2022 年第 8 期（总第 33 期）

国际战略与安全研究报告

INTERNATIONAL SECURITY AND STRATEGY STUDIES REPORT

美欧数字外交研究系列之一：
美欧人工智能政策协调进程、挑战及启示

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

美欧数字外交研究系列之一：美欧人工智能政策协调进程、挑战及启示

张丁、王星懿、汤文君、郑乐锋、董汀、孙成昊
清华大学战略与安全研究中心美欧研究项目

2022年12月5日，美欧贸易和技术委员会（TTC）首次发布《可信赖人工智能和风险管理评估与测量工具联合路线图》（简称“联合路线图”），从术语规范、标准制定以及风险监测等三方面指导双方人工智能风险管理和可信赖人工智能发展，并促进相关国际标准的制定。本报告将主要探讨美欧在人工智能领域政策协调的背景、政策合作路线图、分歧以及相关启示。

一、美欧人工智能领域政策协调进程

2021年6月15日，美欧贸易和技术委员会（TTC）正式启动，并成立技术标准工作组，将人工智能等新兴科技领域的国际标准制定提上合作日程。在2021年9月第一次部长级会议后发布的联合声明中，美欧不仅再次申明双方对于发展可信赖人工智能的愿景，也表示要依照机构议程在部长级会议以外成立10个工作组，分别由美欧各相关机构牵头组织，聚焦开发人工智能等新兴技术领域标准制定的方法，应对技术滥用对安全和人权带来的威胁，其中专门提及在开发可信赖人工智能项目上增进合作。合作内容具体包括两方面：一是双方肯定彼此在人工智能风险管理框架制定的成果外，提出将致力于在可信赖和负责任的人工智能概念和原则层面达成共识，并有意在此基础上建立共同的测量和评估工具。二是双方在可信赖人工智能用于更好的机器学习、隐私保护技术研发及人工智能对未来就业的影响等方面也表

明合作意向。[①]

2022年5月TTC第二次部长级会议之后发布的美欧联合声明显示，双方将工作组的合作深化到各职能机构的协同行动，为12月第三次部长级会议前希望达成的目标制定具体工作计划。工作组重点关注的，是如何在遵循诸如经济合作与发展组织（OECD）有关人工智能发展的原则建议等已有共识基础上，使其适应美欧各自的法律体系，并和《欧盟人工智能法案》、美国《人工智能权利法案蓝图》以及《人工智能风险管控框架》等内部政策倡议相结合。这一过程离不开美欧各有关机构的协作，美国国家标准与技术研究院（NIST）以及欧盟委员会正和欧洲各标准化组织（ESOs）协调行动，尝试在人工智能风险管理、评估与测量工具及发展可信赖人工智能的社会技术要求方面达成一致，以对抗其他“非市场经济体”日益上升的影响力。此外，双方进一步强调开发可互操作的方法，加大信息和资料库共享的程度。[②]

在刚刚举行的第三次部长级会议后，TTC发布了新的联合声明，整理了会议的主要成果。在发展可信赖人工智能方面，美欧发布了联合路线图，为双方在风险管理和可信赖人工智能方面合作提供指引，并推动国际标准制定；另外，双方还达成一项评估医疗卫生领域隐私保护技术应用的试点项目，同时在研究人工智能如何影响就业、极端天气和气候预测、能源等领域增进合作。在技术标准制定合作方面，美欧双方已推出5月联合声明中提出的战略标准信息机制（SSI），这将有助于使双方在标准化及战略性事件上保持信息共享和协同行动。

---

① “US-EU Trade and Technology Council Inaugural Joint Statement”, The White House, September 29, 2021, https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/29/u-s-eu-trade-and-technology-council-inaugural-joint-statement/.

② “US-EU Joint Statement of the Trade and Technology Council”, The White House, May 16, 2022, https://www.whitehouse.gov/wp-content/uploads/2022/05/TTC-US-text-Final-May-14.pdf.

