数据安全视域下的人工智能风险应对研究

吴沈括　石嘉黎

摘　要：当前，人工智能正加速发展，且广泛应用于社会生活的方方面面。与此同时，它在网络安全领域的风险也日益凸显，尤其是数据安全问题，需要引起高度关注。本文从分析数据要素相关的侵害行为出发，选取模式识别这一典型应用场景来具体阐述，并在立法层面探讨现行规则应对与应然规则建构，在产业领域建议相关企业加强风险防控，使各方主体共同助力来确保人工智能安全、可靠、可控发展。

关键词：人工智能；数据安全；刑事立法；风险防控

中图分类号：G210.7　　**文献标识码：**A　　**文章编号：**1009-2447（2019）02-0095-09

一、引言

当今时代，人工智能正在加速发展，在全球范围内掀起日益高涨的浪潮。它依托于大数据和云计算两大基础平台，建立起机器学习、模式识别、人机交互三大通用技术体系，逐渐渗透到各大行业的创新发展之中。2017年，全球人工智能核心产业规模已超过370亿美元，预计至2020年将超过1300亿美元，年均增速达到60%。[①]而且，其产品和服务广泛出现于经济生活的各个方面，如智能机器人、智能驾驶、智能安防、智能家居等。同时，各国也愈来愈重视，积极出台相关政策和规范性文件以构建前瞻性的布局，我国于2017年颁布了《新一代人工智能发展规划》，提出人工智能已成为“国际竞争的新焦点”“经济发展的新引擎”，将其正式提升至国家战略的高度，为这新一代技术的研发、应用和推广筑起坚实的后盾。

然而，在热议人工智能的发展前景和技术红利的同时，挑战和危机也随之而生。不少学者专家对其带来的安全风险有了愈来愈深的担忧：如2017年GMIC大会上，霍金提出人工智能威胁论，称“人工智能可能是人类文明史的终结，除非我们能学会如何避免危险”；特斯拉CEO埃隆·马斯克曾警告说：“人类可能在不知不觉中创造了一个‘不朽的独裁者’”；还有报告指出：人工智能会扩大现有威胁，因其可使攻击成本降低、速度加快，目标扩大化、集合化；会引入新的威胁，人工智能系统可被利用于完成人类无法实现的攻击形态；会改变典型的威胁特征，使攻击目标更为明确、攻击更有效率，且因难以归因而无法对其采取切实的应对措施。同时报告还预测，在未来的五至十年里，人工智能会催生新型网络犯罪、实体攻击和政治颠覆。[②]上述论断说明了人工智能所带来的威胁，不仅仅是物理层面的实体攻击，还包括在伦理、道德、就业、政治等方面的恶性影响。需要特别指出的是，在网络安全领域，针对系统安全、在线内容安全、组织管理安全，人工智能也正在引发一系列安全风险，其中与在线内容安全密切关联的数据问题，成为影响人工智能安全、可靠、可控发展的一大关键，有必要引起高度重视并尽早采取应对策略。

在具体分析之前，这里对于本文所研究的人工智能范围做一个限定。国际上有一种较为常见的

基金项目：国家社会科学基金项目“新一代信息技术与个人信息刑法保护”（15CFX035）。

作者简介：吴沈括，男，浙江宁波人，北京师范大学刑事法律科学研究院暨法学院副教授，硕士生导师，法学博士后，研究方向为网络安全、数据维护、网络犯罪；石嘉黎，女，浙江宁波人，北京师范大学刑事法律科学研究院暨法学院硕士研究生，研究方向为网络安全。

分类方式，即将人工智能分为弱人工智能、强人工智能和超级人工智能，标准在于能否实现人类具有的认知功能、思考功能、推理功能等，或是否拥有类似人类的自我意志和自主思维能力。可以认为，从弱人工智能到超级人工智能，分别达到了“类人”“人类”和“超人”三个级别。从目前发展情况看，我们仍处于弱人工智能这一初级阶段。未来，对于强人工智能和超级人工智能实现的可能性已无争议，只是实现时间的远近尚未确定。[③]因此，本文不停留于泛泛空想，也不过度忧虑未知的风险，仅将弱人工智能纳入研究的范围，讨论当前阶段人工智能数据相关的风险及刑事犯罪问题，并结合包括刑法在内的网络安全法治体系，探讨立法规范的建构，并为产业领域的风险防控提供建议。

