



东方法学
Oriental Law
ISSN 1674-4039, CN 31-2008/D

《东方法学》网络首发论文

题目: “脱离算法自动化决策权”的虚幻承诺
作者: 唐林垚
DOI: 10.19404/j.cnki.dffx.20200929.001
网络首发日期: 2020-09-29
引用格式: 唐林垚. “脱离算法自动化决策权”的虚幻承诺. 东方法学.
<https://doi.org/10.19404/j.cnki.dffx.20200929.001>



网络首发: 在编辑部工作流程中,稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定,且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件,可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定;学术研究成果具有创新性、科学性和先进性,符合编辑部对刊文的录用要求,不存在学术不端行为及其他侵权行为;稿件内容应基本符合国家有关书刊编辑、出版的技术标准,正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性,录用定稿一经发布,不得修改论文题目、作者、机构名称和学术内容,只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约,在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版,以单篇或整期出版形式,在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z),所以签约期刊的网络版上网络首发论文视为正式出版。

“脱离算法自动化决策权”的虚幻承诺

唐林垚*

内容摘要: 欧盟《通用数据保护条例》GDPR 第 22 条应作权利而非禁令解,由此产生的“脱离自动化决策权”存在与“更正权”竞合之表象、与“获解释权”补强之曲解以及与“被遗忘权”混同之误判。界定该权利应把握两条主线:(1) GDPR 序言第 71 条赋予数据主体的三种基本权利乃相互并列而非互为因果;(2) 权利内涵随技术发展嬗变,在基于数据库编码的计算机自动化场景为脱离数据采集,在基于机器学习的算法自动化场景为脱离用户侧写,在基于神经网络的超级自动化场景为脱离平台交互。破除权利失语的关键在于适用条件去场景化以及自甘风险式的豁免进路让位于实现公共利益所必须的法定许可。若不囿于既定框架,陈情权和离线权的重构思路各有合理之处,但应防范可能产生过高的社会成本。

关键词: 脱离自动化决策权 更正权 获解释权 被遗忘权 豁免规则 自甘风险

自动化决策是与自然人决策相对立的概念,意指利用计算机技术(1980-2000)、算法程序(2000-2020)、深度学习或神经网络(2020+)替代自然人处理关键数据,凭借其自动生成对数据主体具有法律效果的决策的行为。欧盟《通用数据保护条例》(以下简称 GDPR)第 22 条第 1 款允许数据主体“脱离自动化决策”,国内主流学者将其翻译为“反对自动化决策”⁽¹⁾或“不受自动化处理的决定的限制”⁽²⁾,但依照原文“有权不成为自动化决策支配对象(not to be subject to a decision based solely on automated processing)”之表述,将其译为“脱离自动化决策”更为妥当。第 22 条第 1 款虽然在文义上赋予了数据主体脱离完全依靠自动化处理对其做出的具有法律后果决策的权利,⁽³⁾可是具体该如何脱离、能够脱离到何种程度,仅凭第 22 条无从知晓。

我国虽未进行相关法律移植,但在实践中以行胜于言的方式,一定程度上实现了上述“脱离自动化决策权”的精神内涵。一个贴切的例子是:政府部门允许个人在健康码的结果同实际情况不符时,拨打 12345 热线发起纠错申请,经县级防控指挥部核实确认后,即可修改风险判定结果。该实例向我们充分展示了三个深刻的现实:其一,在疫情防控常态化的背景之下,数据主体与自动化决策“脱钩”将寸步难行,即便行使了“脱离自动化决策权”,也不过是对不公正结果的暂时纠偏,不能、也不可能完全脱离自动化决策;其二,即便效果有限,“脱离自动化决策权”也是一种值得保护的法律效力,因为无论自动化技术有多么先进,出错在所难免,此时,适当的人工介入乃数据主体脱离自动化决策失误的补救之道;其三,“脱离自动化决策权”是一种颇为特殊的权利,有着不同于一般请求权的行使要件和适用场景。例如, GDPR 第 16 条规定了数据主体的“更正权”,允许数据主体在不违反处理目的的前提下,完善或更正不充分的个人数据。我国网络安全法也有类似的规定,且被最新颁布的民法典所吸纳,成为人格权编中的重要条款。⁽⁴⁾似乎,拨打 12345 发起健康码纠错申请,是当事

* 中国社会科学院法学研究所助理研究员,法学博士、博士后。

本文系国家社科基金青年项目“自动化应用提升现代化治理的法律保障研究”(项目批准号:20CFX006)阶段性研究成果。

(1) 腾讯研究院、中国人民大学法学院未来法治研究院:《GDPR 中文版全文》第 22 条第 1 款规定:“数据主体有权反对此类决策:完全依靠自动化处理——包括用户画像——对数据主体做出具有法律影响或类似严重影响的决策。”载腾讯研究院, <https://www.tisi.org/5029>, 2020 年 5 月 1 日访问。

(2) 中国政法大学互联网金融法律研究院:《GDPR 中文版》第 22 条第 1 款规定:“数据主体有权利不受一个仅仅依靠包括分析的自动化处理的决定的限制,这会产生关于他/她或仅仅影响他/她的法律后果。”载博客园, <https://www.cnblogs.com/ostin/p/9295762.html>, 2020 年 5 月 6 日访问。

(3) 在数据搜集、处理、利用的法律关系中, GDPR 总共界定了五类利益关涉方,分别是数据主体(生产个人数据的用户、受众或普罗大众等)、数据控制者(决定个人数据搜集处理目的和方式的自然人、法人、公共机构、监管部门等)、数据处理者(为数据控制者处理个人数据的实体)、数据接收者和第三方。自动化决策的实现,多数涉及用户侧写,即为了评估数据主体而对其个人数据进行的自动化处理,尤其是为了评估数据主体在工作表现、经济收入、健康状况、个人偏好、兴趣爱好、可靠程度、行为方式和位置行踪等进行的自动化处理。GDPR, Article 4(1), (4), (7), (8), (9), (10).

(4) 网络安全法第 43 条,民法典第 1028 条。

人在行使法定的更正权。对此，需要厘清的是，无论是欧盟还是我国，数据主体行使更正权可以更正的对象，只能是个人的数据而非数据经自动化处理后产生的决策结果，两者有着指向性的区别。就拟取得的法律效果而言，“更正权”甚至算不上最低程度的“脱离自动化决策权”。

在自动化应用结构性嵌入社会运营的大趋势下，“脱离自动化决策权”绝非请求人工介入、修改决策结果这么简单，否则 GDPR 第 22 条大费周章、甚至有些迂回曲折的立法尝试将毫无必要。“脱离自动化决策权”的内涵和外延究竟是什么？与 GDPR 保障数据主体的其他权利有何关联和区别？正确理解“脱离自动化决策权”背后的核心价值与理论基础，将对我国未来人工智能立法带来重要启示。

一、从 DPID 到 GDPR：历史沿革与理论争议

“脱离自动化决策权”并非 GDPR 首创，始见于 1995 年欧盟《个人数据保护指令》（以下简称 DPID）第 15 条第 1 款：“数据主体享有免受基于用户侧写的完全自动化决策的权利”。该权利的行使以两个必要条件为前提：（1）自动化决策必须对数据主体造成具有法律后果的重大影响；（2）用户侧写完全由自动化处理实现。

（一）DPID 第 15 条的历史遗留问题

DPID 是 GDPR 的前身，两者前后间隔 23 年，在此期间，有两个事实需要注意。其一，自动化决策实现了技术层面的飞跃，逐步改变了探讨“脱离自动化决策权”的语境。在制定



DPID 并推行的 20 世纪 90 年代，自动化决策主要指基于数据库编码的计算机自动化，即利用计算机替代自然人实现唯手熟尔的重复性基础工作和流程性工作。2000 年之后，基于数据库编码的计算机自动化逐渐让位于基于机器学习的算法自动化，其实质是作为编程理论典范的透过规则本原寻求逻辑。大数据加持机器学习模型，在范围不确定的环境中进行规律挖掘或模式识别，替代自然人实现温故而知新的基础性预测工作与规范性鉴别工作。GDPR 自 2018 年 5 月开始生效，基于机器学习的算法自动化正逐步向基于神经网络的超级自动化迈进。随着数据平台的开源共享以及机器学习模型的重叠交互，今天的自动化决策已经渐次突破立普斯基所称“自然人可以、机器人不行”的推理和演绎的智能上限，⁽⁵⁾ 其获得识辨特定时期的法律和社会运行宏观规律、以超乎常人的洞察力来提供制度解决方案或进行价值判断的奇点临近。在这个阶段，自动化决策将是“以任何信息技术设备为载体、以持续控制形式干预和引导日常社会互动的高度自主的精细化治理秩序”。⁽⁶⁾ 括而言之，虽然用语均是“自动化决策”，但 DPID 的自动化决策概念仅仅涵盖了基于数据库编码的计算机自动化和基于

(5) See Lipsky, M., *Street-Level Bureaucracy: Dilemmas of The Individual in Public Service*, Russell Sage Foundation, 2010, p.1—25.

