专题文论·人工智能与数字治理

人工智能与全球治理：模式、理据和紧张关系

［英］迈克尔·维尔 基拉·马图斯
［德］罗伯特·戈尔瓦 张涛 译

【编者按】近年来，人工智能成为一个引人注目但又充满争议的议题。全球的行动者都在参与构建围绕人工智能的治理机制。但被治理的“对象”究竟是什么，如何治理，由谁治理，以及为什么治理等问题尚不明确。《俄罗斯学刊》编辑部选译该文，介绍国外学者在参考人工智能、计算治理以及更广泛的监管和治理方面文献的基础上对这些问题的阐释。原文刊登于《法律与社会科学年刊》（*Annual Review of Law and Social Science*）2023 年第 19 卷，经作者授权在《俄罗斯学刊》以中文发表。

【中图分类号】F490.2 【文献标识码】A
【文章编号】2095-1094(2024)03-0049-0024
【关键词】人工智能监管 算法监管 全球治理

【作者简介】迈克尔·维尔（Michael Veale），英国伦敦大学学院法学院副教授；基拉·马图斯（Kira Matus），中国香港科技大学公共政策学部教授（国籍不详）；罗伯特·戈尔瓦（Robert Gorwa）德国柏林社会科学研究中心研究员。
【译者简介】张涛，黑龙江大学信息管理学院教授。

【基金项目】国家社会科学基金一般项目“数智环境下情报分析算法风险治理路径研究”（项目批准号：22BTQ064）阶段性成果。

一、引言

人工智能（AI）议题引起全球高度关注。一些人认为人工智能是经济发展的希望，另一些人则视其为一种商业威胁，还有一些人认为它是一些突出的社会和环境问题的起因或催化剂。上述议题引发了有关全球治理的呼吁，而全球治理本身已成为一个颇具争议的话题。批评者认为，全球治理被工业界把持或完全不起作用，而支持者和密切参与者则认为自己正在建立引导未来的机制。本文列举并审视了各种全球治理倡议及不同的人工智能框架，将它们视为监管竞争的关键点。

本文首先确定了人工智能和全球治理的含义。然后，我们对新兴的人工智能全球治理的不同模式进行了批判性评估，如伦理委员会、行业治理、合同和许可、标准、国际协议以及具有外部影响的国内立法。最后，我们评估了支撑这些模式的特定理据和紧张关系，并关注到推动这些不同模式的利益和观念。

（一）我们在谈论人工智能哪些方面的内容？

人工智能已成为一个宽泛的、包罗万象的术语，越来越多地被炒作、误导和混淆所困扰。近年来，特别是在有关治理的讨论中，人工智能作为一个新术语，指的是以前被称为数据挖掘、大数据或机器学习的技术，而在某些人看来，其几乎等同于任何现代软件。这让那些寻求人工智能具体定义或强调模式识别和统计之外方法（如符号推理）的人感到困扰。本文关注的不是人工智能是什么（或应该是什么），而是当我们谈论人工智能的全球治理时，需要治理哪些技术和实践。

出于上述目的，人工智能一词通常被理解为一种实践，一门“应用科学和工程学科”[①]，旨在将人类认为智能的品质赋予特定的软件。因此，对于作为一种实践的人工智能治理，我们也必须从更广泛的视角来看待。它应该涵盖这一实践所使用的工具和流程，包括其可用性以及社会和物质影响。我们更愿意谈论人工智能模型、开发工具包、框架、数据集或其他人工智能产品，而不是谈论如何对“人工智能”进行治理，因为后者可能包含误导性的含义。这与即使是人工智能技术从业人员也无法就“算法”作为人工产物的定位达成一致的情况十分类似，

① Bryson，J.J. *The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation*.Oxford，UK: Oxford Univ. Press，2020，pp.2-25.