二、解读人工智能数据的风险及犯罪问题

（一）数据要素相关的侵害行为

数据是人工智能进行学习和决策的源头，人工智能通过采集和存储、管理和分析、可视化计算等技术手段来处理具备规模大、种类多、产生速度快、时效性强、价值密度低等特征的数据，这一环节的重要性便使数据要素成为人工智能犯罪的主要侵害对象。数据安全强调三性：数据的保密性、完整性和可用性，在人工智能应用场景中也是如此，下面笔者就从三性入手，分析数据要素上发生的侵害情形。

对数据保密性的侵害。这种情形通常涉及利用用户的个人隐私信息作为实施犯罪行为的手段，比如通过网络窃取用户的交易账单、理财情况来了解用户的资产状况和基于在线行为的支付意愿，从而快速有效地定位金融诈骗的受害者目标群体；再如非法获取受害者的个人隐私信息，组成多维度、海量化、定制型的训练数据，并构建恶意算法用于自动生成非法网站、链接、电子邮件等，向受害者定向推送，诱使其点击；或者通过学习这些信息伪装成受害者的好友，在社交网络聊天中实施犯罪行为。

对数据完整性的侵害。这种情形往往导致人工智能系统获取信息失败，因缺少关键数据而无法进行深度学习，影响下一步工作的效率和准确性。比如，Bot驱动的大规模“Information Generation”攻击，利用分散信息等手段以破坏信息通道的正常运行，从而使获取真实信息变得极为困难。

对数据可用性的侵害。这种情形一般包括对数据的篡改、造假、干扰等，比如对人工智能系统输入非真实的训练数据以获得与目标背道而驰的结果；通过更改、拼接、制作高度逼真的假视频或音频，借名人效应发表煽动性言论，扰乱政治和社会稳定；在人工智能应用于图像识别的过程中，通过制造恶意图片，使系统触发相应的安全漏洞，改变程序正常执行的控制流或数据流，对于标签、索引等一些关键数据进行修改，从而使人工智能系统输出攻击者指定的错误结果。[④]

上述三种情形在未来将越来越常见，且会对人工智能系统本身造成较大的负面影响。总的来说，人工智能算法的训练需以庞大数量的数据为基础，系统会通过终端收集范围更广、数量更多、内容更多样化的信息，这也使得数据所具有的实质性价值更为凸显，有经济价值之处就易出现人为的负面性，随之而生的是安全风险日益加剧；而且，人工智能系统并非独立的、排他性的，而是与外部环境、各种要素形成密切的联系。比如，产品或服务研发、生产、提供和应用的过程中，系统设计者、平台提供者、功能管理维护者以及应用者等多个主体均参与其中，且根据自己的身份具备相应的存储、修改、传输、访问、使用数据的资格，这种数据在多主体之间的流动会增加不稳定性和权属争议。此外，人工智能的智能化还会带来外部风险，犯罪分子可利用其技术升级自己的侵犯个人信息安全的犯罪行为，可能造成数以万计的公民隐私信息的泄露，也可能为已形成明显上中下游关系的网络黑灰产业推波助澜。因此，基于诸多原因，人工智能数据相关侵害行为日益增多，产品或服务的研发者、制造和提供者、应用者等相关人员应当加强对数据保护的重视，在储存、清洗、整合数据的各个环节都注意加大安全保障力度，从而防范风险。

（二）典型应用场景——模式识别的数据相关犯罪分析

模式识别是指“模拟人的感知过程，通过分

析图像、语音、视频等感知数据，对数据中包含的模式（物体、行为、现象等）进行判别和解释的过程”。[⑤]目前，其被广泛应用于生物特征识别、语音识别、文字识别、遥感及医学诊断等领域。下面，笔者从几个子应用场景入手，对模式识别中涉及数据相关犯罪的情形进行具体分析。

