

# 主权的嬗变

## ——数字化“魔兽世界”与法律秩序创新

季卫东\*

### 目次

- |                       |                        |
|-----------------------|------------------------|
| 一、引言：国内秩序与国际关系的现代化标志  | 四、因“算法利维坦”而复兴的国家主权     |
| 二、作为利维坦的主权国家及其制约      | 五、结语：算法利维坦与数字怪兽互动的程序公正 |
| 三、20 世纪末以来推动主权蜕变的四股力量 |                        |

**摘要** 本文通过研究最近三十年间经济和信息的全球化、巨型网络平台崛起、隐匿在区块链中的主权个人、摆脱特征设计的大模型怪兽等重要的现象和趋势，分析了国家主权的相对衰退与强势复兴并行和相反相成的悖论以及数字化转型过程中利维坦与各种数字怪兽之间展开的互动关系。正是主权衰退这一侧面，导致了欧洲的“数字主权”概念和中国的“主权区块链”构想的提出；而主权复兴这一侧面，导致了北美的“监视社会”理论和中国的“电子点穴”隐喻。无论如何，在人工智能驱动的时代，“算法利维坦”正在成为国家和法的理论不得不面对的极其强大的支配性权力。为了防止其失控和滥用，需要通过全球数字契约等方式凝聚基本共识，并通过法律的和技术的正当程序对其进行制度化制约——这正是数字法治的要义。

**关键词** 数字主权 平台治理 大模型怪兽 主权个人 算法利维坦 全球数字契约

## 一、引言：国内秩序与国际关系的现代化标志

西欧式现代化的一个重要象征是形成了主权国家体制。

回顾相关的历史背景，正是以 1494 年爆发的意大利战争为契机，神圣罗马帝国与法兰西王国之间开始围绕欧洲霸权进行了长期的频繁战争，并使英格兰等众多国家卷入其中。在混战之中，各国为了充分调度军费和兵器，不得不加强征税能力、提高行政效率、完善科层制组织机构，进而推进国内市场的统一化以及支配权的集中化。为此，需要划清国境边界、确立只有君主才能代表国家的原则和秩

\* 上海交通大学文科资深教授、凯原法学院讲席教授、教育部“长江学者”特聘教授、法学博士。

序。在这个过程中关于主权的概念逐渐产生,而这种主权就构成近代民族国家的基础以及标志。

复数的主权在激烈竞争中都采取了以富国强兵为首要目标的重商主义政策,并且出现了被称为“绝对王政”的强有力的国家统治体制。正是在列强争霸和资本积累的大变局之下,法国的让·博丹创立了关于主权国家具有不受外部控制的独立地位、主权者对公民享有凌驾于法律之上进行决断的最高权力、统治权力不可分割和转让的主权理论;〔1〕英国的托马斯·霍布斯以让自由的个人摆脱“所有人对所有人的战争”那种自然状态为理由,特别强调了国家主权对内的绝对性、至高无上性;〔2〕荷兰的胡果·格劳秀斯则以这种国家主权观为前提,主张把自然法理论适用于国家与国家之间的关系(国际冲突其实更接近所谓“所有人对所有人的战争”的状态),以缓和对外残杀的烈度,建立一种合乎理性、追求和平的国际法秩序乃至海洋秩序。〔3〕

因此,在现代化的全球地壳运动中,主权成为一个民族国家存续的基础,也成为统治者的排他性权利以及最高权威的判断标准。

## 二、作为利维坦的主权国家及其制约

国家因主权理念而享有几乎无限的权力,俨然成为能够支配一切的魔神。霍布斯把这种主权国家称为“利维坦”——至尊的巨灵。他是这样阐释主权国家的底层逻辑的:

把大家所有的权力和力量托付给某一个人或一个能通过多数的意见把大家的意志化为一个意志的多人组成的集体。……像这样统一在一个人格之中的一群人就称为国家,……这就是伟大的利维坦(Leviathan)的诞生,……用一个定义来说,这就是一大群人相互订立信约、每人都对它的行为授权,以便使它能按其认为有利于大家的和平与共同防卫的方式运用全体的力量和手段的一个人格。〔4〕

显然,如此主权国家宛若所有权力的集装箱。在这里,法律可以理解为主权者的命令,其也构成社会共同的行为准则。霍布斯甚至还认为,要防止回到互相争斗的状态,同时也为了落实黑死病疫情防控的应急方案,国家的权威应该是不可动摇的、绝对的,为此不能承认宗教、思想以及言论上的自由。〔5〕

不言而喻,霍布斯展示的是一幅“无限政府”的设计图。在这里,契约原理事实上只是存在于个人相互行为当中,并不适用于国家与所有个人之间的关系。因此,霍布斯关闭了通往弗雷德里克·梅特兰后来所阐述的那种所谓双重信托的思路。〔6〕当然,他更没有洛克的“有限政府”的理念,没有承认人民对暴政的抵抗权。与霍布斯的国家主权观非常类似的,是中国战国时期秦国法家商鞅(B.C.

〔1〕 具体论述参见[法]让·博丹:《主权论》,[美]朱利安·富兰克林编,李卫海、钱俊文译,北京大学出版社2008年版。

〔2〕 有关文献不胜枚举,例如汪仲启:《绝对主权的逻辑和裂缝——论霍布斯的国家学说》,载《学术月刊》2015年第12期,第109—119页。

〔3〕 参见熊光清:《限制战争——格劳秀斯主权理论新解》,载《太平洋学报》2014年第9期,第20—27页。

〔4〕 [英]托马斯·霍布斯:《利维坦》,黎思复、黎廷弼译,商务印书馆2013年版,第131—132页。

〔5〕 同上注,托马斯·霍布斯书,第18章;另外,关于医疗行政与近代主权国家“利维坦”之间的关系,参见蓝江:《疫病、生命政治与现代主体的诞生——从霍布斯到福柯的治理体系》,载《求是学刊》2020年第3期,第1—10页;关于近代国家权力通过防疫和医疗进行规训及社会统治的具体阐述,参见[法]米歇尔·福柯:《临床医学的诞生》,刘北成译,译林出版社2011年版。

〔6〕 参见[英]F. W.梅特兰:《国家、信托与法人》,[英]大卫·朗西曼、马格纳斯·瑞安编,樊安译,北京大学出版社2008年版。

390—B.C.338)等关于“君权至上、治国作一”的思想。正如福山指出的那样,中国从秦代开始就形成了具有强大执行力的国家,并且具有某种早熟的现代性。他认为,早在秦朝时期,中国就独自创造了韦伯意义上的现代国家,即中国成功地发展出了一个中央集权的、统一的官僚政府,终结了封建制。<sup>〔7〕</sup>

