Open Journal of Legal Science 法学, 2022, 10(4), 444-454
Published Online July 2022 in Hans. http://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2022.104058

Hans汉斯

我国商业自动化决策的算法解释权研究

张炜鑫

华东政法大学，上海

收稿日期：2022年5月17日；录用日期：2022年6月2日；发布日期：2022年7月12日

摘　要

《中华人民共和国个人信息保护法》的颁布为我国提供了商业自动化决策算法解释权的法律依据，但在适用范围、解释标准和适用时间阶段等方面仍存在不明确之处。商业自动化决策算法解释权作为一项由被决策者行使的权利有其存在的必要性。根据个人信息自决权理论以及算法规制理论，商业自动化算法解释权的制度设计既要能够保护被决策者的个人利益，又不能过度限制决策者对数据的开发和利用。就优化的具体方案而言，被决策者的算法解释权应当存在于商业自动化决策做出的全过程，并且以重大影响为标准，区分情形地设定商业自动化决策算法解释权的适用范围、解释标准以及权利实现的方式。

关键词

商业自动化决策，算法解释权，个人信息自决权，算法规制

Research on Algorithmic Interpretation Right of Business Automation Decision in China

Weixin Zhang

East China University of Political Science and Law, Shanghai

Received: May 17th, 2022; accepted: Jun. 2nd, 2022; published: Jul. 12th, 2022

Abstract

The introduction of the Personal Information Protection Law of the People's Republic of China provides a legal basis for the interpretation right of commercial automation decision algorithms in China, but there are still some uncertainties in the scope of application, interpretation standard and application time stage. As a right exercised by the decision-maker, the interpretation right of

文章引用: 张炜鑫. 我国商业自动化决策的算法解释权研究[J]. 法学, 2022, 10(4): 444-454.
DOI: 10.12677/ojls.2022.104058

commercial automatic decision algorithm has its necessity. According to the theory of personal information self-determination and algorithm regulation, the institutional design of business automation algorithm interpretation rights should not only protect the personal interests of decision-makers, but also not excessively restrict the development and utilization of data by decision-makers. As far as the specific optimization scheme is concerned, the algorithm interpretation right of the decision-maker should exist in the whole process of the commercial automation decision-making, and the application scope, interpretation standard and realization mode of the algorithm interpretation right of the commercial automation decision making should be set differently according to the major impact standard.

Keywords

Business Automation Decision, Algorithm Interpretation Right, The Right to Personal Information Self-Determination, The Algorithm of Regulation

1. 引言

随着计算机技术的不断发展，商业自动化决策算法技术也迎来了发展的高峰期，越来越多地影响着人们的生活。从短视频的个性化推送，到广告的精准投送；从外卖平台的自动化定价，到搜索引擎的自动排序；从个人信用的自动化评分，到金融产品的自动交易，自动化决策算法正在广泛地影响着人们的生活，我们仿佛正逐步进入一个规模空前的“评分社会”或者说等级化的“排序社会”[1]。然而，这些自动化决策在为人们带来便利的同时，也引发了一系列社会问题。大数据杀熟的歧视性定价、[1]被困在系统里的外卖小哥、[2]搜索引擎对被遗忘权的忽视[3]等等，这些现象的出现引发了人们对自动化决策领域中更深层次的问题——算法问题的探讨。面对着自动化决策越来越广泛的普及，普通大众对自动化决策算法的不信任也越发强烈，为打破自动化决策算法与被决策者之间的信任鸿沟，自动化决策算法解释权(以下简称算法解释权)应运而生。

算法解释权是自动化决策领域当中一项重要的权利，自信息技术方兴未艾之际即被学者关注和研究。欧盟更是通过《欧盟通用数据技术条例》直接将这一权利写入法律条文当中。[4]我国近期颁布的《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》)也将这一权利写入条文当中，成为了我国算法解释权法律规制的基础。然而，算法解释权对于我国来说仍然是一个全新的法律概念，它既包含信息与编程技术等复杂的科学技术，又凸显了个人信息权利和信息产业发展之间的内在张力，还表露出诸如解释的内容、解释的程度、解释的时间等具体的适用性问题。因此，有必要结合当下我国学术界对算法解释权的学术研究，以《个人信息保护法》的颁布为背景，研究算法解释权在我国的适用。

[1]大数据杀熟杀的是互联网经济的未来》，载《光明网》2021 年 7 月 16 日，https://m.gmw.cn/2021-07/16/content_1302410071.htm。
[2]《外卖骑手，困在系统里》，载《人物》，2020 年 9 月 8 日，https://baijiahao.baidu.com/s?id=1677231323622016633&wfr=spider&for=pc。
[3]任甲玉诉北京百度网讯科技有限公司名誉权纠纷案，北京市第一中级人民法院(2015)一中民终字第 09558 号。
[4]参见《欧盟通用数据保护条例》第十五条第一款：“数据主体应当有权从控制者那里得知，关于其的个人数据是否正在被处理，如果正在被处理的话，其应当有权访问个人数据和获知如下信息：……(h)存在自动化的决策，包括第 22(1)和(4)条所规定的数据分析，以及在此类情形下，对于相关逻辑、包括此类处理对于数据主体的预期后果的有效信息。”