生物特征识别发展至今已有五十余年的时间，逐渐成为身份认证中较为重要的一种方式。比如商业化发展最早的指纹识别，在开关类系统验证身份时具有便捷、精确、安全的优势；在近些年深度机器学习快速发展的背景下，人脸识别通过建立海量的人脸图像数据，开发一系列高效的分析模式，成为安防监控中关键的手段。此外，随着智能手机、可穿戴设备大规模进入民众的生活，生物特征识别的发展趋向于便携式应用和高速产业化。由此可见，生物特征识别对于用户和环境这两个要素有较大的依赖性，体现出更大程度的人机交互特性，且其应用多涉及与个人身份相关的隐私数据，对安全性自然也有更高的要求。实践中，生物特征识别常常与金融类、财产类犯罪相关联，具体的表现形式包括：一是数据泄露问题，犯罪分子可通过攻击漏洞窃取大量用户隐私数据，包括指纹、虹膜、脸、身形等身份信息，而这些信息将被利用于侵入安防系统、盗取保险设备中的钱物、实施财产诈骗行为等；二是数据篡改或破坏问题，犯罪分子可采用变形、模糊、真假调换等处理技术对数据进行伪造，进一步干扰以数据为基础的系统的正常运行。

语音识别技术在20世纪末和21世纪初取得了较大进展，在模型结构、深度神经网络训练、多语言训练、参数学习等技术的支持下，逐渐实现大规模产业化，在声讯服务、语音评测、安全监控、语言翻译等方面发挥着重要作用。而同时，技术的提升也带来了安全隐患。在以往的诈骗案件中，犯罪分子常通过假冒身份的方式迷惑受害者，博取其信任，身份造假多使用文字信息，手段较为单一、平面化，许多警惕性高的人并不容易掉入陷阱。而现在，犯罪分子通过学习语音识别技术，开发语音合成系统，模仿特定对象的声音，达到以假乱真、难以辨别的程度，且在没有其他身份验证措施的情况下，很容易引导受害者走入诈骗迷局。

与语音识别技术相似，视觉识别技术目前也有较广泛的应用领域，尤其在安防监控上发挥着巨大作用，如对人群中特定目标的识别、车辆识别、物体追踪、异常事件检测等。视觉识别依靠图像视频分析技术，而分析的对象便是以物联网为基础的海量数据，包括来自监控相机、移动设备、互联网的各类多媒体数据等，通过整合不同场景的不同数据，系统才有可能分析出目标更复杂的行为、目标之间的关联和群体目标间的事件级演变。[⑥]因此，数据是视觉识别的关键所在，其安全性直接影响着视觉识别的有效性、精确性和保密性。认识到这一点，攻击者可以通过干扰或破坏数据的方式影响视觉识别技术发挥功用。典型的如实施逃逸攻击，即在不改变目标机器学习系统的情况下，通过构造特定输入样本以完成欺骗目标系统的攻击。[⑦]此种对抗样本在不同程度上侵害了在线内容的安全性，干扰了视觉识别的系统运行，为实施进一步的犯罪行为提供了帮助。

三、人工智能数据相关的立法规范建构

（一）立法模式的确定

确定立法模式，即如何对待我国现有的相关立法规范，以及在何种程度上另外制定新规范的问题，有学者针对网络犯罪立法模式提出了两种维度的选择视角：一是静态维度，考虑一元的刑法典模式，或与单行的网络犯罪相关法结合的多元模式，又或颁布一部综合性的网络法；二是动态维度，考虑围绕网络犯罪应采取“渐进式”或“前瞻式”的立法形式。[⑧]而对于人工智能数据相关的立法规范建构，其实也可以借鉴上述视角。从静态维度看，笔者认为在现阶段继续采用一元的刑法典模式更为合理。因为人工智能已逐渐应用于各个场景之中，与各个学科、各个行业领域也有了愈来愈紧密的联系，出台一部具有总括性的、系统化的人工智能法律确有一定的现实需求。目前，美国和欧盟已进行了尝试，美国国会提出《人工智能未来法案》，而欧洲议会通过了“关于制定机器人民事法律规则的决议”，但这些立法实践还是比较初步的，主要针对小范围的某一应用场景，着力于未来进一步研究

和施行对人工智能的监管工作。从当前情况来看，制定综合性的人工智能法律尚缺成熟的时机，而刑法自身有很大的包容性和发展性，可通过修正刑法及出台立法、司法解释的方式来应对人工智能时代的治理需求。而从动态维度看，建议采用“渐进式”和“前瞻式”相结合的模式，一方面视人工智能风险问题为一种客观现象，立法主要通过修正现行法律、适当增加条款的形式来进行；另一方面应考虑未来一段时间内犯罪行为、侵害法益、规制范围、主体责任等的变化，预留给司法解释和相关法律法规一定的空间。