更不用说那些其工作对这些技术的开发和指导非常重要且作用更广泛的学科了[①]。人工智能还包括人工智能从业者及其所在的组织、人工智能技术的使用目的，以及围绕其使用的社会、经济和政治结构。它还应包括人工智能产物本身的特点，这些特点在不同背景下的表现形式，以及如何影响与之接触的人、环境和机构。本文专注于现有和明显新兴的、而非假设的人工智能技术。学术界、工业界以及慈善机构支持的某些智库对于人工通用智能或生存风险的高度推测性讨论高度关注，但这些内容不在本文的讨论范围内。

我们所讨论的某些方面的全球治理是人工智能特有的，而有些方面则是人工智能所不具备的。这与围绕计算治理或社会分类和划分的更久远问题重叠[②]。我们专注于将这些相关领域的讨论集中在人工智能上，同时强调人工智能全球治理的一些最重要方面可能确实是计算技术全球治理（或缺乏计算技术全球治理），人工智能加速了这一进程，并使其比以往任何时候都更加突出。

（二）什么是人工智能治理？

人工智能治理以及人工智能全球治理意味着什么还远未明朗，我们首先概述具有明显全球相关性的各种人工智能治理模式，包括公共、私人、非正式、正式和混合政策工具，从行业标准到涉及多个领域的国际制定的原则和法律。然后，我们会考虑他们不得不应对的一些紧张关系，这些紧张关系在未来可能会更加明显和重要。

理解人工智能全球治理的一种方式是着眼于监管的概念目标。我们可以根据它们试图塑造的人工智能实践的各个方面来区分规则：人工智能开发、使用和基础设施。

人工智能开发的治理包括在系统设计和维护过程中尝试应用各种要求，以实现一系列政策目标。这些目标可能包括更广泛的概念，如安全性或网络安全，具体的统计目标，如狭义的非歧视定义等。这些要求也可能试图提供透明度和监督机制，例如在 AI 系统部署前要求对其进行审计，或在其销售前要求在数据库中列出。这些开发要求可能是国家层面的而非全球性层面的，但就其国际市场的活

① Seaver，N. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." November 9，2017.

② Bowker，G.C.，Star，S.L. *Sorting Things Out: Classification and Its Consequences*. Cambridge，MA: MIT Press，1999.

动而言，可能产生更广泛的影响[①]。将政策目标设计到技术中，就会引出这些目标从何而来的问题。例如，在国际上销售的招聘系统被发现嵌入了美国的非歧视规范，如五分之四规则，这可能无法解决其他司法管辖区的问题[②]。旨在检测仇恨言论或恐怖主义内容的系统必须处理这些在法律上具有国家特定定义的内容。此外，一些与人工智能相关的问题本质上是全球性的，例如，当具有特定功能（如文本或图像生成）的模型在国际范围内发布时，或者当训练过程涉及跨国重要碳排放以及提取材料及劳动力（包括数据标注的心理影响）的供应链时，这些都会不同程度地影响到价值链中的某些司法管辖区[③]。

人工智能使用的治理涉及部署具有政治、社会和经济后果软件的手段和目的。在实践中，这一类别可以涵盖广泛的范围：国际条约中关于自动决策的法律，如《个人数据自动化处理中的个人保护公约》（也称《108 号公约》）；涉及人工智能在医疗设备、警务和情报等国内部门中使用的跨国制度；规范通用人工智能固有的跨国使用制度，如内容审查、国际冲突、人道主义和国际警务。

人工智能基础设施的治理涉及试图超越上述开发与使用二分法的政策努力。这一视角促使人们关注苹果、谷歌、亚马逊和微软等纵向一体化公司对跨国技术堆栈的影响，这些公司销售硬件、操作系统、传感技术、云计算、网络基础设施，甚至专门训练的模型，所有这些都可能成为某些类型人工智能系统开发和使用的前提。这一领域尚不成熟，但正日益引起竞争监管机构的关注，有关数字主权的问题也正以各种形式进入国际谈判。人工智能系统还可被视为组织内部决策和控制的重塑点，这将全球政治经济问题与国内组织经历的实际选择、约束和影响联系起来[④]。

本文从全球治理的突出模式的简要概述开始，这些模式涉及上述一个或多个领域。在随后的部分中，笔者对这些模式凸显的紧张关系进行批判性讨论，从理论目标和实际应用两个方面来看，笔者认为人工智能的全球治理仍然是一个颇具

① Newman，A. "Watching the Watchers: Transgovernmental Implementation of Data Privacy Policy in Europe." *J. Comp. Policy Anal*，2011，no.2.