二、美欧联合路线图主要内容

目前，以美欧各自已推出的相关法规、政策文件为蓝本，双方将基于共同价值观指导新兴技术发展，力求未来在风险监管方式等层面达成一致。

（一）在充分尊重既往成果的基础上谋求共识

在联合路线图发布之前，美欧双方各自已发布人工智能相关监管政策文件。美国出台了《人工智能风险管控框架》《人工智能权利法案蓝图》，欧盟则出台了《欧盟人工智能法案》。这些政策文件是双方进行协商的重要基础。联合路线图所构想的工作将与包括国际标准制定组织（ISO）、OECD和电气与电子工程师学会（IEEE）在内的国际组织的各项全球性工作保持一致，并参考美欧双方的既往成果和理念，综合推进可信赖的人工智能和风险管控。

谋求概念和术语的一致性。联合路线图指出，推进概念、术语在理解和应用层面相一致在有效风险管控方面具有重要意义，在基本术语层面达成一致将在制定标准和明确责任、实践和政策时提供可互操的分类法。TTC将着力推进包括但不限于以下概念和术语的一致性：风险、风险管理、风险容忍度、风险认知、可信赖的人工智能的社会技术特征。此前，欧盟曾指出成员国内部存在人工智能风险评估标准不一致、需要引入明确的人工智能风险标准等问题[③]；美国国家标准与技术研究院则有明确的人工智能风险和可信赖度术语概念[④]。联合路线

③ “Regulatory divergences in the draft AI act: Differences in public and private sector obligations”, European Parliamentary Research Service, May 2022, https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729507/EPRS_STU(2022)729507_EN.pdf.

④ “AI Risk Management Framework: Initial Draft”, National Institute of Standards and Technology, March 2022, https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf.

图对概念、术语的强调，反映了双方对这一问题共同的重视。

（二）寻求人工智能国际标准规则制定

国际技术标准制定、可信赖的人工智能和风险管理工具的开发和使用以及人工智能风险监测和评估，是影响全球人工智能领域规则解释、标准应用、新兴技术赋能、市场开发的核心要素，直接影响利益相关方和人工智能透明化、公开化、公正化、包容化发展进程。在这方面，联合路线图提出的具体举措如下：

在国际技术标准制定层面，联合路线图提出应共同遵守：一是人工智能国际标准中的需求、规范、测试方法和指南应与可信赖特征相关；二是人工智能国际标准的全球领导；三是美欧共同支持和领导国际技术标准制定工作；四是在与各自法律体系相一致的前提下，充分遵守世贸组织《技术性贸易壁垒标准》；五是鼓励广泛的利益相关方参与国际技术标准制定工作。

在开发可信赖的人工智能和风险管理工具层面，依据OECD对“工具”一词的定义，联合路线图提出：一是建立美欧共享指标和方法的数据库，评估人工智能可信赖度和存在风险；二是支持利益相关方针对可信赖的人工智能相关工具开展分析，分析结果用以支持国际技术标准制定。

在监测、评估已出现的和潜在的人工智能风险层面，联合路线图提出，TTC将推动建立人工智能前沿科学及其相关风险研究的知识共享机制，并明确具体实施步骤：一是基于现有背景条件、具体案例、影响和伤害数据，持续跟踪已出现的和潜在的风险和风险类别；二是测试和评估人工智能风险的可互操性。

