事前合规为主的合规机制转型,企业在合规标准构建中发挥主导作用。这种合规义务生成方式颠覆了传统治理模式的科层制结构,并在结构和重组中实现了企业合规的“再中心化”。

### (二)从行为合规到全生态合规

就未来 AGI 的基本构造而言,企业合规范式革新将是一个生态级的系统工作,涵盖企业的全部“社会关系”。AGI 输出产品的可信性必须基于大数据所获得信息的全面性。因此,企业运营全生态数据的汲取尤为关键,这关系到企业行为的因果建构。传统合规以企业内部治理体系规范化为手段,以规避企业或企业人员违规意识的生成为目的,通过经验性的构建手段实现合规整改,具有一定的滞后性。而 AGI 时代企业合规义务识别是一种对真实世界风险的规避,其涵盖了强调先例的经验性及强调对企业违规风险的预警,具有超然的预测性。

利益相关者理论认为,公司的行为建构不只基于股东利益最大化,还与所有公司利益相关者的利益相关,公司天然服务于所有能够与公司收益相连接的相关者。利益相关者理论强调包括“包括雇员(职工)、客户(消费者)、债权人(供应商)社区、政府等在内的所有利益相关者”<sup>④</sup>对企业动因的影响。换言之,企业动因的特质性是基于利益相关者的叠加。在“单体治疗”策略的引导下,企业合规更加关注企业内部治理体系与宏观的合规标准的适配关系。而

① 谢澍:《互联网企业刑事合规义务识别:分层、复合与技数赋能》,《云南社会科学》2023 年第 3 期。

② 杨丰一、孙萍:《数字社会政府治理的逻辑梳理与矛盾纾解》,《中国特色社会主义研究》2023 年第 2 期。

③ 马长山:《数字法治概论》,法律出版社,2022 年,第 16—17 页。

④ 田虹:《从利益相关者视角看企业社会责任》,《管理现代化》2006 年第 1 期。

在 AGI 时代,企业通过生活、工作场景的智能化供给,在对社会个体的全方位智能支持过程中形成相应数据,合规策略从“单体治疗”转向“全生态合规”。

企业全生态运营合规,有三个关键的数据支撑:其一,企业全生命周期数据。AGI 推动新产业、新技术、新业态的快速生成,企业发展的各个阶段和环节均有可能产生难以估量的风险。特别是如果不能及时、有效地发现企业犯罪行为,在社会变革时期可能会产生重大损失。只有对企业持续观测,才能及时发现、预警企业犯罪行为,把握治理全局。其二,企业的“健康状态”与对外“适配环境”。结构功能主义认为“社会是由诸多相互依存的单位组成的统一系统”,各单位具有特定功能,相互依存也相互制约,这种并存的“吸力”与“斥力”形成了平衡的整合系统。而对外“适配环境”则是是否会引发企业状态变化的影响因素。通过对企业自身状态及企业与外界联系的监督,能够分析、发现企业内部治理体系与企业犯罪风险之间的现实关系,进行预测性的合规。其三,关注宏观合规义务。企业数据是企业生产经营管理中所生产、收集的各种数据。<sup>①</sup> 企业数据引导着企业对现实世界的认识,决定了企业的行为导向。宏观合规义务,尤其是商业伦理、行业规范以及实践案例等变化,是对企业数据不足之处的弥补。

### (三)从人工治理到技术治理

传统治理模式是科层制的组织结构,依赖于职能的分工与分层,通过自上而下的专家统治形成等级制的权力矩阵。这在合规中体现为国家主导,监督机关、企业等标定不同职能,并且人类在治理中发挥主体功能。目前,数据化及智能化尚是有限的。我们所掌握的智能工具智能程度不高,例如,ChatGPT 就存在信息滞后、信息载体受限、人工智能幻觉等多种问题。因此,以人为治理主体,强调人对技术工具的支配依旧是传统治理范式。AGI 作为治理主体,在极大程度上代理了人类智能,其将不同主体的职能相融,使科技产品深入内嵌于治理活动的全过程,人类与技术成为共同的治理主体。当科技产品开始具有独立的品格时,社会治理便进入技术治理模式。<sup>②</sup>

其一,合规风险识别日渐依赖企业自身。如前文所述,AGI 生成的差异化合规计划的数据主要为企业运营过程中所产生、接触、收集到的数据。庞大的个人数据、商业数据使数据安全问题更加严峻。公共领域与私人领域的划分是现代文明的重要标志。在 AGI 时代,大量的个人数据、商业数据被 AGI 卷入。企业运营中所获取的数据均有可能成为预测或发现合规风险的样本,涉密数据与其他数据在用途上难以区分。管理部门使用其他主体生成、收集的数据将更为谨慎。在事前合规能够有效预防企业犯罪的前提下,管理部门对企业的干预也更为科学。以数据使用主体为分界,以有效合规为合规成果,形成了以企业为主体的合规机制。