② Sánchez-Monedero，J.，Dencik，L.，Edwards，L. "What Does It Mean to 'Solve' the Problem of Discrimination in Hiring? Social，Technical and Legal Perspectives from the UK on Automated Hiring Systems." In *Proceedings of the 2020 Conference on Fairness，Accountability，and Transparency*，2020.

③ Matus，K.J.M.，Veale，M. "Certification Systems for Machine Learning: Lessons from Sustainability." *Regul. Gov*，2022，no.1.

④ Balayn，A.，Gürses，S. "Beyond Debiasing: Regulating AI and Its Inequalities." 2021. https://perma.cc/4UAV-3UFB

争议的概念，目前主要由行业利益驱动，因此值得仔细拆解和审查。最后笔者提出了一些政策制定者和研究人员在参与这些工作时应该考虑的几点问题。

二、人工智能的全球治理模式

任何文章想要列出一份当前详细、全面的人工智能领域全球治理倡议的清单都是徒劳的。制度环境瞬息万变，几乎每天都有新的倡议出现，以至于这样的清单还没开始制定就已经过时了。尽管如此，笔者还是可以确定并举例说明一些广泛的、理想类型的类别，以显示它们在多个问题领域的发展情况。正如笔者所展示的，这些治理模式都具有高度的政治性，并且越来越成为行业、各国政府、国际组织和公民社会之间日益重要的政治竞争场所。笔者（大致）按照制度复杂性递增顺序来组织这些内容。

（一）伦理准则和委员会

近年来涌现了大量人工智能伦理文件、人工智能伦理委员会和多方利益相关者机构。其中许多都是由大型科技公司或与之有密切联系的组织建立的，有些是公司内部机构，其产品的影响力赋予其全球重要性，如微软公司的人工智能道德（Aether）委员会、负责任的人工智能办公室和人工智能工程战略，国际商业机器公司（IBM）的人工智能伦理委员会，以及谷歌昙花一现的外部伦理委员会，即先进技术外部咨询委员会（ATEAC）。其他倡议由技术公司资助，但至少名义上是外部的。其中最典型的例子或许是亚马逊、苹果、谷歌、脸书（Facebook）、IBM 和微软于 2016 年成立的人工智能合作伙伴关系（PAI）。其他由行业资助的论坛则处于外围，如世界经济论坛或由财政支持的研究实体，如未来生命研究所或人类未来研究所。

这些实体通常在名义上寻求协调开发人工智能系统的行业参与者之间的行动，并旨在公司内部和公司之间塑造未来负责任的人工智能治理议程。然而，在实践中，它们的努力喜忧参半，更严重的是深陷争议。例如，PAI 最初吸引了一系列民间社会成员，但自成立以来，人们对其影响力的热情已经逐渐减弱。著名的非政府组织（NGO）“立即访问”（Access Now）于 2020 年退出，其在退出时表示，“PAI 没有影响或改变成员公司的态度，也没有鼓励它们系统地回应民间社

会或接受其咨询”[①]。行业伦理委员会制定的原则因表达含糊且实际上毫无意义而受到批评——没有执行力度或展示合规的机制。一些人认为，如果这些自愿性努力确实有效的话，那么其在处理商业模式和公司文化的核心伦理问题时应是游刃有余的[②]，或者侧重于工程师和设计选择，并有效地被归入企业逻辑和激励体系[③]。一项对数十个人工智能伦理框架的审查指出，这些伦理原则在如何解释伦理原则，重点涉及哪些问题、领域或行动者，以及它们应如何实施方面存在“实质性分歧”，同时在伦理框架方面，全球多数国家没有做出努力[④]。

