

新网络安全态势下网络安全治理体系的再调整

王璐瑶

中国政法大学, 北京 100091

摘要: 网络安全态势因快速迭代的数字技术进程而呈现复杂多变的发展特征。尽管《网络安全法》已于2025年部分修订,但这并非网络安全治理体系优化的“终点”。作为治理对象和治理工具的人工智能技术使得网络安全态势发生变化,既有网络安全体系与网络安全态势在治理逻辑、治理工具及治理机制层面存在不适配问题。由于这些变化并未整体性变革网络安全态势,回应方式应以“制度细化”代替“专门立法”。在绝对安全无法实现已成共识的当下,需调适网络安全治理理念:区分人工智能引发的网络安全风险和技术滥用产生的衍生风险,并在敏捷治理和协同治理理论指引下,确保网络安全治理体系具备灵活性和可延展性。具体而言,网络安全治理体系应完成治理主体、治理目的和治理工具三个层面的调整,从传统意义上的治理主体各自履行法定义务转向特定场景下不同治理主体相互配合治理。

关键词: 网络安全法; 网络安全治理; 技术标准; 网络安全保险; 市场调节

中图分类号: D922.29 **文献标识码:** A **文章编号:** 1672-3805(2026)02-0009-11

作者简介: 王璐瑶,中国政法大学比较法学院博士研究生,研究方向为比较民商法。

一、问题的提出

自2017年《网络安全法》正式施行以来,我国网络安全治理体系得到长足发展:不仅完成了以《网络安全法》《数据安全法》《个人信息保护法》为核心的框架建构,而且细化了与网络安全相关的配套法律制度实施细则,如《网络数据安全条例》《个人信息保护合规审计管理办法》等。然而,人工智能等数字技术重塑了原有技术路线和业务模式,使得网络安全态势处于难以预料的复杂状态。网络安全攻击的技术门槛因人工智能技术的辅助功能而显著降低,国外甚至出现KawiiGPT这样一款为网络犯罪提供生成钓鱼邮件、勒索软件说明文档和攻击脚本的人工智能,网络安全防护体系面临更严峻的挑战。此外,网络安全问题的表现形式也不再局限于传统意义上的技术漏洞、系统瘫痪,还表现为针对大模型训练活动实施的“数据投毒”。国家安全部专门为此发布安全提示,这类行为除造成削弱大模型性能等损害结果外,还会诱发输出错误诊疗建议等有害违法内容^①。

在此背景下,2025年《网络安全法》迎来首次修订,专门提及“加强风险监测评估和安全监管”“国家支持创新网络安全管理方式,运用人工智能等新技术,提升网络安全保护水平”。此次修订回应了人工智能领域的网络安全问题,但在网络安全治理层面仍存在一个核心问题:面对新网络安全态势,我国网络安全治理体系应否以及如何进行调整。具体而言,该问题可拆解为三个子问题,一是人工智能时代网络安全态势究竟发生何种变化促使网络安全治理体系亟须调整;二是人工智能时代我国应遵循何种网络安全治理逻辑;三是在这种治理逻辑指引下,我国网络安全治理体系应按照

^① 数据来源:《警惕人工智能“数据投毒”》,参见https://tech.cnr.cn/gstj/20250821/t20250821_527330779.shtml。

何种方式调整和优化。

二、网络安全态势的新趋势和新治理需求

(一) 新趋势之一：作为治理对象

人工智能技术正在改变传统代码编写方式和效率，直观表现之一是人工智能成为恶意代码生成的“效率工具”。随着 Claude、Gemini 等技术工具功能迭代，即使网络黑客技术领域的新手也能通过与这些技术工具的高效互动，实施较为简单的网络黑客攻击活动。不同于以往代码生成服务，这些技术工具可以根据用户提出的具体需求生成更具针对性的技术路线，既有信息系统面临更加频繁的“个性化”网络安全攻击。部分学者提出，利用人工智能技术生成的恶意代码具有较强可扩展性、更加复杂且难以被检测^[1]。在现有技术条件下，直接要求人工智能针对关键信息基础设施网络安全防护体系生成难以被检测和阻拦的恶意代码仍存在一定难度，但该技术能够推动网络黑客攻击呈现高频次、规模化、集团化发展趋势。攻击者不再需要花费相当多的时间成本定位安全漏洞并设计相应黑客技术路线，只需要利用人工智能技术批量化生成恶意代码，并实施勒索攻击、钓鱼邮件等网络攻击措施，即可实现“低成本、高收益”的攻击目标。也有学者认为，人工智能时代的网络安全风险主要表现为网络攻击越来越智能化、大规模网络攻击越来越频繁、网络攻击的隐蔽性越来越高、网络攻击的对抗博弈越来越强、重要数据越来越易被窃取或破坏等五个特点^[2]。例如，在传统僵尸网络攻击中，网络黑客控制大量计算机、物联网设备等组成网络集群，这些设备被植入恶意软件，形成规模化攻击。时至今日，这种网络攻击方式在人工智能技术的加持下具有高度可扩展性，特别是在人工智能与物联网技术结合的场景下，此种网络攻击方式可导致更大范围的网络通信瘫痪和信息系统宕机。

