AI时代的“个保法”失灵与制度重塑

高富平

- 现行个人信息保护制度通过一系列程序性与防御性权利，确保个人信息的利用不致损害其尊严与自由。无论域外还是我国现行个人信息保护法，在本质上均可视为前网络时代的制度产物。
- 人工智能对网络数据抓取及感知数据的大量依赖，削弱了个人信息收集制度的有效性。人工智能对数据的使用目的和结果难以预期，使知情和同意制度失灵。人工智能算法的不可解释性，对透明性与问责机制构成直接挑战。人工智能对数据规模化聚合的需求，与个保法保护目的相悖。
- 人工智能已经将既有的个人信息保护制度缺点暴露无遗，因此，突破既有路径依赖，在尊重个人自主、尊严等基本权利的前提下，探索促进人工智能创新发展的新型制度平衡方案，已成为不可回避的时代命题。

个人信息保护制度的诞生与发展

数字技术的产生及其持续演进，重塑了个人数据（个人信息）的处理方式。自20世纪70年代起，社会各界开始系统关注计算机技术在处理个人信息过程中对个体可能造成的影响。早期探索多置于广义隐私保护框架之中，重点探讨数字技术对个人自主、人格尊严及隐私等权利的影响，并逐步发展出“信息隐私”（亦称“数据隐私”）的概念，其核心在于确认个人对其数据处理活动应享有一定程度的控制能力，以防止数据处理者滥用技术手段，从而侵蚀个人对自身事务的自主支配。“信息隐私”与传统意义上的“通信秘密”及民法意义上的“个人隐私”共同构成广义隐私保护体系。我国《个人信息保护法》（以下简称“个保法”）本质上即属于信息隐私保护法，其所维护的核心价值在于数据/信息主体的人格尊严与自由。个人在数据处理活动中所享有的受保护地位，在欧盟法律语境中被概括为“个人数据受保护权”，在我国《民法典》中体现为“个人信息受法律保护”，而在“个保法”中则表述为“个人信息权益”。尽管名称各异，但规范内涵具有高度一致性，即并非赋予个人对其信息以物权式的支配权，而是通过一系列程序性与防御性权利，确保个人信息的利用不致损害其尊严与自由。这种控制机制具体表现为对处理行为的事前同意、事中限制以及事后撤回、拒绝、删除等权利配置。

经合组织（OECD）于1980年发布的《隐私保护和个人数据跨境流动指南》，以及欧洲委员会于1981年通过的《个人数据自动化处理中的个人保护公约》，标志着发达国家在个人数据保护领域逐步形成基本共识，并对全球立法进程产生了持续而深远的影响。欧盟既是上述国际规则的推动者，也是将相关理念法律化的先行者，分别于1995年和2016年完成了两次具有里程碑意义的立法。尽管在2016年《通用数据保护条例》（GDPR）出台之时，大数据与人工智能已初现端倪，欧盟亦试图借助统一数据市场战略推动数字经济发展，但该条例在规范理念与制度结构上，仍主要继承并强化了20世纪80年代形成的个人数据保护原则与范式。我国自2012年起逐步引入个人信息保护制度，并通过多部法律的协调适用推动其规范化发展，最终于2021年制定并施行专门的个人信息保护法。从整体设计来看，我国相关立法在理念与结构上亦明显借鉴了欧盟GDPR的逻辑，其所遵循的仍是国际社会长期形成的传统保护原则。因此，无论域外还是我国现行个人信息保护法，在本质上均可视为前网络时代的制度产物。

然而，在过去四、五十年间，数字技术经历了持续而深刻的变革，当下以万物互联为特征的智能网络逐步成形，人类社会迈入全面数字化生存阶段。与此同时，计算机处理数据的能力不仅在速度和规模上显著提升，而且在逻辑上从“程序驱动数据”转向“数据驱动程序”，这一转向正是以机器学习为核心的人工智能技术的根本特征。机器智能补强人类智能以进一步地认知与改造世界，社会发展由此进入可称为“人工智能时代”或“数智时代”的新阶段。相应地，上世纪形成的个人信息保护理念与方法在新阶段正面临根本性挑战，包括我国在内的现行个人信息保护制度，均呈现出不同程度的治理失灵倾向，这也为个人信息保护法理与制度的进一步重构提出了新的时代命题。

人工智能的数据驱动特征触发“个保法”制度失灵

人工智能在运行机理上具有鲜明的数据驱动特征，其模型训练高度依赖大规模数据资源，在实际应用中还需对人机交互、机机通信及实时感知场景中的数据进行持续处理。随着人工智能的广泛部署与深度应用，现行个人信息保护法在多个关键环节呈现出制度失灵的风险。

首先，人工智能对网络数据抓取及感知数据的大量依赖，削弱了个人信息收集制度的有效性。

“个保法”规定处理者直接收集时需获得个人同意，但是人工智能训练与使用的黑箱却为公开数据的获取开了“天窗”。这不仅对守法主体形成不公平约束，也未能因直接收集环节的合规而实质性提升个人保护水平，致使公开个人信息的保护在实践中趋于空转。