总的来说，诚然，我们不能否认人工智能的发展影响了现有的法律关系，带来了一系列崭新的法律问题，使立法、执法、守法等方面都受到了挑战。但倘若就此高呼立刻制定一部人工智能法，构建特定化、专门化的法律体系，未免有些为时过早了。从目前人工智能的技术水平、应用场景及产业现状来看，还远远达不到人们所担忧的失控程度，一概要求以新立法应对新犯罪的做法并不妥当。[⑨]基于此，对于已经或可能出现的安全问题，一部分可以通过梳理和分析我国现有的相关立法规范，修正现行规定或增加新的司法解释将其纳入；另一部分则可以通过构想和设计未来相关的立法规范，从国际治理和国内治理两个角度制定新的立法条款而加以规制。

（二）梳理和分析我国现有的相关立法规范

与数据问题相关的立法规范，本文是从广义范畴理解的：一是针对数据的保护，主要涉及系统中存储、使用、处理、传输的数据。目前我国刑事立法对此尚欠缺保护力度，仅体现于刑法第二百八十五条第二款，将非法获取数据的行为入罪化。而伴随更多人工智能应用场景的出现，视觉识别领域的人体面部数据、语音识别领域的声纹信息等将成为犯罪分子使用、篡改、传输的对象，其他恶意处理数据的行为也会愈来愈多，刑法有必要考虑在规制非法获取行为之外增设其他条款，突显对于数据保护的重视程度。二是针对个人信息的保护，更强调隐私的内容。此处先说明一下不同国家的用语区别：涉及隐私的相关立法中，我国采用“信息”一词，美国则常用“隐私”一词，而欧盟自1995年《个人数据保护指令》出台起就使用“数据”一词涵盖包括隐私在内的内容，不同的用语对应着不同的保护倾向和保护范围。关于个人信息保护的立法规范，下面进行具体分析。

随着大数据时代的到来，人们日常网络活动会产生海量的个人信息，与此同时，偷取、滥用、恶意披露等行为也越来越严重。对此，我国相关立法活动正在不断推进。不仅刑法确立了侵犯公民个人信息罪，《网络安全法》也将网络信息安全设为了独立的一章；再如2012年《关于加强网络信息保护的决定》的颁布确立了个人信息保护的相关原则，规范了网站和企事业单位收集、使用、处理公民个人信息的行为；2013年出台的《电信和互联网用户个人信息保护规定》完善了电信和互联网行业的个人信息保护相关制度；此外，在地方立法上，贵州省于2016年颁布了《贵州省大数据发展应用促进条例》，作为我国第一个与大数据相关的地方性法规，它体现了政府在规范数据使用及保护个人信息方面的极度重视和积极作为。

由上述规定可知，个人信息保护已成为我国立法进程中极为重要的部分，而在人工智能日渐发展的时代，确保信息安全则有了更多的要求和更高的标准。为避免数据风险成为限制人工智能发展的短板，当前的个人信息保护法律框架，尤其是刑事立法的规定，有必要且需尽快依据人工智能的特殊性对难以规制的部分进行调整。

我国刑法对个人信息的保护可追溯至“刑七”颁布时，它增加了第二百五十三条之一，将主体限定为国家机关或重要单位的工作人员，个人信息来源于单位履行职责或提供服务，重在打击出售、非法提供及获取的行为。该次修正是个人信息立法保护事业的重大起步，但并未详述罪状，未明确界定个人信息的内涵，使得在司法应用中出现不少认定争议，而且所规制的主体范围较小，并不能覆盖产业领域常见的信息泄露问题。而《刑法修正案（九）》（下文简称“刑九”）完善并正式确立了“侵犯公民个人信息罪”。因现实中实名制的施行导致个人信息数量暴增，而买卖个人信息又在利益驱使下逐渐形成完整的犯罪链，对此刑法将主体从特殊扩大为一般，使更多处于犯罪中上游的行为入

罪化。同时，这次修正明确了本罪的客观要件，将非法获取的任何公民个人信息均纳入定罪范围内，解决了原条文中的“上述信息”是否仅指特殊单位工作所获信息的歧义。而且，“刑九”增加了三至七年量刑档和从重处罚情形，使法定刑的设置更能体现罪责刑相适应原则，至此围绕公民个人信息已构建起基础的刑法保护体系，有效回应了数据相关犯罪产业链发展日益猖獗的问题，而人工智能犯罪所涉行为也可基本纳入此种体系。进一步分析侵犯公民个人信息罪在人工智能时代的适应情况，需着重关注以下问题：