2. 自动化决策与算法解释权

自动化决策算法解释权这一概念包含“自动化决策”和“算法解释权”两个组成部分，对这两部分应当分别来理解。

2.1. 算法支配下的自动化决策

何谓算法，目前并无统一定义。严格地说，它也不是法律上的概念。从技术角度看，可将算法理解为一种数学结构以及这一结构转化而来的计算机程序、或将其视作为实现特定任务的技术应用。它具有形式化特征，包括大规模收集个人或环境数据，且借由数据分析从而揭示特征与结果之间的相关性[2]。而所谓自动化决策，我国《个人信息保护法》第七十三第二项规定：“自动化决策，是指通过计算机程序自动分析、评估个人的行为习惯、兴趣爱好或者经济、健康、信用状况等，并进行决策的活动。”从其定义可知，自动化决策以数据收集、分析和决策做出的自动性为核心特征，而这一核心特征则完全依赖于计算机技术，更准确的说是算法技术。因此可以说，自动化决策就是拥有相关技术能力的企业、平台，依托其底层算法的计算能力而作出的自动化商业决策。

从上述定义可见，算法与自动化决策具有十分紧密的关系，可以说算法是商业决策之所以能够实现自动化的核心和基础，自动化决策智能化、科学化、合理化的提升都离不开算法技术的支持。甚至可以说，当下社会生活中涌现的，诸如个性化推荐、自动化定价、自动化评价等自动化决策，从根源上讲即是相应算法的外化表现。

2.2. 回应自动化决策的算法解释权

在数据时代，数据控制者基于海量的数据，利用算法技术实现了其对社会资源的配置，数据控制者正在逐渐掌握一种新兴的“算法权力”[3]。这种权力往往与算法黑箱、算法歧视等问题相伴而来，为自动化决策的发展盖上了一层阴云。为对抗这种复杂而又难以理解的技术权力，被决策者的算法解释权应运而生。“算法解释权”就是通过增强算法决策的透明度进而提高其可理解性，以此来实现对“算法权力”滥用的适度制约的权利[4]。

算法解释权按其解释方式来看可以分为系统解释与个案解释。系统解释指个人可以要求自动化决策者对算法的系统功能和运行逻辑进行解释，强调对自动化决策者解释自动决策系统的逻辑、意义、预期后果和一般功能。个案解释则指个人可以要求对个案决策进行解释，强调向被决策者解释个案结果是如何得出的，自动化决策考虑了那些因素以及其权重是多少。算法解释权按解释的时间和阶段来看可以分为事前解释与事后解释。根据事前解释的要求，被决策者可以要求自动化决策者在算法决策前就说明其算法规则。而根据事后解释，则个体只能在算法决策作出之后提出请求[5]。

3. 我国立法中的算法解释权

2021 年 8 月 20 日，十三届全国人大常委会第三十次会议表决通过了《个人信息保护法》。作为一部个人信息保护领域的法律，它规定了被决策者的算法解释权。[5]这为被决策者维护自身权益，对抗不公正的自动化决策提供了法律依据。但是不得不承认，这部法律所设定的自动化决策的算法解释权尚且存在诸多问题。

[5]《个人信息保护法》第七条：“处理个人信息应当遵循公开、透明原则，公开个人信息处理规则，明示处理的目的、方式和范围。”第二十四条第三款：“通过自动化决策方式作出对个人权益有重大影响的决定，个人有权要求个人信息处理者予以说明，并有权拒绝个人信息处理者仅通过自动化决策的方式作出决定。”第四十八条：“个人有权要求个人信息处理者对其个人信息处理规则进行解释说明。”

3.1. 适用范围不明确

有观点认为我国仅包含以《个人信息保护法》二十四条第三款为法律依据的限定算法解释权，即算法解释权仅被规定在一定限定范围之内，个人只有在个案中满足了相关限定条件以后，才可以行使这一权利；另一种观点则认为我国既包含二十四条第三款所规定的限定算法解释权，也可以从第七条以及第四十八条，被决策者对决策者利用其个人数据的解释说明权当中，推导出自动化决策的算法解释权，因为究其根本，自动化决策也是个人信息的一种处理方式，因此可以被囊括在这之中。

虽然上述两种观点存在争议，但是它们对于限定算法解释权的存在是具有统一认识的。然而，仔细分析我国《个人信息保护法》第二十四条第三款的规定可以发现，这一限定算法解释权也存在适用范围上的不明确。该条文规定，只有当通过自动化的方式作出的商业决策，有可能会对个人被决策者产生“重大影响”时，被决策者才有权利要求决策者进行解释。