（三）联合路线图的短期和长期目标

联合路线图明确了短期和长期目标，从标准制定、工具开发和风险监测与评估等方面建立双边合作机制开始，逐步拓展到在国际层面发挥引领作用：

	短期目标	长期目标
推动建立具有包容性的合作渠道	建立3个专家组，分别负责国际技术标准制定、可信赖的人工智能和风险管理工具的开发和使用以及人工智能风险监测和评估工作； 制定每个专家组的工作计划； 制定利益相关方与专家组之间的协商计划，包括举行专家研讨会。	为国际标准的使用和领导提供信息，举办相关专家研讨会； 审查和评估进展，必要时更新路线图； 确定合作契机，分享经验成果。
推动术语和分类法共享	实现包括但不限于以下概念和术语的一致性：风险、风险管理、风险容忍度、风险认知、可信赖的人工智能社会技术特征。	进一步加深对术语和分类法的共同理解，并开展修订工作。
推动国际人工智能技术标准制定	开展美欧双方关心的标准的前景分析，评估各方对标准制定的参与度和贡献度； 完成双边关心的国际标准的制定； 加强与国际标准制定机构的沟通交流。	组织国际标准制定论坛，并就已确定的项目开展合作； 在共同关心的标准的制定和应用方面与专家开展合作，并提供相应支持。
开发可信赖的人工智能和风险管理工具	建立工具选择、纳入和修正程序； 为相关工具建立评估标准。	明确增加到共享中心/数据库指标和方法； 更新和维护共享中心/数据库。
推动监测、评估已出现的和潜在人工智能风险	根据工具使用情况，标记已出现的人工智能风险，建立跟踪办法，并尝试建立风险分类标准； 确定对新出现的人工智能风险测试和评估方法。	通过对人工智能伤害事件的实证研究，建立风险分类标准和评估机制； 针对新出现和潜在的风险进行理论预判。

三、美欧人工智能合作面临的挑战

美国过去许多分散和渐进式的人工智能监管政策在近几年正汇聚出积极变化。2022年2月，美国布鲁金斯学会发布的报告称欧盟和美国开始在人工智能监管领域趋同对标。[⑤] 自2021年9月召开的首次TTC会议将人工智能政策问题纳入重点讨论范围后，美国已开始启动监管引擎，招募人工智能治理专业人士，与欧盟开展更加积极主动的监管举措。这显示出拜登政府正趋向与欧盟相类似的人工智能监管目标，客观上将加强美欧在人工智能治理政策上的一致性。

此次TTC会议提出的路线图旨在为美欧跨大西洋人工智能风险管理和发展可信赖人工智能方法提供信息，并推进与人工智能相关的国际标准机构的协作方法。[⑥] 为使欧盟和美国在基于风险的方法上保持一致，路线图提出的推进术语共享、领导国际标准和新型风险监测三项具体方案，是美欧推进人工智能合作监管的重要共识。然而，路线图也指出，对于风险管理制度的建立以及监管和自愿措施之间的适当平衡，欧盟和美国在如何应用监管技术方面仍有差距。具体表现为在风险评估的责任分配、建立风险管理系统的可能法律责任，还有风险评估的法律责任等监管问题上存在潜在分歧。此外，人工智能作为美欧跨大西洋技术合作的关键领域，双方合力推动可信人工智能和风险管理制度仍存在以下阻力。

第一，战略与理念冲突。一方面，美欧人工智能战略并不一致。

---

⑤ “The EU and U.S. are starting to align on AI regulation”, Brookings, Feburary 1, 2022, https://www.brookings.edu/blog/techtank/2022/02/01/the-eu-and-u-s-are-starting-to-align-on-ai-regulation/.

⑥ “EU-US Joint Statement of the Trade and Technology Council,European Comission”,December 5, 2022，https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_22_7516.

对于开发人工智能及其监管工具，美国安全和外交政策机构将之视为大国竞争的重要国家安全资产，要使其成为拓展技术影响力的工具⑦，但欧盟基于经济发展与民主价值观更关注人工智能技术的伦理挑战。此外，欧盟对减少技术依赖性与脆弱性的寻求，使之渴望在人工智能领域拥有一定程度的战略自主，这包括事前技术监管上的独特优势。自2020年以来，美欧在人工智能监管领域出台法案并成立相关工作组。相比之下，美国推动的人工智能法案中对人工智能的定义更窄，通用人工智能的豁免范围更广，并在提出自身的个性化风险评估。

在今年10月美国就TTC合作提交欧盟委员会的非正式反馈文件中，美国反对通用人工智能提供商必须与其用户合作以帮助拥护遵守《人工智能法》的观点，包括披露机密商业信息或商业秘密。美国更侧重商业利益的追求一定程度上与欧盟重视人权与隐私的技术发展观相悖。此外，基于双方的不同发展战略与研发创新基础，欧盟人工智能发展相对美国总体落后。当前通用人工智能系统的主要供应商多来自微软和IBM这样的大型美国公司，而据生命未来研究所（FLI）11月发布的报告显示，欧洲通用人工智能模型研发明显滞后，欧盟企业越来越依赖Meta、谷歌、微软和百度等美国和中国公司的通用人工智能系统来开发其他人工智能应用工具。⑧这种依赖可能阻碍欧盟参与制定人工智能全球标准的努力，也将加大美欧人工智能互操作性的阻力。⑨