最令人担忧的是，有观点认为，人工智能伦理原则主要被企业用于阻止监管行动[⑤]，其中“伦理”成为新的“行业自律”[⑥]，并且可能为面向公众的公关活动提供助力。例如，昙花一现的谷歌人工智能伦理委员会的成员似乎受到了更广泛政策策略的影响，该策略旨在赢得美国政界的支持，并回应共和党内对谷歌等公司偏袒民主党价值观和观点的担忧。对于那些认为伦理哲学作为一个领域对人工智能系统讨论有很多有益贡献的人来说，针对工具化伦理的强烈反对是令人遗憾的[⑦]。

（二）行业治理

当前，伦理准则以外的行业自律参差不齐，但仍然具有影响力。这在很大程度上取决于企业在人工智能领域设置的瓶颈。少数企业控制着促进人工智能系统的培训资源，包括物理资源（如连接廉价电力的大型图形处理单元集群）、认知资源（接触尖端研究人员群体的机会、向大学教授支付高薪的能力）和信息资源（访问数据集、支付大量标签费用的资源以及用于收集和实验的实时系统）。行业参与者赞助了大量关于技术工具的研究，尤其是通过高知名度的学术会议，一些

① "Access Now Resigns from the Partnership on AI." *Press Rel.*, October 13, 2020. https://www.accessnow.org/access-now-resignation-partnership-on-ai/

② Munn, L. " The Uselessness of AI Ethics. "2022. https://doi.org/10.1007/s43681-022-00209-w

③ Green, B. "The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice." *J. Soc. Comput.*, 2021, no.3.

④ Jobin, A., Ienca, M., Vayena, E. "The Global Landscape of AI Ethics Guidelines." *Nat. Mach. Intell.*, 2019, no.9.

⑤ Nemitz, P. "Constitutional Democracy and Technology in the Age of Artificial Intelligence."*Philos. Trans. R. Soc.* A. 2018, no.2133.

⑥ Wagner, B. *Ethics as An Escape from Regulation: From "Ethics-Washing" to Ethics-Shopping? In Being Profiled: Cogitas Ergo Sum.*, Amst. Univ. Press, 2018, pp. 84-88.

⑦ Bietti, E. "From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics." *J. Soc. Comput.*, 2021, no.3.

学者将其描述为围绕“公平”等政治概念“制造共识”[①]以及推广浅薄的、非语境化的、以工程师和实验室为中心的方法，这些方法可以顺利而低成本地推广[②]。这种行业治理可被视为一种跨国政策创业形式，包括“专家话语的管理和沟通，而不是数据、证据或研究成果”[③]。

全球人工智能治理迅速趋向的一个具体方向是利润丰厚且公众关注的通用系统领域。为通用分析形式设计的人工智能系统是典型的双用途技术，既支持相对良性的目的，如生成库存图片或模板文本，也支持有争议的目的，如犯罪活动或破坏民主。一个重要的问题出现了，即如何在促进良好目的的同时限制或防止不希望出现的目的？

市场上许多性能最好的人工智能系统都是通过基于云计算的新型平台式商业模式以服务形式销售的[④]。这些模式在文本、图像分析或生成等领域被作为通用的基础能力进行销售，用户可以根据具体应用进行集成和定制。尽管用户经常带来自己的数据集来微调模型以适应他们的使用案例，但他们很少能够完全复制最终模型，通常只被允许通过应用程序编程接口将它们作为问答系统使用。

这种人工智能能力的分布形式使平台有能力充当重要的治理决策者。只要理想的模型是专有的，它们的使用就可以被限定在某些用途上。例如，谷歌仅允许其媒体行业的白名单客户使用其名人识别面部分类系统[⑤]。一些评论家表示担心，如果模型是开源的，治理就会变得困难或不可能，例如生成的文本没有隐写水印，无法辨别其是否为人造文本[⑥]。此外，为支持没有明显危害的产出（有明显危害的产出，如生成儿童性虐待图像和叙述）所需的标签资金，即使是单个模型也需要数十万美元，使用的外包劳动力每小时工资不到 2 美元，并提供有限的心

① Young，M.，Katell，M.，Krafft，P.M. "Confronting Power and Corporate Capture at the FAccT Conference." In *2022 ACM Conference on Fairness，Accountability，and Transparency*，2022.