不得不承认,拥有无限权力的主权利维坦(the Sovereign Leviathan)对个人而言其实是非常危险的,至少很容易窒息创新的活力。为了制约那个日益强大的主权利维坦,需要特别加强对自由和权利的制度保障。借用伊曼努尔·康德的表述,就是需要通过契约的理念把个人的天赋权利作为正义的基石并将其嵌入国家权利体系的框架之中。<sup>〔8〕</sup>因此,在西欧,在启蒙思想和社会革命运动的影响下,1689年英国通过了“权利法案”,宣告议会主权的诞生;<sup>〔9〕</sup>1789年法国颁布了“人权宣言”、美国让联邦宪法正式生效,二者分别明确了“人民主权原则”和“三权分立”的政治框架设计。<sup>〔10〕</sup>

在这个过程中,主权优越性的诉求与主权合法性的诉求互相碰撞和交融,形成了现代民族国家的一个带有辩证法色彩的悖论:如果国家把公民当作政治主体(subject),那么公民将乐于服从(subject);反过来,如果国家把公民当作支配客体(object),那么公民就会选择抗议(object)。<sup>〔11〕</sup>总之,通过以神圣不可侵犯的产权和人权为双轮运行的现代法治来限制国家权力、实现平等的自由并以此作为国家权力的正当性根据——这种理念的逐步践行就成为西欧乃至各国制度变迁的基本趋势。尽管如此,从17世纪后期到20世纪后期的三百年间,现代法治秩序在整体上还是以主权国家体制为基础。

但另一方面,基于历史的原因,当西欧式现代法治秩序出现在中国面前时,其所具体呈现出来的物象却是不平等条约和治外法权,其并不是价值中立、公正不偏的。因此,与日本类似,中国法制现代化的动因并非来自内部自发的需求,而是基于外部冲击的压力,这就很容易引起非常强烈的受害者意识以及坚决捍卫国家主权的倾向。实际上,第二次世界大战结束后摆脱殖民地统治的大批新兴国家,同中国一样,都具有极其浓厚的独立自主观念,把推进现代化与维护国家主权密切联系在一起。换个角度来看,在国际力量对比关系失衡的状态下,也只有主权才能构筑民族认同感和文化传统多样性的坚固护墙。主权国家体制一旦分崩离析,弱势国家主张的所谓价值观多元共生或者文明之间的对话云云就很容易流于空谈,甚至反倒会促成萨缪尔·亨廷顿所预言的那种“文明冲突”。<sup>〔12〕</sup>因此,中国以及第三世界的很多国家,维护国家主权的诉求与通过法治来驯化主权利维坦的诉求之间存在某种微妙而复杂的紧张关系,二者未必总能保持适当的平衡。在很多场合,主权都会显出压倒一切的优势和力量。

### 三、20世纪末以来推动主权蜕变的四股力量

德国纳粹政权残酷虐待乃至大规模屠杀本国犹太人的事实及有些国家出现的内部种族净化

〔7〕 参见[美] 弗兰西斯·福山:《政治秩序的起源——从前人类时代到法国大革命》,毛俊杰译,广西师范大学出版社2014年版,第七章:战争和中国国家的兴起。

〔8〕 参见张恒山:《由个人意志自由到公共意志自由——康德的权利学说》,载《环球法律评论》2013年第3期,第5—17页。

〔9〕 关于英国议会主权原则的嬗变,参见项焱:《英国议会主权研究》,中国社会科学出版社2010年版。

〔10〕 关于法国和美国主权在民和国家制度设计的异同,参见刘瑜:《两种权利观与美法革命道路——基于美法权利宣言的文本分析》,载《中国人民大学学报》2022年第1期,第82—94页。

〔11〕 这种辩证关系在日本把臣民转化为公民的现代化过程中也有典型的表现,相关内容参见[日] 牧原宪夫:《在“客体”与国民之间——近代民众的政治意识》,吉川弘文馆1998年版,第15—16页。

〔12〕 关于亨廷顿对当代全球政治背景下跨文明冲突的论述,参见[美] 萨缪尔·亨廷顿:《文明的冲突与世界秩序的重建(修订版)》,周琪等译,新华出版社2018年版,第九章:多文明的全球政治。

和政治肃清的现象,充分揭示了主权压抑人权的可能性以及苛政的危害。因此,第二次世界大战结束之后,人权保障国际化的意义就得到了广泛认知,这在一定程度上埋下了主权相对化的伏笔。从20世纪70年代起,美国就开始把人权作为外交的重要主题之一,到90年代其还进一步提出了“人权无国界”的口号,积极干预别国内政问题。<sup>[13]</sup>另一方面,从1989年的苏联东欧剧变开始,投资、贸易、金融以及信息等方面的全球化洪流迅速席卷各国,与此相应主权也渐次进一步相对化和转化,出现了以WTO(世贸组织)和WWW(World Wide Web,万维网或互联网)为两根支柱的世界法治秩序构想。立足当下来回顾历史,可以发现推动主权蜕变的实际上主要有四股力量,即1990年代的积极人权外交、2000年代的网络平台崛起、2010年代的电脑算法助推、2020年代的大语言模型和生成式人工智能。本文在这里把讨论的射程仅限于后面三种情形,即互联网、计算机以及人工智能,这三者都与数字信息技术的发达和应用性普及有关。因而也可以说,数字化转型正在动摇现代主权国家体制。

### (一) 来自“网络平台怪兽”的挑战

以1990年诞生的互联网为基础,随着个人电脑和手机功能的强化,人们之间的交流越来越活跃、相互关系也越来越密切。为了促进人们在互联网上的沟通和其内容丰富化,美国国会在1996年通过了新的《通信规范法》,以第230条规定免除了互联网平台对用户发布信息的责任(诽谤责任豁免框架);<sup>[14]</sup>千禧年之交美国又制定了《美国数字版权法》,为搜索引擎、网络储存、在线图书馆等提供了版权责任避风港(通知加移除)。<sup>[15]</sup>美国法院则通过法律的宽泛解释,把责任豁免框架扩张到了其他领域,使美国网络平台在初期获得了极其广阔的发展空间。把主权豁免原则适用于数字信息技术,特别是网络平台上,似乎暗示了这些平台具有类似历史上东印度公司或当代巨型跨国公司那样的“准主权性”。正是在这样的法律制度条件下,网络上的互动非常活跃并导致各种数据不断形成、积累进而产生应用价值,终于引发了真正的信息革命。与此同时,通过“互联网+”的方式又不断发展出物联网、移动通信、大数据、云计算、人工智能系统、机器学习、机器人技术、自动化等等,这一切又造成了产业结构和社会形态的深刻变化,带来数字经济的各种创新。