---

⑦ Schmidt, Eric. “AI, Great Power Competition & National Security.” Daedalus 151, no. 2 (2022): 288–98. https://www.jstor.org/stable/48662042.

⑧ “Emerging Non-European Monopolies in the Global AI Market”, Future of Life Institute, November 2022, https://futureoflife.org/wp-content/uploads/2022/11/Emerging_Non-European_Monopolies_in_the_Global_AI_Market.pdf

⑨ Luca Bertuzzi, “The US unofficial position on upcoming EU Artificial Intelligence rules”, October,26,2022,https://www.euractiv.com/section/digital/news/the-us-unofficial-position-on-upcoming-eu-artificial-intelligence-rules/.

另一方面，美欧对人工智能风险管理理念亦存差异。美国更鼓励人工智能技术创新与发展，强调监管的科学性和灵活性，致力于确保和增强美国在该领域的科技和经济领导地位。欧盟的人工智能方法强调以人为本，其监管风格兼顾发展与规制，期望通过高标准立法和监管来重塑全球数字发展模式。美国极力限制监管范围以促进创新与发展，欧盟意图加强监管以强化个人权利保护。对欧洲而言，自身的数字脆弱性正在成为一种地缘政治安全问题，而美国人工智能的发展影响欧洲防务及更广泛领域，当前美欧的战略合作可能会导致跨大西洋数字鸿沟。[⑩]因此，双方在开展人工智能领域的治理行动难免产生摩擦与分歧。

第二，技术监管差异。人工智能作为数字技术的重要领域，其监管模式将受到整体数字监管环境的影响。在监管方式上，美国与欧盟在人工智能监管方面遵循两种不同思路，**美国侧重于技术应用和流程监管，而欧盟侧重组织管理与风险管理。**2020年1月，美国政府发布的《人工智能应用监管指南》代表美国在人工智能监管上科学审慎监管、不监管、非监管措施多管并举下共促创新与发展的总体思路。同年2月，欧委会在《人工智能白皮书》提出投资和监管并举的思路，包括基于风险路径对人工智能建立新监管框架，目的在于使各种风险和潜在损害最小化。相比之下，欧盟人工智能监管框架延续GDPR的立法初衷，更强调个人权利保护和人工智能应用的负面影响，而美国更强调促进人工智能创新与发展，因而为人工智能应用创设“安全港”、监管例外、监管豁免等制度。欧盟为人工智能应用进入监管领域设置较低门槛，其监管要求具体入微，监管和执法覆盖全过程，而美国所倡导的科学审慎监管、风险评估与管理、成本效益分

⑩ Soare, Simona R. “DIGITAL DIVIDE?: Transatlantic Defence Cooperation on Artificial Intelligence.” European Union Institute for Security Studies (EUISS), 2020. http://www.jstor.org/stable/resrep25027.

析、灵活敏捷等理念，会很大程度压缩监管空间。美国也并不希望完全采用欧盟自上而下的人工智能监管方式影响自身技术创新优势。⑪

由于欧盟市场的规模和美国在数字监管方面的相对不作为，世界各地的国家和公司目前多在该领域采用一些欧洲标准，这常使欧盟与美国科技公司和政府发生直接冲突。虽然美欧双方曾就欧盟技术监管对美国的影响，以及共同监管的跨大西洋联盟的可行性进行讨论，但美国表示不会完全接受欧盟式的监管。