② Gansky，B.，McDonald，S. "CounterFAccTual: How FAccT Undermines Its Organizing Principles." In *2022 ACM Conference on Fairness，Accountability，and Transparency*，2022.

③ Stone，D. "Transnational Policy Entrepreneurs and the Cultivation of Influence: Individuals，Organizations and Their Networks." *Globalizations*，2019，no.7.

④ Cobbe，J.，Singh，J. "Artificial Intelligence as a Service: Legal Responsibilities，Liabilities，And Policy Challenges." *Comput. Law Secur. Rev.*，2021.

⑤ BSR. "Google Celebrity Recognition API Human Rights Assessment." 2019. https://www.bsr.org/reports/BSR-Google-CR-API-HRIA-Executive-Summary.pdf

⑥ Aaronson，S. "My AI Safety Lecture for UT Effective Altruism." November 28，2022. https://scottaaronson.blog/?p=6823

理支持或咨询[①]。

目前尚不清楚平台在未来提供人工智能方面有多么不可或缺。人工智能模型的多边平台正在兴起，例如人工智能公司（Hugging Face），它为分发和使用他人制造的系统提供了基础设施。在当前的数字环境中，选定的实体是极其强大的监管者。应用商店是在许多移动设备上安装某些类型软件的唯一途径，它们通过技术限制和合同限制来规范软件的内容和数据使用等问题[②]。这些规则的一致性和质量一直备受关注，如优图（YouTube）按级别对其用户进行不同的监管[③]，而 Facebook 曾多次违反苹果公司的规则，其行为很可能会导致不太知名的软件被完全禁止[④]。

未来，人工智能全球治理似乎有可能与平台治理高度融合。如果重要的中介机构仍然存在，它们将既是强大的管理者，也是立法者监管人工智能系统的目标[⑤]。这将反映出此前信息技术法律领域的无数经验，即从社交媒体到加密货币等领域的中介机构成为“监管切入点”[⑥]或“瓶颈点”[⑦]。目前，欧盟部长理事会已经考虑在人工智能法案草案中针对通用人工智能系统的提供商规定一些义务，这进一步表明，他们可以通过真正地禁止特定用途，并在检测到“市场滥用”时采取行动来避免某些法律义务[⑧]。

（三）合同与许可

另一种新兴的、重要的、私有的人工智能系统跨国治理形式是使用合同条款

① Perrigo，B. "OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make Chatgpt Less Toxic." January 18，2023. https://time.com/6247678/openai-chatgpt-kenya-workers/

② Cowls，J.，Morley，J. *App Store Governance: The Implications and Limitations of Duopolistic Dominance.* In the 2021 Yearbook of the Digital Ethics Lab，2022，pp. 75-92；Marsden，C.T.，Brown，I. "App Stores，Antitrust and Their Links to Net Neutrality: A Review of the European Policy and Academic Debate Leading to the EU Digital Markets Act." *Internet Policy Rev.*，2023，no.1；Van Hoboken，J.，Fathaigh，R.Ó. "Smartphone Platforms as Privacy Regulators." *Comput. Law Secur. Rev.*，2021.

③ Caplan，R.，Gillespie，T. "Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy." *Soc. Media Soc.*，2020，no.2.

④ Carman，A. "What Would Happen if Apple Fully Banned Facebook from the App Store?" 2019. https://www.theverge.com/2019/2/1/18205291/apple-facebook-developer-bancertificate-app-store

⑤ Cobbe，J.，Veale，M.，Singh，J. "Understanding Accountability in Algorithmic Supply Chains. " In *Proceedings of the 2023 ACM Conference on Fairness，Accountability，and Transparency*，2023.