(6) John Danaher, *The Ethics of Algorithmic Outsourcing in Everyday Life*, Algorithmic Regulation, Oxford University Press, 2019, p.101.

机器学习的算法自动化,而 GDPR 的自动化决策概念囊括了迄今为止所有的自动化类型。这是笔者研究“脱离自动化决策权”需要顾及的权利适用范围变化。DPID 和 GDPR “脱离自动化决策权”的微妙差异如图一所示。

其二,在 DPID 的整个生命周期内,第 15 条几乎完全处于休眠状态,但 GDPR 第 22 条依旧毅然决然地承袭了 DPID 第 15 条的规定,欧盟部长委员会于 2018 年通过的《关于个人数据处理的现代化保护公约》第 9 条也吸收了 DPID 第 15 条的相关规定。⁽⁷⁾这或许意味着,在基于机器学习的算法自动化和基于神经网络的超级自动化大行其道的场景中,“脱离自动化决策权”更有适用和推广的必要。此前,“脱离自动化决策权”一直被戏称为 DPID 中的“二等权利”,因为在欧盟法院和各成员国法院的审判实践中,对抗双方从未就该权利的行使方式和法律效果展开有实质意义的辩论:“脱离自动化决策之诉”客观存在,但是双方总是聚焦于有争议的决策是否完全由自动化处理作出。例如,德国联邦法院在 SCHUFA 案中裁定,资信考察系统输出的信用评价不属于 DPID 第 15 条所界定的完全无人工介入的自动化决策范畴,因为银行对其客户的信用评价实由自然人在自动化决策的“辅助”下完成;⁽⁸⁾法国最高法院认为,执法者采取算法自动化系统进行裁决辅助,也不受 DPID 第 15 条的约束。⁽⁹⁾前欧盟个体保障局在工作报告中指出,DPID 第 15 条不是针对数据保护可以推而广之的一般性原则,而是针对特定自动化用户侧写的例外原则。⁽¹⁰⁾伴随早期立法尝试铩羽而归的新问题是,GDPR 第 22 条相对于 DPID 第 15 条的修改,能否让“脱离自动化决策权”焕发新的活力呢?

(二) 规则演进与责任豁免

相较于 DPID 第 15 条, GDPR 第 22 条的进步主要体现在以下三个方面:

其一, GDPR 第 22 条对于特别敏感的个人数据给予了更多关注。第 22 条第 4 款规定,数据主体有权脱离依据涉及种族、民族历史、政治观念、工会、基因、自然人健康和性生活取向等数据形成的自动化决策。当且仅当数据控制者取得数据主体的明确同意或为实现公共利益所必须的法定许可时,才能不受干扰地处理此类敏感信息。由此可见, GDPR 在个人数据类型化处理方面,具有比 DPID 更高的“颗粒度”。

其二, GDPR 第 22 条拓展了“脱离自动化决策权”的行权范围,将未成年人的个人数据纳入特别保护对象。GDPR 序言第 38 条指出,对未成年人个人数据的具体保护“尤其应适用于可能供机器学习模型生成用户侧写的数据”。序言第 71 条规定,基于用户侧写的自动化决策不作用于未成年人。相较之下, DPID 并未就未成年人数据保护作出相关规定,也难从相关条款中推定未成年人是否应当享有“脱离自动化决策权”。对未成年人个人数据予以特殊保护,彰显出法律顺应市场需求与时俱进:随着互联网低龄时代的到来,九零后逐渐成为被舆论抛弃的“后浪”,零零后甚至一零后接过泛娱乐时代的大旗称雄网络。研究报告表明,零零后在餐饮、颜值经济等领域已经成为消费主力。⁽¹¹⁾数据控制者深谙“得零零后者得天下”“营销要从娃娃抓起”“先入为主培养消费习惯”的道理,在巨大商业利益的诱惑下,社交媒体、电商平台、视频网站对未成年数据趋之若鹜,竞相获取新生代个体的浏览历史、个人轨迹、消费记录、点赞和收藏列表等——精准针对未成年人的内容推送和价值观引导接踵而至,个中风险不言而喻。

其三, GDPR 第 22 条拓展了“脱离自动化决策权”的适用范围,不再局限于 DPID 第 15 条“用户侧写完全由自动化处理实现”的行权限制。依照 GDPR 第 22 条“有权不成为自动化决策支配对象,包括用户侧写”的表述,⁽¹²⁾似乎“自动化决策和用户侧写”共同构成了“脱离自动化决策权”的适用场景。也即是说,自动化决策无论是否以用户侧写实现,均可以引发数据主体行使“脱离自动化决策权”。反过来,用户侧写无论最终是否形成自动化

(7) *Modernised Convention for the Protection of Individuals with regard to the Processing of Personal Data*, Article 9(1)(a).

(8) *Urteil des VI. Zivilsenats vom 28. 1. 2014 - VI ZR 156/13*.

(9) *Cour de Cassation, Chambre criminelle, Audience publique du 24 septembre 1998, No de pourvoi 97-81.748, Publié au bulletin*.

(10) *Article 29 Working Party, Transfers of Personal Data to Third Countries: Applying Articles 25 and 26 of the EU Data Protection Directive (WP 12, 24 July 1998)*.

(11) 奥美:《00 后人群研究报告》,载《中国连锁》2016 年第 4 期。

(12) *Article 29 Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/ 679 (WP 251rev.01, 6 February 2018)*.

决策,也终将受到GDPR第22条的约束。多名学者对此表达了异议,例如,门德萨和拜格雷夫认为,上述理解方式“违背了第22条的基本原理和立法背景,原文中的‘包括(including)’应当被解释为‘涉及(involving)’”。布尔坎认为,现今自动化决策必然涉及用户侧写,因此围绕第22条的争议可谓毫无必要。^[13]在笔者看来,反对派观点的学者至少犯了两个方面的错误:(1)未能深入考察自动化决策从最初基于数据库编码的计算机自动化,到近来基于机器学习的算法自动化,再到未来基于神经网络的超级自动化的跃迁过程,理所当然地认为自动化决策就是机器学习模型依照特定算法进行用户侧写形成的决策。实际上,自动化决策若由计算机检索数据库编码作出,就根本不涉及用户侧写的过程;若由神经网络的实现,用户侧写只是超级自动化决策的一个环节。(2)将GDPR第22条视为DPID第15条的单纯延续,未能深究措辞变化背后可能蕴藏的政策转向。概括适用到局部适用的条件变化,使得本来共同构成DPID第15条必要条件的自动化决策和用户侧写,在GDPR第22条中转变为充分条件。这也从正面证实了,DPID第15条仅允许数据主体在基于机器学习模型的算法自动化情形中行使“脱离自动化决策权”,而GDPR第22条可以同时适用于基于数据库编码的计算机自动化、基于机器学习的算法自动化乃至基于神经网络的超级自动化的所有情形,权利的适用范围被大大拓宽。

遗憾的是,虽有上述大刀阔斧的修改,自GDPR生效至今,丝毫未见“脱离自动化决策权”的复苏迹象。作为一项值得被保护的法律效力,为何“脱离自动化决策权”经常被遗忘、极少被行使、几乎与现实相“脱离”?这个问题的答案首先在于GDPR第22条相对于DPID第15条并未发生实质性修改的“适用条件”——数据主体有权请求脱离的,必须是“完全依靠自动化处理”产生的决策,这无疑极大地提高了“脱离自动化决策权”的适用门槛。无论是基于数据库编码的计算机自动化、还是基于机器学习的算法自动化,最低限度的人工介入实为不可避免,无论是出于维护机器运转之必须、还是确保数据结构化处理的一致性、亦或是确认机器学习模型未脱离“算法代码的缰绳”。针对适用门槛的权威解释长期缺位,进一步加剧了法条本身的模糊性,使得各参与方只能完全依照字面理解来调整自身的行为——既然一丁点自然人参与的“蛛丝马迹”便能推翻数据主体对GDPR第22条的信赖利益,“脱离自动化决策权”被完全架空的结果并不出人意料。