需要说明的是，人工智能技术改变网络安全态势并不意味着该技术应成为网络安全治理对象，而是人工智能技术应用方式应成为网络安全治理对象。在大模型开源成为人工智能产业的常见业务生态时，网络安全治理体系更需要关注相应的技术滥用风险。法律从来不是数字技术创新的枷锁和负担，更不可能因潜在技术滥用风险直接限制开源生态，所以网络安全治理实践真正需要干预的是人工智能技术应用方式，既包括主体层面的技术使用者、技术研发者及服务提供者等，也包括具体的技术应用行为，如禁止人工智能技术应用于恶意代码生成、屏蔽违法违规的提示词等。这也随之产生一个新问题：人工智能技术本身并不能有效识别和确认所生成的代码是否属于恶意代码，“白帽子”可能会利用恶意代码辅助检测和识别是否存在系统安全漏洞，要求技术研发者和服务提供者在生成环节直接切断相应代码生成，可能存在“法律过于干预技术创新实践”的问题。换言之，单纯从信息生成环节限制技术滥用的治理逻辑已经失效，我国网络安全治理体系应进行更深层次的结构调整。这种调整方式不仅涉及从技术研发到技术使用、更新维护等整个技术应用生命周期进行全方位干预，还涉及扩展既有治理工具和治理方式，不再单一依赖法律或者某类技术治理工具控制网络安全风险。实际上，网络信息内容治理同样存在类似问题，尽管《人工智能生成合成内容标识办法》《互联网信息服务深度合成管理规定》等法律法规已对人工智能生成内容进行全方位约束，但现有立法对于使用者自行删去 AI 生成标识、使用国外不受限制的人工智能产品等问题客观上力有未逮。可以说，人工智能作为治理对象时，改变了网络安全治理原有的治理逻辑和治理方式。

(二) 新趋势之二：作为治理工具

此次《网络安全法》的修订专门将人工智能技术作为网络安全治理的技术工具之一，鼓励和支持利用该技术提升网络安全防护水平。人工智能作为一项中立性数字技术，能够为网络攻防双方提供不同程度的辅助功能，这也是当下网络攻防对抗博弈日趋激烈的客观原因之一。在部分学者看来，人工智能作为治理工具的主要表现是以“个性化”“定向传播”以及各类算法推荐系统为支

撑,实现短时间、大范围、规模化地传播特定信息,特别是深度伪造技术的升级,使网络信息内容的真伪难以辨识,信息竞争呈现“易攻难守”的特点^[3]。此种观点实际上也反映出当下网络安全基本内涵的扩张,即从传统意义上的网络运行安全和网络信息安全扩展至网络信息内容安全。《网络安全法》第78条将“网络安全”界定为“通过采取必要措施,防范对网络的攻击、侵入、干扰、破坏和非法使用以及意外事故,使网络处于稳定可靠运行的状态,以及保障网络数据的完整性、保密性、可用性的能力”。这种界定方式与治理实践存在一定脱节:网络安全事件的表现形式逐渐超出狭义层面的技术安全观念范畴,开始呈现技术安全、信息内容安全、业务生态安全等诸多安全要素相互结合的特点。按照这种逻辑,“网络安全”这一概念似乎将陷入内涵和外延无限扩张的理论困境,但倘若置于总体国家安全观框架下,这种概念扩张现状也能够圆满解释:网络安全本身是一种网络空间治理的基本目标,指向网络空间在特定治理体系下所能够维持和优化的稳定可靠状态。在《网络安全法》中,“网络安全”实属狭义概念,更加侧重通过技术安全指标判断网络空间是否满足既有治理目标;在总体国家安全观中,“网络安全”则属于广义概念,与经济安全、政治安全等传统安全相互交织重叠,涉及国家主权竞争和博弈^[4],故而网络信息内容安全也被纳入“网络安全”的概念范畴,更确切地说,网络信息内容安全是网络安全与其他传统安全重叠领域的安全形态。

人工智能的创新应用正在改变整个社会的运行模式,但这种特殊性并不是促使人工智能技术治理成为网络安全治理特殊环节的实践根源所在。在整个数字生态系统中,人工智能技术与其他数字技术并无实质性差异,均服务于数字经济社会的转型升级。该技术确实具有广阔的市场应用前景,但更为重要的是,其对传统的网络安全保护逻辑提出全新挑战。审视《网络安全法》的立法过程,学者曾争论过该法的功能定位应如何设置。彼时学界对于法律是否应将网络安全纳入调整范畴存在分歧,主流观点一度认为网络安全治理问题属于网络安全攻防、计算机系统安全等技术领域的研究议题。随着网络安全成为全球治理的共同议题,学者也开始意识到网络安全不可能仅凭技术类治理工具解决,而是需要在技术无法解决的环节介入法治模块补全整个网络安全治理体系。时至今日,治理观念再次发生变化:在法律无法有效“兜底”网络安全治理实践时,又需要再度回归技术、标准等治理工具。客观而言,在“人工智能+”的业务模式导向下,该技术能够自由扩展至多元化应用场景,这也使得《网络安全法》立法过程中所采取的“法律责任驱动行为选择”治理策略逐渐失效。即便义务主体为避免承担行政处罚等法律责任而选择履行义务,也会因各种不确定因素而导致网络安全事件的发生。因此,人工智能作为治理工具的内在逻辑实际上是拓展网络安全治理方式,借助“技术治理技术”逻辑尽可能降低潜在的网络安全风险。