由此首先会产生第一个问题，即何为“自动化决策”，它是否意味着决策的“完全自动化”。《个人信息保护法》第七十三条第二项解释了自动化决策的一般定义，但却并没有回答自动化决策是否必须完全排除人工干预。如果某项商业决策的做出既有计算机的自动化参与，又有一定程度的人工参与，那么经过该程序的决策是否就是这里所说的“自动化决策”。换句话说，是否只要有人工参与，而不论这种参与是否必要或是否起决定性作用，均一概不再认定是“自动化决策”？这一问题关系到自动化算法解释权的适用范围问题，需要进一步明确。

另一个问题是，“重大影响”这一不确定法律概念应当如何理解。有学者认为，判定“重大影响”应结合当事人的具体情况，如拒绝批准贷款对经济条件较差的人可谓重大影响，对相对经济条件较好的人则可能不构成重大影响，除此之外，学者还认为评价类自动决策具有相比其他类型算法更为重要的地位，因此应当认定为具有“重大影响”[6]。也有学者既认为重大影响的判断应当结合当事人的主观感受来判断[7]。由此可见，对重大影响的认知尚且存在诸多差异。

3.2. 解释标准不明确

《个人信息保护法》第二十四条第三款规定，在特定情形下，被决策者有权要求决策者进行解释，但是却并未指出该种解释应当达到何种程度。除此之外，第七条与第四十八条指出个人信息处理者应当公布其处理个人信息的规则，但是应当解释到何种程度亦未明确。换句话说，《个人信息保护法》虽然赋予了被决策者算法解释权，但是却并没有明确决策者需要解释到何种程度才算合格，这将会给算法解释权的行使带来极大的困难。

这种困难首先来自于被决策者的维权难。在自动化决策的场景之下，被决策者面对的往往是具有极强经济实力和科技实力的企业、平台。通常情况下，被决策者并不具备与之对抗的手段，而算法解释权对于被决策者来说，算得上是一种难得的维权工具，因此对被决策者来说意义重大。然而，自动化决策的运行逻辑往往隐藏在屏幕背后的“算法黑箱”当中，被决策者根本无从获知，只能依靠决策者的解释和说明加以理解。在此背景之下，《个人信息保护法》的算法解释权标准不明，给了决策者逃避解释义务的空隙，将会直接影响被决策者解释权的实现，《个人信息保护法》所确立的算法解释权也恐将成为没有牙齿的老虎。

与此同时，算法解释权标准不明也会给决策者带来合规压力。出于对良好经营状态的维护，企业、平台也并非总是主动充当“恶人”形象，很多自动化决策者也想履行法律规定的解释说明义务，实现企业经营的“合规化”。但如何向被决策者透露算法而不会危及自身的商业秘密，如何将复杂的算法理论解释给知识背景各异的普通消费者，这对于平台企业来说也是一个很难解决的问题。对于决策者来说，是笼统的介绍自动化决策算法的运行逻辑，还是详尽地列出其底层算法；是从整体上进行说明，还是就

个案给出运算的逻辑，这些都莫衷一是。在算法技术急速发展的今天，某一个超大型平台的自动化算法往往是由几十、上百位工程师联合编制的，其复杂程度已经超出常人的想象。更何况，在自动化决策场域当中，算法程序已经具备了一定的自主学习能力，可以通过对数据的收集和分析自主调整自身运行的方式和逻辑[8]，在这样的背景之下，如何准确地向被决策者解释算法将成为一个极其困难的合规难题。

3.3. 适用时间不确定

《个人信息保护法》并没有明确规定算法解释权行使的具体时间。有学者主张应当在自动化决策之前行使算法解释权，即算法的事前解释权。其理由为：在自动化决策的场域之下，个人的人格权极易被不合理的算法所侵害，而在网络时代，侵害人格权的损害后果具有不可逆性，损害一旦发生，就难以恢复原状[9]，所以给予那些可能被自动化决策影响的人以请求解释说明的权利，使得他们在事前即被充分告知这些自动化决策工具潜在的利害关系，便成为了算法解释权的应有功能。也有学者认为应当区分事前告知与事后解释之间的区别，算法解释权是自动化决策中信息不对称的有效纠偏工具，其作用在于使信息从信息优势方，向信息劣势方流动，从而达到双方信息状态的衡平，因此事后解释才是算法解释权的核心。在此基础上，学者认为可以将事后解释进一步分为事后的具体解释和事后的更新解释[10]。还有学者认为，单纯的事前解释往往局限于一般性解释，单纯的事后解释不能有效应对算法风险，也将会使自动化决策的选择权形同虚设，为解决这一困境应当将事后解释和事前解释相结合起来适用，化解风险和选择困境，随后将事后解释作为一种救济途径，为后续维权提供机会[11]。

学术界的争论也从侧面表明了我国算法解释权实施时间的不确定性。《个人信息保护法》第二十四条第三款规定：“通过自动化决策方式作出对个人权益有重大影响的决定，个人有权要求个人信息处理者予以说明。”从条文本身似乎很难判断算法解释的事前事后。一方面，有些具有重大影响的自动化决策在作出之后就会产生相应的后果，因此有必要事前加以解释说明，以确保被决策者充分知悉其中的风险并同意作出决定，比如医疗健康、金融信贷等关系个人重大利益的领域。但是另一方面，有些自动化决策只有在做出之后才会被发觉是否是具有重大影响的，此时被决策者又需要从事救济的视角具有算法解释权，比如搜索引擎的自动排序、大数据杀熟等。由此可见，算法解释权适用的时间阶段是一个复杂问题，需要进一步梳理。

