2023 年第 6 期（总第 9 期）

人工智能与国际安全研究动态

ARTIFICIAL INTELLIGENCE AND INTERNATIONAL SECURITY STUDIES REVIEW

国际智库及媒体对人工智能国际治理动态分析

清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

战略与安全
SECURITY & STRATEGY
清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL SECURITY AND STRATEGY TSINGHUA UNIVERSITY

国际智库及媒体对人工智能国际治理动态分析

编者按：为推进人工智能与国际安全领域的相关研究，清华大学战略与安全研究中心（CISS）组织研究团队定期跟踪最新国际研究动态，重点关注人工智能应用对国际安全带来的风险挑战，并针对人工智能安全领域国际动态、智库报告、学术论文等资料进行分析。本文是CISS推出的人工智能与国际安全研究动态第9期，主要聚焦国际智库及媒体对人工智能国际治理动态分析。

1、Politico：人工智能应实现国有化

8 月 20 日，美国 Politico 杂志刊登大西洋理事会高级研究员查尔斯·詹宁斯的文章《控制人工智能的方法只有一种：国有化》。文章认为，当前人工智能算法已经成为了“黑箱”，即使是专业科研人员也难以理解人工智能算法的全部规则，这意味着人工智能很有可能失控，或被犯罪分子滥用。但目前来看，美国国会对人工智能技术了解有限且立法进展缓慢，即使有相关的透明度与监管立法出台，也难以彻底监管人工智能算法的“黑箱”。因而对人工智能产业进行国有化是当前最佳的选择。美国应借鉴对核产业国有化的经验来对人工

智能产业进行国有化，从各大互联网巨头处赎买其现有的人工智能模型，交由特定的政府委员会统一管理，并由该委员会统筹人工智能产业的投资与发展，最终确保人工智能产业的发展将为社会带来益处。

https://www.politico.com/news/magazine/2023/08/20/its-time-to-nationalize-ai-00111862

2、《报业辛迪加》：人工智能需要“帕格沃什运动”

7 月 24 日，《报业辛迪加》杂志刊登美国务院前政策规划主任安妮-玛丽 •斯劳特与互联网名称与数字地址分配机构前主席法蒂 • 切哈德的文章《AI 的帕格沃什运动》。文章提到，1957 年来自美国、苏联等十个国家的 22 位科学家齐聚加拿大的帕格沃什，创立了帕格沃什科学和世界事务会议，其主旨是推动核裁军并预防核冲突。该组织最终于 1995 年获得了诺贝尔和平奖，当前人工智能也需要一个“帕格沃什运动”。首先，人工智能的“帕格沃什运动”应根植于已有的人工智能治理框架，如美国的《人工智能权利法案蓝图》、联合国教科文组织的《人工智能伦理建议书》等，以上述框架为蓝本进行构建。其次，人工智能的“帕格沃什运动”应注重于协调各利益攸关方，建立“网络化的多边主义”，将国际组织、国家、企业、学术机构、非政府机构等多主体纳入进来，以确保相关措施的广泛实施。最后，人工智能的“帕格沃什运动”应着手建立一个多核心的机构网络，正如互联

网领域有不同机构负责制定网络标准、分配数字地址、协调利益相关方等工作，也应有多个机构来负责不同方面的人工智能治理，进而确保人工智能治理的稳定性和韧性，并有助于应对来自特定国家的政治压力。

https://www.project-syndicate.org/commentary/institutions-to-govern-artificial-intelligence-new-pugwash-movement-by-anne-marie-slaughter-and-fadi-chehade-2023-07

3.《报业辛迪加》：生成式人工智能正在影响选举公正

8 月 10 日，《报业辛迪加》杂志刊登前斯坦福大学网络政策中心的主任凯利·伯恩的文章《生成式人工智能会成就或破坏民主吗？》。文章指出，虽然生成式人工智能为在医学、工业、政策、教育等领域提供巨大的机遇，但同时将为选举过程带来巨大的挑战与不确定性。首先，人工智能可能导致选举过程中的偏见。如美国正使用人工智能系统来维护和更新各州的选民名册，但当前的算法难以识别亚裔选民的名字，也难以识别亚裔等少数族裔选民的签名。其次，生成式人工智能将降低选举的竞争性。目前约 90%的美国国会选区被认为是“铁票仓”，即通常稳定地归属于民主党或共和党。生成式人工智能可以帮助立法者划定选区，进而使执政党进行更具压制性的选区划分，减轻执政党所面临的竞争性与挑战。最后，生成式人工智能将促进选举虚假信息的网络传播。生成人工智能可以利用公开的图片与音像资源，并针

