湖南农业大学学报(社会科学版)　2024 年 4 月　第 25 卷第 2 期
Journal of Hunan Agricultural University(Social Sciences), Apr. 2024, 25(2):79–87

DOI: 10.13331/j.cnki.jhau(ss).2024.02.009

算法解释权适用的不同区分及其实践路径

——基于公私两域划分的视角

刘佳明 [1,2,3]

（1.浙江财经大学 法学院， 浙江 杭州 310018；2.浙江大学 光华法学院，浙江 杭州 310008；
3.浙江湖州市南浔区人民法院，浙江 湖州 313009）

摘　要：《个人信息保护法》第 24 条规定，个人认为自动化决策对其权益造成重大影响的，有权要求个人信息处理者予以说明，此条被认为是对算法解释权的明文确定。但是，对于权益主体的确定、权益影响的程度以及解释内容的要求，第 24 条并未进一步做出具体规定，考虑到公私领域对算法解释规范对象和规范目标的不同，在私人领域中，算法解释的权益主体应为确定型主体，其形式权益遭受影响才能提出算法解释的主张，而信息处理者只需对算法进行事物性解释，在公共领域中，算法解释的主体应为确定或相关型主体，无论是形式还是实质权益遭受影响都可以提出算法解释的主张，而信息处理者需对算法进行事物或事理性解释，此举有利于促进技术发展与权利保障、公共利益与个人利益之间的价值平衡。

关　键　词：算法；算法解释权；个人信息保护法

中图分类号：D922　　**文献标志码**：A　　**文章编号**：1009–2013(2024)02–0079–09

The different distinction and practice path of the application of algorithm interpretation right: from the perspective of public-private domain division

LIU Jiaming[1,2,3]

(1.Law School, Zhejiang University of Finance and Economics, Hangzhou 310018, China;
2.Guanghua Law School, Zhejiang University, Hangzhou 310008, China;
3.Nanxun District People's Court, Zhejiang, Huzhou 313009, China)

Abstract: According to Article 24 of the "Personal Information Protection Law", if an individual believes that automated decision-making has made a significant impact on his rights and benefits, he has the right to request an explanation from the information processor. This article is considered to be a clear text for the algorithm decision explanation rules. For an algorithm decision explanation requirement, there must be a definite of rights and benefits in the subject, the reason must be that it has a significant impact on its rights and benefits, and there must be an objective interpretation standard in the content. However, Article 24 does not provide a further specific provisions for the determination of the subject of rights and benefits. Considering the differences in the normative objects and objectives of algorithm interpretation in the public and private domain, in the private domain, the subject of the interest of algorithm interpretation should be a deterministic subject, and only when its formal interest is affected can it propose the proposition of algorithm interpretation, while the information processor only needs to interpret the algorithm in a physical way. In the public domain, the subject of algorithm interpretation should be a deterministic or relevant subject. In addition, no matter its form or substantive rights and interests are affected, the claim of algorithmic interpretation can be put forward, and information processors need to explain the algorithm by things or things, so as to promote the value balance between technological development and rights protection, public interests and personal interests.

Keywords: algorithm; algorithm decision explanation right; personal information protection law

收稿日期：2023 – 11 – 16

基金项目：杭州市哲学社会科学常规项目（Z23JC079）；国家社会科学基金一般项目（21BFX126）

作者简介：刘佳明（1993—），男，江西吉安人，讲师，博士，浙江大学光华法学院博士后（在站），南浔区人民法院研究室副主任（挂职），主要研究方向为数字法学。

《中华人民共和国个人信息保护法》（以下简

称《个人信息保护法》）第 24 条规定已对算法解释权进行了确认①，但是，该项权利内容不完整、适用范围不明确、行使程序未建立等问题[1]，导致实践中难以进行适用。主要原因在于公私法的规范对象与规范目标不同，使算法解释权在不同应用场景中有不同的表现[2]。而目前学界关于算法解释权的研究主要集中在两个方面：一是肯定算法解释权的价值并提出相应的制度建构路径，包括内部解释和证明解释，或“全局解释”“局部解释”，或“透明化”“事后归因”等；二是对算法的可解释性进行讨论，认为算法可解释性不仅需要凭借不同的解释路径透视算法黑箱，还需要满足受众对算法解释的不同需求，从而构建一套合理的制度。此外，也有学者从公私领域出发，认为算法解释权的适用不要求决定“仅”由自动决策算法做出，从而将该权利的适用范围扩张到决策支持型算法、分组型算法与总结型算法等类型，全面规范公私领域中的算法应用②。考虑到公私领域规范价值和规范目标上的不同，在私人领域，算法解释的权益主体应当是确定型主体，其形式权益遭受影响才能提出算法解释的主张，而信息处理者对算法只需进行事物性解释即可；在公共领域，算法解释的权益主体包括确定型主体和相关型主体，无论是形式还是实质权益遭受影响都可以要求算法解释的主张，而信息处理者必须进行事物性和事理性解释。但是在生成式人工智能算法中，算法技术隐层复杂的程度对大模型可解释性构成很大限制，因而无论是公共领域还是私人领域，要完全实现对算法解释十分困难。对此，除了需要建立分层分级的监管应对策略外，还可以建立“可控制”的监管能力与规则框架，从而实现对大模型算法的责任规制。

