VOICE OF AMAC

（2024 年第 5 期，总第 188 期）

中国证券投资基金业协会　　2024 年 5 月 13 日

大数据应用对社会的影响及算法伦理

——专题征文之四

【编者按】本文首先梳理了大数据技术的发展现状，对算法伦理在国际组织、监管、企业的共识性框架进行梳理，并归纳提出了算法、应用、长期社会影响三个层面下，透明度与可解释性、数据隐私和保护、数据安全、维护公平公正社会责任和可持续五个主要维度的算法伦理原则。其次，基于上述原则，从算法、应用、长期社会影响三个层面讨论了大数据的潜在风险，并结合实际案例梳理了大数据技术的积极影响。最后，本文提出充分发挥大数据技术积极作用的对策建议。包括完善相关立法，规范行业发展，以此作为保障大数据技术规范应用的最后一道防线；加大技术管控，提升技术水平，保护隐私及规范算法使用；加强公众教育规范主体行为，形成对法律监管的有效补充。

一、研究背景

当前，全球大数据技术产业与应用的创新不断迈向新的高度。大数据技术日益渗透到生产、分配、流通、消费和社会服务管理等各个领域，深刻改变着人们的生产、生活方式以及社会治理方式。自“十三五”至今，中国的大数据产业高速发展，政策体系日益完善，产业基础日益巩固，产业应用不断丰富，产业生态持续优化。

我国数据量持续高速增长和频繁交互推升了大数据技术和服务需求，大数据产业规模持续高速增长。2022 年我国数据产量达 8.1ZB，占全球数据总产量 10.5%，位居世界第二。截至 2022 年底，我国存力总规模超 1000EB，数据存储量达 724.5EB，同比增长 21.1%，占全球数据总存储量的 14.4%。数据量的激增和数据的频繁交互带动了大数据产业的蓬勃发展。据测算，大数据产业规模年均复合增长率超过 30%。在技术发展和政策鼓励的双重推动下，大数据已成为我国经济转型发展的重要支点。据统计，2020 年我国数字经济增加值占 GDP 比重近四成[1]。

随着数据经济的繁荣发展，大数据产业逐渐成为推动经济转型发展的新力量。一方面，大数据产业对数据存储、计算、网络、终端信息采集等方面的需求不断增长，为计算机通信和其他电子设备制造业的发展提供了有力支撑，推动了数字产品制造业和数字产品服务业的繁荣发展。另一方面，大数据产业对数字技术的集成创新为数字技术应用业的蓬勃发展提供了坚实后盾，助力软件、互联网、信息技术服务

[1] 国家互联网信息办公室，数字中国发展报告（2022 年）[R]，2022.

等数字技术应用业的繁荣。此外，互联网、数字内容等大数据的“数据原生”行业的发展持续为数字经济带来新模式、新业态的灵感，大数据技术与传统行业的融合加速了传统行业的转型升级进程，同时也提升了政府社会治理的能力。

与传统技术不同，大数据技术的发展依赖于数据这一新型生产要素。数据通过算法不断生成新知识、形成新服务，并赋能传统行业，随着大数据技术在众多领域的深度应用，人们的思维方式和行为方式正经历着深刻的变革。然而，这种转变在数据、算法及应用层面不仅孕育了提升生产率的巨大机遇，同时也引发了全新的社会挑战。

那么，大数据技术的发展对经济社会发展的影响如何？大数据技术的发展应遵循哪些伦理原则？如何通过顶层设计解决大数据所带来的伦理问题，从而充分发挥大数据技术对经济增长的驱动作用？为回答上述问题，本文首先梳理了大数据技术的伦理原则及其应用的潜在影响；其次深入分析各方面影响与案例；最后提出大数据技术规范发展的对策建议。

二、大数据应用的伦理原则

（一）大数据伦理原则的来源

国际各类机构对大数据伦理的关键性原则表述和分类有一定差异，但本质上有共通之处。大数据伦理原则旨在引导大数据支持社会可持续发展。联合国宪章和可持续发展目标为原则提供了基础[2]，各国根据自身情况明确了大数据伦理原则要点与监管框架，而企业则基于自身业务与技术情况形