(三) 现有网络安全治理体系与新趋势的不适配性

在人工智能技术作为治理对象和治理工具两种新趋势的作用下,既有网络安全治理体系逐渐表现出与这些新趋势不匹配的现象,具体表现在治理逻辑、治理工具及治理机制三方面。

治理逻辑方面,既有网络安全治理体系无法有效回应人工智能应用方式的不确定性。网络安全态势虽未发生根本性变革,但数字技术的创新进程使得现行立法与科技创新实践发生脱节。需要强调的是,此处所指“脱节”并非简单意义上的法律空白或者监管缺位,而是指向法律介入数字技术治理的时机不再明确。作为治理对象的人工智能技术应用所引发的网络安全风险并非传统意义上在事前阶段可以有效干预,其风险产生与技术创新保持同步。过去,网络安全治理的基本逻辑是在事前阶段尽可能设置多个防护环节以预防风险发生;但当下的部分网络安全风险由技术创新本身导致,特别是智能体技术应用一旦成熟,不排除网络安全攻击自动化的新兴风险产生,而这类风险治理显然不可能在事前阶段完全预测并预防。在网络安全治理体系趋于细化的背景下,网络安全、数据安全、网络信息内容安全、数字技术安全等安全概念的逻辑关系尚待理清。学界普遍认为,网络安全治理应当与网络安全态势保持同步发展状态,但并未细化“同步发展”如何实现,大多停留于“出现

新问题就需要新立法”的笼统逻辑。在实现高质量发展和高水平安全的治理实践中,首先需要明确网络安全治理目标究竟是什么,狭义、广义的网络安全观也意味着不同的治理逻辑和治理架构。

治理工具方面,新修订的《网络安全法》专门增加诸如国家支持运用人工智能等新技术提升网络安全保护水平,背后的原因正是旧法未能给人工智能技术赋能网络安全治理预留足够的法律解释空间。此次修订虽然对此予以增补,但更依赖通过技术标准、开源实践以及产业试点扶持等方式加速推动人工智能技术在网络安全领域的创新应用,故而需要将《网络安全法》的鼓励性条款延伸至具体场景。例如,《网络安全法》第24、25条明确网络产品、网络服务应当符合国家标准的强制性要求,针对网络运营者在采购人工智能驱动的网络安全产品或服务场景,除了需要进一步明确该场景中的国家标准外,还需要增补有关人工智能驱动的网络安全产品或服务的监管要求。实践中,部分厂商可能在已有产品基础上进行“套壳”服务,又或者直接采用开源产品导致实际的网络安全防护存在重大漏洞,需要在网络安全治理体系层面对这些问题进行专门回应。

治理机制方面,人工智能引发网络安全风险的成因多元化与既有网络安全治理体系的单一化不适配。虽然人工智能技术创新并未产生完全意义上的“新风险”,却使网络安全风险的成因不再以线性方式逐层发生。例如,在人工智能技术被应用于生成黑客代码场景下,网络安全风险的形成不能再简单归因为技术服务提供者提供技术工具,而是用户的使用行为超出了技术研发者和技术服务提供者所能实际控制的范畴。客观而言,在相当长的一段时间内,强制要求技术研发者、技术服务提供者在生成环节限制黑客代码的生成无可操作性。黑客代码根据用户指令生成,人工智能产品并不能自动生成这类代码。在网络安全攻防领域,黑客代码也会用于测试网络安全防护系统,其本身并不具有违法性,真正具有违法性的是技术滥用行为。法律或许可以要求技术研发者、技术服务提供者在生成环节以弹窗、加粗字体等方式提醒用户遵守《网络安全法》《刑法》等法律法规,但缺乏足够的正当性依据限制人工智能产品生成具有技术中立性特征的信息内容。

解决人工智能技术创新引发的网络安全新趋势与既有网络安全治理体系的不适配性,依赖于治理逻辑和治理理论层面的“微调”。新网络安全态势终究未产生“全新问题”,网络安全治理体系的调整方向则应以“修缮”和“优化”为核心导向。

三、面向新兴风险的网络安全治理理论

(一) 治理体系再调整的基本前提: 风险概念的再确认

网络安全治理面临的核心挑战便是“科林格里奇困境”,即如何在保障技术创新进程的同时实现对技术风险的有效管控。在技术发展早期,管控技术要求较为简单,但也无法形成对技术风险的全面认知,过早干预反而会限制技术创新;在技术相对成熟之后,管控措施则会对业已稳定的产业生态施加全面影响,管控成本难以控制^[5]。这种困境产生的原因之一是如何较为全面地识别和认知潜在技术风险,在技术早期以最小干预方式达成合理的技术管控效果。

面对人工智能技术创新引发的网络安全态势变化,治理体系再调整的关键是厘清和区分人工智能引发何种网络安全风险及其与传统意义上的网络安全风险有何区别。总结现有研究,有关人工智能技术风险的讨论可分为两个阶段:第一阶段是在ChatGPT问世之前,学界有关人工智能的讨论大体围绕“人工智能能否成为民事法律主体”“人工智能犯罪”等领域,这类研究通常被视为“超前预判”“科幻主义式的风险研判”。当时的人工智能技术水平仍停留于“智能客服”等浅层应用,故而此类质疑观点反而切中研究痛点。第二阶段则是在ChatGPT问世之后,其取得的创新突破使得研究重心发生改变,过去停留于理论层面的各类技术风险正在成为可能。但这一阶段也产生了另一现象,即泛风险化研究。该类研究以总结和梳理各类技术安全风险为论证起点,套用以往理论范式推导出相应治理对策。事实上,关注技术风险本身并无问题,但在《网络安全法》《数据安全法》《生成式人工智能服务暂行办法》等法律法规实施之后,再总括性地讨论人工智能技术风险并不能解决不同领