一、公私两域权益主体的不同认定

（一）公私两域算法解释主体不同认定的缘由

在《个人信息保护法》第 24 条关于算法解释规则的规定中，对于遭受自动化算法决策影响，而有权要求信息处理者对自动化算法决策进行解释的“个人”，是权益直接相关人还是间接相关人，这在法律实践中会存在诸多疑问，并会对信息处理者算法解释规则的履行造成实际困难。对此，往往涉及对算法与受影响权益主体之间因果关系的判断。而因果关系的判断，是每一个案件都必须面对的问题，对此，立法常常只是提出一个因果关系的要求，至于应当如何判断，其往往沉默不语[3]。从形式上，因果关系可以分为两种类型，一为相关型因果关系，即满足两个条件可构成相关性的原因，算法须是权益遭受影响的客观因素之一，不能完全排除权益遭受算法影响发生的客观可能。二为确定型因果关系，即算法须是权益遭受影响的确定原因，这种原因有明确的法律关系和以事实行为为依据。无论是相关型因果关系还是确定型因果关系，都将算法视为权益遭受重大影响的客观因素，只不过在对算法与权益影响之间关系的判断程度上有所差别而已。

因此，从算法与权益主体遭受影响之间因果关系的角度分析，算法与权益影响之间或存在相关因果关系，或存在确定因果关系。前者是指信息处理者的算法对权益主体的影响是相关的，后者则是指算法对权益主体的影响是确定的。对此，根据算法与受影响权益之间的关系，可以将算法解释的权益主体分为确定型权益主体和相关型权益主体。确定型权益主体，其权益影响之所以可以确定，主要是因为受到算法影响的权益主体和信息处理者之间的法律关系是确定的，以至于可以直接根据相关法律行为来确定二者之间的关系。而相关型权益主体，其权益之所以是相关的，主要是因为受到信息处理者自动化算法影响的主体是不确定的，以至于不能直接根据确定的法律关系来决定信息处理者与个人之间的算法解释法律关系，这时候就需要考虑算法与受影响主体的权益之间是否存在相关因果关系，来判断或确认该主体是否有权要求信息处理者对算法规则进行解释。因果关系是一种带有很强确定性的事实，法律中的因果关系分析应该以这种确定性的事实为基础，法律分析的理性在一定程度上来自对事实的认同[4]。因此，对受影响的权益与算法之间相当因果关系的判断，必须依赖于特定的事实，即信息处理者所使用的自动化算法须是影响个人权益的必要条件以及增加影响权益的客观可能。

（二）公私两域算法解释不同主体认定的路径

其一，在私人领域，个人之所以有权要求信息处理者对其使用的自动化算法进行解释说明，往往

是因为权益主体与信息处理者之间基于特定的法律行为而产生解释关系，或契约，或侵权。此时，算法解释的权益主体为确定型权益主体。以杭州“人脸识别”一案为例[5]，在该案件中，市民甲与公园管理者之间是基于契约而产生相关法律关系，那么，针对公园管理者使用特征识别算法来进行入园身份验证的方式，作为契约当事人一方的市民甲就有权要求公园管理者对其进行解释说明，而其他非契约关系的市民就无权要求公园管理者对算法进行解释。当然，在私人领域中，信息处理者与权益主体之间并非一定是基于明确的法律行为而产生的算法解释法律关系，也有可能基于一些确定的事实行为而产生。例如不当得利、无因管理，但由于此类法律关系的主体相对比较确定，仍然可以将算法解释的主体归为确定型权益主体。此时，信息处理者虽然与权益主体没有发生直接的法律关系，但仍需要承担相应的算法解释义务。

其二，在公共领域，行政法律关系的主体除了行政主体和行政相对人之外，还有行政相关人（或第三人）。由于行政行为的作用效果不是单一的作用于相对人，而是对相对人具有双重乃至多重的影响与效果，它不仅指向其本来意图作用的相对人，同时也辐射至其他公民或组织，其共同构成行政行为作用的对象[6]。在这里，行政相关人与行政主体之间并不存在直接或确定的法律关系，只是基于行政主体的行为对相对人的权利义务也构成了实质性影响的客观事实，故而将其纳入行政法律关系的保护范围。在公共领域，有些算法不仅会对特定主体的权益产生实质性影响，同时也会对不特定主体的权益产生实质性影响。因而，该算法与受影响权益主体之间或存在着确定的法律关系，或存在着相关的法律关系，而在算法解释的权益主体上，就包含了确定型和相关型权益主体两种类型。无论怎样，这些主体都应当有权要求信息处理者对算法进行解释。