在2004年的一次头脑风暴论坛上,奥莱利传媒公司创始人和总裁蒂姆·奥莱利首倡信息双向交流的“互联网2.0”概念,试图与此前的那种信息单向通行的互联网进行切割,强调互联网2.0具有“用户不仅可读而且还可写”等七种基本特征,并且还有一种新的功能:促进企业借助网络计算平台来与用户加强互动,根据大数据对用户行踪和偏好的分析、预测以及量体定制的算法来改进营销和服务。这种算法助推实际上也构成一种强大的权力,在有些场合甚至还能独断专行并大幅度提高经济的效率。显然,这是关于万物互联互通的企业视角。

但是,随着数据价值被发现和其商业化应用,互联网企业开始有意识地运用基于电脑代码和技术规格的规制权来增强自身的商业竞争力,其运营模式也从原来的信息中介机构转换成了线上

[13] 中国政府和学界都对“人权无国界”论持批判态度,强调主权是人权保障的基础。相关文献例如,徐崇温:《人权与主权》,载《中国社会科学院研究生院学报》1992年第6期,第42—54页;中国人权研究会:《论人权与主权:兼驳“人权高于主权”论》,当代世界出版社2001年版。有些有代表性的专家还强调生存权、发展权是首要人权,例如李步云:《什么是人权——根据李步云教授讲座录音整理》,载《东南法学》2019年第1期,第3—7页;李步云:《论人权》,社会科学文献出版社2010年版。关于主权、人权以及霸权的三元互克关系,参见季卫东:《从主权的双重结构看中国与世界的互动关系》,载《二十一世纪》2000年2月号(总第57期),第87—96页。

[14] 参见黄宇帅:《美国网络治理追踪:〈通信规范法〉第230条的历史、现状与未来》,载《网络信息法学研究》2021年第1期,第203—218页。

[15] 参见万勇:《著作权法强制性过滤机制的中国选择》,载《法商研究》2021年第6期,第184—196页。

社群管理机构；其结果是巨型网络平台迅速崛起，终于导致数字经济出现了力量对比的悬殊和市场垄断现象。面对这些电商和科技的“平台怪兽们(platform behemoths)”，在网络空间反垄断的呼声一浪高过一浪。另外，数据的广泛采集、分析以及营销，也有威胁个人信息安全以及侵害隐私的风险，并且引起了关于数据所产生经济价值的所有关系如何厘清、分配是否公正、如何保护个人信息安全和隐私等法律问题的讨论——欧洲以及其他国家都开始酝酿制定数据治理和人工智能治理的规范，从2016年起该类规范渐成体系；在美国，从2008年起法院开始对网络沟通鼓励机制进行收缩解释，在2019年国会众议院司法委员会又针对四大巨型网络平台科技公司“GAFA”进行反垄断调查。法律界进行了相关的热烈讨论，矛头直指《通信规范法》第230条的责任豁免框架。到2020年，关于修改法律的舆情达到沸点，并逐步影响到后来的立法提案和司法判断。<sup>〔16〕</sup>

在互联网2.0阶段，中国也采取了鼓励数字化沟通和互动的法律政策取向，使大数据不断积累并产生经济价值，从而形成了以腾讯、阿里巴巴、抖音、百度、携程等为代表的网络平台巨头。网络平台甚至还能行使自治权并代替政府对平台上的交易活动进行监管，进而通过平台治理改变既有的国家权力结构及权力组成部分（例如货币发行权、信用评价和排序、制裁方式等）之间的关系。例如网络平台阿里巴巴旗下的“芝麻信用”评分体系，与“支付宝”使用的信用履历挂钩，不仅会对使用者租车、住酒店、买保险、购房、出国旅游等产生影响，甚至执法和司法机构都来寻求合作。最高人民法院系统与芝麻信用签署合作协议，目的就是要把资产分布大数据以及消费行踪大数据与诉讼保全、判决执行等司法举措的实效性联系起来。<sup>〔17〕</sup>但是，从2018年起，这些平台市场支配地位的影响、平台权力化的倾向以及算法助推带来的个人信息安全问题开始受到广泛关注，法律政策也开始调整，试图限制“平台怪兽们”的市场行为和权力化倾向。2020年，中国政府对大型网络平台采取了断然的反垄断举措，加强了对平台的法律控制。但是，如何借鉴美国的经验和教训适当重构权力化平台的相应责任与技术避风港——在某种意义上也不妨称之为“技术豁免(technology immunity)”——之间的关系，对中国而言仍然是一个重要的法学课题。<sup>〔18〕</sup>

总而言之，回顾“互联网+”的发展过程，可以发现平台治理方式对既有的国家权力结构以及法治的制度安排形成了挑战和修正，自然而然脑海中会浮现出一个主权利维坦与一群富可敌国、以各种方式积极分担国家职责的平台怪兽们互相博弈、进行管控力竞赛的意象。但是，为了促进创新与发展，国家也会支持平台经济，试图形成一种合作共赢机制。<sup>〔19〕</sup>

## （二）区块链、个人主权以及“算法权力怪兽”

实际上，如果主权国家与平台企业联手合作，还会形成空前强大的“算法利维坦(Algorithm Leviathan)”联盟，通过社会交往中枢与经济交易中介之间的合力作用来极大地压缩个人自主选择的空间。为了避免这种事态，特别需要寻找一种个人本位的对等性互联网形态，以严格保护作为各种自由基础的隐私。

2008年11月1日，一个号称“中本聪(Satoshi Nakamoto)”的神秘人物通过加密邮件列表发

〔16〕 参见《美最高法院开审大科技共识命运：是否取消27年互联网“保护伞”？》，载澎湃新闻网2023年2月24日，[https://www.thepaper.cn/newsDetail\\_forward\\_22061908](https://www.thepaper.cn/newsDetail_forward_22061908)。

〔17〕 参见《最高法联手芝麻信用网络惩戒失信见成效》，载最高人民法院网2015年12月20日，<https://www.court.gov.cn/zixun-xiangqing-16351.html>。

〔18〕 参见沈伟伟：《技术避风港的实践及法理反思》，载《中外法学》2023年第4期，第1—17页。

〔19〕 例如《总理召开的这场座谈会，再向平台企业释放积极信号》，载澎湃新闻网2023年7月13日，[https://www.thepaper.cn/newsDetail\\_forward\\_23827347](https://www.thepaper.cn/newsDetail_forward_23827347)；《坚定信心练好“内功”，平台经济大有可为》，载中国经济网2023年7月14日，[http://views.ce.cn/view/ent/202307/14/t20230714\\_38631259.shtml](http://views.ce.cn/view/ent/202307/14/t20230714_38631259.shtml)。

表了关于点对点电子通货系统和交易媒介——比特币的论文,展示了区块链技术的基本原理。<sup>[20]</sup> 2009年1月3日,比特币从构想变成了现实。从此以后,互联网上的交易可以借助区块链技术的应用而在有意愿的客户之间直接进行,而不必让特定的第三方来独占数据进行处理。这样一来,互联网就转换成无需中介和中枢的分散化结构,因而每个人就都有可能收回自己的隐私控制权,个人终端就构成出发点和中心点并与整个信息关系空间进行沟通和互动。<sup>[21]</sup> 在中国,如果十四亿人都构成终端,都编织自我中心的蜘蛛网,都塑造一个个“小宇宙”,那将是何等壮观而又复杂的景观!