当前在与数字监管相关的关键点上美欧存在三大分歧：美国反垄断立法对政府的支持程度低于其欧洲同行；缺乏真正的美国数字监管框架；美国宪法第一修正案对在线平台的保护程度不足。⑫在平台监管与数字审核方面，欧盟正采取事前监管，而美国为事后监管。⑬在监管落实速度上，参考《通用数据保护条例》（GDPR）的生效过程，《欧盟人工智能法案》主导下的重大监管决定出台仍有待时日，而在通过TTC加强与欧盟治理合作的背景下，美国无论是现阶段国内监管政策的调整还是未来几年政策变化的趋势，都有可能与欧盟逐渐靠拢，甚至在决策和执行速度方面领先于欧盟。值得注意的是，美欧的不同监管方式也将影响各自技术和产业发展，过度强调监管存在拉大欧盟与美国人工智能总体发展差距的风险，并限制自身标准的全球

⑪ AdamThierer,“Why is the US following the EU’s lead on artificial intelligence regulation?”, The Hill, July 21, 2022, https://thehill.com/opinion/technology/3569151-why-is-the-us-following-the-eus-lead-on-artificial-intelligence-regulation/

⑫ “The EU as a digital regulatory superpower: Implications for the United States”, European Council on Foreign Relations,April8,2020, https://ecfr.eu/article/commentary_the_eu_as_a_digital_regulatory_superpower_implications_for_the_u/.

⑬ “The US-EU Trade and Technology Council (TTC): State of Play, Issues and Challenges for the Transatlantic Relationship”, European Council on Foreign Relations, January 2022, https://www.esade.edu/ecpol/wp-content/uploads/2022/12/AAFF_EcPol-OIGI_PaperSeries_ENG_def_jan22.pdf.

影响力[14]，双方监管差异也将阻碍双方的规制整合。

第三，数据治理分歧。此次路线图特别提到隐私在推进负责任的人工智能发展中的重要性。此前，美欧在技术政策中存在的摩擦尤其集中在数据治理的主权与隐私问题上。[15]在数据隐私方面，技术竞争政策和数字监管是跨大西洋技术政策中最具争议的问题。在立法层面，欧盟早已将数据治理和隐私保护纳入法律，2020年的数据战略目标是在欧盟内部建立单一的数据市场和相应的数据治理战略，但这实际上会阻碍美欧之间的数据流动。而美国在数据保护和隐私问题上缺乏全球影响力，法律上的参差不齐难以缓解相关的国家安全风险。2022年6月，美国众议院和参议院发布《美国数据隐私和保护法案》（ADPPA）讨论稿，但该法案离正式成法还有一定距离。关于人工智能监管与算法决策，ADPPA将为公司收集的个人信息制定国家标准和保障措施，包括旨在解决算法潜在歧视影响的保护条款。[16]虽然法案中诸多条例都与欧盟GDPR类似和趋同，但法案出台的重要意图之一便是制衡欧盟在全球隐私保护领域的影响，以推广美国隐私保护理念，例如“选择退出”机制、有限私人诉讼权等。[17]

---

⑭ Alex Engler, “The EU AI Act will have global impact, but a limited Brussels Effect”, Brookings, June 8, 2022, https://www.brookings.edu/research/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/

⑮ “Lighting the Path Framing a Transatlantic Technology Strategy”, Center for A New American Security, August 30, 2022, https://www.cnas.org/publications/reports/lighting-the-path.

⑯ Niketa K. Patel and Tori K. Shinohara,“The American Data Privacy and Protection Act: Is Federal Regulation of AI Finally on the Horizon?”, Mayer Brown, October 21, 2022, https://www.mayerbrown.com/en/perspectives-events/publications/2022/10/the-american-data-privacy-and-protection-act-is-federal-regulation-of-ai-finally-on-the-horizon.

⑰ American Data Privacy and Protection Act Draft Legislation Section by Section Summary, https://www.commerce.senate.gov/services/files/9BA7EF5C-7554-4DF2-AD05-AD940E2B3E50.