域的技术应用监管难题。从社会实践来看,人工智能技术引发的安全风险是多方面的,既包括生成恶意代码、用户数据泄露等风险类型,也包括技术滥用引发的各类侵权问题。这些安全风险本身在治理逻辑和治理方式上存在较大差异,在调整和优化网络安全治理体系时,应当结合技术实践区分网络安全风险。技术层面,人工智能安全评估指标通常包括信息系统的完整性、可用性、私密性、公平性等,故而也有观点将人工智能安全体系划分为通用人工智能安全和扩展人工智能安全两大板块,前者包括通信安全、数据安全、系统安全,后者则包括应用安全、数据隐私、模型算法安全^[6]。治理层面,这些安全评估标准可根据治理方式的差异性转化为人工智能系统安全和人工智能应用场景安全。人工智能系统安全主要是指人工智能信息系统本身所存在的网络安全风险,这类风险以传统观念中的网络安全技术风险为主,如漏洞供给等。因此,这类风险的治理方式除沿用既有网络安全治理机制外,还需要根据技术创新动态增补相关网络安全技术标准。人工智能应用场景安全则是指人工智能技术在特定场景下与其他信息系统交互应用所产生的综合性网络安全风险。这类风险的核心特征是风险成因多元化,使得基于链式因果关系设置的法律认定机制失效。其他风险类型则可以置于广义上的技术滥用风险范畴。微观层面,人工智能系统安全风险新趋势表现为因技术创新突破导致的网络安全漏洞、外部新兴网络安全攻击等,人工智能应用场景安全风险新趋势表现为人工智能技术与特定业务场景的创新结合产生了全新的网络安全风险,如医疗人工智能产品存在数据污染导致大模型功能稳定性安全风险。

(二) 治理理论之一: 敏捷治理

面对人工智能等新兴数字技术引发的新型网络安全问题,不少学者认为应将现有治理模式调整至敏捷治理,进而更好地适应瞬息万变的产业实践需求。敏捷治理主要是为回应数字时代外部环境互联互通、风险不确定等要素所产生的社会治理新需求,旨在“提高组织在复杂和动态环境中的应变能力和决策效率”^[7]。最常被引用的概念界定是2018年世界经济论坛白皮书所提出的“适应性强、以人为本、包容性和可持续性的决策过程”^[8]。也有学者认为敏捷治理在概念内涵层面指向的是一种治理价值。总结学界现有观点,大体包含回应性与适应性、开放性与结构性、灵敏性与系统性、持续性与调适性等核心价值^[9]。而在人工智能治理领域,部分学者则将“敏捷治理”以治理框架形式予以展开,提出治理理论的适应性、治理对象的包容性、治理主体的多元性以及治理工具的灵活性四项核心特性^[10];又或是将网络安全治理框架建构遵循的基本原则细化为遵循整体方法、全面覆盖组织、确定利益相关者需求、适应企业需求、创建动态治理体系以及治理与管理分离等六项内容^[11]。尽管具体治理理论及其相应治理方案不尽相同,但普遍认同当下的网络空间治理需要更为多元化、灵活化的治理方式,以便填补传统治理模式与复杂技术实践之间的“鸿沟”。

敏捷治理之所以在人工智能时代备受追捧,是因为该理论突出了网络安全治理的“快捷性”。这种快捷性并非治理政策、法律法规出台速度快,而是能够快速响应各种超出预期的网络安全问题。其一,敏捷治理理论延续了风险与收益平衡的基本逻辑,网络安全等级保护、数据分级分类治理等现有制度便是敏捷治理的直观体现。人工智能时代,网络安全态势再度发生变化,在敏捷治理导向下,网络安全治理体系需要根据网络安全生态系统进行必要的机制延伸和拓展。其二,敏捷治理填补了传统事后追责的治理局限,强调以更为灵活的方式构建各类治理主体共同参与的治理机制。《生成式人工智能服务管理暂行办法》等法律法规固然能够通过法律责任约束各类主体的行为选择,但人工智能相关的网络安全事件成因具有复杂性特征,传统的因果关系认定方式难以继续适用。在敏捷治理导向下,网络安全治理体系所需要延展的新机制则应当是引导各类治理主体采取整体性治理措施,各个治理环节相互关联,进而更为灵活地管控安全风险和减少实际损失。其三,敏捷治理突破单一化治理模式,更强调治理体系本身的可发展性。客观而言,人工智能技术的创新发展超出了社会公众预期,当然也超出了监管机构的预见范围。因此,强行要求监

管机构及时采取合理的治理措施解决新兴网络安全问题未免过于苛刻。监管机构更需要提升适应能力,即在首次发生类似网络安全事件后,能够有效避免同类网络安全事件的再次发生,进而不断提升网络安全防护能力。

也有学者注意到作为国外社会治理理论的敏捷治理存在“水土不服”问题,进而提出数据安全领域的“中国化”方案:建构战略指导层、管理层、参与层和监督层四个架构,其目的是形成一个正式的、分层级的权利和控制结构^[12]。事实上,敏捷治理的理论价值主要在于提供了有别于以往的治理逻辑,而非提供具体可操作的治理措施。现有研究多将其与协同治理混同论证,忽视了该理论的核心内容是发现和回应快速变化的社会实践,而不是论证如何具体治理。在该理论的导向下,网络安全治理应当以“尽可能预防”取代“全面管控”,达成事前预防和事后减损的双向治理目标。