与此同时，图像识别、生物特征识别以及物联网技术的广泛应用，使机器能够持续感知和处理外部环境变化，实时采集来自家庭、工作场所和公共空间的数据。这类信息足以描绘个人的生活轨迹、行为习惯甚至心理状态。即便设备安装或初始使用阶段曾履行告知与同意义务，该种一次性授权亦难以覆盖后续复杂、持续的数据处理活动。再加之模式识别与预测分析技术的应用，看似无害的数据片段经聚合后即可揭示个人的敏感细节，使个人既无法掌控影响其决策的算法处理行为，又面临潜在的隐私侵害风险。以同意为核心的制度安排，在人工智能语境下已难以发挥实质性保护功能。

其次，人工智能对数据的使用目的和结果难以预期，使知情和同意制度失灵。目的限定原则是个人信息保护法的基石，它既是同意的基础，也是判断数据处理是否符合最小必要原则的依据。然而，人工智能训练天然要求尽可能广泛的数据输入，而数据收集阶段往往无法准确预见未来的具体使用方式，更难预判模型训练可能产生的衍生效果。规模效应驱动下，模型提供者普遍倾向于最大化数据获取，以实现能力跃升，这不仅与最小必要原则相悖，也使得同意机制在事实上难以发挥约束作用。为应对合规要求，处理者往往在隐私政策或用户协议中罗列大量、高度概括甚至穷尽式的处理目的，反而使数据主体的同意沦为空洞的形式确认，失去实质意义。

第三，人工智能算法的不可解释性，对透明性与问责机制构成直接挑战。人工智能系统的推理路径与决策依据具有高度复杂性和不透明性，在算法缺乏可解释性的情况下，个人无法获知其信息被如何处理，一旦出现错误、歧视或不当输出，也难以及时发现并追责。透明原则与问责原则既是“个保法”的核心规范，也是负责任人工智能治理的基本要求。诚然，在一定范围内容忍算法不透明性，可能是提升系统性能的现实需要，但这一容忍应当以风险分级与利益平衡为前提，在风险可控的场景中适度放宽要求，同时确保人工智能生成结果具备可核验性与可追责性。

第四，模型记忆导致个人对其个人数据控制能力的丧失。用于训练模型的基础数据集，往往会转化为模型内部的“知识结构”，一旦特定信息被模型“记忆”并固化，便可能在特定提示下被原样输出，从而造成敏感信息泄露。此外，人工智能通过向量化方式对数据进行嵌入处理，这些向量在特定条件下存在被反向还原的风险，进而重构原始的文本或图像内容。当模型在训练过程中“记忆”诸如私人地址等信息时，便可能在无意中向其他用户披露。更为关键的是，机器学习并非止于训练阶段，而是在持续交互中不断吸收新数据，用户往往并不知晓其输入信息被存储或再次用于训练。由此，人工智能对数据的利用方式已从根本上改变了人与计算机之间的被动数据处理关系，使个人信息事实上脱离个人的控制范围。

最后，人工智能对数据规模化聚合的需求，与个保法保护目的相悖。高性能人工智能模型的构建离不开大规模、高质量的数据集，这要求打通不同主体之间的数据流通渠道，实现数据的快速汇聚与整合。这亦是我国推动数据要素市场建设的重要现实动因。在实践中，数据不仅可以通过物联网实现机器间的交换与共享，经整理的数据集及基于大数据形成的分析结论亦可能在合作主体之间流转。然而，现行“个保法”主要以规范直接数据使用关系为目标，对个人数据的对外提供设置严格限制，要求再次取得个人的单独同意。这一制度设计一方面旨在保障个人对数据流向的知情与控制，另一方面也是目的限定原则的必然要求。显然，上述制度安排与人工智能对数据流通和共享的现实需求之间形成了直接冲突。

突破既有路径依赖重塑个人信息保护制度

人工智能能够从多元渠道持续获取数据，万物互联的社会基础设施已不断为其生成并输送涉及个人、组织、自然环境及社会事件的数据，其中亦包括用户在使用数字终端过程中留下的足迹。传统以个人信息自决为核心的保护模式，建立在数据主体对信息收集目的、处理范围与存储方式等具有充分知情与理解的前提之上。然而，在非接触、无感知的自动化采集广泛存在的情形下，我们只好宣称已经进入“默许数据收集”时代，此时该拿什么捍卫我们的尊严？万维网发明者蒂姆·伯纳斯-李曾说过，“数据是一种宝贵的东西，它将比系统本身持续更长的时间。”在人工智能时代，如果不加干预，那么数据的存在期限就可能变为永久。囿于没有好的方法可以让人工智能“遗忘”其非法习得的东西，那种建立在个人干预或控制处理者使用，用完即可要求删除的制度安排将完全沦为虚置。

人类文明的演进始终以信息处理能力的提升为基础。计算技术不断迭代，大数据+人工智能正在改变人类的认知能力和行动能力，改变社会运行结构和效率，但也深度影响着个体、社会和国家。半个世纪前形成的个人信息保护理念并未充分预见社会运行过程中所生成的海量数据将普遍进入机器学习与处理的现实。尽管早在2010年，OECD已开始反思30年前确立的个人信息保护基本原则，但彼时各国普遍不愿挑战既有的人权保护框架，甚至需要援用它构筑本国数字经济的“护城河”。今天，人工智能已经将既有的个人信息保护制度缺点暴露无遗，它既不能给个人有效的保护，也不利于人工智能造福人类社会。因而，放下历史包袱，突破既有路径依赖，在尊重个人自主、尊严等基本权利的前提下，探索促进人工智能创新发展的新型制度平衡方案，已成为不可回避的时代命题。

【作者系华东政法大学法律学院教授、博士生导师，互联网法治研究院（杭州）常务副院长】