4. 算法解释权的理论分析

为解决上述诸多问题，本文认为有必要对自动化算法解释权的相关理论问题进行梳理和分析，并以理论分析为基础，利用我国《个人信息保护法》现有的法律规范，开拓出我国自动化决策算法解释权的优化路径。

4.1. 算法解释权的必要性

自动化算法解释权从其产生之日起就伴随着非常激烈的必要性争论。有学者指出自动化决策算法解释权本身的权利属性、权利构造均存在问题，就算法解释权的本土化来说也存在着可行性、可欲性、必要性的困境，因此学者建议应当取消自动化决策的算法解释权代之以有效的政府监管[12]。有学者认为，关于是否存在算法解释权在欧盟尚且存在争议，因此学者认为并不存在算法解释权，而仅存在于算法解释说明义务[13]。还有学者认为，算法解释权并非国际立法之通例，其可行性尚无定论，难以实现被期望的目的，重要性和必要性也不足，况且在既有法律体系下，其旨在实现的功能和目标，均可为其他法律规制手段所实现，无须另行创设算法解释权[14]。但本文认为，自动决策算法解释权有其存在的必要性，理由如下。

4.1.1. 自动化决策内部存在固有风险

自动化决策自其产生之日起就被打上了客观、准确、高效、公正的标签，然而现实情况却是，由人工编写的算法作为自动化决策的核心，其自身亦存在着诸多固有风险。

第一，自动化决策算法缺乏自我纠错能力。自动化决策算法较之于人类决策，不仅缺乏自我纠错能力，而且更具系统性影响[15]。算法决策看似不受人类判断中非理性因素的影响，然而，算法归根结底产生于人类的设计，因此不可避免地会将人类的认知缺陷及感性偏见融入到自身的决策体系当中。因此，自动化决策算法极有可能会忽视数据中的系统性偏见或结构性歧视[16]。算法解释权可以在一定程度上增加算法问题被发现的可能性，及时控制风险，同时也为被决策者行使自动化决策的拒绝权提供了条件[17]。

第二，算法解释权的存在关乎人格尊严。在当下的社会生活中，自动化决策算法的影响力与日俱增，人们对自动化算法发自内心的担忧也逐渐涌现。人们对自动化决策算法的过分依赖，已经逐渐使算法成为了分配社会资源的重要因素。[6] 在算法面前，人类往往会体会到一定程度上的无力感和脆弱感，人的尊严会受到极大挑战。因此，赋予被决策者算法解释权，使其对算法的运行逻辑和工作方式有一定理解有助于缓解这种担忧。

第三，算法解释权是实现算法正义，增进决策者与被决策者之间相互理解的桥梁。如上文所述，基于算法的运行逻辑和自身特征，自动化决策算法往往是缺乏自我纠错机制的，因此从外部为被决策者设置一个算法解释权，一方面可以促进决策者及时发现自动化算法的问题，另一方面也有助于促进决策者与被决策者之间的沟通，增进相互之间的理解，为算法技术手段赋予伦理[18]。从长远来看，这种沟通将会促进企业、平台和消费者之间的信任度，不仅有助于企业、平台的进一步发展，也有助于帮助提高消费者的商品服务体验实现双赢。

4.1.2. 外部监管手段不能替代算法解释权

除了自动化决策算法的内部固有问题以外，外部监管的不足也难以完全替代算法解释权的作用。

第一，自动化决策往往直接指向被决策者，因此被决策者才是决策最直接的受众，是对自动化决策体验最直接、最真实的主体，也是自动化决策结果的承担者，是最直接的利害关系人。因此，赋予他们算法解释权，使其具有相应手段对抗强大的自动化决策主体，对于维护其权利具有十分重要的意义。

第二，政府监管可行性、效率性不高。自动化决策算法是一项带有非常强烈技术色彩的商业模式，面对这种新兴的商业模式，政府部门往往缺乏必要的技术条件，更进一步，作为一种最近几年新兴的商业模式，政府部门对其监管的方式和限度也远尚未形成体系。与此同时，相比于被决策者自发主张的解释权，由政府部门发现问题并开展监管的效率也很难能够保障[18]。

第三，政府监管与算法解释权并不存在先天的矛盾，完全可以相互促进、相互补充。政府监管是一种以国家强制力作为后盾的算法规制路径，具有不可替代的兜底作用。而算法解释权则从被决策者的角度，充当促进自动化决策算法发展的柔性工具。一方面，算法解释权可以为政府监管提供契机和线索；另一方面，政府监管则可以为算法解释权提供国家强制力保障。从这个角度上讲，政府监管和算法解释权本来就不是相互冲突的，两者完全可以在自动化决策算法的规制场域中“互演共生”。