“脱离自动化决策权”的式微,还源于过于宽松的豁免条件。DPID第15条规定,数据主体主动要求并同意接受自动化决策服务时,将受到类似“禁止反言原则”的约束;在签订合同时,只要数据控制者或处理者采取了合适的措施来维护数据主体的合法权益,那么纯粹自动化决策的效力将不受挑战。GDPR实际上扩大了DPID的豁免范围,在第22条第2款中明确规定了“脱离自动化决策权”不适用的三种情形——当事人同意、法律授权以及合同约定。在上述三种豁免情形之外,GDPR第4款额外增加了需要和其他法条相互参照的行权限制条件。第22条第3款虽然对豁免条件进行了适当限缩,要求“数据控制者应当采取充分措施保障数据主体的权利、自由和正当利益,允许数据主体对数据控制者进行人工干涉,以便表达其观点和对决策表达异议的基本权利”。但是深究该款措辞不难看出,在豁免情形中,数据控制者依法应当保障数据主体的,只是“数据主体进行人工干涉”而非“数据主体请求数据控制者进行人工干涉”的权利,并且,数据主体进行人工干涉是为了表达其观点和对决策表达异议,具有“反对权”而非“脱离自动化决策权”的外观。易言之,“脱离自动化决策权”豁免条件生效,只需以对反对权的保障作为前提,而反对权已在GDPR第21条中得以单独规定:“出于公共利益、官方权威、控制者或第三方追求正当利益所进行的数据处理,包括根据相关规定进行的用户侧写,数据主体有权随时反对。”第22条第3款对第21条的简单重复并未实质性提高数据控制者的豁免门槛。

(三)“权利”与“禁令”的实质之争

在适用条件苛刻、豁免门槛过低的双重制约下,“脱离自动化决策权”赋予数据主体的权利在法律上并非以“可执行的状态”存在,数据执法机关在实践中也很难将第22条作为切实可行的执法依据。实践中,对GDPR第22条的内在属性的认识割裂,在立法者和执法者之间渐次成型:立法者认为自己为数据主体创设了一种在特定情形中可以行使的权利,但执法者只将其视为针对特定类型自动化决策的禁令。与GDPR相对应的《欧盟执法指令》(下

[13] Mendoza I., Bygrave L., *The Right not to be Subject to Automated Decisions based on Profiling in Synodinou T. (eds), EU Internet Law: Regulation and Enforcement* (Springer 2017).

简称 LED) 第 11 条针对特殊类型的自动化决策连续使用了三个“禁止”，显然是将 GDPR 第 22 条视为一条禁令，⁽¹⁴⁾ 招来了多数成员国的质疑。从执法者的角度来看，禁令思维确实更易于执法活动的开展，也在表面上维护了 GDPR 与 LED 的和谐统一。但此种“为了执行而执行”的粗浅认识罔顾了欧盟立法者在个人数据保护方面的大局观，强行以 LED 第 11 条的规定去统合 GDPR 第 22 条的实现方式反倒会破坏欧盟整体数据保护框架的一致性。

在笔者看来，以权利思维而非禁令思维理解 GDPR 第 22 条，至少有三点好处：其一，承认 GDPR 第 22 条为数据主体可以行使的“脱离自动化决策权”，同第 22 条的字面表述相吻合，同时也符合立法者以法律规则“钳制”自动化决策过程的主观想象；其二，权利思维契合当下自动化决策被广泛应用于私营和公共部门的现实，对特定类型的自动化决策不宜一概禁止而应当考察其应用场景；其三，权利思维更符合辩证法所崇尚的自然科学观，即自动化决策完全可以给社会整体带来可观的利益，而非总是因外部性扩散催生社会成本。总而言之，以禁令思维理解 GDPR 第 22 条的做法过于简单，这其实是 LED 其他条款也或多或少存在的共性问题，反映出人工智能领域立法者与执法者难以消磨的思维偏差以及由此产生的释法断层——可解释的法律不可执行、可执行的法律不合解释；如何尽可能缩小两者之间的差距，是在规则的应然和实然争辩之外，值得学术界和实务界上下求索的基础问题。

作为一项披着禁令外衣的权利，“脱离自动化决策权”本质上是一项请求权，是自动化决策关系中数据主体请求数据控制者“为或不为一定行为的权利”，数据主体不能对自动化决策的权利标的进行直接支配，而只能请求数据主体予以配合。由此引出了本文的关键问题，“脱离自动化决策权”究竟赋予了数据主体哪些请求数据控制者“为什么”与“不为什么”的权利？

长期以来，有关 GDPR 第 22 条所界定的权利性质、正当性与适用范围的争论从未停止，由此形成的学说千姿百态，其中存在两种常见的误解。一种误解是将 GDPR 第 22 条视为算法可解释性要求的圆心，与 GDPR 中“获解释权”的多个条款互为补强；另一种误解是将“脱离自动化决策权”与“被遗忘权”混为一谈，将数据主体脱离自动化决策的尝试等同于向数据控制者行使擦除个人数据的请求权。在应用层面上，将“脱离自动化决策权”视为数据主体请求“获解释”或者“被遗忘”的权利，具有一定的可操作性，实为将 GDPR 第 22 条作权利解的大前提下，权利思维向禁令思维有限度的靠拢的折衷之举。这两种不同的理解进路，正是导致“脱离自动化决策”被翻译为“反对自动化决策”或“不受自动化处理的决定的限制”的根本原因，也同时反映出各版本译者绝非生硬干涩地对原文进行单纯直译，而是在极高的人工智能法学造诣之上融入了自身对 GDPR 各条款的深入理解，缜密的法律思维可以从别具匠心的意译表达中窥见一斑。那么，“脱离自动化决策权”同“获解释权”“被遗忘权”之间的区别是什么？各权利之间的区别是虚是实？从权利思维出发，GDPR 为即将到来的超级自动化时代建立了怎样的制度防火墙？“脱离自动化决策权”在自动化治理中的地位和功用是什么？

二、必也正名乎：三权并立的联动体系

本部分将从“脱离自动化决策权”同“获解释权”“被遗忘权”的差异入手，一则反思“脱离自动化决策权”的立法初衷；二则探讨“脱离自动化决策权”的实质内涵。

(一) “脱离自动化决策权”非“获解释权”之补强

大数据、云计算和人工智能技术的突飞猛进，“革命性地改变了网络空间内主体的能力差异和关系结构”，⁽¹⁵⁾ 传统的法律制度难以应对技术黑箱掩映之下受众操控、信息寻租和监管套利的三重失控，对算法可解释性的合规要求应运而生。虽有学者反复指出，旨在提升算法透明度的硬性规定“既不可行，也无必要”，⁽¹⁶⁾ 但在漫长的监管实践中，算法可解释性的合规要求还是成为了世界各国人工智能法律法规的核心要旨，⁽¹⁷⁾ 即便是那些高度依赖事后问责机制的国家，⁽¹⁸⁾ 也不否认算法可解释性实乃人工智能时代“对抗数据个体的主体性和

(14) Law Enforcement Directive, Article 11.

(15) 陈璞：《论网络法权构建中的主体性原则》，载《中国法学》2018 年第 3 期。

(16) 沈伟伟：《算法透明原则的迷思》，载《环球法律评论》2019 年第 6 期。

(17) 参见丁晓东：《算法与歧视：从美国教育平权案看算法伦理与法律解释》，载《中外法学》2017 年第 6 期。

(18) 例如，美国算法问责法案第 3 部分、法国信息和自由法第 82 条、印度个人数据保护法案第 34 条、新加坡防止网络虚假信息和网络操纵法案第 32 条等均规定了数据控制者的“记录义务”，以确保对算法决

自治性沦陷和丧失的内在之善”，⁽¹⁹⁾是明确自动化决策主体性、因果性和相关性以确定和分配算法责任的重要前提。⁽²⁰⁾究其根源，在公众问责空前加强的年代，任何新技术的批量部署都必然被置于传媒与社会持续性互动的场域之中，以逐渐打消公众质疑、充分回应社情民意以及成功取得多数人同意为其合法性基础。

令人感到疑惑的是，在制定过程中反复强调算法可解释性有多么重要的 GDPR，只在条例第 5 条第 1 款笼统地要求对涉及数据主体的个人数据，应以合法、合理、透明的方式进行处理。显然，公开透明的数据处理方式并不足以确保算法可解释性，顶多被视为算法可解释性众多构成要件中不可或缺的一环。实践中，人们不得不结合 GDPR 的序言，为算法可解释性要求寻找依据。GDPR 序言第 71 条指出，接受算法自动化决策的数据主体应当享有“适当的保护”，因为潜在的机器错误和数据歧视极有可能给整个社会带来“偏见和不公”。第 71 条随即将“适当的保护”分为三类：一是获得人类干预的权利；二是表达异议和质疑的权利；三是获取相关决策解释的权利。就此而论，GDPR 是从数据主体“获解释权”的进路出发，来反向构建算法可解释性的合规要求。⁽²¹⁾佐证这一现实的依据是，GDPR 第 13 条、第 14 条、第 15 条分别规定了直接从数据主体获得个人数据、间接从数据主体获得个人数据以及因各类原因访问个人数据的数据控制者向数据主体履行详细信息披露义务以及就特定事项进行进一步解释说明的义务，尤其是“决策过程所涉及的逻辑程序及其对数据主体的重要意义和可能影响”。