其二,监督机关的监督对象为 AGI。<sup>③</sup> AGI 在社会生活中的广泛应用使其数据的积累远超传统治理模式的上限。在 AGI 建构的全景社会中,我们必须,也只能依赖 AGI,使其成为治理主体。在传统治理模式中,人类行为的可预期性源自法律法规、伦理道德等规则,而在 AGI 时代,人工智能行为的可预期性源自数据及算法。在数据与算法可信的前提下,AGI 对数据精细

<sup>①</sup> 程啸:《论数据权益》,《国家检察官学院学报》2023 年第 6 期。

<sup>②</sup> 宋辰熙、刘铮:《从“治理技术”到“技术治理”:社会治理的范式转换与路径选择》,《宁夏社会科学》2019 年第 6 期。

<sup>③</sup> 马亮:《良术善用:政府如何监管新一代人工智能技术》,《学海》2023 年第 2 期。

化的分析使其生成结果较之于人工操作的生成结果更为稳定可信。因此,监督机关无须再监督合规计划,仅需监督数据与算法,并依照合规计划整改成果,即可确保有效合规。

其三,依托 AGI 进行事前合规成为企业义务。合规生成于对企业行为的不信任,<sup>①</sup>是依靠外在力量而推进的治理机制。合规的理论依据在于,在治理资源有限的时代,以企业违规行为为基点的传统合规模式因治理成本无法覆盖所有的合规监管而不得不作出的妥协。合规能够以事前合规的方式深入每一个企业,其不再是企业运营中的“节点式”治疗,而是从企业设立之初就伴随企业共同成长的“终身守护”。明确企业在成立及运营过程中使用 AGI 的规则,强制实现 AGI 在企业中的“内嵌式”治理,能够确保企业进行有效事前合规。

### 基于 AGI 路径的企业合规发展规划

如前文所述,AGI 技术将会带来企业合规机制的重大变迁。面对逐渐显现的企业合规途径,应当提前布局。AGI 会给人类带来丰厚的社会福祉,但其也会带来新的风险,典型的就是 AGI 合规的技术标准、风险问题以及人机关系问题。如何对这些问题进行防控,是 AGI 路向的企业合规研究所需特别关注的论题。

#### (一) AGI 合规的技术标准:算法可解释性

当前人工智能存在缺乏可解释性的算法黑箱问题,用户仅能够看到生成结果,而无法了解生成的原因及过程。作为治理主体,AGI 必须得到人类的充分信任。国家可以通过建立专门审查机制和秘密保护规定,实现合规 AGI 在审查上的安全性。可解释的人工智能研究使 AGI 具有自我解释功能。

当然,算法可解释性还需要同步的制度构建。其一,算法透明,即 AGI 应当向专门审查机关公开其算法。算法在 AGI 中起着系统设计、实现功能和行为的基础性作用,但是算法透明不等于算法可知,由于干扰性披露的存在,大量冗余干扰性数据与关键数据被混同,妨碍了对关键数据内容的解释,形成了公开性与解释性之间的悖论。<sup>②</sup>但是,算法透明可以作为计算机工程师免责机制,推进 AGI 的事前规制,辅助 AGI 提高可解释性。其二,建立 AGI 合规反馈机制。无效合规始终以企业违规行为为结果,具有外显性。对此,应当建立企业相关利益者的合规反馈机制,在合规无效时向专门机关进行反馈,由专门机关改进引领 AGI 合规的技术问题,协助 AGI 的进一步发展。

#### (二) AGI 合规的风险问题:数据出境安全

数据是人工智能的基础性资料,优质数据是人工智能的驱动力量。在 AGI 时代,深度的人机交互关系必然使企业运营数据被深度卷入。企业经营是一个社会过程,其与政府、社会生活的天然亲近性决定了其会全面收集国家、社会及个人的数据。<sup>③</sup>基于 AGI 对真实数据的现

<sup>①</sup> 李本灿:《刑事合规的制度史考察:以美国法为切入点》,《上海政法学院学报(法治论丛)》2021年第6期。

<sup>②</sup> 沈伟伟:《算法透明原则的迷思——算法规制理论的批判》,《环球法律评论》2019年第6期。

<sup>③</sup> 印波、肖湘漓:《数据权的法律界定与权能构建——基于数据纠纷典型案例的分析》,《三峡大学学报(人文社会科学版)》2024年第5期。



实需求,人工智能成为治理主体的未来,秘密数据被人工智能大范围获取并成为企业经营数据,都是无法避免的。企业作为主要的跨境主体,也会面临数据出境所涉及的国家安全问题。

AGI时代涉及安全的数据更具隐蔽性,极可能藏匿在日常的经营数据之中。对AGI时代数据安全的思考应当基于数据审查的立场。其一,应当建立数据分类制度。<sup>①</sup>数据分类是规范审查的典型方法,通过事前的规制避免敏感数据出境,也是基础性的方法。其二,应当重视行业自我监管。行业自律相对于公权力介入更加具有规范适配意义。根据本行业的特点与主流企业的具体形态进行规范约束,对增强企业数据安全意识、充分发挥规范规制功能更加具有现实意义。具体而言就是通过成立各行业领域的专门的数据行业协会,建立适应本行业的重要数据安全机制,引进激励和问责机制提升数据安全。<sup>②</sup>其三,应当充分重视AGI在数据安全中的作用。既然人工智能能够通过数据挖掘、机器学习进行信息挖掘,那么AGI也能通过对涉及国家安全的数据进行机器识别,进而精准实现对敏感数据使用、流动的限制。其四,应当建立“以数据本地化存储为原则、数据出境安全评估为例外的数据跨境治理模式”。<sup>③</sup>境外AGI与境内AGI分离使用的规定可确保境外AGI只能使用境内AGI所生成的结论性内容,而不能使用具体数据,以实现数据流动的有效限制。