此外，欧盟对数字主权的追求，双方对地缘政治中利益优先事项的差异，还有美国政治愈发极化下的不稳定政策而造成跨大西洋隔阂，都将影响美欧人工智能监管合作的实质效果。

四、美欧人工智能政策协调启示

近年来，人工智能发展更为强大复杂，其技术成熟度已被认为是国家实力的核心组成部分，但人工智能系统的脆弱、偏见、不安全和不透明等问题使之无法在高风险用例中使用，如何构建可信赖的人工智能系统成为主要大国规制竞争的关键问题。[⑱]美欧在TTC联合路线图中所提出的核心愿景，便是合力打造基于风险的监管方法和可信人工智能系统。虽然目前尚未看出双方具体的监管调整进展[⑲]，但美欧技术联盟开展的政策协调行动，对他国推动人工智能风险监管与治理合作具有启示意义。

首先，参考国际标准开展国际合作，积极参与国际人工智能治理标准体系。阐明与可信特性相关的要求、规范、测试方法或指南的人工智能标准，有助于确保人工智能技术和系统满足互操作性等关键目标，也能促进人工智能准确性、可靠性和安全性等性能特征。因此，人工智能国际标准的全球参与和合作对于实现互操作性至关重要。[⑳]这要求政府在制定内部人工智能治理政策时与国际方法保持一

⑱ “International Competition over Artificial Intelligence.” Strategic Comments, vol. 28, no. 3, 2022, pp. vii–ix, https://doi.org/10.1080/13567888.2022.2091878.

⑲ David Matthews,“EU and US set out plan to create rules of the road for artificial intelligence”, December6, 2022, https://sciencebusiness.net/news/eu-and-us-set-out-plan-create-rules-road-artificial-intelligence

⑳ “US-EU Joint Statement of the Trade and Technology Council”, The White House, December 5, 2022, https://www.whitehouse.gov/briefing-room/statements-releases/2022/12/05/u-s-eu-joint-statement-of-the-trade-and-technology-council/.

致。当前，美国和欧盟寻求领导人工智能国际标准化工作，并通过在国际标准组织目前正进行的人工智能技术标准开发方面的合作来实现，这些标准将影响可信赖人工智能和风险管理的设计、操作、评估和测量。酌情支持和使用国际标准，可作为技术法规、合规评估（conformity assessment）和区域标准的基础。与此同时，欧盟和美国的TTC合作经验表明与利益相关者及相关机制合作非常重要，从中能找出现有国际人工智能标准制定活动中的关键差距。

其次，推进人工智能治理共享术语和分类，在国际合作中协商统一内涵以扩大框架的互操作性。TTC路线图提出提供一个可互操作的词典来交流风险和适当处理风险，这将能够反过来促进人工智能风险和影响的可互操作测量和评估。在人工智能国际监管合作中共同开发工具，例如共享指标存储库，同样可以促进风险衡量的透明度、互操作性和统一性。TTC路线图里提出要共享术语和分类的核心基于国际标准化组织（ISO）和国际电工委员会（IEC）联合发布的人工智能基础标准，中国企业也可以兼顾在美国国家标准技术研究所（NIST）、电气电子工程师学会（IEEE）和欧洲标准委员会（CEN）、欧洲电工技术标准委员会（CENELEC）等相关标准框架下积极参与。

最后，建立人工智能风险监测机制与风险评估审计机制，完善人工智能风险治理工具包。TTC路线图里提出建立风险监测体系和知识库，能够让美国与欧盟更加全面深入的掌握人工智能治理方向。中国可考虑通过风险监测和对策的高效反馈来建立人工智能治理领域的专业领导力。人工智能产品的市场份额是未来掌握技术话语权的重要基础，但会受到人工智能安全认证的影响。中国可考虑积极通过双边或多边机制构建互认的人工智能风险评估审计机制。此外，应完善人工智能风险治理工具包，为中国企业特别是中小企业应对人工智能治理提供可负担、易操作的工具选择，在加强人工智能治理的同时辅助企业符合法规和技术标准的要求。

发表日期：2022年12月26日

审编：肖茜　　签发：达巍

扫码关注我们

清华大学战略与安全研究中心编印

办公地点：北京市海淀区清华大学明理楼428房间

联系电话：010-62771388

http://ciss.tsinghua.edu.cn　邮箱：ciss@tsinghua.edu.cn