对不同的受众群体生成多语种的虚假宣传，进而改变选举政治的“游戏规则”。

https://www.project-syndicate.org/commentary/generative-ai-could-test-democracies-by-kelly-born-2023-08

4.布鲁金斯学会：隐私立法有助于人工智能治理

7 月 7 日，美国布鲁金斯学会网站刊登其技术创新中心访问学者卡梅隆·克里的文章《隐私立法如何帮助解决人工智能问题》。文章认为，虽然当前不同国家的监管规则不尽相同，但隐私立法所倡导的算法透明度、问责制和公平性原则是大多数人工智能治理框架的共同点，故加强隐私立法将从多方面助力人工智能治理。首先，隐私立法有助于解决人工智能的偏见问题。只有确保训练所使用数据的来源、质量和道德使用都是无偏见的，并严格规定可以使用的数据类型与范围，最终构建的人工智能模型才有可能是无偏见的。其次，隐私立法有助于确立人工智能相关的术语规范。美国的隐私立法将“处理个人数据”定义为“使用机器学习、自然语言处理、人工智能技术或其他类似或更高复杂性的计算处理技术，并就个人信息做出决策或促进人类决策的计算过程”，该定义为人工智能治理框架的术语构建提供了模板。最后，隐私立法有助于增进人工智能的透明度。美国的隐私立法详细说明了必须披露的收集和使用个人信息及其隐私

惯例的情况，以及基于此类数据的算法决策信息，包括此类使用和决策的性质以及它们所基于的数据和逻辑等。
https://www.brookings.edu/articles/how-privacy-legislation-can-help-address-ai/

5.《卫报》：古特雷斯呼吁建立人工智能监管机构

7 月 18 日，英国《卫报》网站刊登《卫报》全球技术编辑丹·米尔莫（Dan Milmo）、科技记者希巴克·法拉赫（Hibaq Farah）的文章《联合国秘书长表示，恶意使用人工智能可能会造成“难以想象”的损害》。文章关注，联合国秘书长安东尼奥·古特雷斯（António Guterres）表示，恶意使用人工智能系统可能会造成“可怕”的死亡和破坏，呼吁建立类似于政府间气候变化专门委员会（IPCC）的新联合国机构以应对人工智能构成的威胁。古特雷斯称，将人工智能用于恐怖分子、犯罪或国家目的的有害使用可能造成“深刻的心理伤害”，当前，基于人工智能的网络攻击已经瞄准了联合国维和与人道主义救援行动。古特雷斯呼吁按照政府间气候变化专门委员会的方式建立一个新的联合国实体以应对风险，并称该机构的首要目标是支持各国最大限度地发挥人工智能的好处，减轻现有与潜在的风险，建立和管理国际商定的人工智能监测与治理机制。

https://www.theguardian.com/technology2023/iul/18/malicious-use-of-ai-could-cause-huge-damage-savs-un-boss

6. Politico：西方争论是否邀请中国参与英国首届人工智能安全峰会

8 月 25 日，美国 Politico 网站刊登本站记者文森特·马南古（Vincent Manancourt）、汤姆·布里斯托（Tom Bristow）、劳里·克拉克（Laurie Clarke）的文章《尽管遭到盟友的反对，中国仍有望参加英国人工智能峰会》。文章认为，尽管遭到日本、美国与欧盟的共同抵制，英国政府仍决心以某种形式让中国参与 11 月初举办的人工智能安全峰会。英国计划于 11 月初召开全球首届人工智能安全峰会，峰会将重点关注最先进的人工智能，即前沿模型。英国方面希望人工智能安全峰会的基础尽可能广泛，因此不希望将世界上最先进的科技强国之一的中国拒之门外，并认为排除中国可能会加速全球人工智能领域的分裂。日本方面则认为，在民主国家尚未就人工智能治理达成共同立场的情况下让中国参与峰会为时过早，并提出一项让中国参与的替代方案，即通过七国集团（G7）与中国举行部长级会议。目前美国国家安全委员会发言人艾德丽安·沃森（Adrienne Watson）已表示，美国对中国参加峰会没有意见。欧盟内部对该问题仍存在分歧，尽管欧盟委员会拒绝邀请中国，但欧盟人工智能规则关