(三) 治理理论之二:协同治理

在社会治理实践中,多元主体协同治理已成共识,大部分学者在论及网络安全治理时也习惯性地协同治理作为核心结论,遗憾的是,“如何协同”“按照何种方式协同”等核心问题并未得到真正解决。部分学者认为,当下网络安全治理“缺乏多方协作共同治理的经验”,过度依赖政府的强制措施^[13]。究其根本,部分学者所主张的协同治理本质上是不同主体履行不同的法定义务,而非理论范式层面的协同式治理。事实上,现有研究对协同治理概念本身也存在诸多分歧:治理工具协同论是以促成“政策与法律的协同”治理为目标^[14];多元主体论则是认为协同治理理论起源于西方,主要包含“多元主体基于共同目标的联合行为过程”和“多元主体在规则支持基础上自愿、平等协作”两层内涵^[15];生态治理理论倾向于将协同治理描述为笼统的治理方向,如“法治是基石,多元协同是关键”^[16],其核心是以网络空间生态为基础,强调系统性治理理念。诚然,也有学者提出多元主体协同治理应当建立“多元主体关系相关规则”和“多元主体协同规则”^[17],但未能释明应如何治理。因此,在治理理论层面,首先需要明确的是协同治理究竟归属独立的理论范式,还是笼统的治理措施。倘若将协同治理作为一种具体的治理理念,其核心是公私部门之间通过各种正式或者非正式互动机制实现单一组织无法实现的公共目标,故其理论功能所要解决的核心问题是如何推动不同主体达成并执行集体化的治理行动。倘若将协同治理作为一种笼统的治理方向,其与敏捷治理则存在交叉重叠,即均强调不同治理主体共同参与社会治理实践这一状态,其核心则转变为如何推动单一治理向多元治理转变。

在网络安全治理领域,笼统层面的治理主体范围扩张无助于解决现实问题,反而会削减中心化社会治理架构的实际效能。换言之,协同治理应当作为一种具体的理论范式用于指导网络安全治理体系的设置与调整。“协同”应当是主体与主体、治理机制与治理机制之间的协调互动,既表现为社会主体相互配合,共同推动建构更为全面的网络安全防护体系;也表现为治理机制的彼此衔接与嵌入,填补单一治理环节单一治理机制的不足。部分学者认为,多元主体共同参与且充分履职尽责是协同治理的内涵^[18],鼓励社会公众自觉加入网络安全治理实践亦是协同治理的直观表现^[19]。也有部分学者认为,除多元主体参与外,协同治理的实质在于“通过多个主体的参与、各种规制方式的运用以实现各项治理要素的协同”^[20]。不过,这些观点仍然没有触及协同治理的深层内涵,即如何促成协同,而这需要回归到网络安全治理实践。尽管网络安全态势严峻,但并不意味着所有网络节点均处于被频繁攻击状态。网络安全事件大体包括外部网络安全攻击和内部系统安全漏洞两大类:前者以攻击者为导向,通常会选择经济价值较高的网络节点作为攻击对象,此时的“协同”需侧重被攻击对象、监管机构、用户等治理主体之间的相互配合,如网络安全信息共享机制便是最佳选择之一。后者则以网络运营者为导向,如内部员工窃取内部数据、安全漏洞报告不及时等风险因素,网络安全防护不再是一种负担,而是网络运营者为了避免经济损失自发选择的防护策略,治理重心则是内部管理机制能否有效落实。此时的“协同”则需要网络运营者采取更为主动的安全管理策略,在

内部设置多种保障机制确保一旦某项网络安全管理机制被“空置”，其他安全保障机制能够迅速替代，尽可能降低网络安全风险。同时，监管机构、用户出于各自不同的目标，形成外部监督机制，进而达成内外部协同治理状态。

四、面向新兴风险的网络安全治理体系调整方向

（一）架构调整：治理环节、治理主体与治理工具

人工智能创新对网络安全治理实践的影响主要是网络安全治理理念变化，不再局限于某一时间节点的绝对安全，而是强调在特定时间周期内的相对安全。因此，网络安全治理体系的架构调整需从治理环节、治理主体以及治理工具三个层面切入。

治理环节层面，网络安全治理体系需在事前风险预防的基础上增补“事前风险警示”和“事后快速恢复”两大内容。“零风险”偏好已经脱离了产业实践，会导致企业创新成本增加。网络安全事件导致的损害结果往往是规模化、系统性的，难以通过事后经济补偿、技术修复等方式回溯至之前的状态，但网络安全事件成因具有多样性和多变性特征，立法、执法乃至司法均不可能全面阻断网络安全风险，也不可能强行要求各类信息服务提供者达成超出自身能力范围的“零风险”目标。所以，网络安全治理核心应当是尽最大可能预防各类网络安全风险和最大化网络安全事件的事后恢复能力。一方面，风险预防不是全面预防，而是以义务主体的能力范围为边界，尽可能在事前控制各类风险因素。风险预防并不局限于义务主体充分履行网络安全保护义务，还包括以各种方式提示用户事前警惕潜在的网络安全风险，尽可能减轻网络安全事件导致的损害结果。另一方面，网络安全事件的损失类型主要分为两类，一类是事件发生时产生的经济损失，另一类则是在服务中断、信息系统故障期间，该事件产生的持续性经济损失。既然网络安全事件的发生不可能完全预防，那么就理应提前做好相应的应急处置措施，尽可能确保发生网络安全事件后在最短时间内快速恢复。这种事前风险预防和事后快速恢复的治理目标也意味着网络安全治理能力能够跟随技术创新进程和产业实践一并变化，否则只会因过时的网络安全策略导致无法在网络安全事件发生后快速排除故障、恢复服务。所以，网络安全治理的核心要素便是提升网络安全治理方式的可延展性，在面对复杂多变的网络安全态势时，各类网络安全防护机制不仅能够内部循环互动，还能够结合人工智能等数字技术的创新发展而延伸出多元化、个性化的治理策略。