此外，由于公共领域和私人领域立场的不同，其权益主体的认定有所不同。除了信息处理者与信息所有者之间在法律关系的确定程度和权益是否遭受实质性影响上存在差别之外，还有一个重要的缘由，那就是私人领域和公共领域在规范目标上有所区别。其背后所依靠的理念是，在公共领域，必须严格规范和限制公权力利用算法进行自动化决策，以保障公民的基本权益，哪怕受算法影响的是相关型主体；而在私人领域，法律除了必须保障公民权益之外，同时还必须尊重当事人之间的意思自治，并鼓励技术创新。因为对算法的解释不仅是意思自治原则的必然推论和合同信息不对称的矫正工具，同时也是对合同风险的一种合理分配[7]。它是在信息处理者违背当事人意思自治的情形下，通过国家力量对其进行干预和矫正的一种手段。因此，在私人领域中，对权益主体的认定比较严格，只有在信息处理者与权益主体之间具有明确法律关系之时方可进行。而在公共领域，对权益主体的认定则比较宽松，除了具有明确法律关系之外，与算法具有相关关系的，同样有权要求信息处理者进行算法解释。因为“与一般决策相比，那些可能对个人自由或权利产生重大影响的决策，如法庭裁决、雇佣或解雇、入学、升职、金融服务供给等，它们需要更高水平的问责制”[8]。

二、公私两域算法解释的不同来源

（一）公私两域算法解释不同来源的缘由

信息处理者之所以被要求对算法进行专门解释，除了因为算法技术或具有专业性和复杂性，或该要求来源于法律的明文规定之外，还往往是因为信息处理者没有遵循法律规定的要求，利用算法对个人权益造成重大影响。而信息处理者之所以被要求对算法进行解释，主要原因也正在于算法技术具有专业性和复杂性，这种专业性和复杂性一旦被不法利用，并被视为侵害个人权益的正当理由，对个人权益的保护而言往往是具有重大影响的。因此，当信息处理者利用算法对主体的权益造成重大影响之时，个人才有权要求信息处理者对算法的基础规则进行解释说明，而权益主体才能了解其权益何以遭受影响，并以此获取相应的申诉或救济途径。

算法对个人权益造成重大影响中的“影响”，既可以是形式上的影响，也可以是实质上的影响，具体到个人的权益之上，则可以分为形式上的权益和实质上的权益两种类型。从概念上理解，权益是指因享有某种权利而包含的利益，它是权利的主要内容，从“权益”向“权利”的转化需要类型化的利益，并经立法程序方可成为一个专门的权利类型[9]。

例如个人信息保护法中规定的知情同意、算法解释、数据修改、数据管理等新兴(新型)信息权益就属于形式上的权益，这些属于通过成文法的形式已经成为一种新的权益保护类型。形式上的权益是立法者认为值得保护的利益，并以此作为受保护的举止规范之对象，它是建立在实在法基础之上，强调的是“规则之权益”。而实质上的权益则除了包含形式权益的要求之外，还将一些特定的东西、能力和状态视为权益，例如能够维护个体的尊严和促进个体的自由发展。它体现的是超实在法的要求，是通过超实在法的规范或者权益保护政策的抽象理念和内容，来确定实在法的规范内容，以及为立法者确定实在法在目的上所要保护的范围[10]。人们相信，一旦形式上的权益受到算法的影响，可以要求信息处理者对算法进行解释，但如果是实质上的权益受到算法的影响，是否还可以要求信息处理者承担算法解释义务，这就需要根据公共领域和私人领域对信息处理者算法解释的来源要求不同而有所区别。

主要原因在于，相较于私权力主体而言，公权力主体承担着保障公民基本权利的义务，同时也受到更多的限制。而我国宪法以限制和规范公权力作为主要目标，这种限制既包括了形式上的限制，同时也包括实质上的限制，即宪法不仅规定了公权力的行使要符合法定的形式要件，同时也规定了公权力的行使必须要符合宪法确立的某些目标和对象，而不能有所偏离。例如，宪法中规定公民享有的一些基本权利条款就构成了对公权力的实质限制，这意味着在公共领域，公权力主体使用算法决策侵害公民基本权利的行为是遭到禁止的，故而应当承担相应的算法解释义务。

《个人信息保护法》第 24 条规定并未对个人权益造成重大影响中的“权益”是一种形式上的权益还是一种实质上的权益进行严格区分，因而，根据算法是发生在公共领域还是私人领域，来划分信息处理者的算法解释要求没有存在的必要。相反，正是因为法条没有对个人权益是一种形式上的权益还是实质上的权益进行专门规定，无论算法是发生在公共领域还是私人领域，信息处理者都需要进行相应的算法解释，这反而有利于对信息所有者的权益进行保障。毋庸置疑，无论算法是对个人的形式上权益还是实质上的权益造成重大影响，信息处理者都需要进行相应的算法解释，对于保障个人权益而言，无疑是最好的，而对于立法而言，也是最为便捷的。但是，一项法律的规定还必须考虑它的实际运行效果，以及它对社会整体运行效率产生的影响。而当前《个人信息保护法》第 24 条关于算法解释规则的规定并没有区分公私领域中关于不同权益要求，只是笼统地要求对个人权益造成重大影响，并未对影响的是何种权益进一步详细说明，难以在实践中直接适用。