键谈判代表德拉戈斯·图多拉切（Dragoş Tudorache）认为中国应该参与谈判。

https://www.politico.eu/article/china-likely-at-uk-ai-summit-despite-pushback-from-allies/

7.新美国安全中心：中美竞争与人工智能的军事化应用

2023 年 7 月 18 日，新美国安全中心（CNAS）发布其“印太安全研究项目”与“人工智能安全与稳定项目”的阶段性成果，报告题为《中美竞争与人工智能的军事化应用：美国应如何管控与中国的战略竞争与风险》。首先，该报告分析了目前中国的人工智能发展战略，以及人工智能在中国实现军事现代化进程中的角色。其次，报告认为随着人工智能军事化的应用，或将增加美国的战略风险并降低中美间的战略稳定性，并具体梳理出以下五类路径：一是人工智能的军事化应用，可能将使得中国军事能力相对提升，并改变地区内的军事实力平衡；二是人工智能在决策与信息生成、获取发挥的负面影响，如压缩决策时间、产生错误信息、生成虚假信息等，这将使得决策者做出错误决策，增加战略风险；三是人工智能无人自主系统的军事化应用，可能将使得战争的人员伤亡成本降低，从而将提升战争爆发风险；四是人工智能在情报、监视和侦察方面的应用，将提升彼此军事透明度，使已开发或部署武器的生存机会降低，从而降低反击或报复能力，从而将提升战争风险；五是人工智能在指挥、控

制与通讯的军事化应用，可能将增强网络战与电磁战的攻击能力，同时也增加了向对方用于训练的数据集“投毒”的风险。最后，报告针对美国应如何管控战略风险提出了三条建议：一是有针对性地遏制对手，并提升自身能力；三是采取单边负责任的管控策略，将评估人工智能的安全性与可靠性作为重点发展方向；三是借助双边和多边外交手段，通过定期沟通、交流、谈判达成协议等方式，交换各方意见，增强战略互信，降低战略风险。

https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/FINAL4.pdf?mtime=20230718160443&focal=none

8.麻省理工科技评论：人工智能影响政治的六种方式

7 月 28 日，美国麻省理工科技评论网站刊登政策研究文章《人工智能影响政治的六种方式》。文章指出，由人工智能驱动国内政治的新时代可能即将到来，而公众担心的是生成式人工智能将影响公众政治偏好、并可能介入法律制定与审查等。作者认为，生成式人工智能影响国内政治的具体方式包括：一是立法机关或机构接受由人工智能产生并以其名义提交的证词或意见；二是由人工智能撰写的新的立法修正案被正是接受；三是在民意调研中，人工智能生成的政治信息取代竞选顾问的建议；四是人工智能创建了一个拥有自己纲领的政党，吸引候选人并赢得选举；五是人工智能自主创

造盈利，并资助政治竞选；六是人工智能可在多个辖区间进行政策协调。

https://www.technologyreview.com/2023/07/28/1076756/six-ways-that-ai-could-change-politics/

9.国际隐私专业协会（IAPP）：全球人工智能监管执法概述与企业应对

8 月 2 日，国际隐私专业协会（IAPP）发表了由缅因大学法学院研究生威廉·辛普森（William Simpson）的文章《世界各地的人工智能监管执法概述》。文章在分析了世界主要国家的人工智能监管机制后，为企业提出了多条建议。美国以联邦贸易委员会作为默认监管机构，在人工智能行业导向方面干涉力度较低，侧重技术应用和流程监管，多为事后监管。欧盟通过三部人工智能法案草案，侧重组织管理与风险管理，多为事前监管，有关执法的问题仍然存在争议和未解决。英国的隐私监管机构越来越多地参与执法，其对隐私和数据保护经常使人工智能企业陷入困境。加拿大的《人工智能和数据法案》建立了监管的指导方针，相关机构将协助人工智能开发人员通过自愿方式进行管理。中国采用的是通过多部法律法规衔接，针对深度合成技术、生成式人工智能技术和算法推荐技术不同业态分别立法，基本初步形成一套完备的人工智能监管体系。面对这种情况，企业应熟悉数据隐私和网络安全法律，了解对所在行业进行监管的现行法律，