第一，“公民个人信息”这一概念是否需重新界定？2017年公布的司法解释第一条即对公民个人信息做了明确解释。[10]范围除涵盖身份识别信息外，对反映特定自然人活动情况的信息例如行踪轨迹作了特别的提示性规定，体现了专属性、可识别性、价值性的特征，即通过信息可直接或间接识别出唯一对应的主体，且该信息具有刑法保护的法益价值。这一解释顺应了打击犯罪的实践需求，也助益于各单位主体正确判定合规对象的注意范围。[11]结合人工智能技术特征来看，其背景下的公民个人信息特征并无不同，但在内容上可考虑有所扩充，比如可列举更多的记录载体或形式，还可列举一些与人工智能相关的信息形态。《网络安全法》在定义个人信息时提及了“个人生物识别信息”。[12]类似的，在生物特征识别、语音识别、视觉识别等应用场景的推广下，人们日常活动会产生更多样化的身份数据，如人脸图像、指纹、声纹等，这些数据的使用范围和频率甚至可能逐渐超过传统的账号密码、身份证件号码等信息，因此可以考虑在解释个人信息定义时有所体现。

第二，人工智能技术支持下个人信息流转的多向性、频繁性、复杂性，是否会影响犯罪客观行为和刑法保护路径的设定？有学者提出，目前的立法主要是从公民个人信息的“来源”和“去向”两个方面来类型化界定犯罪行为，体现了立足于信息横向流动过程的保护路径，但较少触及信息的纵向使用过程。[13]从现实案例可知，人工智能为个人信息犯罪提供便利条件以及促进个人信息的流转，并非仅局限于一个方向上的各环节，而是遍布于整个网状环境中，这就要求立法需扩展保护路径，将新的侵犯手段纳入客观行为之中。

第三，应用人工智能产品和服务时，被害人承诺是否面临新的情形？个人信息刑法保护中被害人承诺的问题其实就是个人信息主体知情同意权的问题，即主体享有知悉本人信息被收集、使用等权利，和享有除例外情况任何人处理本人信息均需由本人同意的权利。司法解释第三条明确了非法提供公民个人信息的前提就是“未经被收集者同意”，《网络安全法》第四十一条、第四十二条也分别强调了网络运营者收集、使用和提供个人信息时的告知和经同意义务，并在第四十三条进一步规定了个人的要求删除或更正权。而人工智能发展环境下，如何平衡信息主体的知情同意权和运营者在开展业务活动中的数据处理自由？首先，目前知情同意权的对应范围主要是收集、使用和提供个人信息的行为，而随着人工智能领域相关主体的增多，接触个人信息的行为方式也更多样复杂，可能会出现部分新类型的信息处理方式，或原本无权处理的方式可能因应用场景的特殊性而纳入经个人同意即可处理的范围内。其次，知情同意权的保障程度和实现途径方面，将此权利规定于用户条款是当前的常见方式，且大多是注册使用的前提，未赋予用户选择的权利，这样在司法认定中可能存在被害人消极承诺的问题，尤其当越来越多的智能产品和科技化功能进入民众生活，有必要在立法上尽早明确这一点。

（三）构想和设计未来的相关立法规范

在构想刑事立法路径前，首先需确定立法的原则和态度：其一，以安全与发展两大价值为导向，实现刑法对人工智能涉及的各个法益和技术产业创新发展的均衡保护。公权力介入产业领域时应恪守技术中立原则，对于所发生风险未超过社会的相当性、控制度和容忍度的，不应过度干预和打压；[14]其二，刑法应坚持谦抑性，在科技带来的新奇和狂热中保持理性，严守“最后一道防线”的地位。同时，风险社会下刑法也应具有前瞻性，关注潜在的风险问题，积极探索适当的早期介入模式，如预备行为的入罪化；其三，注意刑法的延续适用和后续衔接问题。前文已阐明，目前人工智能犯罪并不会从根本上动摇现行刑法的罪名体系，许多行为仍能