利用区块链技术,不仅可以界定和保障个人对数据经济价值享有的合法权益,还可以创建各种各样的目标多元化泛用平台。例如2013年末公开设立的以太坊,就是最广为人知的以区块链为基础的智能合约平台,其可以发挥多层多样的功能。因此,嘎文·伍德在2014年4月17日推送的一篇文章中,重新发掘并创新了互联网3.0的概念。<sup>[22]</sup> 在他看来,互联网3.0是一种以区块链为基础的分散型互联网,也是用户享有自主权和相互信任的互联网,因而还是一种身体化的人人互联网。实际上,又过了六年,直到2020年,在分散金融(DeFi)方兴未艾、非同质化代币(NFT)出现爆发式增长之后,人类才真正迈进互联网3.0的时代。

更值得重视的是,点对点的区块链技术、智能合约平台、加密资产以及网络信任等因素叠加起来,实际上在元宇宙或者多重宇宙中确立了在线用户的自主管理身份(Self-Sovereign Identity, SSD),或者说形成了“具有主权的个人(The Sovereign Individual)”。<sup>[23]</sup> 在19世纪末,尼采说“上帝已死”、信仰崩溃。在20世纪后叶,拉康和福柯以可谓“后现代主义”的名义宣告了“主体的死亡”、个人成为在虚无中飘动的无根浮萍。但是,到2010年代,启蒙运动所创造的那种主体、那种具有先验本质的个人又活过来了,而且是以一种喧嚣的、肆意的方式。在很多场合也不妨借助法学界已经确立的“消费者主权”的概念来表述这种复权。归根结底,这还是属于一种向他者开放的柔性个人主权,需要以自我为中心不断进行权力的试行并编织出各种社会关系,而且还反过来受到这种与他者之间关系的制约。但无论如何,从国家主权到个人主权,这是区块链、互联网3.0以及元宇宙在社会秩序原理上引起的质变,势必带来国家治理模型和法律范式的创新。<sup>[24]</sup>

有必要指出的是,在区块链游戏中,人工智能作为虚拟空间的非玩家控制角色(non-player character, NPC),或者海量数据的分析师和问题解决者,或者作为某些用户的代理机器人(BOT,即自动玩游戏的智能工具),都发挥着非常重要的作用。不过在大多数链游中,代理机器人是被禁止使用的。然而在那些获准使用的场合,会有一部分人出于炫耀或营利的目的把代理机器人安装到很多台电脑里,通过它们进行管理操纵,就可以达到支配和垄断数字资源的目标。在这个意义

[20] See Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, <https://bitcoin.org/bitcoin.pdf>, last visited Aug. 15, 2023. 据马斯克于2022年3月上旬在推特爆料,这个神龙见首不见尾的作者其实就是四家国际知名企业(三星、东芝、中道、摩托罗拉)英文名称开头字母合写的虚构人物。

[21] 参见[加]唐·塔普斯科特、亚历克斯·塔普斯科特:《区块链革命》,凯尔等译,中信出版集团股份有限公司2016年版,第一章:可信的协议。

[22] Gavin Wood, *DApps: What Web 3.0 Looks Like*, <https://gavwood.com/dappsweb3.html>, last visited Aug. 15, 2023.

[23] 关于互联网3.0的基本特征和优势,参见姚前:《Web 3.0: 渐行渐近的新一代互联网》,载《中国金融杂志》2022年第6期,第14—17页。关于个人的主权化的预言式论述, see James D. Davidson & William Rees-Mogg, *The Sovereign Individual: Mastering the Information Age*, Touchstone Books, 1997.

[24] 参见季卫东:《元宇宙的秩序——虚拟人、加密资产以及法治创新》,上海人民出版社2023年版。

上,应用人工智能其实成为那一小部分玩家的特权,成为绑架和控制其他用户的工具,形成人类与机器(人工智能)相对峙和相抗争的格局。<sup>[25]</sup> 在这里,可以发现类似格尔茨通过巴厘岛斗鸡活动等深层游戏描述的“剧场国家”现象,<sup>[26]</sup>也不妨称之为“算法权力怪兽”或者所谓“私人化的算法利维坦”——它在虚拟世界的博弈中激活荣耀的象征性意义,让所有资源服从于闪亮的仪式和游戏规则。

实际上,NPC也意味着算法成为一种强大的权力,使虚拟空间的运行符合基于信息的计算逻辑,使人工智能的一系列指令都得到有效执行。归根结底,在互联网3.0上展开的创世故事其实就是:计算机语言,或者说作为交互机制的沟通过程,或者说控制游戏的程序,建构了整个世界。由此可见,算法不仅是逻辑或机器的计算次序,也是一种由人类语言创造的全新生态系统,一个不断产生千姿百态状态的智能物种,在虚拟空间里提供海量数据和无限选项,并通过机器学习形成网络结构中的“长尾效应”:绝大多数选择由算法进行,但最终的抉择权以及最好的选项仍然保留在人的手里。基于这样的技术条件,实际上从2012年开始,美国总统大选就开始应用大数据的机器学习进行模拟和预测,并据此决定选举策略;2020年的总统选举结果据说也受到高明的AI技术和代理机器人影响。在这个趋势的延长线上,可以预测某种算法权力怪兽将很有可能反过来支配民主政治。

### (三)“大语言模型怪兽”与智能的“波兰尼悖论”

自谷歌在2017年发布“变形金刚”(Transformer)网络结构以来,在仅仅五年多的时间里,全球就已经迅速出现了一支庞大的大模型群,进而其衍生出各种技术架构、各种模态、各种场景。从全球已发布的大模型分布来看,中国和美国大幅领先,数量合计超过全球总数的80%,美国在大模型数量方面始终居全球最高。<sup>[27]</sup> 2022年11月底ChatGPT一经公布就因对话能力强和应用范围广而席卷全球,仅用两个月就使月活跃用户规模达到一亿,增长速度极其惊人。此后这类大语言模型相继问世,分别从赋能个人和减负业务两个方面深刻影响了包括法律运行在内的各种社会实践场景,留下一道生成式AI物种大爆炸的数字“寒武纪”风景线。据不完全统计,到2023年5月,中国科技企业和网络平台推出了79个各种类型的人工智能语言大模型,其中34个是通用类大模型。<sup>[28]</sup>