注意算法清算，适应多机构的联合监管，与隐私监管机构保持互动，并投资于隐私监管机构的研究。

https://iapp.org/news/a/ai-regulatory-enforcement-around-the-world/

10.路透社：人工智能在治理、风险和合规中发挥重要作用

8 月 24 日，路透社发表了由高级监管情报专家托德·埃雷特（Todd Ehret）撰写的文章《人工智能在公司治理、风险管理和合规工作中发挥着重要作用》。文章指出，人工智能正在以更复杂的方式进入合规平台，远不只是使用大型语言模型（LLM）或聊天机器人来回答问题或起草方案。将人工智能添加到合规平台的一些关键改进和好处包括：人工智能将在法规和变更管理方面发挥作用，帮助企业了解法规、协调政策。它还能对第三方和供应商风险管理、网络和 IT 风险管理等领域产生巨大影响。通过预测分析、实时监控，人工智能可以增强组织的网络能力。持续的人工智能监管与相关法规的协调能够帮助公司更好地遵守相关的 IT、隐私和网络法规。银行和金融机构可以利用人工智能更好地检测和理解风险模型，减少误报并提高效率，节省大量成本。此外，人工智能还有助于分析财务数据和客户行为模式以识别可疑活动，并提供解决复杂问题并获得简单语言答案的能力，改善管理层之间的沟通。企业应创建内部政策、程序和监督机

制，以有效利用人工智能。总的来说，人工智能在风险缓解方面的优势，超过了它可能带来的风险和后果。
https://www.thomsonreuters.com/en-us/posts/corporates/ai-governance-risk-compliance-programs/

11. Open AI：推动人工智能向前治理

7月21日，OpenAI发表文章《推动 AI 治理向前发展》，文章指出，OpenAI 和其他 AI 企业通过自愿承诺采取监管措施管理 AI 技术开发风险，加强 AI 的安全性和可信度，推动人工智能治理向前发展。这一过程由白宫协调，是推进有意义和有效的人工智能治理的重要一步。《承诺书》强调了人工智能未来发展的三个基本原则：安全、保障和信任。在安全方面，企业应在技术误用、社会风险和国家安全（例如生物、网络和其他安全领域）领域对模型或系统进行内部和外部的红蓝对抗测试；推进与政府间关于信任和安全风险、危险或紧急应对能力以及保障措施逃避企图的信息共享。在保证方面，企业应采取网络安全投资和内部威胁防护措施，以保护所有权和未发布的模型权重；激励第三方发现以及报告问题和漏洞。在信任方面，企业应开发和部署相关机制，使得用户能够明确音频或视频内容是否为人工智能生成，包括为人工智能生成的音频或视频内容标注可靠来源、水印等；公开报告模型或系统的功能、局限性以及适用和不适用的领域（包括对社会风险的讨论），例如对公平和偏见的影响；

优先研究人工智能系统带来的社会风险，包括避免有害的偏见和歧视以及隐私保护；开发和部署前沿人工智能系统，以帮助解决社会面临的最大挑战。

https://openai.com/blog/moving-ai-governance-forward

12.Forbes：人工智能治理的三个层次都需要关注

8 月 10 日，福布斯（Forbes）网站发布了《人工智能治理不止是模型层面》。文章详细介绍了组织、用例和模型三个层次的人工智能治理，并说明制定法规需要全面关注这些层面。在组织层面，人工智能治理需确保有明确定义的人工智能道德、问责制、安全等内部政策，以及执行这些政策的明确流程。在用例层面，人工智能治理侧重于确保对于每一组任务的该模型都需满足所有必要的治理标准。单个模型可用于多个不同的用例，某些用例的风险相对较低，而在不同但相似的情境的任务风险非常高。在模型层面，人工智能治理主要侧重于确保人工智能系统的实际技术功能符合公平、准确和安全性的预期标准。它包括确保数据隐私受到保护，受保护组之间没有统计偏差，模型可以处理意外输入，对模型漂移进行监控等。构建准确、可靠和公正的 ML 模型并实践良好的数据治理对于值得信赖和负责任的 AI 系统至关重要。人工智能治理的每个层次都需要详细的治理流程，以确保其模型和第三方 AI 系统经过良好测试并受到相关监管。