（二）公私两域算法解释不同来源的要求

其一，在私人领域，算法对个人形式权益造成重大影响的，即可要求算法主体对算法决策进行解释。主要原因在于：一方面，在私人领域中，当个人认为算法对其实质性权益造成重大影响，而要求信息处理者对算法进行解释时，一旦信息处理者以实质性权益无法律明文的规定或不在法律的保护范围之内为由而予以拒绝，那么个人的实质性权益保障将很难实现，这无疑会危害信息处理者算法解释的有效实现；另一方面，作为法律规范的主要内容，虽然法律义务是被设定的，设定法律义务的过程也通常被视为通过立法确立义务规则的过程[1,11]。因而，法律义务具有“应为性”和“必为性”的特征，前者强调法律义务是作为被人们期待的行为模式而存在的，是“应当为”或“应当不为”的规范性行为模式[12]。而后者则强调个体承担义务意味着其自由要受到限制，其必须根据法律义务规则作或不作一定的行为，否则将要承担相应的法律责任[2,11]。无论法律义务是“应为性”的还是“必为性”的，作为立法设定的法律义务还必须考虑义务主体的实际履行能力，因为法不能强人所难。对于立法者而言，一个超过义务主体承受能力的法律义务安排是不正当，或者至少说是不正义的。因此，作为立法设定的法律义务还必须具有“能为性”，它必须考虑现实性，要求的是信息处理者能为之的行为，作为算法解释的义务应当同样如此。例如在私人领域中，当个人认为算法对其实质性权益造成重大影响，而要求信息处理者对算法进行解释，如果要求其承担超出实在法规范范围的算法解释义务，显然不具有现实性，这同样会危害信息处理者算法解释的有效实现。

其二，在公共领域，算法对个人形式或者实质性权益造成重大影响的，都可要求算法主体对算法决策进行解释。理由在于，虽然有关个人实质权益的认定在公共领域中也会存在困难，但算法在公共领域的运用对公民权益造成影响范围往往比较大，它并不能成为信息处理者逃避算法解释的正当理由。同样以杭州“人脸识别案”为例③，如果信息处理者没有获得个人许可而使用其人脸作为身份识别的验证方式，或者虽然获得了许可，但却对个人权益造成重大影响，那么，个人就有权要求信息处理者对使用算法做出的决策予以解释说明，而信息处理者也就需要承担相应的算法解释义务。在这里，即使当时的法律并未明文规定人脸是个人权益的重要保护内容，当信息处理者利用算法进行决策时，个人认为算法决策内容对其实质性权益造成重大影响的，也应当有权要求信息处理者予以说明，并有权拒绝信息处理者仅通过自动化算法的方式做出决策。这既是公民寻求自我保护的需要，也是维护个体尊严的基本要求。因此，在公共领域，无论算法是对个人形式权益还是实质性权益造成重大影响的，都可要求算法主体对算法决策进行解释。

其三，对于影响个人实质权益的算法是发生在公共领域还是私人领域，有时认定会存在困难，有关对信息处理者算法解释的要求并不一定能够获得法律支持。在哈贝马斯看来，公共领域是一个被描述为关于内容、观点的交往网络，在那里，交往被以一种特定方式加以过滤和综合，从而成为根据特定议题集束而成的公共意见或舆论[13]。按照他的理想模型，公共领域是价值中立的、单一的、超越生活世界的[14]。但是，随着社会和市场的逐渐平台化，越来越多的社会交往行为在网络空间产生，国家与社会、公共空间与私人空间以及网络世界与现实世界之间的界限逐渐被打破，传统建立于政治国家与市民社会二元基础上公私二分法，已越来越难以适应互联网时代社会发展的新变化。因此，认定影响个人实质权益的算法是发生在公共领域还是私人领域，主要还是取决于该算法决策是具有公共性还是私人性。如果是私主体所使用的，那么就发生在私人领域，其解释义务的理由仅限于形式上的权益受到重大影响，但如果私主体使用的算法会对公共利益造成影响，那么也应当被视为发生在公共领域，其算法解释的理由就不仅包括自动化算法对个人形式上的权益受到重大影响，还包括实质上的权益受到重大影响。