⑥ Finck，M. *Blockchain Regulation and Governance in Europe*. Cambridge，UK: Cambridge Univ. Press，2018.

⑦ Goldsmith，J.L.，Wu，T. *Who Controls the Internet? Illusions of a Borderless World*. New York: Oxford Univ. Press，2006；Tusikov，N. *Chokepoints: Global Private Regulation on the Internet*. Berkeley: Univ. Calif. Press，2016.

⑧ Counc. Eur. Union. "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (Preparation for COREPER). "2022. https://perma.cc/564K-RCKR

来限制人工智能的使用及其产出。这种机制似乎受到了计算机治理史上重要创新的启发，即围绕开源软件出现的知识产权制度。

针对许多人工智能开发的专有的、由企业控制的本质，以及至少在一定程度上由于对数字主权和国家对某些重要人工智能系统和功能的依赖与日俱增的担忧，近年来人们对创建开源通用人工智能系统的兴趣有所增长。为了促进开源人工智能系统的查询、分发和开发，出现了一些平台，如 Hugging Face 等。然而，这些系统的易于访问性带来了一些政治问题，因为很少有制衡措施来防止它们被用于有争议的任务。对此，一些模型设计者转向合同法，保留模型的知识产权（即不将其发布到公共领域），而只提供有条件的使用许可[①]。长期以来，此类许可一直是软件和数字知识产权（IP）治理的核心部分，著作权许可系列，如通用性公开许可证（GPL），要求任何基于许可代码的衍生作品都必须按照相同或等效的许可条款进行发布，而知识共享许可则允许在一系列条件下重复使用内容，如署名、不篡改或用于非商业目的[②]。

在人工智能领域，负责任的人工智能许可（RAIL）倡议声称要超越这些倡议，增加“行为使用限制”。知识产权持有者将这种许可应用于稳定扩散模型，该模型被试图制作成类似于美国人工智能研究公司（OpenAI）专有的 DALL-E 2 图像生成模型。该许可证禁止用户使用模型对他人进行诽谤、提供医疗建议、用于执法或类似目的、用于意图歧视或具有广义的歧视效果的目的、用于某些完全自动化的决策（这是基于数据保护法的规定），以及以可能引起身体或心理伤害的方式剥削个人——这些规则似乎是直接取自欧洲委员会的《人工智能法案》草案[③]。尽管这些合同条款并不完美，而且其中一些条款措辞含糊不清（例如，从法律中提取相关定义部分，或未明确规定诽谤或一般非法行为的管辖权），但有效落实这些条款的最大障碍在于只有版权持有者才能执行版权许可。尽管软件公司可能相当了解该领域中数量有限的竞争软件公司，并能够收集足够的信息进行有效执行，但如果问题是围绕技术的使用而不是技术的开发或所谓的知识产权盗

① Contractor，D.，McDuff，D.，Haines，J.K.，Lee，J.，Hines，C.，et al. "Behavioral Use Licensing for Responsible AI." In *2022 ACM Conference on Fairness，Accountability，and Transparency*，2022.

② Guadamuz，A. "Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses. " *Eur. Intellect. Prop. Rev.*，2004，no.8.

③ Rombach，R.，Esser，P. "CreativeML Open RAIL-M. Hugging Face. "2022. https://huggingface. co/spaces/CompVis/stable-diffusion-license

窃，那么成功防止社会滥用似乎不太可能。与此相关的是，以这种方式授权开发人工智能系统的开源或公益团队很可能没有能力发现和监管通用人工智能系统的大量用户，也没有必要的法律资源来实现大规模的治理。我们或许可以设想另一个制度层面的执行机制——许多开源软件版权的基金会所创造的“以社区为导向”的执行原则[①]，但据我们所知，目前还没有此类提议出现。