但是，从 GDPR 第 13 至 15 条引申出来的“获解释权”完全局限于数据的搜集和处理阶段，如此“事前解释”只能肤浅地让数据主体得知自己个人数据的处理概况和潜在影响，并不能让数据主体真正知晓最终形成的自动化决策与其当下际遇之间的因果关联。很明显，处于信息极度弱势地位的数据主体，主动行使“获解释权”的理想预期，是要获得其在承受自动化决策阶段的事中和事后解释，而 GDPR 对此语焉不详。在此种背景下，部分学者将 GDPR 第 22 条视为同第 13 至 15 条的衔接，补足了事中和事后“获解释权”的法律真空。⁽²²⁾

诚然，GDPR 字面规定的“获解释权”确实存在事前、事中和事后脱节，导致权利难以覆盖自动化决策全过程的问题，但是，衔接断裂的问题完全可以通过直接补充规定的方式予以修复，实在没有必要在距离 13 至 15 条“至少 7 条开外的”第 22 条中以如此晦涩、隐蔽的方式对事中和事后的“获解释权”进行界定。欧盟立法者之所以不在 GDPR 中建立完整的“获解释权”链条，极有可能源于两个层面的思考和顾虑。其一，批评者们看到的是一个设定一项“获解释权”就可以让自动化决策可解释的世界，而这种理想愿景同当今技术运作的现实情况存在巨大鸿沟。尤其是在事中和事后阶段，随着机器学习模型的日益复杂、卷积神经网络的广泛适用、数据搜集源头的交汇重叠，要真正就自动化决策的运作原理与裁判机制作出详细且外行人也能理解的解释，可谓难上加难。⁽²³⁾事前的各类解释或许有助于数据主体作出是否同意自动化决策的判断，但考虑到人工智能是以技术手段对自然人的深度仿生，以预期效果为导向的、事中和事后的因果关联尝试极易将物质表象的经验规律同内在直觉的感性判断混淆起来，最终只能得到故弄玄虚、毫无根据的假设而非解释。其二，算法可解释性的合规要求，必然伴随相当可观的合规成本。“获解释权”的规定越是完整无遗漏，身处信息弱势地位的数据主体前期缔约和谈判的成本就越低，但因此增加的其他成本则完全由数据控制者或处理者概括承受。长期以来，对于此类成本转嫁是否合理的争论从未平息。一方面，正如上文所指出的那样，数据控制者针对自动化决策进行的事中和事后解释极有可能属于徒劳无功的解释，经常是“解释了也是白解释”；另一方面，完整意义上的“获解释权”至少在一定程度上有助于消弭算法歧视、降低数据控制者和数据主体之间的信息不对称。对于此类“或有或无”的法律规则，现代法经济学给我们的教义是：应当衡量法律实施的成

策中责任主体链条的追踪识别。

(19) 张欣：《算法解释权与算法治理路径研究》，载《中外法学》2019 年第 6 期。

(20) See Pasquale, Frank, *The Black Box Society*, Harvard University Press, 2015, p. 1—16.

(21) Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a ‘Right to an Explanation’ is Probably Not the Remedy You are Looking For*, *Duke Law & Technology Review*, Vol.16, p.18(2017).

(22) See Casey, Bryan, Ashkon Farhangi, Roland Vogl, *Rethinking Explainable Machines: The GDPR’s ‘Right to Explanation’ Debate and the Rise of Algorithmic Audits in Enterprise*, *Berkeley Technology Law Journal*, Vol.34, p.145—187(2018).

(23) 参见刘洪华：《论人工智能的法律地位》，载《政治与法律》2019 年第 1 期。

本与潜在社会收益之间孰轻孰重，即那些看似公平的原则是否同时也是那些符合效率的原则。^[24]在全球人工智能竞争白热化的阶段，说服各国政府和跨国大型公司建立事前、事中和事后的“获解释权”闭环实属痴人说梦。

（二）“脱离自动化决策权”不与“被遗忘权”混同

在“脱离自动化决策权”难同“获解释权”等量齐观的情况下，另有一些学者试图将 GDPR 第 22 条视为“被遗忘权”相关条款的延伸乃至重复。^[25]因为在历史上，“被遗忘权”的创设同“脱离自动化决策权”的发展有着千丝万缕的联系。

在 2010 年 *Google Spain v. Costeja Gonzalez* 案中（以下简称谷歌案），原告要求被告谷歌公司删除 12 年前因为房屋网络拍卖而在搜索引擎中陈列的、已经过时的个人信息。欧盟法院在裁判时，从 DPID 中推导出了数据主体应当享有的“被遗忘权”，依据是第 6 条“个人数据的处理应当适当、相关、同被处理（以及进一步处理）的目的吻合，在必要时进行更新以保证准确性”以及第 12 条允许数据主体“更正、擦除以及屏蔽不符合 DPID 所规定的处理的数据”。^[26]2013 年欧盟《信息安全技术公共及商用服务信息系统个人信息保护指南》吸收了欧盟法院的判决和 DPID 的前述规定，直截了当地建议：“当数据主体有正当理由要求删除其个人数据时，应及时删除个人数据。”GDPR 第 17 条对“被遗忘权”进行了直接规定：“数据主体有权要求数据控制者擦除关于其个人数据的权利。”

“被遗忘权”与“脱离自动化决策权”的混同，源于谷歌案引发的有关搜索引擎性质的争议。一方认为，搜索引擎只是数据的媒介，而非严格意义上的数据控制者或者处理者，一如谷歌的首席法律顾问沃克在接受《纽约客》采访时所言：“我们不创造信息，只是让信息变得可访问。”^[27]另一方则针锋相对地指出，搜索引擎在信息处理方面并非全然被动，结果的呈现方式与先后次序，即是搜索引擎处理数据后形成的自动化决策。^[28]从搜索引擎公司“竞价排名”的常见盈利手段来看，^[29]显然后者的观点更有说服力。将搜索引擎确认为数据控制者，可以推导出一个能够达成共识的认知，即谷歌案中原告向谷歌公司行使“被遗忘权”，是希望谷歌公司能够对特定搜索结果进行人工干预，实现“脱离自动化决策权”的行权效果。应原告诉求进行人工干预的结果并非真正让原告在网络世界中“被遗忘”，而是让谷歌的搜索界面不再将原告已经过时的信息纳入排序算法。也即是说，作为数据主体的原告行使“被遗忘权”拟达到的效果实际上是脱离搜索引擎的自动化决策。此种做法自欺欺人的地方体现在，搜索引擎没有办法、也没有权力真正删除所有包含数据主体希望被遗忘信息的页面，而只能将这些页面在搜索结果中隐去，如果在地址栏中输入这些页面的网址，仍将能够打开这些页面。需要注意的是，有不少热门网站专门记录谷歌删除或移除掉的链接，例如 *Hidden From Google*、*Wikimedia* 和 *Reddit* 等，使得希冀行使“被遗忘权”的数据主体常常陷入越描越黑的境地，即使脱离了搜索引擎的自动化决策也脱离不了“好事者竟趋焉”的无辜生非。在这个意义上，“脱离自动化决策权”似乎是行使“被遗忘权”的主要手段，也是行使“被遗忘权”能够实现的最低目的，部分学者将 GDPR 第 22 条视为对第 17 条“有益而有必要的重申”就不足为奇了。

但是，“被遗忘权”并不能在所有情形中都等同于“脱离自动化决策权”。只需在谷歌案的基础上稍微做一点延伸，就能理解“被遗忘权”的独特之处。原告向谷歌公司行使“被遗忘权”，最终只能达至“脱离自动化决策权”的效果，这是由于搜索引擎的特殊性质所决定的。如果希望达到被彻底遗忘的效果，原告应该一一向刊登原告个人数据的页面所有者行使“被遗忘权”。这些被搜索引擎直接抓取的页面，是纯粹的信息刊载或转载方，几乎不涉及自动化决策，只需简单地从服务器中删除页面，即可履行数据主体的被遗忘请求。这个过程虽然也涉及人工干预，但同行使“脱离自动化决策权”有着天壤之别，是以删除内容、移除

[24] [美]波斯纳：《正义/司法的经济学》，苏力译，中国政法大学出版社 2002 年版，第 6 页。

[25] Jeffrey R., *The Right to be Forgotten*, *Stanford Law Review*, Vol.64, p.88(2012).

[26] *Directive on The Protection of Individuals with Regard to The Processing of Personal Data and on The Free Movement of Such Data*, Article 6(c), 12(b).

[27] Jeffrey T., *The Solace of Oblivion*, *New Yorker*, 29 Sep 2014.

[28] *Google Spain SL, Google Inc v. Agencia Espanola de Proteccion de Datos es Mario Costeja Gonzalez*, Case C-131/12.