不得不承认,大语言模型在给国家和社会带来便利和效益的同时,也带来了令人不安的风险甚至威胁。择其荦荦大端可以举出如下四种。首先,在类似ChatGPT的大语言模型提供网络对话服务之际,其可以比既有的互联网搜索引擎收集到更多的个人信息和隐私。因此,在“知道太多、利益相悖”的情况下,大语言模型及其运营者也许会通过控制沟通的方式对用户进行诱导,使之做出有违自己本意和利益的选择。其次,现阶段的大语言模型在对话中会把训练数据中没有也不可能存在的事物当作真实的存在,并且以不可置疑的语气进行描述。这就是用户们经常抱怨的

[25] 见前注[24],季卫东书,第163—166页。人工智能带来的算法权力和算法独裁问题,被概括为“塞维坦(Seviathan)”现象,需要人类通过治理机制和法律进行驯化。关于这个方面的思考,参见高奇琦:《人工智能——驯服塞维坦》,上海交通大学出版社2018年版,第三部分:人工智能与未来。

[26] 参见[美]克利福德·格尔茨:《文化的解释》,韩莉译,译林出版社2014年版,第十五章:深层游戏;[美]克利福德·格尔茨:《尼加拉:十九世纪巴厘剧场国家》,赵丙祥译,上海人民出版社2018年版,第四章:政治表述。

[27] 参见中国科学技术信息研究所:《中国人工智能大模型地图研究报告》,2023年5月28日发布,https://k.sina.com.cn/article\_6380588872\_17c500f4801902pmz9.html。

[28] 参见何喜军、张惠娜:《营造人工智能大模型产业生态》,载《经济日报》2023年6月14日,第5版。

“一本正经的胡说八道”现象。从科学技术层面上解释,这当然只是一种类似人类记忆失误的“幻觉(hallucination)”现象,与机器学习追求的能够“以小运大”、凭借有限的训练数据来适当处理无限的未知数据的那种泛化能力密切相关。但是,在应用场景中这种幻觉却会引起虚假信息蔓延,对用户或者社会造成致命性影响。再次,大语言模型在利用各种数据进行学习,或者人工智能自动生成各种内容之际,都有可能引起知识产权认定和保护方面的复杂问题。为了确保 AIGC 的可信度,同时明确责任,应该发明、应用以及推广各种数字水印防伪技术。最后,大语言模型还有可能有意无意地获取企业或政府机关的机密信息、操纵舆情,导致国家中枢系统的安全保障体系出现漏洞、信息社会发生功能障碍,甚至因为恶性事故和犯罪而引发动荡。

人类对语言的处理、对智能的驾驭实际上大部分是在无意识的状态下进行的。科学哲学家迈克尔·波兰尼在 1964 年曾经指出:“人类所知远胜于其所能言传。”<sup>[29]</sup>换言之,知识体系还应包括这种没有明确意识到,或未被社会常识所认可,或不能言说的默会知识。该命题被表述为“波兰尼悖论”,并成为人工智能理论的基础。<sup>[30]</sup>这也意味着人工智能对无意识语言的处理,根本就无法设计那种获得和应用所有语言的算法,也很难为机器学习设定明确的训练目标。现在通行的利用神经网络的机器学习,通过误差反向传播算法不断调整神经元权重和更新网络参数,逐渐减少误差,寻找训练数据的正解。人们发现,当神经网络的规模被大幅度扩张之后,人工智能接龙预测的精确度就会突然得到显著改善。这个发现及其有意识的应用使机器学习进入了深度学习阶段:无需复杂的规则和学习方法,只要让网络规模倍增就可以使许多难题迎刃而解,迅速提高泛化能力——不言而喻,这种神奇效果也证明了大语言模型的重要意义。其实质是多层网络的自我学习和进入语境(in-context)的学习,以及在此基础上实现学习方法的学习——元学习。这样一来,人类的特征设计也就变得没有意义,人工智能实际上开始进行自我塑造,进而形成一种自动化的生态系统。

正是在这里,“大语言模型怪兽(LLM behemoths)”纷纷崭露头角,并且有可能因为放弃事先给定的特征设计、转而自我设置次级目标而脱离人类控制,进而引起治理方面的严重问题。这意味着大语言模型将为一种非人类或超人类的新型智能助产,其会渐行渐远并产生与人类截然不同的价值观。这也意味着除了前面所说的平台怪兽们之外,主权利维坦还将面临来自几十个乃至几百个功能强大的大模型怪兽的挑战,即国家主权在数字领域,或者说“数字主权(digital sovereignty)”<sup>[31]</sup>正面临“百模大战”和失控的挑战。当然,不仅主权国家,也包括网络平台以及自主个人,甚至还有人类文明的操作系统都面临大语言模型不同程度的威胁。<sup>[32]</sup>但是“数字主权”这个概念的提出和强调,则清楚地反映了主权国家对社会数字化转型的回应以及自我防卫的立场。

为了防止上述各种风险演化成不可逆的灾害,专家和业界领袖们提出了暂停大模型开发、实

[29] 波兰尼悖论的中文表述引自《哈佛商业评论》(中文版)2017 年第 10 期,第 1022 页。

[30] 参见潘斌:《重思人工智能的辩证法:从人工智能到人类未来》,载《天津社会科学》2020 年第 3 期,第 23—29 页。

[31] See EPRS, *Digital Sovereignty for Europe*, July 2, 2020, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS\\_BRI\(2020\)651992\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf); Frances G. Burwell & Kenneth Propp, *Digital Sovereignty in Practice: The EU's Push to Shape the New Global Economy*, Europe Center of Atlantic Council, October 2022.

[32] 尤瓦尔·赫拉利(Yuval Harari)在 2023 年 4 月 29 日 Frontiers 论坛的演讲“AI 与人类未来”中提出了大语言模型可能破坏人类文明操作系统的观点,参见《尤瓦尔·赫拉利:人工智能对我们的操控,可能比你想象的更严重》,载澎湃新闻网 2023 年 6 月 1 日, [https://www.thepaper.cn/newsDetail\\_forward\\_23297386](https://www.thepaper.cn/newsDetail_forward_23297386)。

现价值对齐、加强 AI 监管等各种对策和建议。<sup>〔33〕</sup> 仅就价值对齐而言,例如美国布鲁金斯学会在 2022 年 12 月 8 日发表了本杰明·拉森的文章《人工智能的地缘政治和数字主权的崛起》,作者认为 AI 的不均衡发展将会让国家之间的不信任日益增长,进而导致数字主权的上升,出现技术脱钩;国家之间的意识形态差异或道德原则差异可能会对管理人工智能和信息技术产生更广泛的地缘政治影响;因此确保 AI 在国际层面的价值一致性,可能是本世纪最重大的挑战之一。<sup>〔34〕</sup> 无论如何,这是一场前所未有的大变局,势必形成崭新的国家与法律存在形态,促进秩序范式的创新。