一种在法律上更加微妙的尝试可见于一些平台试图对模型的输出而非模型本身进行授权，以此来管理其全球使用。OpenAI 是一家以生成新颖文本和图像模型而闻名的公司，它实施了一项内容政策，要求必须将生成的媒体内容披露为人造的，并且禁止某些主题，例如描述“非法活动”、涉及政治人物或宣扬“重大阴谋论”的主题。如果违反上述准则，那么声称在法律允许范围内拥有生成媒体内容所有权的 OpenAI 将保留撤销向用户提供许可的权利。然而，基于知识产权的有效执法程度可能受到两个因素的限制，即司法管辖权和制作系统用于生成媒体的提示所需的有限创造性技能或劳动力[②]。精心设计的提示可能会提供一些保护，但这似乎是一种有限的治理模式，因为提示的复杂性与潜在危害之间并没有可靠的联系。

（四）标准

计算机的全球治理历来在很大程度上依赖于自我监管组织创建的工程标准[③]。这些标准对于网络技术一直至关重要，因为网络技术要求各组件遵循相同的规则来实现功能。互联网工程任务组、万维网联盟和电气与电子工程师协会（IEEE）等机构管理着重要标准，如 TCP/IP、HTML 和 802.11（WiFi）等。然而，这些组织的产出具有政治维度，不仅仅促进功能性，还产生了与人权相关的、富于价值的设计理念和成果，例如与隐私和自由表达有关的理论和成果[④]。统一网络标准

① Ballhausen，M. "Copyright Enforcement. " In *Open Source Law，Policy and Practice*，2022.

② Guadamuz，A. "Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works. "*Intellect.Prop. Q.*，2017，no.2；Guadamuz，A. "DALL•E Goes Commercial，But What About Copyright? "July 25， 2022. https://www. technollama. co. uk/dall% c2%b7e-goes-commercial-but-what-about-copyright

③ Harcourt，A.，Christou，G.，Simpson，S. "Global Standard-Setting in Internet Governance. "In *Global Standard Setting in Internet Governance*，2020.

④ Braman，S. "The Framing Years: Policy Fundamentals in the Internet Design Process，1969 - 1979." *Inf. Soc.*，2011，no.5；Cath，C. "The Technology We Choose to Create: Human Rights Advocacy in the Internet Engineering Task Force. " *Telecommun. Policy*，2021，no.6.

在国际上的普及使其发挥了重要的全球治理作用[①]，实质性的政策决定和价值观可以有效地利用其功能的必要性并在技术层面上得到体现。

标准流程可以增强资源丰富的现有企业的能力。在网络领域，大型企业采用的标准可能会迫使其他参与者也做出相同的变更[②]。参与任何标准制定，机构都需要投入大量人力资源，而那些拥有相关经验的人更有能力对其施加影响。成本会进一步限制问责制。许多与计算相关的标准都是通过专有标准机构制定的，在这些机构的标准制定过程中，其仅对付费成员开放，最终产品也是专有的，通常需要花费数百美元才能获得一整套相互关联的规则。

尽管如此，网络标准的成功使一些组织自愿寻求以同样的方法来治理人工智能系统。最早出现的标准之一是 IEEE P70xx 系列，这些标准包括已发布的透明度标准（7001-2021）、在设计中考虑道德问题的流程标准（7000-2021），以及关于偏见和“伦理驱动的推动”标准（7003TM，7008TM）。国际标准化组织（ISO）也有一系列标准，其中大部分正通过 2017 年成立的 ISO/IEC JTC 1/SC 42 委员会制定中。这些组织通常采用订阅模式，保留所发布标准的版权，并通过直接授权访问或经典型的私营国家标准机构，例如美国国家标准协会（ANSI）、英国标准协会（BSI）或德国标准化协会（DIN）授权本地化和翻译许可来赚钱。与大多数标准机构不同的是，国家测量机构通常是公共实体，它们也一直致力于人工智能标准化工作，其中包括美国国家标准与技术研究院，它建立了一个人工智能风险管理框架[③]，以及英国国家物理实验室，它是英国人工智能标准中心的合作伙伴。

并非所有标准的功能都是所必需的，有些标准可能只是向能力较低的组织传递知识，使其了解如何按照当前的最佳做法构建系统，类似于简单的知识输出，而不是协调机制。另一个重要角色是传递信号，计算标准可以向其他市场参与者传递最佳实践信号，例如向保险精算师发出有关保险目的的信号[④]。当立法者考

① DeNardis，L. *The Global War for Internet Governance*. New Haven，CT: Yale Univ. Press，2014.