[29] 参见陶乾：《论竞价排名服务提供者注意义务的边界》，载《法学杂志》2020 年第 5 期。

链接作为权利行使的主要法律后果。^{〔30〕}实际上,搜索引擎要想完全履行数据主体的被遗忘请求,除了以人工干预的方式将数据主体从自动化决策中剔除出去之外,还需要在其服务器中一一手动删除相关页面的快照,且删除过程的工作量甚至不低于脱离过程的工作量。在上述认知的基础上,还要强行将“被遗忘权”与“脱离自动化决策权”混同,就显得不合时宜了。

(三) GDPR 权利体系重构:回归序言第 71 条

归根结底,“获解释权”和“被遗忘权”各自具备深刻的内涵和外延,是与“脱离自动化决策权”截然不同的两种权利。然而,行使这两种权利确实在一定程度上也能让数据主体“脱离自动化决策”,由此引发的问题是,“获解释权”和“被遗忘权”是否可以被反向视为“脱离自动化决策权”的被动权能和主动权能?易言之,GDPR 第 13 至 15 条以及第 17 条,是否是第 22 条的两种不同行权方式?答案无疑是否定的。

1. “获解释权”的补强条款:GDPR 第 35 条

首先来看为什么“获解释权”并非“脱离自动化决策权”被动权能。理论上,火力全开的“获解释权”,确能为“脱离自动化决策”结果的实现,提供较强助力,因为只要数据控制者或处理者不能向数据主体提供合乎情理的行为解释,那么数据搜集和处理将失去正当和合法性——作为程序运行“燃料”的大数据一旦断供,自动化决策将不复存在,数据主体便自然实现“脱离”。但正如本文第一部分所指出的那样,GDPR 第 13 至 15 条只规定了数据主体的事前“获解释权”,数据主体在行使这种权利的时候自动化决策尚未发生,自然也就没有“脱离自动化决策”可言。同时,出于对效用的追求,欧盟立法者具有维持弱式“获解释权”的激励,不会贸然赋予数据主体全过程链的“获解释权”。

在第 13 至 15 条规则圆满性难以为继的情况下,毋宁以 GDPR 第 35 条数据控制者和处理者的“数据保护影响评估义务”作为其事中和事后补强:当用户侧写可能对数据主体的权利和自由带来高风险时,数据控制者或处理者必须在“充分考虑数据处理性质、范围、语境和目的的基础上”,评估自动化决策将对个人数据保护造成的影响。第 35 条对事中和事后算法可解释性要求的补充体现在三个方面:其一,数据保护影响评估要求对数据处理的可能操作和计划目的进行系统性描述,形成文字的系统性描述以数据控制者所追求的正当利益为主要内容;其二,数据控制者必须对数据的搜集和处理的必要性与相称性进行阐述,即用户侧写应以数据控制者的正当利益为限,未经数据主体同意不得牺牲其权利;其三,当数据主体的正当权利和数据控制者的正当利益存在此消彼长的张力时,数据控制者必须在数据保护影响评估中表明可以采取的应急性风险应对措施,包括但不限于 GDPR 已经规定的法律保障、安全措施和激励机制。从条文结合可以产生的效果来看,第 13 至 15 条主要是要求数据控制者和处理者就事前数据挖掘的信息来源、典型特征和分类办法以及算法程序的运作机理、代码逻辑和预期效果进行事前解释,第 35 条则是要求其就事中的系统偏差、运行故障和矫正机制予以说明,这实际上只是在最低限度实现了算法可解释性的显化,并且由于事后解释的要求仍然孱弱,数据主体行使“获解释权”最多只能得到一个看似将人工智能载体的硬件、软件 and 数据处理之间的相互作用如何导致自动化决策之间因果关系阐明的解释。如果心怀善意,数据控制者或处理者给出的是“自以为正确的解释”。如果别有用心,解释和说理的方式将完全取决于拟实现的经济或政治目标。在这个意义上,GDPR 第 13 至 15 条与 35 条的结合,勉强构建了多重弱化但相对全面的“获解释权”。

2. “被遗忘权”的上位条款:GDPR 第 21 条

上文指出,“脱离自动化决策权”并非“被遗忘权”的行权方式,因为“被遗忘权”在人工干预之外,还包含了数据控制者删除原始页面、擦除特定内容以及移除相关链接的责任规定。更重要的是,“被遗忘权”的行使,不以自动化决策存在为前提。但在自动化决策已经进行的情况下,“被遗忘权”是否是“脱离自动化决策权”的主动权能呢?这一点将更难证否,因为数据主体“脱离自动化决策”意味着个人数据至少在处理的某个环节得以擦除。

要彻底回答上述问题,我们必须换个角度进行梳理。“被遗忘权”同“获解释权”一样,是一种独立的权利,两者均在 GDPR 正文中进行了专门规定。略显突兀的是,为何“被遗忘权”没能像“获解释权”一样,位列序言第 71 条所规定的的数据主体应当受到保障的三种基本权利——获得人类干预的权利、表达异议和质疑的权利、获取相关决策解释的权利——

〔30〕参见李爱君:《数据权利属性与法律特征》,载《东方法学》2018 年第 3 期。

之中呢？这也牵扯到另一个问题，如果 GDPR 第 17 条不是第 22 条的行权方式，那么是否存在另一条款作为第 17 条的上位条款呢？“被遗忘权”是一种请求权，既可作为单独权利，也可作为实体权利的内容，但不能脱离基础权利而存在。在笔者看来，“被遗忘权”的权利基础，是 GDPR 第 21 条所规定的“反对权”。从立法技术来看，如果第 21 条只是赋予数据主体反对数据处理或用户侧写的权利，而不辅之以必要的反对措施的话，所谓的反对只能是纸上谈兵。将第 17 条与第 21 条结合，不难看出，行使“被遗忘权”是数据主体反对数据处理或用户侧写的有力手段。质言之，以“被遗忘权”作为手段和目的，“反对权”才有用武之地。如此，“被遗忘权”方能同 GDPR 序言第 71 条中规定的三种基本权利之一——表达异议和质疑的权利相吻合。

GDPR 第 17 条是第 21 条行权方式的最直接证明，莫过于两个条款在文义和结构上的重复。第 17 条第 1 款列举了可以行使“被遗忘权”的 6 种情形：（1）个人数据对于实现其被收集或处理的相关目的不再必要；（2）数据主体撤回此前同意；（3）出于公共利益或基于官方权威进行的数据搜集和处理的压倒性原因不复存在；（4）存在非法处理个人数据的行为；（5）数据控制者需要履行特定法律责任；（6）搜集的社会服务相关信息中包含不满 16 周岁未成年人的个人数据。与之相对应的是，第 21 条的 6 个主要条款，几乎完全是对第 17 条第 1 款中 6 种情形的机械重复，即数据主体可以反对数据处理和用户侧写的 6 种情形。虽然，第 21 条还略微增加了用户有权反对以营销为目的的用户侧写，以及反对用于科学发明、历史研究、统计回归目的却不必要于公共利益的数据处理，但这两种情形又同第 17 条第 3 款所规定的情形竞合。总之，GDPR 序言第 71 条中数据主体“表达异议和质疑的权利”，对应的是 GDPR 正文第 21 条所规定的“反对权”，而“反对权”则是经由第 17 条所规定的“被遗忘权”予以实现。

3. 从“不是”到“是”：揭开“脱离自动化决策权”面纱

至此，“获解释权”和“被遗忘权”已经同 GDPR 序言第 71 条的两种基本权利规整对应，序言第 71 条余下的“获得人类干预的权利”就成为了理解“脱离自动化决策权”的关键所在。“获解释权”和“被遗忘权”都需要不同程度的“人工介入”，前者是人工对自动化决策予以合理解释、后者是人工对过时或错误数据进行清理，但“人工介入”与“人类干预”在程度和性质上都存在着诸多不同。结合三种权利之间的显著差别，笔者认为，“脱离自动化决策权”是数据主体请求数据控制者对不公正的自动化决策结果进行人工干预的权利，人工干预的目的是让数据主体脱离自动化决策错上加错的恶性循环。

在上述定义的基础上，还需从以下三个方面来精确界定“脱离自动化决策权”。其一，“脱离自动化决策权”要求比“获解释权”和“被遗忘权”更高的人工介入程度，以干预不公正的自动化决策为根本目的。从指向性来看，“获解释权”和“被遗忘权”面向过程，而“脱离自动化决策权”面向结果，这也从侧面证明了为何针对算法透明度的“事后解释”在 GDPR 中几乎暂付阙如。其二，“获解释权”“被遗忘权”和“脱离自动化决策权”共同构成了数据主体在自动化决策的各个环节寻求保护的手段，三者之间存在层层递进的行权诉求：当数据主体对自动化决策存有疑义时（一般是在事前），可以行使“获解释权”要求合理解释。在自动化决策过程中（事中），数据主体可以行使“被遗忘权”，请求删除不准确的个人数据。当不公正的决策结果产生外部性时（事后），数据主体方才行使“脱离自动化决策权”，消弭自动化决策的负面影响。其三，GDPR 第 25 条规定了数据作业的基本原则——影响最小化原则和目的适当性原则，要求个人数据始终保持“精简和锁定”的被动状态。在这两个原则的约束下，GDPR 第 22 条应结合第 25 条要求数据控制者在数据处理过程中“整合必要的保障措施，以符合 GDPR 要求保护数据主体的权利”的规定进行理解，即保障数据主体能够行使“脱离自动化决策权”是自动化决策程序设计之初，数据控制者必须充分考虑并预留相应安排的基础工作。