#### 四、因“算法利维坦”而复兴的国家主权

中国的现代化运动在经过一百多年各种各样的社会实验并且遭受诸多挫折之后,似乎在最近二十年间找到了一条现实可行的新路径,即借助数字信息技术进一步提高行政效率和法律效力,实现国家权力不依赖强制就能进行充分治理的目标。其结果是主权利维坦越来越明显地呈现出“算法利维坦”的特征。也可以说,主权正在借助数字信息技术重新抬头,并获得前所未有的强大力量。因此,“数字主权”这个概念其实具有双重结构,一方面是指国家主权在数字空间面临被削弱的风险,另一方面则暗示主权采取数字形态,转化为一种“算法利维坦”,获得新生和前所未有的魔力。这两个方面相映成趣,构成奇妙的悖论。

从一般理论上讲,科学技术本来就是现代化的重要标志和内涵。回溯历史可以发现,从莱布尼茨对罗马法的数学改造,到边沁对权力建筑和幸福计算器的设计,整个西欧国家与法制现代化进程的本质在于成本效益以及概念的计算,以便适当调整主权的优越性与正当性之间的关系。特别是现代法学,试图通过公理体系和形式逻辑来克服主观任意性,确保判断的客观公正,实现市场和社会的可计测性,从而使国家对暴力的垄断性行使得以正当化,以实现韦伯所揭示的“正当的强制”这一目标。从 19 世纪后期开始,权力结构理性化的思路通过法律数学构想、司法统计分析、法律社会工程、计量法学、田野调查、实验主义法学、法律和案例的信息检索、涉法涉讼大数据建模、网络空间的代码框架设计等不同的中介环节,一直延伸到科学技术条件更加成熟的现阶段,形成了计算法学和数字法治欣欣向荣的景象。<sup>〔35〕</sup>

从中国实践看,首先是审判裁量权的滥用、司法腐败等问题,极大地动摇了人们对人脑判断的信心,同时也就助长了社会对电脑量刑的客观性、中立性以及确定性的期待。因此,互联网、大数据以及人工智能从 20 世纪末开始在中国司法界迅速得到略显草率的承认和广泛应用,甚至还出现了世界罕见的“智慧法院热”。一般而言,人工智能在本质上是规则嵌入后的系统,可以造成法律规范刚性化和技术代码化的结果,并通过计算机系统全程留痕的方式防止外部对司法判断的干预,因而有利于加强对国家权力的制约、陶冶全民守法的行为方式。在这个意义上可以说,人工智能以及更广泛的数字化处理是有利于贯彻现代法治的精神的。因此,2014 年之后的法治中国建设,实际上在相当程度上呈现出数字信息技术驱动的法制现代化的特色。

但在另一方面,数字化的法治政府和法治社会,也存在一些不得不正视的风险和问题。例如

〔33〕 参见季卫东:《强人工智能的治理与法律挑战》,载《凤凰周刊》总第 833 期(2023 年),<http://www.ifengweekly.com/detil.php?id=19376>。

〔34〕 参见清华大学《人工智能国际治理观察》第 127 期(2022 年 12 月 18 日),<http://aiig.tsinghua.edu.cn/info/1442/1815.htm>。

〔35〕 参见季卫东:《计算法学的疆域》,载《社会科学辑刊》2021 年第 3 期,第 113—126 页。

人工智能以数据为养料,通过收集、分析、学习、预测数据而不断提高其性能,因而不免在竭力扩大数据规模之际突破个人信息和隐私的各种屏障,从而在不同程度上动摇自由的根基。另外,以对大数据的机器学习为背景,尤其是在智慧网络化的社会条件下,算法在演变成强大权力的同时,人类对算法的理解和说明将变得很困难。这种算法黑箱,实际上会在不同程度上妨碍对权力问责,同时也会助长回避责任和转嫁责任的不良倾向,甚至会培育出一个无所不在、无所不能的“算法利维坦”。在当下这个数字时代,与算法黑箱化相对应的是社会透明化、个人透明化。社会到处布满了电子探头和感应器,仿佛有千万只眼睛在闪烁,无声地不断扫描和分析一切现象,并对人和物进行全景无死角、常时无间断的监视。这种视线的陷阱,真正实现了边沁关于作为“一种新形式的通用力量”的环形一览装置(Panopticon)的设计构想,也充分印证了福柯提示的对现代国家和法的独到见解。<sup>[36]</sup>

在福柯的理论视域里,我们的确可以观察到现代化权力和法制内涵的另一侧面:监视和规训,一览无余的观察、记录、分析以及绵密的计算。按照他的观点,对理性与非理性进行区别会产生一种排除的空间,但权力却可以不依赖强制而对个人进行驯化和管教,对社会进行治理和助推;法律秩序在本质上就是这种关于统治性(governmentality)的多样化权力技艺及其复合体,不仅包括政治机构的运营,也包括广义的治安管理,还包括决定个人与政府之间互动关系的整体框架乃至各种调整机制。福柯认为,现代法律是精心筹谋、持久运作的,既包含统治的策略,也具有不完备性以及失败的契机;因此,法律的运行离不开权力,但也因而伴随着抵抗和政治化因素。但在另一方面,法律作为操作技艺是以包括科技在内的知识谱系为前提的,有赖于真理体制和可变的文化一权力关系。<sup>[37]</sup>从无所不在的“大监察”这个维度来理解国家与法的现代化进程这种论述虽然不无偏激,但的确包含对现代化及其法治内涵某些侧面或维度的深刻洞察和反思,值得我们含英咀华、重新体会其中的醍醐味道。

在三年新冠疫情防控期间,通过数字信息技术进行社会治理的做法更加司空见惯,并且不断花样翻新。在发布“禁足令”、“电子点穴式”的弹窗以及给部分人群“赋红码”等场合中,主权国家俨然重新抬头,甚至显得更有力量。但与此同时,借助区块链进行分布式社区治理、一些地方的社区在2022年封控期间出现各种自组织机制(例如自发负责调度生活必需品的“团长”和组织核酸检测的“楼长”),主权相对化的侧面也在特定进路上延伸。主权复兴与主权退隐似乎在同时进行,形成软硬兼施、相反相成的辩证关系。借助互联网、大数据以及人工智能进行的这种严格监管与自治、共治相交错的复杂状况,让我们不禁联想起莱昂法社会学的发现。