治理主体层面，网络安全治理体系的延展方向应当是扩展网络安全防护的实际参与主体。当然，如果仅在数量层面强调监管机构、企业、社会公众、行业自律协会等主体应当积极参与网络安全治理实践，这种宽泛结论不啻于再将治理问题以另一种形式重新表述。故此，有必要以更为具体的方式描述治理主体如何实现多元化。一方面，“多元化”侧重强调网络安全治理应当采用非单一化的治理架构。无论是依靠监管机构全方位强化网络安全监管要求，还是企业等市场主体出于维护商誉、避免经济损失等目的采取自发式网络安全防护，客观上均难以有效管控潜在的风险因素。强调治理主体多元化的底层逻辑是通过治理主体之间的相互监督、相互协助，将网络安全防护范畴进行实质性扩张。另一方面，“多元化”侧重强调网络安全治理机制的场景化设置。如若仅在宏观层面解释治理主体的多元化，势必会推导出“监管机构+企业+社会公众+行业自律协会”的宽泛结论。在高水平安全的治理目标导向下，网络安全治理体系的延展应当是以新出现的网络安全风险为基础，而这些新风险、新问题也与特定的应用场景、行业领域相关，需要在中观层面的特定场景、特定领域中解释治理主体的多元化。一方面，网络运营者作为直接的责任主体，需按照《网络安全法》《关键信息基础设施安全保护条例》等要求严格落实网络安全等级保护、网络安全审查、网络安全漏洞报告、监测预警等法定义务；另一方面，由于关键信息基础设施与国家安全密切相关，监管机构需以更为主动的方式介入网络运营者的运营活动，确保其内部安全管理制度充分落实。此外，治理主体的多元化也意味着其他治理主体的“必要参与”。在监管资源有限的情况下，则需引入其他治理主体

督促网络运营者履行义务,如第三方认证机构评估和认证网络安全防护体系的有效性、审计机构对网络安全业务合规情况进行合规审计等^[21]。在其他场景下,治理主体多元性则会有所变化。如在健身场馆等线下非以网络信息服务为主营业务的场景中,常见的网络安全风险以客户个人信息泄露、会员认证信息服务中断等为限,在不增加额外经济成本的情况下,治理主体则以经营者、监管机构和消费者为主,其治理重心是规范内部员工的业务活动,预防其操作不当或者恶意窃取消费者个人信息等情况的出现。

治理工具层面,人工智能嵌入网络安全治理体系的方式应当是以《网络安全法》第24、25条规定的网络安全设备、产品或服务的采购标准为切入点。既要明确该场景下人工智能产品相关的技术标准,还要明确人工智能网络安全设备的应用范围。需在网络安全治理体系中细化三项内容:一是细化《网络安全法》第25条规定的网络安全认证、检测机制在人工智能领域的适用方式;二是细化《网络安全法》第24条第2款规定的网络产品、服务持续性安全维护义务履行标准;三是细化《网络安全法》第27条规定的网络安全事件应急预案、补救措施在人工智能领域的适用方式。鉴于人工智能技术对治理实践的辅助功能,也有观点提出“人机协同治理”的新型治理模式。该模式主要是指“人与机器在治理过程中共同参与、相互协作”,本质上是一种“合作—竞争—合作”的动态协作关系^[22]。此时的“协同”与“分工合作”并无差异,只不过这种治理模式混淆了机器在治理实践中的功能定位。在“作为治理对象”和“作为治理工具”的发展趋势下,人工智能更多是以技术工具而非治理主体的方式参与网络安全治理实践。人工智能可能被应用于网络安全风险监测、内部业务合规审计等不同场景,但这些应用方式均以人工智能提升治理效率为内核。

(二) 机制延展:公众参与、治理透明与协同式监督

敏捷治理和协同治理的另一理论价值在于重新释明了网络安全治理的基本目标,即从结果目标导向转向过程实施导向。在传统观念中,网络安全指向的是网络信息系统能够预防和抵御各类网络安全风险以及网络数据的保密性、完整性和准确性,描述的是一种采取相应网络安全措施之后所达成的静态结果目标。但是,在相对安全取代绝对安全后,这种结果目标导向的治理理念显然无法契合网络安全态势,无论采取何种网络安全防护措施,均会指向相类似的网络安全治理目标,所以更重要的是关注作为过程的网络安全。作为过程的网络安全指向的是各类治理主体不仅需要充分履行各自的法定义务,还需要构建链条式网络安全防护体系。类似于全生命周期治理,这种链条式的网络安全防护体系同样强调各业务环节应当配置相应的安全保障机制,但此模式以网络空间生态体系为实践基础。实践中,不同规模的企业存在差异化网络安全防护能力,其网络安全保障措施既囊括自身内部安全管理措施,还涉及网络安全产业的安全防护服务。因此,链条式网络安全防护体系更加侧重“内部为主、外部协作”的基本治理模式,力求在治理过程中实现各类风险因素的应对处置。

强调过程层面的网络安全是“尽最大可能采取合理措施保障网络安全”的应有之义。特别是在人工智能技术广泛应用的当下,网络安全的另一层内涵表现为通过过程层面的网络安全实现技术可信状态。伴随大数据、云计算等数字技术创新,各类网络安全事件层出不穷,社会公众对于网络信息服务的安全可靠性持有质疑,不利于数字技术的长久发展。强调过程层面的网络安全则是将网络安全防护过程透明化,社会公众知晓并了解各类网络安全防护措施。这不仅有利于引导社会公众对网络安全风险形成正确认知,也有利于以相对透明的方式形成充足的外部监督机制,进一步强化网络安全防护体系的可靠性。在部分学者看来,安全化过程涵盖了“安全化主体、指涉对象、存在性威胁以及受众”^[23],此种观点揭示了网络安全评价的核心要素,即各类治理主体的义务履行和风险应对情况、特定网络安全治理需求、主要网络安全风险类型以及网络安全的受众群体。在现有研究中,“受众”是一个常被忽视的核心要素,导致网络安全治理体系的延展方向局限于网络安全态势的