综上所述，以序言第 71 条为核心展开，GDPR 并非以单一、个别权利为依托，而是齐头并进构建了充分保障数据主体利益的三权并立的联动体系，贯穿自动化决策的事前、事中和事后全流程，实现了数据把控和程序治理的系统联结。以笔者思路重新解读 GDPR，得到的权利体系与条文组合机理如下图所示。

图二 GDPR 三权联动体系组合机理

基本权利	主要条款	补强条款 行权方式	权利内容	主要 阶段	行权诉求程度	原则体现
获得人类干预的权利	第 22 条	第 25 条	请求实质性人类干预，让数据主体能够脱离自动化决策失误的多米诺效应。	事后	高 (为脱离而行权)	人类自主性原则 目的限制原则
表达异议和质疑的权利	第 21 条	第 17 条	对自动化决策表达反对或质疑，并请求擦除错误的个人数据、移除过时的链接。	事中	中 (为纠正而行权)	数据最小化原则
获取相关解释的权利	第 13 条- 15 条	第 35 条	请求数据控制者对自动化决策相关信息进行详细说明、对潜在危害进行信息披露。	事前	低 (为知晓而行权)	合法性、合理性和透明性原则

三、破除权利虚幻承诺：去场景化适用与法益衡量

数据主体主动置身事外的权利设计并不意味着对自动化决策的立场中立，恰是在技术发展势不可挡的滚滚洪流中以退为进、以静制动的高阶智慧。到目前为止，我们只是将“脱离自动化决策权”作为一种应然权利进行探讨，明确其内涵并没有能够解决“脱离自动化决策权”长期难以实施的窘境。本文第一部分指出，造成“脱离自动化决策权”长期尘封搁置的状况，主要源于相关法条的适用门槛过高而豁免门槛过低。解决的方案是从权利的行使要件入手，反其道而行之，向上回溯不同的应用场景探讨“脱离自动化决策权”的内涵和外延，向下检视切实可行的制度框架和行权途径，为“脱离自动化决策权”的适用奠定良好的制度基础。

（一）去场景化适用：破除传统人工智能二分法

一直以来，人工智能系统被人为划分为“智能系统”（Automatic System，狭义的人工智能）和“辅助系统”（Augmentation System，广义的人工智能），两者的差别在于关涉数据主体切身利益的决策，是完全由自动化处理作出，还是由自然人在自动化系统的辅助下作出。DPID 第 15 条与 GDPR 第 22 条显然认可了上述二分法，不仅不对“辅助系统”生成的自动化决策予以限制，还一度将任何程度的人工介入都视为决策非纯粹自动化处理的证据，使得基于数据库编码的计算机自动化和基于机器学习的算法自动化被一揽子归于“辅助系统”的范畴，而必然属于“智能系统”的基于神经网络的超级自动化在短期内尚难以普及，“脱离自动化决策权”便沦为了一项“只有在未来可以行使的权利”。

对“辅助系统”的自动化决策网开一面，并非欧盟专有。2016 年，美国威斯康辛州最高法院在 State v. Loomis 案中认可了法官借助辅助系统进行量刑裁判的合理性与正当性，⁽³¹⁾ 长期引领工会斗争之风的威斯康辛州的这一举动，旋即激发了学界对公共治理领域自动化决策滥用的担忧。连法官司法量刑裁判都可以借用辅助系统完成，人工智能技术的应用是否还存在边界？大量的研究表明，自然人极易对机器产生过度依赖，并沉醉其中。辅助系统会降低自然人的职业警惕，直至其彻底丧失批判性分析的能力。⁽³²⁾ 更极端的情况是，为了减少自身的思考与付出，自然人决策会自然而然地倾向于同自动化决策保持一致，原本正确的判

(31) State v. Loomis, WI 68 S (2016).

(32) Pasquale F., *Secret Algorithms Threaten the Rule of Law*, MIT Technology Review (2017), 1023.

断最终被计算机扭转，导致失误叠加的“自反性悖论”。例如，算法工具根据人群聚集密度和高危人员流动轨迹推算出某一区域可能存在大量的新冠病毒无症状感染者，从而提醒自然人监管者对该区域进行更高频次的核酸检测。高频次的主动检测必然会查出更多的感染者，又将反向刺激公共卫生部门对该区域施行更高强度的算法监管。在这个过程中，辅助系统可能存在的判断谬误给自然人监管者带上了有色眼镜，裹挟其对特定区域进行“有罪推定式调查”。

为辅助系统开脱的学者指出，自动化系统虽远非尽善尽美，但在多数领域的出错概率远低于自然人。即便该观点得到了实证研究的支持，万无一失的自动化系统受制于自然人禀赋的情况在现实中屡见不鲜。以当下基于机器学习的算法自动化为例，机器学习模型定然具有指数级优于自然人的算力，但由于大数据的结构化处理难以自动完成，数据处理环节的人力资源上限就成为了阻碍机器学习模型发挥实力的瓶颈。在自动化决策环节，自然人与生俱来的情感偏见和价值负载，经常成为机器决策的拖累，这点在业务审计、欺诈防范、暴力冲突避免等辅助系统的应用场景中已经得到反复证实。⁽³³⁾由此可见，恰是由于自然人介入，才使得部分原本可能滴水不漏的自动化系统纰漏频出，就此而论，“脱离自动化决策权”的无差别适用只会更有必要，而不是相反。

在广泛的质疑声中，欧盟数据保护部门发布了《为实现监管目的的个人自动化决策指南》，要求利用辅助系统的自然人必须“有能力和有能力质疑自动化决策且经常表达对自动化决策的异议”。⁽³⁴⁾在笔者看来，强行提升自然人参与度的漏洞填补方式过于理想化，不具备操作性。一方面，鉴于自然人所固有的“工具控”本性，自动化决策过程中的自然人介入，哪怕是充分、饱和式介入，也难以真正对自动化决策形成有效钳制；另一方面，自动化决策失灵的后果固然耸人听闻，但部分失误恰是由于自然人的不当介入产生，所谓的“经常表达异议”要么在日复一日的重复决策场中流于形式，要么将进一步扰乱自动化决策的应用规划，造成适得其反的效果。考虑到披着辅助系统外衣的自动化决策应用已经结构性嵌入社会运营的各个方面，数据主体“脱离自动化决策权”不应再以自动化决策完全由机器做出为行权前提。亦即是说，没有必要再对“脱离自动化决策权”进行基于数据库编码的计算机自动化、基于机器学习的算法自动化和基于神经网络的超级自动化的场景划分。不过，权利的实际意涵还是应当根据行权效果进行区辨：数据主体行使“脱离自动化决策权”，在计算机自动化场景为脱离数据采集，在算法自动化场景为脱离用户侧写，在超级自动化场景为脱离平台交互。三种场景的共同点是，一旦进入“脱离自动化决策”的过程，较高度的人类干预将覆盖或者附议已经作出的自动化决策。

（二）有限豁免：自甘风险让位于社会责任

审视“脱离自动化决策权”的豁免条件设计，无论是 DPID 第 15 条，还是 GDPR 第 22 条，均采用了类似早期美国产品责任领域“明示的自甘风险”的免责进路，以合同约定、当事人同意和特定法定事由作为受侵害方的行权阻却事由，其理论依据是法经济学家科斯在《社会成本问题》一文中有关权利的安排和界定必然影响经济资源配置的论述。科斯指出，经济的外部性或不效率可以通过当事人的谈判得以纠正，从而实现社会效用的最大化。本来，美国的产品生产者受制于 1916 年联邦法院以长臂规则取代当事人原则判例的束缚，不仅要对与之具有直接合同关系的经销商负责，还要对任何因其产品缺陷受到伤害的消费者（及其家庭成员）和旁观者负责。⁽³⁵⁾为了规避责任，生产者刻意将自己列为产品销售合同或用户协议的相对方，并就特定产品瑕疵和潜在风险以条款约定的方式免除自身责任。一旦消费者因产品瑕疵受到损害将生产者告上法庭，生产者可以立即以消费者知情同意或合同约定主张消费者自甘风险。在近代法经济学理论的影响下，美国法院一开始近乎盲目地认可了当事人之间“明示自甘风险”约定的效力，⁽³⁶⁾甚至不在意当事人的约定是否违背了强行法的规定。

(33) Monteiro J., *Behavioral Investing: A Practitioners Guide to Applying Behavioral Finance*, John Wiley & Sons, 2009, p. 1—6.

(34) Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679*, February 6, 2018.

(35) *MacPherson v. Buick Motor Co.*, 217 N.Y. 382, 111 N.E. 1050 (1916).