三、公私两域算法解释内容的不同要求

（一）公私两域算法解释何以有不同要求

个人因自动化算法对其权益造成重大影响的，有权要求信息处理者予以说明。但是，信息处理者对算法的解释或是具体的，或是抽象的。因此，根据算法解释内容是具体或抽象上来看，可以将信息处理者对算法解释分为两种类型，一是事物性解释，即要求信息处理者对算法这一具体事物进行解释说明，通过对自动化算法的数据收集、算法程序的设计、网络架构的部署以及算法的性质、特点、用途等做客观而准确的解释，使个人能够对受到算法影响的运行逻辑有基本的认识和了解。二是事理性解释，即将算法决策的成因、关系、原理等说清楚，使个人不仅知其然，还能知其所以然。不管是事物性解释还是事理性解释都要求信息处理者对算法进行真实的介绍，使个人能够对其有基本的了解和认识，只不过后者对算法解释的要求更高，也更为严格。而信息处理者对算法是进行事物性解释还是事理性解释，取决于该算法是发生在公共领域还是私人领域。

其一，人们之所以需要对算法进行解释，主要是因为算法技术具有专业性和复杂性。当前，算法被广泛应用于社会各个领域，其核心主要是算法决策[15]。得益于机器学习技术的研发和改进，算法在一定程度上还具备相应的自主决策能力，计算机通过借助以深度神经网络为架构的最大似然算法反复进行网络推理，每一次网络推理都试图捕捉给定数据集内数据之间的关系，并反复推理学习数据特征的层次结构，从而赋予算法一种处理高层次抽象的技术特征[16]。此类算法被广泛运用于现实生活中的各种商业场景和公共用途。就算法技术的专业性而言，算法是为实现特定行为而设计的，它们被放置在虚拟的轨道之上，必须按照给定的流程运行，这其中包括构成算法的技术、工具和方法，它们有自己特殊的词汇、语法，以及编译单词、句子和文本的规则[17]。就此而言，普通公众无法理解算法是如何产生的，也不能应用传统的自然语言来描述这

种关系，此时，当公众因权益遭受侵害而要求算法技术使用者对算法进行解释的权利理应得到保护。

其二，算法具有非中立性也成为需要对其进行解释的重要缘由。当算法是可解释之时，一个外部的观察者可以理解算法依靠什么因素来进行决策，以及它给予每个因素多大的权重，这些不同因素的权重直接决定着算法决策的输出效果，并对算法产生重要影响。然而，可解释性又是有代价的，因为一个可解释的模型并不比一个黑箱模型更简单，也通常不太准确[18]。但要求信息处理者对算法进行解释仍是有必要的。因为算法解释不仅是确保法律规制信息处理者运用算法的逻辑基础，还是维系社会公众对利用自动化算法进行决策的信任基础。从概念上来理解，信任被视为一种特殊的社会关系，而社会关系本身又从属于特殊的规则系统[19]。信任具有强化人们现有认识和简化复杂未来之功能，而算法信任正是通过法律、行业规范、技术伦理等制度创建可信的治理环境使得公众增强对算法技术的掌控感和影响力，从而使公众对算法的运行形成稳定预期和信赖[20]。因此，对于一个非中立性的算法而言，如果某种包含复杂逻辑的算法应用可能对个人的合法权益造成风险，设定算法解释的法律义务或明确算法解释请求权，使受影响的主体有机会了解算法设计的原理即为必要[21]。它不仅是实现算法规制的客观需求，还是增强公众对算法信任的重要基础。

其三，与个人相比，作为掌控自动化算法决策的信息处理者凭借技术和资源上的优势，能够利用算法对个人形成较强的控制力和影响力，因而必须要求其承担相应的算法解释义务。具体表现为，在私人领域，智能算法以其精准的解析和控制能力逐渐成为各互联网平台实现商业利益的工具，基于算法工具的权力操纵与滥用会给个人带来潜在威胁[22]。而在公共领域，随着算法和大数据技术日渐成熟，公共管理部门运用人工智能开展信用信息评价、构建人员信息识别系统、辅助行政执法等已成为常态，作为人工智能核心的算法在社会治理和政务服务中发挥着愈发重要的作用[23]。因此，无论自动化算法是发生在公共领域还是私人领域，信息处理者在利用算法对个人行为进行控制，并对其权益造成重大影响之时，都负有不可推卸的算法解释义务，只不过在对算法是进行事物性解释还是事理性解释上所承担的义务有所不同。

（二）公私两域算法解释内容不同要求之判断

一方面，在私人领域，考虑到自动化算法影响的是个人的形式权益，故而信息处理者对自动化算法决策进行解释的范围应当仅限于事物性解释。但这种解释和说明还必须考虑到个人的可理解性，即要求个人对算法的运行逻辑有基本的了解、认识。“理解”是解释学的一个基本概念，作为一种非常普遍的认知活动，不管是人对人的理解，还是人对某事物的理解，理解的核心对象、本质内容都是“信息”，所以，从某种意义上说，理解本身就是信息理解[24]。而算法解释的可理解性本质上也是对算法决策信息的基本理解。对于一个算法的解释是否能够为权益主体所理解，直接决定了信息处理者对算法解释的实际效果。因此，为了保证权益主体的可理解性，对算法进行事物性解释可以概括为以下几个方面要求：第一，信息处理者要以清晰明了、简明扼要的方式向权益主体呈现算法信息；第二，据以解释的算法信息要满足权益主体的理解需求；第三，对算法的解释有助于权益主体进行判断和选择。这些因素都在客观上决定了信息处理者对算法的解释能否为个人所理解，是衡量信息处理者对算法进行事物性解释的重要标准。