被现行刑法和司法解释直接或经修正后进行评价，因此不必盲目追求大刀阔斧地改革立法。在补充立法过程中，既要设计紧紧围绕人工智能犯罪特点的刑事条款，又要考虑与其他刑事规范的衔接问题，尤其是后续出台的网络违法犯罪防治法、相关司法解释等；其四，将技术规则与法律规则相结合，体现科技立法的特色。当前一些通用的技术规则已出现于人工智能技术应用发展的过程中，在立法确定主体、对象、手段、数额等内容时应当予以参考。而且人工智能具有系统性、协同性，其犯罪具有集对象、手段工具、空间于一体的特征，因此不能仅从片面分析其犯罪行为和侵害结果，要尽可能带着整体和联系的思维进行立法构想。有了上述这些原则的指引，下一步设想具体的立法内容，可作如下展开。

从各国保护数据的实践来看，主要有几种路径：欧盟通过制定详备的法案如《通用数据保护条例》（General Data Protection Regulation，以下简称GDPR）对数据施以保护，在国际范围内产生重大影响；美国主要从保护数据产业发展的角度出发，并未单独规定个人数据权利，而是侧重于市场规范，在发生数据侵权时多由联邦贸易委员会处理；德国创设了联邦数据保护专员制度，特别规定在相关公共机构和个人组织中需任命数据保护官；而我国当前的立法多着力于规范网络信息服务和信息处理行为，强调对个人信息的保护，而对数据权利等具体问题的规范还存在较多空白。[15]我国借鉴各国保护思路，应用到人工智能相关立法上，需加快建立关于个人信息和数据的合理保护机制。对于个人信息的保护，考虑重新界定其范围，区分敏感信息和其他特殊信息类型并分别规定保护的形式和限制性内容，进一步明确知情同意原则等，这在上一节已具体说明，此处不再赘述。而对于一般数据的保护，现行刑法“静态有余，动态不足”，[16]所设罪名主要围绕系统、个人信息的安全性，唯一与数据相关的罪名“非法获取计算机信息系统数据罪”实则也是落脚于对系统功能和通信安全的保护。对此，需加强立法层面对数据保护的重视程度，比如明确相关主体的安全管理义务，规范其获取、使用、处理数据的行为，建立可执行的且可用于不同应用场景的标准流程；评估数据权利方面是否有必要增加被遗忘权、可携带权等特殊权利；欧盟法律事务委员会建议，在结合人工智能特性制定数据相关政策时，应进一步完善默认保护隐私、知情同意、加密等概念的标准。[17]且必要时可设立专门性的管理机构或研究机构，如英国政府正与阿兰·图灵研究所合作建立一个旨在研究数据科学的“数据伦理委员会”，以加强数据使用的审查；另外，法律可采取延伸式保护，依靠人工智能技术提升保护能力，如尝试算法代理人模式，可根据不同情形设定不同的数据使用权限，同时管理个人同意和拒绝分享的信息，使控制和使用两者共存，便于数据所涉主体间的意思联络。[18]

四、产业领域防控数据风险的策略建议

（一）开展人工智能安全科学发展的基本工作部署

在产业领域，开展人工智能安全科学发展的基本工作部署，首要的任务就是树立关于人工智能的正确认识，采取辩证且积极的态度对待这一新兴技术。当前，一部分人过分夸大人工智能的好处，在技术应用时有些盲目；与此相反，也有一部分人对科技抱有无知的畏惧，害怕人工智能最终会失控而反对产业继续发展，这都是不可取的。我们知道，人工智能是一把双刃剑，既能推动经济社会发展，也会带来各种类型的安全风险。因此，各企业在开展业务时首先应理解人工智能的双重性质，兼顾两大价值导向——一是发展，二是安全。

以发展为价值导向，企业应当详细了解外部市场情况，包括行业内同类企业的技术布局、价值链各环节的商业应用案例、未来可预估的发展前景，同时评估基于自身业务的应用机会和组织、技术、资源等形成的核心竞争力，并在此基础上制定具体可行的发展规划。而在执行过程中，激励内部创新，善用外部资本，能够加速人工智能发展。

以安全为价值导向，企业应当将安全理念贯穿于设计、研发、生产、销售等各个环节，对风险采取“事先预防为主，事后补救为辅”的规避路径，并制定一系列相关原则与行动指南。如微软公司以“六项原则”为任何部署人工智能驱动解决方案的