[2] 联合国，联合国宪章[EB/OL]，https：//www.un.org/zh/about-us/un-charter/full-text

成原则，约束自身大数据应用。

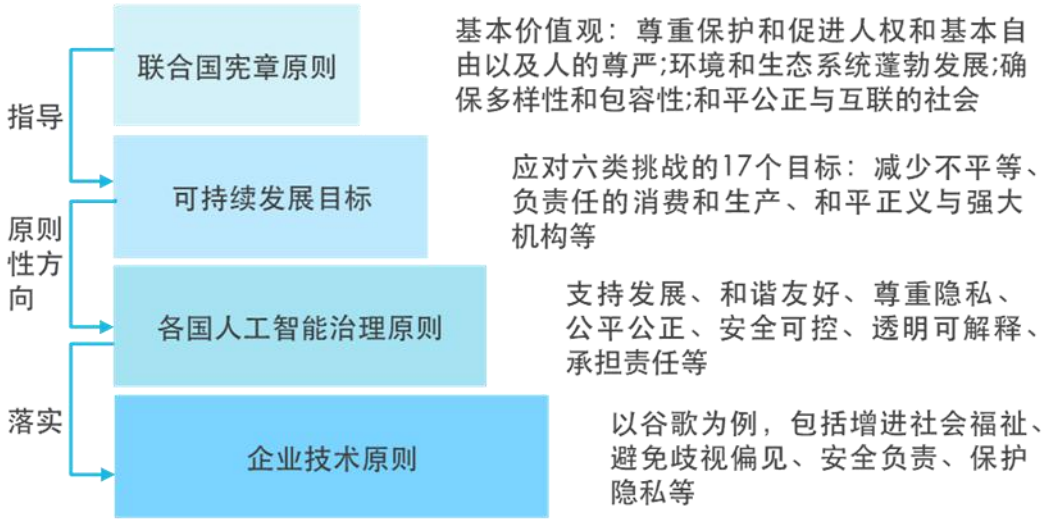

图 1. 国际大数据伦理原则体系概括

1.全球大数据伦理原则有共识基础

国际上多个不同机构都曾提出过大数据的伦理原则，其对关键性原则表述和分类有一定差异，但核心观点有较高的重合度。2020 年哈佛大学的一项研究对比汇总了国际人工智能伦理的关键性原则，研究发现，各机构在保护隐私、支持算法安全可靠、维护公平公正、尊重人的权益、公开透明等方面有很大的重叠[3]。研究对相关原则的进一步趋同持非常乐观的态度。

大数据的伦理原则可以溯源至联合国宪章与全球可持续发展目标。大数据发展本身是为了支持全社会长期可持续发展，因此，现行原则基本上与联合国宪章和可持续发展目标提倡的方向非常一致。联合国宪章是很多大数据伦理框架引用的基本共识，其尊重保护和促进人权和基本自由以及人的尊严、环境和生态系统蓬勃发展、确保多样性和包容性、

[3] Berkman Klein Center for Internet & Society, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI [R], 2020.

和平公正与互联的社会等内容成为大数据的伦理原则[4]。而联合国提出的可持续发展目标，作为指引国际社会可持续发展的关键方向，也提供了可适用于人工智能的原则包括减少不平等、负责任的消费和生产、和平正义与强大机构等等[5]。