② Cohen，J. E. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford，UK: Oxford Univ. Press，2019；Ten Oever，N. " 'This Is Not How We Imagined It': Technological Affordances，Economic Drivers，and the Internet Architecture Imaginary." *New Media Soc.*，2021，no.2.

③ NIST (Natl. Inst. Stand. Technol.). *AI Risk Management Framework (AI RMF 1.0)*. Rep.，2023.

④ Shackelford，S.J.，Proia，A.A.，Martell，B.，Craig，A.N. "Toward a Global Cybersecurity Standard of Care: Exploring the Implications of the 2014 NIST Cybersecurity Framework on Shaping Reasonable National and International Cybersecurity Practices." *Tex. Int. Law J.*，2015，no.2.

虑是否需要更严格的监管时，它们可以被用作一种立法者的信号[①]；或当法院在评估侵权过失时，它们也可以作为一种信号[②]。

在人工智能标准制定方面，各国政府似乎主要采取了两类方法，按照标准和认证文献的分类，可分为混合型和共生型[③]。混合型方法旨在激励私营标准，以允许遵守欧盟《人工智能法案》中提出的法律文书的规定（下文将进一步讨论）。共生型方法认为私营认证体系可以加强其他治理系统的权威性和合法性，欧盟数据保护和网络安全法中的可选行业认证机制就是一个例子[④]。随着标准化激励措施超越功能性和网络一致性，人工智能标准的监管轨迹似乎更类似于可持续产品或产品安全等非计算标准化领域，而不是典型的网络标准。

（五）国际协议

在国际行业自律的同时，围绕人工智能问题出现了一系列政府间标准和组织，其中包括经济合作与发展组织（OECD）的《人工智能建议书》（2019）、联合国教科文组织的《人工智能伦理问题建议书》（2021）以及二十国集团的《人工智能原则》（2019）。在大多数情况下，这些标准和原则可以说与行业文件或行业利益并无明显相悖。例如，原则综述中并未提及数字竞争、权力或对技术的控制[⑤]。尽管数字权力在全球政策中具有重要意义，但它似乎是技术公司不愿意讨论的问题。一系列国家已经建立了一个人工智能全球伙伴关系（关于这一点，不要与上面讨论的由私营部门领导的 PAI 相混淆），以加强在经济合作与发展组织基础上的工作，尽管目前其影响尚不明确。

欧洲委员会（CoE）在人工智能治理方面发挥了主导作用，因为它审查了过去的数字和数据相关法律文书。《欧洲人权公约》是欧洲人权法院以及许多国际监督和分析法律体系的基础，它还牵头制定了《布达佩斯网络犯罪公约》《个人数据自动化处理中的个人保护公约》，后者是国际数据保护标准的基础，以及

① Marsden，C.T. *Internet Co-Regulation: European Law，Regulatory Governance and Legitimacy in Cyberspace*. Cambridge，UK: Cambridge Univ. Press，2011.

② Schepel，H. *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets*. Oxford，UK: Hart，2005.

③ Cashore，B.，Matus，K.J.，Norris，R. "Pathways to Impact: Synergies with Other Approaches." In *Toward Sustainability: The Roles and Limitations of Certification*，2012.

④ Kamara，I. "Co-Regulation in EU Personal Data Protection: The Case of Technical Standards and the Privacy by Design Standardisation Mandate." 2017. https://ejlt.org/index.php/ejlt/article/view/545/725

⑤ Cowls，J.，Floridi，L. "A Unified Framework of Five Principles for AI in Society." 2019. https://doi.org/10.1162/99608f92.8cd550d1

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/056025123241011000