变化，遗漏了社会公众对人工智能技术信任度的考量。

具体而言，从结果目标向过程实施的转变主要涉及三个方面：第一，基于社会公众的充分知情，延展社会公众的参与渠道。社会公众参与网络安全治理实践的方式不应局限于投诉、举报网络安全问题，还应当包括作为受众群体知晓和了解特定网络运营者的网络安全防护能力和措施，进而决定是否接受相应的网络安全服务。第二，借由第三方网络安全保护主体的引入，延展网络安全防护的透明化机制。不公开具体网络安全防护措施固然有利于提升网络安全保护效果，但这种不公开并不等同于不告知、不透明，现有网络安全实践已经证明单一主体的网络安全防护能力终究有限，这也是《网络安全法》等法律法规引入第三方机构安全认证、风险评估的原因所在。借由网络运营者、监管机构以及第三方机构的三方介入，实现更大范围的网络安全防护效果。第三，考量风险控制的全面性，延展网络安全防护的监督机制。过程层面的网络安全还指向各类网络安全保护措施得以充分履行实施，但在经济利润、内部组织机构等各类因素的作用下，网络运营者可能存在虚置网络安全保护措施的问题。此时，治理主体多元化表现为外部主体从不同层面介入网络安全防护措施的监督活动中，以市场调节、强制性规范等方式影响网络运营者自觉履行网络安全保护义务。

（三）工具增补：市场工具的加入

敏捷治理侧重网络安全治理对网络安全态势变化的灵活应对，同时兼顾核心治理架构的稳定性。因此，这种灵活应对不是通过频繁修订网络安全治理机制予以实现，而是通过延展网络安全治理工具，再借由不同治理工具之间的相互组合，实现更为多元化的治理架构。此次《网络安全法》的修订过程亦体现了这一点，如增加人工智能技术应用于网络安全保护等条款，其直接目的便是充实网络安全治理的“工具箱”，以便在出现新问题、新风险时能够提供有效的治理模式。尽管《网络安全法》已将法律、技术标准、科技伦理、行业自律规范等纳入治理体系，但这些治理工具之间如何相互组合仍未得到明确。结合协同治理理论来看，这些治理工具的相互组合与衔接需以具体的治理任务和治理对象为依据。在人工智能生成恶意代码环节，治理任务是预防大批量生成涉及网络安全攻击的恶意代码，单纯让技术研发者、技术服务提供者限制此类功能的提供显然不切实际。所以，治理工具的组合方式表现为“法律明确禁止性事项+科技伦理促成合理使用+技术标准限定违法事项”。

我国在延展网络安全治理工具的过程中，市场工具的组合衔接方式常被忽视。在法律、技术、市场、伦理等治理工具中，市场工具通常以市场资源分配规律为基础，以经济激励效益为作用机制，引导各类治理主体自发采取网络安全防护措施。在宏观层面，一旦发生网络安全事件，企业所面临的实际损失包括商誉减损、业务中断的直接经济损失以及未来的商业缔约机会丧失等间接损失。而在市场调节作用下，为避免这些损失，即便没有相应的强制性义务，以网络信息服务为主业的企业也会主动采取措施预防网络安全损失。市场作为治理工具的实际内涵也不仅限于此，还包括网络安全态势变化影响网络安全产业。特别是在人工智能技术高速迭代的当下，传统网络安全厂商开始尝试将人工智能技术融入网络安全产品和服务，这亦是技术工具与市场工具相互组合的直观体现。市场工具存在固有的调节失灵问题，还需要法律、政策工具在必要时予以引导，规范网络安全产业发展方向，避免将存在安全风险隐患的开源大模型直接应用于网络安全防护体系。事实上，《网络安全法》还专门规定关键信息基础设施运营者在采购网络产品和服务时，一旦涉及国家安全问题，还需要提供国家安全审查。该制度是为避免市场工具可能导致的“优先选择价格低廉产品和服务”之弊端。

此外，网络安全治理体系在市场工具层面的延展内容还包括网络安全保险。工业和信息化部及金融监管总局早在2023年就发布了《关于促进网络安全保险规范健康发展的意见》，提出以该险种提升企业的网络安全防护能力。网络安全保险作为一种新兴险种，主要以特定的网络安全事件为

承保事项,其内在逻辑被部分学者总结为网络安全领域的补充型治理,通过保险业务的市场逻辑促使保险机构和投保人采取同向的网络安全防护措施^[24]。对于保险机构而言,为避免支付保险理赔金,不仅会在承保过程中对投保主体的网络安全风险水平进行全面评估,还会在保险期间内要求投保人严格履行网络安全法定义务;对于投保人而言,为避免因违反《保险法》以及保险合同约定的免责条款而无法获得保险理赔,则会主动采取必要的网络安全防护措施。如此一来,即便投保人自身的网络安全防护能力有限,也能够借由保险机构的外部辅助,进一步优化和完善内部的网络安全管理。

五、结语

数字技术创新发展具有“双刃剑”效应,在社会治理层面更需要持续审视现有治理体系如何与数字技术创新进程保持同步。人工智能所引发的网络安全风险虽非“全新问题”“全新风险”,但超出了现行网络安全治理体系调整范畴。面对这些新旧交织的网络安全风险,“专门立法”显然并非最佳选择,灵活调整和延展网络安全治理体系能更好地适应网络安全态势的动态变化。敏捷治理和协同治理作为两类具体理论范式,其共同点在于能灵活调整网络安全治理架构,敏捷治理侧重根据网络安全态势变化调整治理策略,协同治理则侧重围绕网络安全风险成因补全缺漏的网络安全防护板块。网络安全治理终归不是超前治理或者预测式治理,不可能对当下和未来的网络安全问题“面面俱到”,治理实践的关键仍需回归网络安全态势的可适应性。防止网络安全事件的发生固然重要,但更重要的是通过已经发生的网络安全事件及时调整和扩展网络安全治理体系,避免类似事件重复出现。