(36) 例如，*Get off v. Von Lengerke Buick Co.*, 14 N.J. Misc. 750, 187 Atl. 539 (Sup. Ct. 1936)，法院认可了销售合同中禁止消费者就庞蒂克汽车缺陷提起诉讼的约定；*Sonnenberg v. Nolan Motors, Inc.*, 2 Misc. 2d 185, 36 N.Y.S.2d 549 (Sup. Ct. 1942)，法院支持了道奇汽车销售合同中买方放弃纽约州法律中所有默示保证的条

“明示的自甘风险”以受益方对亏损方的合理补偿作为优于最大多数人的最大幸福的法益，在特定的历史时期具有合理性。但在人工智能时代，这种抑“帕累托最优”、扬“卡尔多—希克斯效率”的做法是十分危险的。因为历史经验表明，技术进步从未促进社会福利的雨露均沾，而是屡屡架空权力专属原则和正当程序原则，轮番算计普罗大众以最大限度实现监管套利。技术黑箱配合用户知情同意，为数据控制者为所欲为大开方便之门，因特定理由搜集的多渠道信息极有可能被别出心裁地用于不可告人的目的。技术成为了少数人牟取暴利的工具，而多数人无力反抗。一言以蔽之，“知情同意原则与人们的风险认知和数据利用的客观规律存在矛盾与冲突。”⁽³⁸⁾更何况，“获解释权”“被遗忘权”“脱离自动化决策权”之间本存在“层层递进的行权诉求”，以合同约定和当事人同意作为“脱离自动化决策权”的豁免事由，无异于直接剥夺了本就身处信息弱势地位的数据主体最有力的救济手段，彻底让数据主体与自动化世界的每一次联结都变为自甘风险。数据主体对合同和协议的不感冒早已成为业内心照不宣的秘密——关涉自身重大利益的条款，数据主体未必看、看了未必懂、懂了也未必有能力进行反驳，部分自动化决策服务动辄对应着成百上千页的《用户须知》，某些银行的人工智能投资顾问服务甚至要求客户对着屏幕朗读经过巧妙包装的特定免责条款并对客户进行录像。

实际上，从20世纪60年代开始，美国法院就已经开始对产品责任领域明示的自甘风险实践进行纠偏。伴随着风起云涌的民权运动，美国产品责任法从疏忽责任和担保责任逐渐转变为严格责任，⁽³⁹⁾以无过错责任的追责框架将产品产销各个环节的相关人员都纳入其中，自甘风险规则只在破除产品严格责任的极少例外情形中得以适用。本质上，自动化决策可以被看作是数据控制者向数据主体提供的一种服务，是聚合数据的一方通过人力、物力、资源和环境所组成的结构系统来销售和实际生产及交付的、能被提供数据的另一方购买和实际接收或消费的产品，是以非实物形式存在的劳动成果。问题是，若不加区别地对此类特殊产品或服务的生产者，即数据控制者或处理者，套上严格责任的枷锁，将必然阻碍科技创新的步伐。数据是现今最重要的生产要素已经成为世界共识，各国为争夺人工智能技术制高点可谓不遗余力，不宜将过高的履责成本强加于肩负更高使命的科技类企业。相比起成本过高的严格责任，过错责任能够更好地平衡个体保护和科技创新的利益冲突——在避免责任缺位的同时鼓励技术发展。⁽⁴⁰⁾笔者认为，衡量数据控制者有无过错的审查标准有二。其一，数据控制者是否通过持续的信息披露反复对数据主体作出真实、准确、完整的“风险提示”。数据控制者在信息披露中应对自动化决策可能对数据主体造成的潜在伤害列出清单，对于清单之外的妨害行为不得以自甘风险为由阻却数据主体行使“脱离自动化决策权”。其二，数据的采集和处理是否被用于经过数据主体同意的正当、唯一目的。为使判定标准进一步明晰，还应当通过专门立法的方式，构建企业数据利用与分享准则，对相对敏感的隐私信息采取加密储存、吞吐限制、访问分级等安全措施。违背相关规定的數據使用行为，将不得享有“脱离自动化决策权”的豁免。

同样有必要关注的是，近年来，公共卫生突发事件频发，以行政手段促进社会共律对社会个体普遍自律不足的必要补充无可厚非，但要时刻警惕由此可能对用户隐私和个人生活带来的变本加厉的骚扰。⁽⁴¹⁾据报道，杭州拟推出“渐变色”健康码，通过整合电子病历、体检信息甚至生活习惯等个人数据，建立个人健康指标评价体系。不仅如此，杭州健康码应用还打算将楼道、社区和企业的群体评价体系纳入自动化决策范围，打造全知全悉的健康生态服务圈。从健康码向“万能码”之拔擢，仁者见仁、智者见智。单就数据主体在此类场景中是否享有“脱离自动化决策权”，笔者认为，这将取决于以提升公共治理为目的的自动化决

款。

(37) 例如，*Dutsu Ford Motor Co. v. Cullum*, 96 F.2d 1 (5th Cir. 1938), cert. denied, 301 U.S. 627 (1938)，法院支持了福特公司在销售合同中将经销商的质保责任限定于更换零件并禁止消费者解除合同的做法，尽管彼时的《统一销售法》允许消费者就重大缺陷解除销售合同。

(38) 姚佳：《知情同意原则抑或信赖授权原则——兼论数字时代的信用重建》，载《暨南学报》2020年第2期。

(39) 参见张骥：《在效益与权利之间——美国产品责任法的理论基础》，载《中国法学》1997年第6期。

(40) 高丝敏：《智能投资顾问模式中的主体识别和义务设定》，载《法学研究》2018年第5期。

(41) 许可：《从规制科技反思“健康码”》，载《FT中文网》2020年4月1日。

策,是否同时为实现公共利益所必须。一般来说,个人数据经过结构化改造后在社会运营网络中流通,便摆脱个人掌控范围成为公共数据库的有机组成部分,对任何自动化决策的脱离都将“牵一发而动全身”。从信息的流动角度来看,如果毫无保留地允许数据主体享有对其数据的绝对权利,无疑将侵蚀公共治理领域自动化决策的全面性和公正性,破坏国家治理体系现代化的完整生态。因此,在涉及公共利益的自动化决策场景中,“脱离自动化决策权”的适用只能是例外而非一般原则。此种认知同 GDPR 第 23 条所囊括的限制情形相符:“出于国家安全、国防、公共安全和公共利益等考量,欧盟各成员国可以通过立法手段限制第 12 至 22 条、34 条以及第 5 条所规定的责任与权利范围。”GDPR 第 89 条也对此作出了相应规定:“因为公共利益、科学或历史研究或统计目的而进一步处理数据…可以对数据主体的防护措施和权利进行克减。”总之,公共利益与个人利益总是存在此消彼长的张力,这是立法者长期努力调和却难以平息的固有矛盾,公共卫生领域尤其如此。但倘若能在变动不居的法益环境中,因时制宜做出彰显社会整体价值的伦理判断,那便是以看得见的方式实现了以个人健康促进社会健康的最大正义。

(三) 超越 GDPR: 保护人类自主性的额外可能

在“获解释权”和“被遗忘权”的基础上层层递进,以更高程度的人类干预取代较低程度的人工介入,GDPR 界定了数据主体在请求获得解释、表达异议和反对之外可以暂时请求脱离自动化决策的权利,并与其他两种基本权利一道构成了保护数据主体权益的、三权并立的联动体系。上述探讨,一直未跳出 DPID 与 GDPR 为“脱离自动化决策权”框定的篱藩,此等认知限制势必带来法律移植方面的困难,因为独立于原框架的“脱离自动化决策权”将再次面临内涵的不稳定性和外延的模糊性。

阿芬那留斯指出,人们惯于在经验中“摄取”思想、感情和意志,导致经验分裂为外在和内在,并被迫对其作主体和客体的划分,造成的结果是“‘实在’被颠倒黑白。”⁽⁴²⁾以消除“摄取”为要领,公共治理领域的学者们试图寻求应对自动化决策的“纯粹经验”,由此生发出数据主体所应当具有的“实在权利”。这些“实在权利”有着显著詹姆斯“实用主义”的特征,剔除一切被概念思维加工的权利属性,把逻辑的一致性和证实包括在实际的功利观念之中,或多或少还原了“脱离自动化决策权”的“清静”和“原初”状态。