表 1. 国际数字与人工智能伦理原则文件

时间	会议文件	主要内容
2022.11	世界互联网大会	人工智能与数字伦理分论坛上相关专家指出人工智能的研究、开发都要以人为本，从公正公平的伦理原则出发，努力建立可解释的鲁棒人工智能理论。2007 年图灵奖获得者约瑟夫·斯发基斯强调要在严格执行风险预防原则和接受创新以提高效率或质量之间取得平衡。
2021.11	《人工智能伦理问题建议书》	教科文组织《人工智能伦理问题建议书》是首个关于以符合伦理要求的方式运用人工智能的全球框架，于 2021 年 11 月获得全部会员国通过。建议书关注 AI 创新所带来的诸多伦理问题，尤其聚焦歧视和陈规定型观念，包括性别不平等问题，同时兼顾打击虚假信息、保障隐私权、保护个人数据、维护人权及环境权。
2019.09	《 The global landscape of AI ethics guidelines》[6]	Jobin 等人对来自不同国家或国际组织发布的 84 个 AI 伦理文件进行了分析，发现目前所发表的 AI 伦理指南在五个关键原则上达成了广泛共识，即透明度、公正和公平、非恶意、责任和隐私等。
2017.01	Beneficial AI 会议	近千名人工智能和机器人领域的专家，联合签署了阿西洛马人工智能 23 条原则，呼吁全世界在发展人工智能的同时严格遵守这些原则，共同保障人类未来的伦理、利益和安全。

2.各国制定大数据伦理与治理原则

全球性可持续发展的共识为大数据技术的设计、规章制定和应用提供了总体的伦理指导。而各国参考国际共识，形成了自己的纲领性政策文件。

欧盟重视大数据与人工智能技术的安全问题。2020 年，欧盟委员会发布人工智能白皮书，重点围绕“卓越”“可信

[4] 联合国教科文组织，人工智能伦理问题建议书[EB/OL]，2022. https：//unesdoc.unesco.org/ark：/48223/pf0000381137_chi

[5] 联合国，联合国可持续发展目标[s]，2020，https：//sdgs.un.org/goals

[6] Anna Jobin, Marcello Ienca & Effy Vayen, The global landscape of AI ethics guidelines [J], https://www.nature.com/articles/s42256-019-0088-2

赖”两个目标强调AI技术应用和高风险人工智能应用识别与监管。2021年4月欧盟委员会发布的《人工智能法[7]》在区分“禁止类AI”和“高风险类AI”的基础上，提出了具体目标。随之9月份的《AI责任指令[8]》重点解决AI引发的补偿问题。通过立法文件，引导大数据支持社会发展、尊重生命、维护、控制风险、公开透明。

在我国，大数据和人工智能的安全性也逐步受到关注，大数据于2014年首次写入政府工作报告，自此，相关政策指引了行业发展和治理。2017年，浙江、辽宁、北京等地也陆续发布新一代人工智能发展规划。在规划指引下，产业布局不断深化。2021年9月，国家新一代人工智能治理专业委员会发布了伦理规范提出和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理八项原则，以发展负责任的人工智能。2023年5月，国家互联网信息办公室通过《生成式人工智能服务管理暂行办法》（简称《管理办法》）。《管理办法》提出，提供和使用生成式人工智能应坚持社会主义核心价值观、防止歧视、尊重知识产权和商业道德、提升透明度和有效性[9]，这些原则性文件与监管规则为人工智能技术健康发展和规范应用提供了重要依据。

3.大数据企业制定自身伦理原则

大数据企业主要参考国家要求的共识性的伦理原则号召，制定自身技术原则规范。以谷歌为例，其遵从的伦理原

[7] European Parliament, Artificial Intelligence Act [S/OL], 2023, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

[8] European Parliament, Liability Rules for Artificial Intelligence [E/OL], 2023.

[9] 国家互联网信息办公室等，《生成式人工智能服务管理暂行办法》[EB/OL]，2023，https：//www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

则包括增进社会福祉、避免歧视偏见、安全负责、保护隐私[10]。腾讯研究院于2019年发布了人工智能伦理报告[11]，报告提出，人工智能的新的技术伦理观包含技术信任、个体幸福、社会可持续三个层面，善用技术塑造健康包容可持续的智慧社会。

（二）大数据伦理的主要原则

基于对上述文件的归纳，大数据的伦理原则包括数据与算法、应用、长期社会影响三个层面的五个维度。

1.层面一：数据与算法层面

第一，透明度与可解释性。大数据与人工智能需要确保算法的运行和数据处理过程是透明的，并能够解释算法的决策过程。这一原则有助于建立用户和利益相关者对算法的信任，有助于发现和纠正错误，确保算法决策的准确性。此外，这一原则对于监管合规非常重要，监管机构需要了解数据处理和决策的细节，以确保其合法性和公平性。这是社会更好地理解、监控和参与数据处理和决策过程的技术基础。