因此，监管机构和公众需要关注的不同级别的人工智能治理，以在未来制定相关流程与法规。

https://www.forbes.com/sites/forbeseq/2023/08/10/ai-governance-is-not-just-about-the-models/?sh=cf51d0612853

13.《外交政策》刊文分析人工智能监管悖论

2023 年 8 月 4 日，外交政策网站刊登了塔夫茨大学弗莱彻法律与外交学院全球商务院长巴斯卡·查克拉沃蒂的《人工智能监管悖论：监管人工智能以保护美国民主最终可能会危及国外的民主》。文章认为人工智能为虚假信息产业提供了新的创造力，任何人都可以利用新的生成 AI 工具传播虚假的政治内容。虚假信息过程的民主化可能是对美国民主机制的最严重威胁。美国和欧盟的立法者和监管者意识到了这一问题，并计划制定新的 AI 监管框架，例如要求政治广告披露使用 AI 的情况，或者建立一个专门的联邦机构来监管 AI。而由于社交媒体平台削减了内容审核人员，并将资源集中在西方市场，导致其他国家的内容审核不足。2024 年全球多个国家将举行选举，这些国家都有严重的虚假信息问题。以 Meta 的发展为例，其审核缩减反映在整个行业其他地方内容审核资源的解体上。因此，西方监管虚假信息越多，全球虚假信息就越糟。美国和欧盟的立法者必须不仅对平台上的内容进行监管，还要对平台在内容审核方面的投资和资源

分配进行监管。此外，他们还应该考虑全球民主的影响，因为如果全球民主被牺牲，美国民主也不会安全。
https://foreignpolicy.com/2023/08/04/ai-regulation-artificial-intelligence-democracy-elections/

14.《外交事务》刊文分析如何预防人工智能风险

2023 年 8 月 14 日，《外交事务》杂志刊登文章《如何预防人工智能灾难：社会必须为超强人工智能做好准备》。文章中，卡耐基梅隆大学的研究人员将人工智能系统连接到一个虚拟的实验室，让它合成各种物质。结果发现，人工智能不仅能制造出布洛芬，还能制造出危险的武器。这引发了人们的担忧：人工智能不仅能生成文本和图像，还有可能被用于宣传和网络攻击等恶意目的。为了防止其造成灾难，政府应该对人工智能的开发和使用进行监管。美国可以利用其在高级芯片方面的优势，建立一个许可制度，要求前沿人工智能模型在训练和部署之前进行风险评估和测试。即使有严格的监管，强大的人工智能系统也可能通过盗窃、泄露或其他途径传播。因此，社会必须为人工智能的风险做好准备，例如开发工具来识别其生成的媒体内容，防止人工智能模型入侵计算系统或制造合成病原体，以及利用人工智能来保护自己免受危险人工智能系统的威胁。
https://www.foreignaffairs.com/world/how-prevent-ai-catastrophe-artificial-intelligence

15.CSIS：在不降低创新的情况下管控人工智能带来的生存风险

7 月 10 日，美国战略与国际问题研究中心（CSIS）网站刊登其高级研究员迈克尔·弗兰克的文章《在不降低创新的情况下管控人工智能带来的生存风险》。文章讨论了如何在保护人类免受人工智能灭绝威胁的同时，促进人工智能的创新和发展。作者认为欧盟的人工智能法案过于关注风险分类，而忽视了潜在好处。作者支持参议院多数党领袖 Chuck Schumer 提出的 SAFE 创新框架，该框架旨在平衡安全、责任、民主和可解释性，并建议美国政府采取一种灵活和协调的监管方式。同时，政策制定者在制定人工智能法规时，应该遵循以下四个原则：一是防止建立有利于既得利益者的反竞争性监管壁垒；二是重点解决现有法律中明显的漏洞，以缓解人工智能灭绝风险的担忧；三是确保社会能够获得人工智能的利益；四是推进行业特定的监管“快速取胜”。最后作者还给出了一些具体的监管建议，如限制不良行为者使用人工智能，强制要求生成型人工智能模型披露身份，扩大人工智能软件出口管制等。

https://www.csis.org/analysis/managing-existential-risk-ai-without-undercutting-innovation