虽然可理解性是对算法进行事物性解释的一个重要特征，但对算法解释的可理解并不等于算法一定能够被理解，或者能够被所有人理解。因为可理解性是一个具有很强的主观性概念，以至于令人怀疑是否还会存在一个完全客观的算法解释。事实上，完全可理解的算法解释不是存在的，或者能够被所有人理解的算法解释是不可能实现的。因为基于人们地域、年龄和教育上的差异，每个人对算法知识的掌握和理解有所不同，尤其是对于一种具有高度专业性和复杂性的算法技术而言，由于所有决策都是由算法程序产生，而算法程序通常由代码编译而成，在一个由代码组成的算法程序中，算法语言与人们通常使用的自然语言有很大区别，因而也就决定了它不能被所有的人理解。即使可以，也并不意味着，就能对算法的运行逻辑进行完全理解，因为，算法的运行还通常包含着随机过程，不同变量之间往往存在复杂的、不可预测的交互作用效应[25]。这

些都在不同程度上限制着信息处理者算法解释义务的实现。事实上，也正是基于算法技术带来的爆炸式知识增长，从而加深了人们对算法相关知识的理解，而理解上的差异和缺失阻隔了人们对算法技术的普及和有效利用。于是，在算法和个人之间存在着一道巨大的数字鸿沟，这恰恰成为要求信息处理者承担算法解释义务的重要缘由。

另一方面，在公共领域，由于自动化算法影响的范围不只是个人的形式权益，同时还包括了个人的实质权益。因而，信息处理者对算法的解释也就不能仅限于事物性层面，还应当包括事理性层面。“事理性”所指向的是，对算法的解释应当以可理解和可接受的方式呈现。对算法决策内容进行事理性解释，不仅涉及算法的一些专业性知识，同时还会包括算法与个人权益遭受影响之间的因果联系，使个人了解算法决策何以产生，何以对其权益造成影响，以及算法决策的正当性需要智能机器人运算和分析过程中的程序透明和因果透明得以支持，并能经得起从结果向数据输入的倒推[26]。因此，在公共领域，信息处理者对算法的解释不仅要考虑到个人的可理解性程度，还需要考虑达到个人的可接受性。可接受性是指人们的内心世界对外在世界的某种因素或者成分的认同、认可、吸纳甚至尊崇而形成的心理状态或者倾向[27]。它是一种理性的接受，是在论证者和受众之间相互影响的辩证条件下予以理解[28]。由于事理性解释是一个认知程度上的问题，信息处理者对算法进行事理性解释最终取决于个人的理解并认同，问题是，一旦个人因为不接受风险而拒绝接受和认同该解释，如果拒绝的人多了，该算法解释就有可能无效，或者该规定将难以执行，因此，对于算法的事理性解释除了需要具备事物性解释的相关要求之外，还需要有更高的标准。具体而言，则表现为：对算法决策运行逻辑的解释要清晰、合理、完整，每个环节都能言之有据，不违背逻辑的基本要求，以及在算法解释程序上能够妥善处理公众可能遭致的质疑。而通过程序来实现可接受性原则不仅在理论上已经有所建树，并且在现实的制度中有所表现。程序作为评价算法解释可接受性的标准在于，能够迫使信息处理者对依靠算法进行决策的解释公开化，从而为公众对基于这些理由而做出决策的有效性和合理性质疑[29]。因此，算法解释技术最核心的贡献是“提供了一套还原模型开发、机器决策的过程和结果的事实材料”。这些模型开发的信息若能得到完整记录，将为规范研究进一步的价值判断、制度设计和责任判定提供事实依据[30]。

但是，在公共领域中，信息处理者对算法进行事理性解释，也不完全是绝对的，还需要根据算法在公共领域中的性质来进行考量。因为同样是在公共领域，对于算法的事理性解释，作为信息处理者的公主体和私主体也有所不同。就前文所言，公共主体使用算法会对公民的权益造成重大影响，这些影响既包括形式上的影响，同时还包括实质上的影响。因此，公共主体在将自动化算法引入社会管理领域之时，必须表现得更为谨慎小心，尤其是在涉及公民实质性权益的领域，需要通过建立健全事前的风险评估以及事后的监督和问责机制，来确保算法的健康、平稳运行。私人主体能够凭借技术优势和资本优势参与公共管理活动，成为一些“准公共主体”，因而也应当承担要求其对算法进行事理性解释的要求，但考虑到私人主体毕竟不是真正意义上的公共主体，如果对其苛以过高的算法解释要求，将不堪重负，这将会抑制私主体参与公共事务管理和服务的积极性。