加拿大女王大学法学院的大卫·莱昂教授从1980年代开始对“电子眼”进行研究,出版了《监视社会》《监视文化》等一系列具有国际影响的专著,被认为是监控理论的先驱者。他从边沁和福柯关于环形一览装置的隐喻出发,考察了监控方式从国家、企业主导到民众参与的演变过程,不久前又提出了基于大数据的“监视文化”概念。<sup>[38]</sup>莱昂教授特别强调数字化信息技术与为了安全、健康以及方便的监控之间的关联性,认为从个人对健康状态的自我追踪式监测(例如计步器、血压心率手环、手机体检软件)到企业对顾客行为履历的监控,再到政府对犯罪的预测性监控,实际上

[36] 参见[法]米歇尔·福柯:《规训与惩罚》,刘北成、杨远婴译,三联书店1999年版,第三章“全景敞视主义”关于边沁圆形监狱建筑设计的介绍和分析。

[37] 同上注,米歇尔·福柯书,第193页。

[38] See David Lyon, *The Electronic Eye: The Rise of Surveillance Society*, University of Minnesota Press, 1994; David Lyon, *Surveillance Society: Monitoring Everyday Life*, Open University Press, 2001; David Lyon, *The Culture of Surveillance: Watching as a Way of Life*, Polity Press Ltd., 2018.

存在监视者与被监视者之间的共谋。在这个意义上,在关照与控制的交织状况里,一种柔性监控社会应运而生。对疫情的恐怖感大幅度加强了人们的风险意识以及增加了对各种监控举措的包容和支持,甚至使监控在由表及里的同时带上了某种美感以及民主参与的色彩。<sup>[39]</sup>

由此可见,作为“算法利维坦”而再次强大起来的国家主权有时的确以高高在上的姿态而君临天下,但更多的时候却融化在数字网络之中,变成互动关系的一段编程代码。换句话说,原有的主权国家已然旧貌换新颜,蜕变成为更加庞大但却多元一体、伸缩自如的“算法利维坦”,在社会治理方面显得更加法力无边。特别是在进入 AI 4.0 阶段<sup>[40]</sup>之后,AIGC 拆除了语言和价值之间的藩篱,大语言模型的多语种、多模态工作机制自然而然会促进国际化、全球化与国家主权之间的相互影响,这样又会反过来刺激民族国家的身份认同感。实际上,“数字主权”的理念就意味着国家权力正在重新寻找本土化、区块化的立足点。鉴于网络平台怪兽、大语言模型怪兽、个人主权化怪兽以及代理机器人等算法权力怪兽的挑战,面对当下“所有模型对所有模型的战争”,迄今为止中国政府的对策是通过严厉的反垄断措施来驯化平台怪兽,通过统一的超算生态体系和投入高昂的基座模型(Base Model)来支撑几十个大模型怪兽,通过加强严格监管以及主权区块链(sovereignty blockchain)来防范主权化个人之间点对点互动的失控风险,<sup>[41]</sup>通过防护配置和对抗策略来制约 NPC 或 BOT 的活动。

例如“主权区块链”的概念是 2016 年由《贵阳区块链发展和应用》白皮书首先提出来的,尽管它与其他区块链同样具有分布式、不可篡改、互相可信任、通过智能合约转移价值等特点,但却向区块链中注入了国家主权意志,加强了对区块链的政府监视、技术干预,因此具有非完全去中心化的特征。按照全国科学技术名词审定委员会在 2017 年做出的概念界定,主权区块链是以分布式账本为基础,以规则和共识为核心,以国家主权为前提的区块链。在这里,规则是指“代码加法律”,共识强调以“人”而不是“数”为导向,倾向于“多中心”而不是“去中心”。<sup>[42]</sup> 贵阳创新驱动发展战略研究院常务副院长龙荣远认为,“主权区块链不仅是一种集成技术、一场数据革命,也是一次秩序重建,更是一个时代的拐点,已经成为全球治理重构的前沿力量”。<sup>[43]</sup>

又例如,针对“百模大战”对大模型算力需求猛增的形势,科技部在 2023 年 4 月 17 日发起成立国家“超算互联网”联合体,把分布在全国的大量超算中心通过算力网络连接起来,构建“大一统”的算力服务平台,通过“超算互联网”来打造国家算力基座乃至基座模型。<sup>[44]</sup> 7 月 7 日,在世界人

[39] 参见季卫东:《疫情监控:一个比较法社会学的分析》,载《中外法学》2020 年第 3 期,第 565—589 页。

[40] 人工智能发展的四个阶段,是指(1) 1950—1960 年代,出现 AI 术语、图灵测试、神经网络模型、LISP 语言;(2) 1970—1990 年代,出现 PROLOG 语言、专家系统开发、AI 产业化、机器学习、认知网络复兴、遗传韵律、强化学习;(3) 2000—2022 年,最新成果包括深度学习、Watson 在答疑竞赛中战胜人类、生成对抗网络(Generative Adversarial Network)、AI 画像精准识别能力超越人类、“阿尔法围棋”系统战胜世界顶级专业棋手;(4) 2023 年之后,生成式人工智能的发达、语言大模型的崛起、通用人工智能曙光在望。

[41] 与此形成鲜明对照,美国国会在 2023 年 7 月 26 日通过了《区块链监管确定性法案》(Blockchain Regulatory Certainty Act),对于那些预防区块链风险的监管提供更加明晰的指导,以防止对数字资产的侵犯。

[42] 关于主权区块链的来龙去脉,相关文献可参见连玉明主编:《主权区块链 1.0: 秩序互联网与人类命运共同体》,浙江大学出版社 2020 年版;连玉明主编:《主权区块链 2.0: 改变未来世界的新力量》,浙江大学出版社 2022 年版;连玉明主编:《主权区块链 3.0: 共享秩序下的全球治理重构》,浙江大学出版社 2023 年版。

[43] 《主权区块链成为全球治理重构的前沿力量》,载当代先锋网 2023 年 5 月 28 日, [http://www.ddcpc.cn/detail/d\\_guizhou/11515116165400.html](http://www.ddcpc.cn/detail/d_guizhou/11515116165400.html)。

[44] 参见孙硕:《科技部启动国家超算互联网部署工作》,载《数字经济洞察周报》2023 年第 15 期,截至顶网 2023 年 4 月 20 日, <https://insights.zhiding.cn/2023/0420/3148891.shtml>。

工智能大会上,国家标准委员会指导的国家人工智能标准化总体组宣布成立首个大模型标准化专题组,由上海人工智能实验室、百度、360、华为、阿里等企业联合负责,正式启动大模型测试国家标准的制订工作。<sup>[45]</sup>显然,其目标是推动大模型与标准化相结合,形成头部科技企业侧重行业大模型、国家推动通用大模型、打造高质量发展新基座的分工格局。包括北京智源人工智能研究院2023年6月发布的AIGC底层基座天鹰·Aquila在内,基座模型一般都需要千亿级以上的数据进行训练,参数量也会达到百亿级以上,只有在这样的基础上才能建立很好的通用大模型,产生涌现能力。可想而知,“大一统”的基座和通用大模型构想可以避免重复投资、盲目发展、产业内卷。但这种思路并非仅仅出于成本效益的合理计算,显然也有利于维护数字主权,加强算法利维坦的权力集中化效应。