4.2. 算法解释权的理论基础

对于算法解释权的理论基础，学界主要有两种认识。一种观点是算法解释权发轫于对公民个人信息自决权的实现，认为算法解释权系现代社会为保护个人信息免受算法侵害的保护工具。个人信息自决权最早产生于美国和德国的司法案例当中，它是个人有权自我决定、透露或使用其个人数据的权利[19]。个

[6]《中科大利用程序算法“隐形资助”贫困生》，载《搜狐网》，2017 年 7 月 18 日，https://www.sohu.com/a/158145778_298038

人信息自决权理论强调个人对其数据的控制，由于自动化决策依赖于对大量被决策者数据的获取和分析，被决策者对数据的这种控制权就成为了算法解释权的正当性基础。同样的观点也体现在欧盟的立法当中。《欧盟通用数据保护条例》第十五条一款(h)项对算法解释权的规定聚焦于实现数据主体的权利，绕开了直接规制复杂的自动化决策算法，转而赋予被决策者获得算法解释的权利，从而把算法的规制问题放在公民数据保护的法律框架内进行。

另一种观点则认为，算法解释权产生于对自动化决策算法的防御和规制，其核心指向防范算法权力的负面影响对个人的危害，防止算法权力将我们彻底地“客体化”[20]。国内学者赞成这一观点的较多。如有学者认为算法解释权的目的在于提高算法透明性，进而制约算法权力[21]；有学者认为算法解释权的配置旨在规制算法权力，其理论正当性在于这一制度内涵平等、正义、自由目标[22]；还有学者提出算法解释权可以实现对算法权力的约束，增强算法的透明度与可理解性[23]。相比较于个人信息自决权理论，算法规制理论更加强调维护信息社会的客观秩序，认为算法解释权产生并服务于，缓解自动化决策算法在使用过程中出现的信息不对称，以及由此引发的潜在的算法权力滥用风险。

上述两种理论的优劣并非本文所讨论的内容，本文试图论证的是上述两种理论在如何塑造算法解释权问题上具有的一致性。第一，从算法解释权的实施阶段来看，算法解释权应当存在于自动化决策的全过程。从个人信息自决权的理论视角出发，个人对其信息的自决权存在于自动化决策的整个周期，而算法解释权作为个人信息自决权实现的重要工具，亦应当存在于自动化决策的事前和事后阶段；从算法规制的理论视角出发，算法的运行是一个动态的、不断变化的过程，因此应当对算法进行全过程的监督和规制，因而作为算法规制重要手段的算法解释权也就理所当然地要在算法运行的全过程发挥作用。第二，从算法解释权应当实现的目标来看，算法解释权应当尽可能多地破解当下决策者与被决策者之间悬殊的信息不对称。信息不对称是信息经济学的一项重要理论，指某些信息的缔约当事人一方知道而另一方不知道，甚至第三方也无法验证，即使验证，也需要巨大的人力、物力和精力[24]。在当下的实践当中，随着算法决策技术的不断发展，算法本身也愈发复杂和高端，被决策者主体面对的信息不对称远高于以往，而其所享有的权利却没有相应改变，这导致了两者之间权利的不匹配，造成了被决策者主体提供数据，却始终处于数据关系中的劣势地位的现象[4]。因此，不论采取上述哪种理论，都应当通过算法解释权破解决策者与被决策者之间悬殊的信息不对称。

4.3. 算法解释权的实施边界

在如今的社会生活当中，数字经济已经成为了经济发展的重要动力，而如果将数字经济比作是一辆正在运行的汽车，那么数据的流通与利用就应当是这辆汽车所使用的汽油，只有允许大量的个人数据在社会经济生活中广泛、快速地流通，才可以促进数字经济的不断发展。

在这一背景之下，算法解释权显然并非是一项绝对的权利。正因如此，算法解释权的个人信息自决权理论和算法规制理论，均强调了被决策者的算法解释权不应当过分行使。就个人信息自决权理论来说，强调个人的信息权利并不能以牺牲社会效益为代价。德国联邦宪法法院在确立了个人信息自决权的同时亦指出：“信息自决权利并非无限。对于其自身信息，个人并不具有任何绝对或无限的控制”，其理由是“个人是在社会共同体之下发展其个性。因此，即使是个人信息，也同样是社会事实的反映，而并非纯粹地与个人相联”[25]。与此同时，坚持个人信息自决权理论的欧盟《通用数据保护条例》也在第一条第三款明确指出，“不能以保护处理个人数据中的相关自然人为由，对欧盟内部个人数据的自由流动进行限制或禁止”。由此可见，在信息自决权理论场景之下，算法解释权的行使不应当阻碍信息的正常自由流通。而算法规制理论则更加强调创造一个平衡、稳定、公平的算法运行环境。因此，如果因算法解释权的过度行使，导致个人数据的流通被严格限制，并使企业的商业机密面临泄露的风险，那么势必就