将信息处理者对算法解释的内容区分为事物性解释和事理性解释，这在法律上不会有太大问题。因为个人信息保护法并未对信息处理者算法解释做进一步的明确规定，一旦涉及算法解释的实践问题，信息处理者也就可以对算法决策的内容做出变通解释，而个人也不会对这是一个事物性解释还是事理性解释存在异议，即使存在，也需要由相关部门对其做出进一步规定或解释说明。但是，无论是在公共领域还是私人领域，对于算法的解释，如果全都要求对其成因、关系、原理进行解释，则可能会面临技术和工作上的困难，因而往往是不必要的，或者是低效率的，最终导致被排斥在算法解释的要求之外。因此，尽管个人信息保护法并未对信息处理者对自动化算法决策的解释是一种事理性解释还是事物性解释做出明确规定，但基于现实因素的考量，根据自动化算法发生在公共领域和私人领域的不同，而对其解释内容进行区分也就显得非常有必要。但是，在生成式人工智能中，算法技术

隐层复杂的程度对大模型可解释性构成很大限制，使得依靠目前技术手段难以对“输入层”如何到“输出层”进行有效解释，在解释技术尚未成熟前，应当限制对生成式大模型算法在公共领域的应用。

此外，不同领域的“解释”要求，对于不同的主体还有不同的社会意义。在私人领域，作为自动化决策的控制者还有强烈动机拒绝披露更多的算法信息，以保守自身商业秘密或避免损害其他当事人的隐私等权利与自由[31]。因此，对私人主体课以算法解释义务之时，必须控制在必要的限度范围之内，否则，这种解释要求对私人主体而言不仅是一种很大的负担，而且其还面临着巨大的风险，即私人信息处理者的商业秘密以及其他当事人的个人隐私可能遭到泄露。虽然在公共领域中，信息处理者在对算法进行解释之时，也会面临国家秘密和公民隐私遭遇泄露的重大风险。但是在公共领域中，对于国家秘密或个人隐私的保护，并不当然构成排斥对算法进行事理性解释要求的正当理由，这就还需要信息处理者依法或依裁量来决定算法解释的范围，并对算法解释进行严格程序限制[32]。通过在遵循法定的基础之上，依照比例原则、正当程序原则以及对决策内容所涉及的公共利益、算法不公开对公共利益的影响程度、信息敏感程度等因素进行综合考量后来做出算法解释的决定，从而保证公民在充分知情的情形下主张自己的权利[33]。

结语

《个人信息保护法》第 24 条关于算法解释规则的专门规定，为信息所有者的权益保障提供了法律依据，具有重要的现实意义。但是该规定由于过于笼统、模糊，以至于信息所有者和信息处理者能够基于不同的利益立场出发对其做出不同的解释，势必会影响该规则的实际执行效果。因此，有必要从公共领域和私人领域的角度出发，根据韦伯的理念类型法对不同领域下的算法解释权的行使进行梳理，从而减少甚至避免在算法解释的实践中对国家利益、商业秘密、个人隐私以及技术创新造成的不利影响，以实现信息所有者和信息处理者之间利益的平衡④。而针对生成式人工智能大模型应用进行解释在当下仍缺乏相应的深度研究，如何建立大模型算法应用的监管体系和治理规则，亟待理论和实务界做进一步研究。

注释：

① “个人信息保护法”第 24 条使用的是说明，而不是解释。根据《新华词典》中的定义，“说明”有两层含义：一为解释清楚、说明原因，二为解释，例如产品使用说明；而“解释”也有两层含义：一为分析阐明，二为说明含义、原因、理由。说明和解释虽然具有很大的相似性，被视为一组近义词，但是二者含义仍有所不同。由于学界通常将此条视为算法解释，本文为避免混淆，也将说明和解释视为同一概念进行使用。

② 解释权的证成和构建路径，包括张凌寒：《商业自动化决策的算法解释权研究》，载《法律科学（西北政法大学学报）》2018 年第 4 期，李婕：《公共服务领域算法解释权之构建》，载《求是学刊》2021 年第 3 期，何新新：《算法解释权的证成与限定》，载《大连理工大学学报（社会科学版）》2023 年第 2 期；算法可解释性的讨论，包括王海燕：《算法可解释性的价值及其法治化路径》，载《重庆社会科学》2024 年第 1 期，周翔：《算法可解释性：一个技术概念的规范价值》，载《比较法研究》2023 年第 3 期；此外，从公私场景出发对算法解释权讨论的学者有，林洹民：《个人信息保护法》中的算法解释权：兼顾公私场景的区分规范策略》，载《法治研究》2022 年第 5 期。

③ 在此之前，对于个人的人脸信息是一种身份权还是财产权并没作出规定，也缺乏相应的法律规范对其予以专门保护。因此，在后来的《民法典》以及《个人信息保护法》制定过程中，人脸作为公民的实质性权益，已经在法律中得以明确确立。

④ 表格如下：

	权益主体	解释来源	解释要求
私人领域	确定型权益主体	形式权益遭受影响	事物性解释
公共领域	确定/相关型权益主体	形式/实质权益遭受影响	事物/事理性解释

参考文献：

[1] 李天助. 算法解释权检视——对属性、构造及本土化的再思[J]. 贵州师范大学学报(社会科学版)，2021(5)：151-160.