核心，其内容是隐私和安全原则、透明度原则、公平性原则、可靠性原则、包容性原则和问责制原则。[19]再如有报告提出各行业人工智能发展基础的评分体系，具体包括：一是组织机构基础，包含人工智能战略视野与方向、人才与技术能力、组织灵活性和驱变力等；二是数据、工作流与技术基础，包含可获取的数据量、数据储存流程成熟度、数据整洁度、数据有良好的记录与说明文档、工作流自动化程度、对人工智能友好的IT系统等；三是实施与应用基础，包含应用场景清晰度、人工智能运用准备的成熟度、解决方案服务机构合作情况等。[20]评分体系中的部分要素也可为企业构建人工智能安全框架提供方向。

在参考上述内容的同时，尤其应当注意责任制和开放性。企业应建立责任文化，通过开展员工教育，强调道德声明和标准的重要性、人工智能良性与恶性用途的区分，以及从技术研发到应用过程的风险防范意识。尤其是研究人员，他们对促进技术的有益应用和防止有害使用负有责任，可通过与决策者合作、提供专业知识、评估项目安全性等来践行责任。而开放性则体现于对技术信息的适度公开、安全制度和经验教训的共享，以及在合理可行的情况下提高公众的参与度、施行公众决策。此外，产业领域还需特别强调巨头企业的引领作用，无论是发展层面的布局还是安全层面的保障，这些企业有着明显的优势地位，应当承担更多社会责任，并为中小企业提供参考范式。

（二）完善对人工智能数据的管理控制

确保人工智能相关数据的保密性、完整性及可用性，要求企业制定明确的数据战略，其内容可包含以下几个部分：

第一，重视数据资产价值。构建服务于企业发展的数据收集、整合与运用的生态系统，在业务部门的支持下，采取程序化的方式为数据资产提供架构，同时优化数据的采集、聚合、使用与后续更新，并保持数据的准确、一致与安全。[21]此外，从发展的视角来看，保证数字系统存在开放的接口也是必要的，以便未来灵活整合技术来适应人工智能的发展。第二，加强个人隐私保护。相关人员在收集、使用、处理数据的过程中，需要采取适当的技术手段，防止个人隐私信息的泄露、篡改及损毁，如进行数据加密，通过采用“同态加密”的手段，有效地提升信息传输的安全性；[22]也可以进行数据的失真处理，这类数据仍可维持一些性质不更改，且使得攻击者难以从中还原出真实原始数据；还可以采取限制发布的方式，通过“数据匿名化”技术，折中考虑披露风险和确保精度之间的平衡，而后有选择地发布敏感数据及相关信息，此时也可能涉及对于数据重要性的等级划分。此外，可以建立认证管理体系，包括用户身份、应用身份、设备身份等内容的认证，提高隐私数据泄露的门槛。第三，建立数据自我保护体系，通过程序设定实时监测数据的整洁度和安全性，使其不受外界环境的干扰，保证核心数据在任何时间、任何地点都有能力抵抗各种形式的恶意攻击。

对于数据安全的保护，还需特别注意GDPR的要求，有报告指出GDPR对人工智能的九大影响，包括：第6条对重新利用数据作了禁止规定，这对人工智能开发和使用的创新性有一定程度的限制；第17条规定了数据的被遗忘权，这同人工智能机器学习需要海量数据相悖，可能会影响人工智能的准确性，甚至产生一定破坏结果；第20条规定了数据的可携权，会推动消费者与人工智能企业分享有关数据，从而刺激企业间的竞争；第44条至第50条要求的数据本地化，会大大提高人工智能处理数据的成本；此外，GDPR对数据去识别标准缺乏明确的规定，也会抑制企业使用未识别的数据，缩小其可能的合法用途范围。[23]由此可见，GDPR对人工智能产业的数据处理行为影响重大，凡在GDPR管辖范围之内的相关企业都应尽早采取措施提高业务合规性，以适应上述影响。

五、结语

与网络犯罪日益加剧时刑法的积极回应一样，伴随人工智能而来的新的刑事风险，也呼唤刑法适时适当地介入，面对刑事归责理论和刑罚体系的动摇、危害行为及因果关系的认定困难，及时进行调整和转型，以实现新时代的刑法价值。[24]当然，人工智能数据风险治理具有复杂性、广泛社会性，并

非仅依靠法律手段就能解决，还需要学术界打通学科壁垒、产业界加强技术沟通、政府和公众积极开展对话，将政府、企事业单位、专家、技术人员、公众等纳入多元一体的人工智能安全治理参与机制中，建立包括法律规制、伦理规制、行业规制、自律规制在内的多元互动的风险规制体系，㉕从而营造人工智能创新发展所需的良好生态环境。

注释

① 中国电子学会，新一代人工智能发展白皮书2017。

② Brundage M, Avin S, clark J.“The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”[EB/OL].（2018-02-20）https://www.repdsitory,cam.ac.uk/hendle/1810/275332.