回归行政法的“自然正义观”,英国学者提出了“陈情权”的概念,主张保障因某项决策致使自身权益受损的个体享有获得公平聆讯以及请求行政复议的权利。尽管人工智能模拟人类情感的能力与日俱增,但短期内仍然无法像自然人一样表达出有温度的怜悯、同情和关爱。行使陈情权的目的是迫使政策制定者以“同为自然人”的立场对被决策者的实际遭遇感同身受,凭借其制定出更人性化的政策和法律。在自动化决策领域,美国一直是陈情权的践行者,经常通过听证会的方式对自动化决策失灵予以纠正,让受害者彻底“脱离自动化决策”。例如,2018 年 1 月,在缺乏科学论证的基础上,美国纽约州率先利用预警保护系统来鉴别具有严重家暴倾向的父母,并对高危儿童进行隔离保护。该系统不仅没有大幅降低虐童事件的发生概率,还造成了上万正常父母同其子女的被隔离,上线不到半年就引发众怒。在数次听证会后,当局撤销了此前的自动化决策,并放弃了预警保护系统的使用。2019 年 4 月,一场针对密歇根州政府利用人工智能技术审查社会救济申请的听证会,引发了逾四万居民对州政府的集体诉讼。据悉,用于反欺诈的米达斯系统出错率高达 93%,致使数万低保申请者被无端惩戒。集体诉讼一旦获胜,全体密歇根居民将彻底脱离米达斯系统的自动化决策,而密歇根州政府将陷入“破产”的境地。⁽⁴³⁾

因此,陈情权虽然能够取得较好的“脱离自动化决策”的效果,但行使该权利必然伴随极高的法律成本,且需以成熟的集体诉讼等制度作为配套,有为求正义矫枉过正之嫌。波斯纳指出:“法律并没有试图在最高程度上实施任何道德准则……全部承诺的法律实施成本会超过其收益。”⁽⁴⁴⁾可以想象,一旦陈情权的实施成本超过社会所能承受的临界值,行使该权利将失去原本的正当性,因为维护一种正义不应以牺牲另一种正义作为代价。更有可能的情况是,在实施成本逼近临界值之前,陈情权的行使就已经自然出现了传播边际效用急速递减的情况。我们身处一个信息爆炸的时代,信息增量无限而注意力极为有限,人们经常出现上

(42) Avenarius R., *Kritik Der Reinen Erfahrung*, O.R. Reisland(1907).

(43) 唐林垚:《遏制人工智能算法的公共妨害》,载《法制日报》2020 年 1 月 7 日。

(44) [美]波斯纳:《法律的经济分析》,蒋兆康译,中国大百科全书出版社 1997 年版,第 341 页。

一分钟还在为某个新闻义愤填膺、这一分钟就已经为某条信息拍手叫好、下一分钟再因某篇软文泪流满面的涣散状态。面对人工智能迅猛发展，人们的心理状态也必然会在“日居月诸，胡迭而微”中从一开始的毛骨悚然逐渐过渡至视若无睹乃至最终安之若素。信息传播的规律表明，只有第一个、第二个受害者的陈情能够引发较大的公众关注，此后相同受害者的再次陈情非但效果将指数级递减，甚至还会因为反复陈情引发受众内心嫌恶的负面效果，就像一句流行歌词中所表达的那样：“谁会不厌其烦地安慰那无知的少年？”

公共治理领域的学者还指出，资讯类的自动化决策进一步将人们本来就相当涣散的注意力彻底碎片化，预测算法和偏好算法更是以“个性化定制”之名试图操纵人们的注意力。弗瑞施曼和塞林格在《重塑人性》一书中指出，如果人类想要在21世纪继续保持代理权和自主意识，就需要在法律层面承认“离线权（the right to be off）”。^{〔45〕}所谓“离线权”，顾名思义，即脱离强加于自身的自动化决策正常生活的权利，和“脱离自动化决策权”有异曲同工之妙。弗瑞施曼和塞林格并没有为离线权指明清晰的行权路径，离线权因此被部分学者从实操层面解读为“注意力保护权”，^{〔46〕}即脱离自动化内容推送和营销类用户侧写的权利。范围限缩的解读方式虽然极大地束缚了离线权本来的综合意涵，但就其为普罗大众所灌输的拒绝被轻易掠夺数据红利的法治观念和自我保护意识而言，意义不可小觑。

自动化决策裹挟着人们自然而然地接受构建于一个个信息茧房之上的“后自主社会”，并以半将就、半威胁的方式推动人们步入“后真相时代”，亚马逊公司前首席科学家维真德在《数据为民》一书的开头给出了一个重要的前提认知：“时间已经认识到，隐私和自主不过是一种错觉。”^{〔47〕}有鉴于此，陈情权也好、离线权也罢，任何一种进路只要能够实现“脱离自动化决策”的效果——哪怕是片刻脱离、部分脱离——也能让面对自动化决策霸凌早已退无可退的人类有固可守，有险可依。从这一点来看，未来数据控制者所需要履行的，恐怕不是一星半点的社会责任那么简单。

结 语

“人生而自由，却无往不在枷锁之中”。针对自动化决策的法律应对重点在于，警惕打着促进自由幌子的技术工具，被用来生成更多难以挣脱的枷锁。马克思指出，法律乃强权意志的实证主义法学认知仅道出了部分真相，法律既是人类理性所理解的道德准则的表达，也是根植于社会共同体的历史价值和规范的习惯的派生物。^{〔48〕}欧盟立法的经验和教训表明，多权并立的联动体系不一定能够确保个人意志可以独立、自主地依据普遍承认的法则和习惯同他人的意志相协调，反而可能会因为部分不那么重要的权利对另一些真正重要的权利的遮蔽和掣肘，阻碍实质公正与有效救济的兑现。在人工智能领域，各国立法者创造出了过多的于事无补的完型与范式，来与广为人知的毫无意义的知情同意这一比喻相配。技术膨胀出的权力越是绝对，人类的尊严就越受到威胁。“获解释权”和“被遗忘权”至多只能争取到局促的看似自主和有限自由，而只有在决定自身利益的关键事务上享有积极参与的权利，人类才有可能维持和延续在知识、经验和道德上的卓越。在这个意义上，“脱离自动化决策权”实乃关系人类共同命运的最后一道屏障。

“法者，治之端也”。回到开头的例子，如果健康码对个人的风险等级判定有误，致使持码人通行受阻，持码人拨打12345发起纠错申请，即可视为“脱离自动化决策权”。此时，县级防控指挥部既非简单地对持码人进行解释或心理疏导（让其接受错误的结果）、也非直接擦除或更正个人数据（对决策结果在所不问），而是通过自动化决策求得的结果进行反向检视，找到出现差错的环节，并对错误结果进行系列纠偏。如果没有环节出现差错，县级防控指挥部也应对判定结果进行自然人审查，并重新形成判定结果。当然，这是高度理想化的情况，但也反映出在自动化决策逐步成为国家治理体系和治理能力现代化重要组成部分的必然趋势下，我国未来《个人信息保护法》中引入“脱离自动化决策权”的必要性。民法典人格权编对个人信息处理确立了合法、正当、必要和不过度的原则，对信息处理者采取技术措施和其他必要措施保障信息主体权益的各项要求和规定，为“脱离自动化决策权”不再只是

〔45〕 Frischmann B, Selinger E, *Re-Engineering Humanity*, Cambridge University Press, 2018, p. 96—102.

〔46〕 Tran J., *The Right to Attention*, Indiana Law Journal, Vol.91, p.1023 (2016).

〔47〕 Andreas Weigend, *Data for the People, How to Make Our Post-Privacy Economy Work for You*.

〔48〕《马克思恩格斯全集》第4卷，人民出版社2006年版，第121—122页。

一种虚幻承诺预留了空间。⁽⁴⁹⁾ 下一步工作是透过解释或针对典型性事件的具体化，将既存规则和立法顾虑之间的可能冲突加以澄清，发展出切实可行、同时又具备弹性的行权路径，让“脱离自动化决策权”成为我国人工智能领域良法推进善治、将制度优势转化为国家治理效能的人权保障。

The Illusory Promise of the Right not to be Subject to Automated Processing

Linyao TANG

Abstract: The very nature of GDPR Article 22 is a right, not a prohibition; yet the right not to be subject to a decision based solely on automated processing conveniently confuses with the right to erase, the right to be provided with an explanation and the right to be forgotten. There are two critical paths to understand this right: (1) Recital 71 of GDPR has stipulated three fundamental rights for the data subject to protect their legitimate interests; in between these three rights there are no reciprocal causation but degree ascension. (2) the right itself varies in different scenes, it may mean not to be subject to data collection, not to be subject to profiling or not to be subject to platform interaction according to different technological contexts. In order to break the illusory promise of the right not to be subject to automated processing, the Assumption of Risks alike exemption conditions should give way to statutory permit aiming to increase public well-being. To think outside the GDPR's box, the right to be heard and the right to attention both revealed practical path for the application of the right, so far as the social cost could be maintained in a proper manner.

Key words: the Right Not to Be Subject to Automated Processing; the Right to Erase; the Right to Be Provided with An Explanation; the Right to Be Forgotten; Exemption Rule; Assumption of Risks

(49) 民法典第 1035 条，第 1038 条。民法典中信息控制者和信息主体的提法是非常先进的，相对于 GDPR 中数据控制者和数据主体的概念更符合人工智能技术发展的现状；如今机器学习模型不仅可以处理数据，还能对文本和图像等更复杂的信息进行“抓取式分析”。参见王利明：《彰显时代性：中国民法典的鲜明特色》，载《东方法学》2020 年第 4 期。