### (三)AGI合规的人机关系:保障企业创新

创新与风险是共生体,任何创新都是对既有存在的突破,给复杂社会结构带来不确定性。在企业服务向人类生活不断深入的AGI时代,未知因素对社会个体的影响更为巨大,对个人法益造成的风险更加严重,可能会导致AGI将创新行为视为企业犯罪的潜在因素而予以限制。必须明确的是,人工智能是人类为实践而制作的工具,不能将其作为人类向未来探索的主体。<sup>④</sup>AGI即使是如在科幻电影中一样产生了自我意识,基于人工智能与人类在认知、身体、社会地位等方面的差别,人工智能也不可能真正地与人具有相同的情感与思想。AGI时代人类的主体地位更多体现于富有人文主义关怀的创新。因而必须继续鼓励人文主义的创新行为,以维系社会的进步与人类的主体地位。这就要求人类在风险承担上必须以决策者的姿态,对创新行为的风险予以判别、认定和限制,构建和谐共生的人机关系。企业作为创新前沿,是人类发展与探索的重要推力,企业合规必须在保障企业创新的前提下进行。AGI时代企业合规的价值锚点就在于智能合规与企业创新之间的平衡。人类在创新与风险的选择中发挥主体作用,是人类社会向未来继续探索的重要保证。

作者简介:印波,中国政法大学刑事司法学院教授、博士生导师;张笑宇,中国政法大学刑事司法学院博士研究生。北京,100088

〔责任编辑:杨笑宇〕

<sup>①</sup> 陈兵、郭光坤:《数据分类分级制度的定位与定则——以〈数据安全法〉为中心的展开》,《中国特色社会主义研究》2022年第3期。

<sup>②</sup> 徐玉梅、王欣宇:《我国重要数据安全法律规制的现实路径——基于国家安全视角》,《学术交流》2022年第5期。

<sup>③</sup> 陈兵:《数字企业数据跨境流动合规治理法治化进路》,《法治研究》2023年第2期。

<sup>④</sup> 钟义信:《人工智能范式的革命与通用智能理论的创生》,《智能系统学报》2021年第4期。

shared social governance. By leveraging the high-level coordination of the Ministry of Social Work in institutional integration, we can improve the social work system through various aspects such as service, management, security and fostering professionals; strengthen problem awareness and action ability in professional training; highlight action research in the construction of localized social work theory, promote the autonomous development of Chinese social work in the service field, knowledge system, and practical mode.

#### **4. New Quality Productivity and Change of Development Paradigms: The Case of Artificial Intelligence**

*Fan Conglai, Zhu Simin · 45 ·*

The development of new quality productivity not only needs to accurately grasp its core connotation, but also needs to form corresponding development paradigms. Based on the detailed discussion that the development of new quality productivity should focus on the disruptive sci-tech innovation represented by artificial intelligence, we discuss the paradigm changing under the development of new quality productivity from three aspects: micro sci-tech innovation paradigm, meso enterprise management paradigm and macro venture capital support paradigm. We propose new paradigms to accelerate the formation of new quality productivity: the sci-tech innovation paradigm of open-source sharing and collaborative creation, the enterprise management paradigm of iterative product commercialization, technology dominance and flat organizational structure, and the venture capital support paradigm of “capital support+technology discovery+equity agreement innovation+strategic guidance”. We provide a theoretical basis and practical guidance for realizing the disruptive sci-tech innovation and accelerating the formation of new quality productivity.

#### **5. Corporate Compliance in the Era of Artificial General Intelligence**

*Yin Bo, Zhang Xiaoyu · 86 ·*

Artificial General Intelligence (AGI) is emerging as a new force in social governance. By means of its advanced data extraction and computational capabilities, AGI creates a “data world” that mirrors the real world. This facilitates a shift from broad, extensive compliance to precise, refined compliance. In the AGI era, artificial intelligence replaces traditional manual compliance approaches, using intelligent processing to directly identify and mitigate corporate risks. This results in a customized compliance scheme that enhances its overall effectiveness. Corporate compliance has evolved from simply identifying obligations to focusing on risk management, from behavior-based to ecosystem-wide compliance, and from manual to technological governance, whilst proactive compliance dominates the whole process. However, the integration of AGI also brings new challenges, such as establishing technical standards for AGI compliance, ensuring data security, and managing human-machine relationships. In China’s context of “intelligent compliance”, it is crucial to research the interpretability of AGI algorithms and address new types of data involving national security. It is essential to ensure accountability for human decision-making for the sake of harmonious human-machine interactions.