③ 卡鲁姆·蔡斯，人工智能革命：超级智能时代的人类命运，张尧然译，机械工业出版社，2017年，第118页。

④ 参见360互联网安全中心编，《互联网安全的40个智慧洞见（2017）》，人民邮电出版社，2018年版，第296页。

⑤⑥ 人工智能学会，《中国模式识别白皮书》（2015），第1页，第26页。

⑦ 参见360互联网安全中心编，《互联网安全的40个智慧洞见（2017）》，人民邮电出版社，2018年版，第296页。

⑧ 王熠珏，《我国网络犯罪治理的回溯与反思》，载《石河子大学学报（哲学社会科学版）》2019年第1期，第55页。

⑨ 张明楷，《网络时代的刑事立法》，载《法律科学（西北政法大学学报）》2017年第3期，第80页。

⑩《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第一条：“刑法第二百五十三条之一规定的‘公民个人信息’，是指以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息，包括姓名、身份证件号码、通信通讯联系方式、住址、账号密码、财产状况、行踪轨迹等。”

⑪ 吴沈括、石嘉黎，《网络安全法背景下个人信息刑事司法保护》，载《检察日报》2017年7月11日第3版。

⑫《网络安全法》第七十六条第五款规定：“个人信息，是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等。”

⑬ 赵秉志、詹奇玮，《现实挑战与未来展望：关于人工智能的刑法学思考》，载《暨南学报（哲学社会科学版）》2019年第1期，第105页。

⑭ 张全印，《机遇与挑战：人工智能带来的刑事风险与刑法应对》，载《中国刑警学院学报》2018年第6期，第19页。

⑮ 姜野，《算法的规训与规训的算法：人工智能时代算法的法律规制》，载《河北法学》2018年第12期，第149-150页。

⑯ 赵秉志、詹奇玮，《现实挑战与未来展望：关于人工智能的刑法学思考》，载《暨南学报（哲学社会科学版）》2019年第1期，第106页。

⑰ 曹建峰，《10大建议！看欧盟如何预测AI立法新趋势》，载《机器人产业》2017年第2期，第19页。

⑱ 参见中国电子技术标准化研究院，《人工智能标准化白皮书》（2018）。

⑲ See Ralph Haupter,“AI is built on trust”<https://news.microsoft.com/apac/2018/04/17/ai-is-built-on-trust/> accessed 5 November 2018.

⑳㉑ 参见中国人工智能学会、罗兰贝格，《中国人工智能创新应用白皮书》（2017），第18页，第35页。

㉒“同态加密”允许对密文进行特定的代数运算，得到的仍是加密的结果，这与对明文进行同样的运算再将结果加密一样。360互联网安全中心编：《互联网安全的40个智慧洞见（2017）》，人民邮电出版社，2018年版，第110页。

㉓ See Information Technology and Innovation Foundation,“The Impact of the EU’s New Data Protection Regulation on AI”.

㉔ 刘志伟，《改革开放40年中国刑法学研究的成就与展望》，《湖南科技大学学报（社会科学版）》2018年第8期，第91页。

㉕ 马长山，《人工智能的社会风险及其法律规制》，《法律科学（西北政法大学学报）》2018年第6期，第53页。

Study on Risk Response of Artificial Intelligence from the Perspective of Data Security

Wu Shenkuo　Shi Jiali

Abstract: At present, artificial intelligence is accelerating development and is widely used in all aspects of social life. At the same time, its risks in the field of network security have become increasingly prominent, especially data security issues, which need to be highly concerned. This paper starts from the analysis of the infringement behaviors related to data elements, selects the typical application scenario of pattern recognition to elaborate, and discusses the current rules and the construction of the rules in the legislative level. In the industrial field, it is recommended that relevant enterprises strengthen risk prevention and control, so that each entity could work together to ensure that artificial intelligence is safe, reliable, and controllable.

Keywords: artificial intelligence; data security; criminal legislation; risk prevention and control