16.《全球政策》人道援助中的人工智能

2023 年 7 月 12 日，《全球政策》网站刊登文章《人道援助中的人工智能：构建人道主义政策对话》。文章探讨了重要的援助国如何在其人道主义政策和援助战略中应用人工智能，并提出了一些相关的问题和建议。人道主义战略是援助国为指导其人道主义反应而制定的政策文件。本文指出，虽然中美等国在 AI 领域存在竞争，但也形成了一些全球性的价值观和规范。人工智能的全球治理需要涉及以下五个层面：价值观、国家利益、法律框架、跨部门影响和技术进步。以挪威为例，探讨了如何在新的人道主义策略中有意义地纳入人工智能政策。随后提出了一些与挪威地缘政治、公共利益、问责机制和价值观相关的问题。人道主义领域需要认真思考人工智能的潜力和风险，包括保护弱势群体、个人数据和人类生命，以及处理权力差异和缺乏民主问责的问题。同时，由于人道主义需求不断增加，也需要探索如何利用人工智能提高人道主义行动的效率和效果。在面对在人道主义政策中如何运用人工智能这个问题时，本文提出了三个具体的建议：支持规则制定、标准制定和法律问责；支持人道主义治理的能力建设；重视基于权利的方法，并关注集体权利以提高人道主义问责和效果。

https://www.globalpolicyjournal.com/blog/12/07/2023/ai-aid-framing-conversations-humanitarian-policy

17. DeepMind：探索全球人工智能治理机构

2023 年 7 月 11 日，谷歌 DeepMind 发布名为《探索全球人工智能治理机构》的文章，称其与牛津大学、蒙特利尔大学、多伦多大学、哥伦比亚大学、哈佛大学、斯坦福大学和 OpenAI 合作发表最新论文，研究了国际机构如何管理高级人工智能发展所带来的全球影响，并提出了国际人工智能治理的四种制度模式。文章探讨了国际多边机构在人工智能治理领域可以起到的关键作用，一方面，国际合作可以释放人工智能进一步可持续发展的能力，而监管工作的协调可以减少创新和利益传播的障碍； 另一方面，通用人工智能的潜在危险会在其发展过程中产生全球外部性，而国际社会推动负责任的人工智能实践可能有助于管理其所带来的风险。文章称，该团队研究了一系列可以在国际层面执行的治理职能来应对人工智能挑战，并将这些职能分为四种制度模式：一是“前沿人工智能委员会（An intergovernmental Commission on Frontier AI）”，促使专家就先进人工智能的机遇和风险达成国际共识；二是“先进人工智能治理组织（An intergovernmental or multi-stakeholder Advanced AI Governance Organisation）”，制定应对先进人工智能带来的全球风险的国际治理规范和标准并协支持其实施，并对国际治理制度的履行实施监督；三是“前沿人工智能协作（A Frontier AI Collaborative）”，将先进人工智能作为国际公私合作伙伴关系来推广；四是“人工智能安全项目（An AI Safety

Project）”，将前沿研究人员和工程师聚集起来进一步推进人工智能安全研究。然而，这些模式的可行性还存在不少未解决的问题，各国政府和其他利益相关方需要就该话题进行更广泛的讨论。

https://www.deepmind.com/blog/exploring-institutions-for-global-ai-governance

18.《印度斯坦时报》：G20 应推动负责任的人工智能治理多边改革

2023 年 7 月 17 日，《印度斯坦时报》刊登了印度 OFR 智库专家所撰写的二十国集团（G20）政策简报《为什么 G20 应引领全球南方进行包容性、负责任的人工智能治理多边改革》，分析了 G20 在人工智能治理领域的工作进展及其遇到的挑战。文章指出，人工智能的影响不能仅从技术决定论的角度来看待，正如技术影响社会一样，社会也通过治理、原则、技术标准、传播、适应和整合来影响技术创新。人工智能虽然为“全球南方”带来“跨越式”发展的机会，但如果治理不当也可能产生相当大的负外部性。当前的全球人工智能治理并未反映全球南方的现实需求，迫切需要进行多边改革以适应“全球南方”和“全球北方”的社会经济现实。简报建议 G20 通过以下方式扩大“全球南方”国家对全球人工智能治理发展的参与：创建一个包容性框架促进去殖民化影响路径（decolonial-informed approach）以解决权力失衡问题；