会引发算法秩序向另一个极端的倾斜，给决策作出者的生产经营带来极大的困难[26]，而这也不是算法规制理论的理论初衷。综上所述，不论是个人信息自决权理论还是算法规制理论，它们均要求自动化决策的算法解释权不能过度行使。因此，算法解释权的设置，除了要考虑如何有效地维护被决策者主体的权利之外，还应当考虑到平衡决策者与被决策者主体之间的关系，实现两者权益的共同维护。

5. 算法解释权的优化路径

综合上述，新的算法解释权制度必须既能够充分发挥算法解释权的作用，维护被决策者的利益，又能够满足保障和促进网络服务行业的正常发展，体现推动科学发展、社会进步的要求。接下来，本文将以算法解释权适用的时间阶段为线索，以自动化决策的类型为基础，设置自动化决策的事前解释权和事后解释权，并将算法解释权的适用范围以及解释程度穿插其中，试图利用我国《个人信息保护法》已有的法律条文，构建出一个新的、更加明确和更具科学性的算法解释权制度。

5.1. 一分为二的事前解释权

在既有的研究当中，所谓“事前解释权”，一般指的是在决策者准备要对被决策者做出具有重大影响的自动化决策时，被决策者主体可以行使的算法解释权。但本文认为，可以根据自动化决策的不同类型，将事前算法解释权一分为二，有针对性地对不同的自动化决策实施算法解释权。

5.1.1. 囊括所有自动化决策的事前解释权

如上文所述，个人信息自决权理论和算法规制理论均要求决策者充分尊重被决策者主体的算法解释权，然而过度的算法解释权也会给决策者带来难以估量的损失，一个比较合理的处理办法是：所有具备自动化决策能力的决策者，在和被决策者签署用户协议书时，应当由决策者对自动化决策算法进行显著而清晰地解释，以此来实现对被决策者算法解释权的实现。这种囊括所有自动化决策算法的解释权，从理论上符合个人信息自决权理论和算法规制理论，也有《个人信息保护法》第七条和第四十八条作为其法律依据。同时，本文认为根据上述两种理论，应当将“自动化决策”的定义作扩大解释，既应当包含完全由自动化算法做出的决策，同时考虑到人工决策者对算法纯属出结果的盲目信任[27]，也应当将自动化算法具有重要影响的决策纳入到此处的“自动化决策”范围当中来，以保护被决策者的权利。

此外，上文提到的解释说明应当具备如下要素。第一，就解释的程度来说，算法解释应当是清晰准确的，可以不涉及算法的具体表达，而应当侧重于向被决策者介绍自动化决策算法如何收集、整理、分析其数据，并对算法所蕴含的价值判断和整体运行逻辑的说明，但不应当仅限于对定义的描述。这种解释说明方式一方面可以完成被决策者协议当中对数据使用的告知义务；另一方面，也从价值判断的层面上完成了对自动化决策算法的解释。值得注意的是，由于这种事前解释并没有涉及到决策者自动化决策的核心算法逻辑，因此不会对企业造成过分负担。与此同时，对于绝大多数并不具有“重大影响”的自动化决策来说，价值和整体层面上的算法解释也已经足够满足被决策者主体的解释权。第二，就解释的范围来说，应当包含个性化广告推荐、自动化定价等所有类型的自动化决策算法。这是因为不论何种类型的自动化决策，决策者均需要完成用户协议中对被决策者主体的告知义务，因此，在用户协议当中设置算法解释权就可以涵盖所有自动化决策，达成算法解释权的周延实现。第三，就实现途径来说，由于这类算法解释权与决策者的被决策者协议通知书一同出现，因此应当由决策者主动提供。

5.1.2. 具有“重大影响”自动化决策的事前解释权

这类事前解释权应当区别于上文提到的“囊括所有自动化决策的事前解释权”，因为这类解释权仅存在于具有“重大影响”这一限定条件之下。之所以要在上述事前算法解释权之后再单独设置一个事前

算法解释权，是因为这类具有“重大影响”的自动化决策对于被决策者来说，具有十分重要的意义，一旦做出就难以变更，因此需要在作出之前授予被决策者主体一个寻求解释说明的权利，以确保被决策者主体能够完全理解其运行逻辑，并接受自动化决策的结果。