## 五、结语：算法利维坦与数字怪兽互动的程序公正

从上述分析可以看到,一方面,基于国家主权的算法利维坦的魔力似乎可以无远弗届了;另一方面,这又必然会强化政府机构与实力雄厚的网络平台、隐匿在区块链中的主权化个体、具有泛用性的大语言模型、影响政治权力游戏的代理机器人之间频繁而持续的相互作用。在这样复杂而流动的状态下,为了通过人工智能系统内嵌的公正程序以及不同人工智能系统之间的分权制衡机制来防止大平台和大模型的滥用,也为了制衡那个独特的算法利维坦及其私人化变种,现代法治的程序正义原则也将被重新定义并发挥更加重要的作用。<sup>[46]</sup>在一定意义上也可以说,寻找法律性正当程序与技术性正当程序的最佳组合方式,就是今后由AI驱动的数字国家继续推进现代化事业的主要法治内涵。

所谓“法律性正当程序”或者符合正义的法律程序,是指确保论证性对话在自由而平等的条件下顺利进行的制度性装置,当然不是指电子计算机的程序或者通过人工智能系统进行的程序性控制。但在数字时代,程序的法律往往需要转换成计算机程序的代码来运行。正如马克斯·蒂菲克所指出的那样,代码“决定了什么样的人可以接入什么样的网络实体……这些程序如何规制人与人之间的相互关系……完全取决于做出的选择”。<sup>[47]</sup>非常有趣的是,正是代码把程序与选择以及人际关系串联在一起。这也昭示了代码具有通过技术性正当过程对算法助推进行制约和改进的可能性。所谓“技术性正当过程”,主要是解决代码及其框架应该如何适当规制、代码的作者究竟是谁、谁能控制代码的作者、对代码的妥当性是否存在事先的论证程序或事后的纠正程序、作为中介机构的网络服务提供商的权力和责任应该如何配置、应用数据流的软件开发受到什么样的规制、数字监视和网络搜索是否设定了限制性条件等一系列涉及过程正当化的问题。一言以蔽之,技术性正当过程的本质是代码规制。如果这种对代码的规制也采取法律的形式,那么就可以发现技术性正当过程与法律程序正义或者“程序性正当过程”之间的交集和组合。

作为正当程序的前提条件,还需要达成社会的基本共识为程序设计提供价值标准。换言之,程序

[45] 参见宋薇萍:《大模型国家队来了! 360、百度、华为、阿里等入局》,载《上海证券报》·中国证券网 2023年7月7日, <https://news.cnstock.com/news/bwqx-202307-5087478.htm>。

[46] 参见季卫东:《探讨数字时代法律程序的意义——聚焦风险控制行政的算法独裁与程序公正》,载《中国政法大学学报》2023年第1期,第83—99页。See Danielle Keats Citron, *Technological Due Process*, 85(6) *Washington University Law Review* 1249-1313 (2008).刘东亮:《技术性正当程序:人工智能时代程序法和算法的双重变奏》,载《比较法研究》2020年第5期,第64—79页。

[47] [美] 劳伦斯·莱斯格:《代码 2.0——网络空间中的法律(修订版)》,李旭、沈伟伟译,清华大学出版社2018年版,第7页。

需要契约的基础。面对数字主权各自为政、算法利维坦无所不在的现实,安东尼奥·古特雷斯秘书长曾经在 2021 年 9 月公布了关于《全球数字契约(Global Digital Compact)》的倡议,联合国还在 2023 年 5 月正式发布了政策简报“全球数字契约——为所有人创造开放、自由、安全的数字未来”并征询各方面的意见。全球数字契约将以普遍人权为基础,通过各个利益攸关方的合作建立一个全球框架,推进开放、自由、安全、以人为本的数字未来,实现可持续发展的目标。按照这个构想,各会员国应该建立数字人权咨询机制;保护互联网的自由和共享性质,使之成为真正的全球公共物品;为网络平台和用户制定强有力的问责标准,加强不同司法管辖区在线安全专员的合作;实现数据治理原则的统一;形成人工智能的敏捷治理框架;等等。<sup>[48]</sup> 中国已经就全球数字契约向联合国提交意见,<sup>[49]</sup> 如果顺利的话,将在 2024 年 9 月举行的联合国未来峰会上就全球数字契约达成国际基本共识。

上述思路不禁让人重新联想到霍布斯关于主权利维坦和社会契约的宏大叙述。在某种意义上也可以说,联合国的全球数字契约仿佛就是数字时代反过来针对主权与算法相结合的那个无所不能的利维坦而缔结新社会契约的一项重要提案,或多或少展示了数字立宪主义和数字法治的普遍性愿景。但数字法治究竟如何制约算法利维坦,能否真正为网络空间奠定新的正当性基础,还是一个今后有待观察和深入探讨的课题。

---

**Abstract** This article analyzes the paradoxical relationship between the relative decline and resurgence of national sovereignty in the context of economic and informational globalization, the rise of mega-platforms, the sovereign individuals hidden within blockchains, and large models behemoths free from the feature design of programming. It highlights how the decline of sovereignty has given rise to concepts such as “the digital sovereignty” in Europe and “the sovereign blockchain” in China, while the resurgence of sovereignty has led to theories like the “surveillance society” in North America and the “electronic forbidding leave” metaphor in China. In any case, with the era of AI driven, “the Algorithmic Leviathan” is becoming an extremely powerful dominating force that countries and laws must confront. To prevent its runaway and abuse, it is necessary to consolidate a basic consensus through global digital compacts and institutionalize it through legal and technical due process—this is the essence of digital rule of law.

**Keywords** Digital Sovereignty, Platform Governance, Large Model Behemoths, The Sovereignty Individual, The Algorithmic Leviathan, Global Digital Compact

---

(责任编辑:方潇)

---

[48] United Nations, *A Global Digital Compact: An Open, Free and Secure Digital Future for All* (Our Common Agenda Policy Brief 5, May 2023).

[49] 参见《中国关于全球数字治理有关问题的立场》,载中华人民共和国外交部官网 2023 年 4 月, [https://www.mfa.gov.cn/web/wjw\\_673085/zjzg\\_673183/jks\\_674633/zclc\\_674645/qt\\_674659/202305/t20230525\\_11083602.shtml](https://www.mfa.gov.cn/web/wjw_673085/zjzg_673183/jks_674633/zclc_674645/qt_674659/202305/t20230525_11083602.shtml)。