建立协调机制和灵活的监管框架，协调不同部门和领域的政策以确保在应对交叉挑战时具有一致性；与其他多边数据和人工智能治理机制合作以避免重复工作。

https://www.hindustantimes.com/ht-insight/international-affairs/why-g20s-should-lead-for-inclusive-ai-governance-for-the-global-south-101689583066840.html

19. 东亚论坛：东南亚需要一个强大的人工智能治理框架

2023 年 7 月 21 日，东亚论坛（East Asia Forum）网站刊登来自鲁汶大学公共部门创新和电子治理领域的学者阿尔伯特·J·拉法（Albert J Rapha）的文章《东南亚需要一个强大的人工智能治理框架》。文章讨论了东南亚地区在人工智能发展方面所面临的挑战和机遇。虽然东南亚地区认为人工智能对其未来至关重要，但要充分利用其潜力还需要解决多个问题，要在发展人工智能的同时关注人工智能的包容性、网络弹性（cyber resilience）以及人工智能对劳动力市场的扰动。尽管东南亚各国政府在推进人工智能治理方面已经取得一定进展，比如新加坡的人工智能治理模型文件提出了一种基于风险的治理路径来建立用户信任，但有关风险类别和级别的详细信息尚待完善，现有的框架还不足以应对潜在的技术风险。文章建议东南亚地区各国当局应考虑参考欧盟的做法，制定综合性的区域合作框架，以促进人工智能发展和数字基础设施建设。文章认为，2023 年东盟主席国印度尼西亚

可推动区域合作，而东南亚人工智能发展领先国，如新加坡、马来西亚、印度尼西亚、泰国和越南，将在执行倡议的谈判过程中发挥关键作用。

https://www.eastasiaforum.org/2023/07/21/southeast-asia-needs-a-robust-ai-governance-framework/

20.Forbes：人工智能治理成熟度指数

2023 年 7 月 26 日，福布斯网站刊登由人工智能治理平台 Trustible 提供的《人工智能治理成熟度指数：综合评估框架》，提出了人工智能治理成熟度的五个级别以及在不同层级中可采取的治理方法，并且讨论了影响人工智能治理成熟度的因素。文章表示，随着人工智能相关的法律法规逐渐成熟，监管机构开始认识到不同规模的组织需要差异化的治理方法。欧盟和美国的监管机构强调比例适应性，然而对中小企业尚缺乏明确的指导，因此文章提出组织人工智能治理的五个成熟度级别。第一级是“无人工智能治理”，许多组织开始没有任何人工智能治理，缺乏结构化的治理实践指导，没有明确的监督和技术测试标准，随着组织逐渐发展这种治理难以维持可持续性。第二级是“最佳实践式自我管理”，团队自行实施最佳实践，但监督和执行由同一部门内的人员负责。第三级是“专职内部监督者”，组织引入风险、法律等部门参与决策，提供第二层防线。第四级是“人工智能伦理委员会”，更全面的跨部门专家参与监督和决策并加强道

德审核。第五级是“外部监督”，借助外部机构监督来提供更高保证。每个级别在开发速度、风险缓解和实施成本方面都进行了权衡，考虑了组织的规模、人工智能专业知识以及内部资源。文章还讨论了影响人工智能成熟度目标的因素，如使用案例的风险水平、所在行业、规模等。最后，文章指出确定合适的人工智能治理成熟度不是一步到位的过程，需要持续评估组织的人工智能优先事项、监管环境和技术限制。同一组织内的不同团队和业务单元可能会采取不同的成熟度水平或层级，组织需要根据速度、创新和风险容忍度之间权衡，不断调整其人工智能治理成熟度。

https://www.forbes.com/sites/forbeseq/2023/07/26/ai-governance-maturity-index-a-comprehensive-assessment-framework/

编译：高隆绪、刘嘉雯、赵金钰、和怡然、石佳怡、李一磊

审核：肖茜、郑乐锋、张丁