[2] 林洹民. 《个人信息保护法》中的算法解释权：兼顾公私场景的区分规范策略[J]. 法治研究，2022(5)：48-58.

[3] 叶金强. 相当因果关系理论的展开[J]. 中国法学，2008(1)：34-51.

[4] 孙晓东，李炜. 法律因果关系分析[J]. 法学杂志，2009，30(10)：28-31.

[5] 朱健勇. 中国人脸识别第一案：杭州一动物园被起诉[N]. 北京青年报，2019-11-04(A7).

[6] 肖金明，张宇飞. 关于行政相关人问题[J]. 政治与法律，2005(6)：62-68.

[7] 张凌寒. 商业自动化决策的算法解释权研究[J]. 法律科学(西北政法大学学报)，2018，36(3)：65-74.
[8] 解正山. 算法决策规制——以算法“解释权”为中心[J]. 现代法学，2020，42(1)：179-193.
[9] 张琴. 智慧城市治理中个人信息的权益解析和权利保护[J]. 南京社会科学，2020(11)：93-98，107.
[10] 马长山. 智慧社会背景下的“第四代人权”及其保障[J]. 中国法学，2019(5)：5-24.
[11] 齐崇文. 法律义务设定原理研究[J]. 东岳论丛，2014，35(10)：175-183.
[12] 钱大军. 法律与法律义务关联研究[J]. 法制与社会发展，2010，16(1)：51-58.
[13] 哈贝马斯. 在事实与规范之间——关于法律和民主法治国的商谈理论[M]. 上海：生活·读书·新知三联书店，2003.
[14] 王晓升. “公共领域”概念辨析[J]. 吉林大学社会科学学报，2011，51(4)：22-30，159.
[15] BRAUNEIS R，Ellen P. Algorithmic transparency for the smart city[J]. Yale journal of law and technology，2018，103(20)：113-114.
[16] 本吉奥. 人工智能中的深度结构学习[M]. 俞凯，吴科，译. 北京：机械工业出版社，2017.
[17] ALEXEY V. Law as a programming language[J]. Review of central and east European law，2012，115(37)：118.
[18] MICHAEL L. Machine learning，automated suspicion algorithms，and the fourth amendment[J]. University of Pennsylvania law review，2016，164(4)：887.
[19] 卢曼. 信任：一个社会复杂性的简化机制[M]. 瞿铁鹏，李强，译. 上海：上海人民出版社，2005.
[20] 张欣. 从算法危机到算法信任：算法治理的多元方案和本土化路径[J]. 华东政法大学学报，2019，22(6)：17-30.
[21] 苏宇. 算法规制的谱系[J]. 中国法学，2020(3)：165-184.
[22] 段鹏. 平台经济时代算法权力问题的治理路径探索[J]. 东岳论丛，2020，41(5)：110-117，192.
[23] 孙清白. 人工智能算法的“公共性”应用风险及其二元规制[J]. 行政法学研究，2020(4)：58-66.
[24] 陈红星. 基于信息理解的数字鸿沟[J]. 图书馆学研究，2008(2)：96-98.
[25] CARY C，DAVID L. Regulating by robot：administrative decision making in the machine-learning era[J]. Georgetown law journal，2017，105(17)：1224.
[26] 唐林垚. 人工智能时代的算法规制：责任分层与义务合规[J]. 现代法学，2020，42(1)：194-209.
[27] 孙光宁. 法律论证中的可接受性原则[J]. 法律方法，2009，8(00)：371-383.
[28] 杨猛宗. 法律论证可接受性的内涵与类型之探析[J]. 湖北大学学报(哲学社会科学版)，2017，44(2)：108-114.
[29] ANDREW D，SOLON B. The intuitive appeal of explainable machines[J]. Fordham law review，2018，114(3)：1122.
[30] 周翔. 算法可解释性：一个技术概念的规范研究价值[J]. 比较法研究，2023(3)：188-200.
[31] 解正山. 算法决策规制——以算法“解释权”为中心[J]. 现代法学，2020，42(1)：179-193.
[32] 刘佳明. 公共决策算法的程序规范——以立法性算法为例[J]. 财经法学，2022(3)：16-28.
[33] 刘佳明. 人工智能立法的运用及其规制[J]. 湖南农业大学学报(社会科学版)，2021，22(1)：56-62.

责任编辑：黄燕妮