与此同时，上述算法解释权需要满足以下条件。第一，就适用范围而言，该决策的所具有的“重大影响”属性应当能够在“事前”即能够被人知晓并普遍认可。在这里需要进一步区分“重大影响”的情形。本文认为《个人信息保护法》第二十四条第三款中的“重大影响”应当分为两类，一类是按照领域划分，这类自动化决策往往存在于事关被决策者的生命健康、财产安全等具有重要意义的领域当中，而这种领域的独特性是事前便能够知悉的，因此具备事前解释的可能性；另一类具有重大影响的自动化决策是自动化决策的结果有可能会使被决策者遭受损失的自动化决策，如“大数据杀熟”的差异化定价等。本文认为，第二种类型的自动化决策也具有“重大影响”属性。这是因为，我们赋予被决策者算法解释权的原因在于，相比于人类本身做出的决定，我们对机器做出的决定具有天然的不信任。这种不信任在自动化决策算法为我们提供个性化广告、自动化推荐时，不会大幅升高，但是当自动化决策的决策结果可能会使被决策者主体遭受额外的损失时，这种不信任就会急速升高，对于被决策者的影响也就显得更为“重大”。为了缓解这种自动化算法与个人之间的不信任关系，算法解释权也就应声触发了。而相比于特定重要场域的自动化决策结果，后者往往不具备事前显现的属性，一般只有在自动化决策做出以后，被决策者的利益面临减损风险，他们才有可能对该决策的公正性、合理性产生质疑，进而引发行使算法解释权的行使需求。综上，本文认为，具有重大影响的事前算法解释权应当仅适用于那些涉及重要领域、可以被事前知悉其重要属性的自动化决策。第二，就解释程度来说，这种事前算法解释权应当着重于向权利人解释自动化决策所考虑的因素及其大致权重，以及自动化决策作出后的结果及其风险。因为，此时被决策者所关心的更多的是，自动化决策是如何处理自己的重要利益，以及处理之后形成的决策具有怎样的风险。因此，这里并不需要对算法本身的技术样态做过多的解释，着重解释风险即可。第三，就实现途径来说，由于这类自动化决策关涉重大，同时数量也并不是很多，因此应当由决策者在做出决策以前，主动以通知的方式向被决策者传达，被决策者可以明示或默示的方式表达是否通过自动化决策作出决定。

5.2. 区分类型的事后解释权

所谓事后解释权即决策者利用自动化决策算法已经做出了决策，此时被决策者要求决策者对相关自动化决策算法进行解释的权利。事后的算法解释权具有明显的个案属性，相较事前解释权将会花费更多的资源，因此出于平衡决策者与被决策之间利益的考量，算法解释权应当仅存在于具有“重大影响”的决策场域之下。上文已经将自动化决策的“重大影响”做了两种不同类型的区分，与之相应的，事后算法解释权也应当具有两种。接下来本文将以此为基础，分别阐述这两种事后解释权应当如何展开。

5.2.1. 重要领域自动化决策的事后解释权

如前文所述，由于重要领域的自动化决策在事前阶段已经经历过“囊括所有自动化决策的事前解释权”和“具有重大影响的自动化决策的算法解释权”两次解释，因此事后的解释权应当与之前的解释权有所区分。从功能上讲，重要领域中的事前算法解释权应当系属于一种风险的提示，意在确保被决策者能够真实、准确、完整地了解相关自动化决策算法的运行逻辑，使其充分预期自动化决策所面临的结果和风险。与之相反，事后的算法解释权则意在为被决策者提供了一个充分质疑自动化决策的机会，是一项事后的救济权利。

以此为基础，本文认为此类事后算法解释权应当具备如下因素。第一，就适用范围而言，其“重大

影响”性应当是来源于其所属领域，并且已经经历过了两次事前解释。第二，就解释的程度而言，由于相关自动化决策算法在做出决策前已经经历过了两次事前解释，被决策者均已明示或默示的方式表达了同意，因此这类算法解释权系属于对自动化决策的一种救济途径，故决策者应当在自查之后，向被决策者解释自查的逻辑过程，以及算法运行过程中是否出现了运算错误等情况。第三，就实现途径而言，由于这一算法解释权在性质上属于是一种事后救济权利，因此应当由被决策者决定是否行使。

5.2.2. 其他重大影响自动化决策的事后解释权

对于在事后才表现出其“重大影响”属性的自动化决策，被决策者对其享有的解释权从其性质上来看应当也属于一种事后救济权利。

本文认为，这一权利的行使应当具备以下条件。第一，就适用范围而言，只有在自动化决策已经做出以后，该决策有可能会对被决策者利益造成损害的情况下，才可以被认定为是具有“重大影响”，进而被决策者才有适用事后算法解释权的可能。在判断被决策者是否会遭受损害方面，应当坚持主客观相结合的原则来确定。客观上，由于个性化推荐等自动化决策不会对被决策者的权益造成直接损害，所以这类自动化算法应当排除在外；主观上，只要被决策者本人认为除个性化推荐之外的自动化决策对其权益造成了不良影响，该被决策者就可以行使算法解释权。第二，就解释程度而言，由于这类自动化决策往往具有很强的个案属性，被决策者关心的焦点往往集中在决策是否公正、是否存在歧视上。因此对其算法的解释则应当突出该个案决策的公正性、合理性，具体来说决策者可以向被决策者解释说明，在输入了与被决策者情况相似的其他人的数据以后，自动化决策算法将会得出相同或相类似的决策结果，比如在大数据杀熟的情境下，决策者可以解释为何被决策者会得到这一决策结果，并且通过输入与被决策者情况相类似的他人的数据，得出相同或相类似的决策结果来解释说明算法的公正合理性。这种解释程度有其优势：一是，被决策者在这类场景中，往往并不真正关心自动化决策算法整体的运行逻辑和具体的算法设计，他们更多地关注的是，自己的权利是否因为自动化算法的自动决策而受到了不公正的对待，因此专注于个案的算法解释，将该被决策者的结果与其他被决策者的算法结果相对照，往往就可以满足被决策者的解释需求；二是，这类解释的数量虽然可能会相对较多，但这并不会给决策者增加额外的算法解释负担，原因在于这种关注个案的算法解释，其实只需要利用同类对照的办法，向被决策者解释其算法的公正性即可，实质上相当于是多运行了一次算法程序，更不会涉及到商业秘密的泄露问题。第三，就实现途径而言，这种事后算法解释权具有很强的个案性和主观性，因此应当由被决策者个人来主张。