参考文献

- [1] 韩陆超,贾梦培.生成式人工智能赋能网络安全:挑战与机遇[J].科学管理研究,2025,43(6):93-99.
- [2] 贾焰,方滨兴,李爱平,等.基于人工智能的网络空间安全防御战略研究[J].中国工程科学,2021,23(3):98-105.
- [3] 杨楠.人工智能时代的国家信息战——以美俄信息战为例[J].俄罗斯东欧中亚研究,2025(4):112-134+165.
- [4] 赵精武.人工智能治理中的技术自主权[J].中国社会科学院大学学报,2025,45(1):94-112+154-155.
- [5] 赵中华,张倩.走出“科林格里奇困境”:生成式人工智能的风险表征及治理探索[J].大连理工大学学报(社会科学版),2025,46(4):8-15.
- [6] 孙宇,王红杰,杜彦辉.人工智能安全标准化综述[J/OL].指挥与控制学报,1-15[2026-02-08].DOI:10.20278/j.c2.2096-0204.2025.0066.
- [7] 刘邦奇,黄倩倩,张金霞.GenAI教育应用伦理风险多主体协同治理框架及实践策略[J].现代远程教育研究,2025,37(5):48-58.
- [8] 世界经济论坛官网.白皮书:敏捷治理——在第四次工业革命中重新思考政策制定[EB/OL].(2018-04-24)[2026-01-07].<https://www.weforum.org/publications/agile-governance-reimagining-policy-making-in-the-fourth-industrial-revolution/>.
- [9] 刘洋溪.迈向敏捷治理:教育数据治理的风险识别与优化路径[J].华中师范大学学报(人文社会科学版),2024,63(6):154-164.
- [10] 郑煌杰.AIGC赋能新质生产力的数据风险及其敏捷治理[J].河海大学学报(哲学社会科学版),2024,26(4):89-102.
- [11] Melaku, Henock Mulugeta. A dynamic and adaptive cybersecurity governance framework[J]. Journal of Cybersecurity and Privacy, 2023(3): 327-350.
- [12] 范玉吉,张潇.数据安全治理的模式变迁、选择与进路[J].电子政务,2022(4):114-124.
- [13] 宋瑞娟.大数据时代我国网络安全治理:特征、挑战及应对[J].中州学刊,2021(11):162-167.
- [14] 张宇帆.论人工智能政策与法律协同治理[J].东方法学,2024(5):187-200.
- [15] 张敏,王玥.美国“受控非秘信息”协同治理路径及对我国“重要数据”立法的启示[J].情报理论与实践,2022,45(10):36-44.
- [16] 文新良.数智化时代网络舆情风险的四维生态及其协同治理[J].吉首大学学报(社会科学版),2025,46(5):117-124.
- [17] 田少帅.数字经济背景下元宇宙空间的法律治理[J].北方法学,2024,18(5):131-145.

- [18] 危红波. 我国数字社会风险治理责任分配[J]. 学术交流, 2021(10): 130-143.
- [19] 林伟. 人工智能数据安全风险及应对[J]. 情报杂志, 2022, 41(10): 105-111+88.
- [20] 张荣刚, 李英楠. 数字社会网络犯罪协同治理模式转型与优化进路[J]. 河南财经政法大学学报, 2025, 40(4): 114-127.
- [21] 赵精武. 个人信息保护合规审计的基本框架与制度衔接[J]. 法律科学(西北政法大学学报), 2025, 43(3): 83-97.
- [22] 詹海宝, 孙江, 郭梦圆. 基于人机协同的网络文化安全治理新模式[J]. 传媒, 2023(15): 75-77.
- [23] 季雨洁. 安全化理论视角下的印度人工智能发展探析[J]. 印度洋经济体研究, 2025(6): 69-98+172.
- [24] 周瑞珏. 数据泄露风险治理中网络安全保险的介入路径[J]. 北方法学, 2024, 18(2): 76-90.

(责任编辑: 徐 源)

Re-adjustment of the Cybersecurity Governance System under the New Network Security Situation

WANG Luyao

China University of Political Science and Law, Beijing 100091, China

Abstract: Cybersecurity posture exhibits complex and dynamic development characteristics driven by the rapidly iterative progress of digital technologies. Although the Cybersecurity Law was partially amended in 2025, this does not mark the "endpoint" of the optimization of the cybersecurity governance system. As both a governance object and a governance tool, artificial intelligence has reshaped the cybersecurity posture, giving rise to mismatches between the existing cybersecurity system and the current cybersecurity posture in terms of governance logic, governance tools and mechanisms. Since these changes have not fundamentally transformed the nature of cybersecurity posture, the response should prioritize "system refinement" over "dedicated legislation". Against the current backdrop where the infeasibility of achieving absolute security has become a broad consensus, the concept of cybersecurity governance needs to be adjusted: cybersecurity risks induced by artificial intelligence should be distinguished from derivative risks arising from technology abuse, and under the guidance of agile and collaborative governance theories, the cybersecurity governance system should be endowed with flexibility and scalability. Specifically, the cybersecurity governance system needs to undergo adjustments at three levels: governance subjects, governance objectives, and governance tools, shifting from the traditional paradigm, where each governance subject independently fulfills statutory obligations, to a model where different governance subjects collaborate for governance in specific scenarios.

Key words: cybersecurity law; cybersecurity governance; technical standards; cybersecurity insurance; market regulation