6. 结语

本文首先以我国《个人信息保护法》中规定的自动化决策算法解释权为切入点，分析了自动化决策算法解释权在适用范围、解释标准以及适用时间段上的不明确，揭示出了算法解释权需要进一步优化的现实需求。在此基础上，本文从算法解释权的固有问题以及政府监管的缺陷这内外两个角度出发，论证了自动化决策算法解释权存在的必要性。接着，本文通过分析个人信息自决权理论、算法规制理论，论证出了自动化算法解释权的优化方向，即必须要平衡被决策者的个人权益和决策者的商业利益之间的关系，力求在实现各自价值的基础上相互促进、协同发展。最后，就具体优化路径方案方面，本文明确了“重大影响”的判断标准，并以“重大影响”为基础，将自动化决策区分出了不同情形，以事前解释权和事后解释权为线索，提出了算法解释权的适用范围、解释标准和权利实现方式等问题的解决方案。

参考文献

[1] 季卫东. 数据、隐私以及人工智能时代的宪法创新[J]. 南大法学, 2020(1): 3.

[2] 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.

[3] 韩万渠, 韩一, 柴琳琳. 算法权力及其适应性协同规制: 基于信息支配权的分析[J]. 中国行政管理, 2022(1): 33.
[4] 梁振文. 算法解释权的构造与法治保障路径——以社会信用体系建设为场景[J]. 吉首大学学报(社会科学版), 2021, 42(1): 152.
[5] 丁晓东. 基于信任的自动化决策: 算法解释权的原理反思与制度重构[J]. 中国法学, 2022(1): 101-104.
[6] 张凌寒. 商业自动化决策算法解释权的功能定位与实现路径[J]. 苏州大学学报(哲学社会科学版), 2020(2): 57-58.
[7] 杨立新, 赵鑫. 利用个人信息自动化决策的知情同意规则及保障——以个性化广告为视角解读《个人信息保护法》第 24 条规定[J]. 法律适用, 2021(10): 24-28.
[8] 张恩典. 大数据时代的算法解释权: 背景、逻辑与构造[J]. 法学论坛, 2019, 34(4): 153.
[9] 王利明. 论人格权请求权与侵权损害赔偿请求权的分离[J]. 中国法学, 2019(1): 232.
[10] 张凌寒. 商业自动化决策的算法解释权研究[J]. 法律科学, 2018(5): 71-72.
[11] 丁晓东. 基于信任的自动化决策: 算法解释权的原理反思与制度重构[J]. 中国法学, 2022(1): 108.
[12] 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): 156-160.
[13] 吕炳斌. 论个人信息处理者的算法说明义务[J]. 现代法学, 2021, 43(4): 92-95.
[14] 辛巧巧. 算法解释权质疑[J]. 求是学刊, 2021, 48(3): 100.
[15] 弗吉尼亚·尤班克斯. 自动不平等[M]. 李明倩, 译. 北京: 商务印书馆, 2021: 66-70.
[16] 马颜昕, 等. 数字政府: 变革与法治[M]. 北京: 中国人民大学出版社, 2021: 157.
[17] 张欣. 算法解释权与算法治理路径研究[J]. 中外法学, 2019, 31(6): 1429.
[18] 丁晓东. 基于信任的自动化决策: 算法解释权的原理反思与制度重构[J]. 中国法学, 2022(1): 112.
[19] 何渊. 数据法学[M]. 北京: 北京大学出版社, 2020: 36-40.
[20] 温昱. 算法权利的本质与出路——基于算法权利与个人信息权的理论分疏与功能暗合[J]. 华中科技大学学报(社会科学版), 2022, 36(1): 54-63.
[21] 张恩典. 大数据时代的算法解释权: 背景、逻辑与构造[J]. 法学论坛, 2019, 34(4): 152-160.
[22] 张凌寒. 商业自动化决策的算法解释权研究[J]. 法律科学, 2018, 36(3): 65-74.
[23] 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): 173-193.
[24] 张维迎, 博弈论与信息经济学[M]. 上海: 上海人民出版社, 1996: 396.
[25] 赵宏. 信息自决权在我国的保护现状及其立法趋势前瞻[J]. 中国法律评论, 2017(1): 149.
[26] 宋华健. 反思与重塑: 个人信息算法自动化决策的规制逻辑[J]. 西北民族大学学报(哲学社会科学版), 2021(6): 102.
[27] 张凌寒. 风险防范下算法的监管路径研究[J]. 交大法学, 2018(4): 58.