第二，数据隐私和保护。大数据使得侵犯个人信息的形式越发多样化、隐蔽化、严重化，需要采用强有力的数据保护措施，以确保敏感数据的安全应用。此外，这一伦理原则在涉及敏感信息的领域，如医疗保健和金融，是确保数据主体权益受到尊重、维护个人隐私和建立信任的至关重要组成部分。

第三，数据安全。保护大数据存储和传输的安全，防止敏感数据外流。这要求采取严密的安全措施，以确保数据的

[10] Google Public Policy, Responsible AI[EB/OL], 2023. https：//publicpolicy.google/responsible-ai/

[11] 腾讯研究院，智能时代的技术伦理观——重塑数字社会的信任[R]，2019.

完整性、可用性和保密性，尤其是在金融机构中，大量敏感的客户财务信息需要采用强大的数据加密、访问控制和监控措施，以确保客户数据的安全性和隐私性。

2.层面二：应用层面

第四，维护公平公正。大数据的应用需要确保其算法公平对待各类人群，维护社会公平。因此，要求算法的设计和应用公平对待不同特征的个体，确保大数据和人工智能成为推动社会进步的工具。这一原则是确保大数据算法负责任和可接受、促进社会公平和个体权益的保护的重要基石。

3.层面三：长期社会影响

第五，社会责任和可持续性。在大数据技术的发展中，应考虑社会和环境的可持续性。采取可持续性措施可以降低能源和资源成本，提供长期经济效益。社会可持续性也有益于社会公益，包括改善生活质量、降低城市交通拥堵和改善资源分配。

三、大数据发展的潜在风险

大数据从生成到应用的各个环节，会涉及诸多的伦理问题，进而对理论和管制政策有着更高的要求。算法环节包括安全、可解释性、幻觉问题；应用环节包括信息的保护、版权的划分争议和责任归属等监管问题。从更宏观的视角来看，大数据在长期发展过程中可能对就业市场、社会价值观产生深远且广泛的影响，这要求我们必须审慎对待并寻求平衡之道。

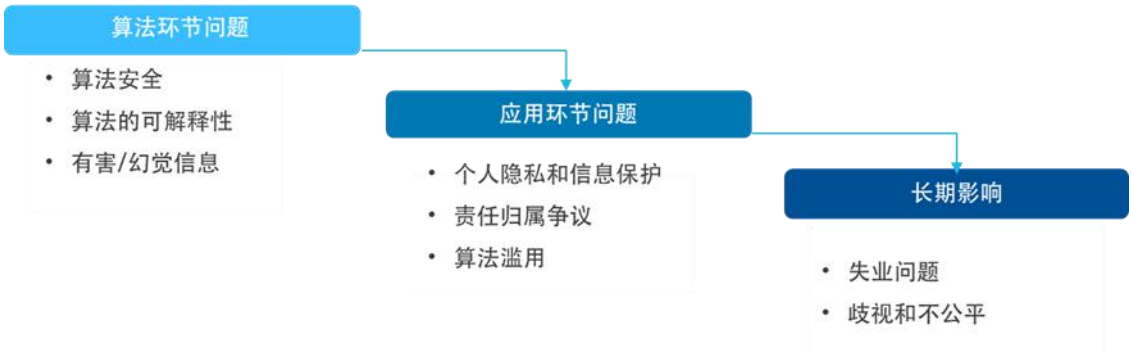

图 2. 大数据发展的潜在风险

（一）算法环节：安全、可靠、可解释性风险

1.安全风险：大模型自身安全维护问题

以深度学习为核心的 AI 技术天然地容易受到竞争对手和使用者的数据攻击、模型攻击、提示语攻击。为保障大数据技术的安全与稳定，必须采取有效措施，防止数据泄露和模型被窃取等风险。同时，还需警惕负面样本攻击所可能导致的算力和能源浪费等问题。

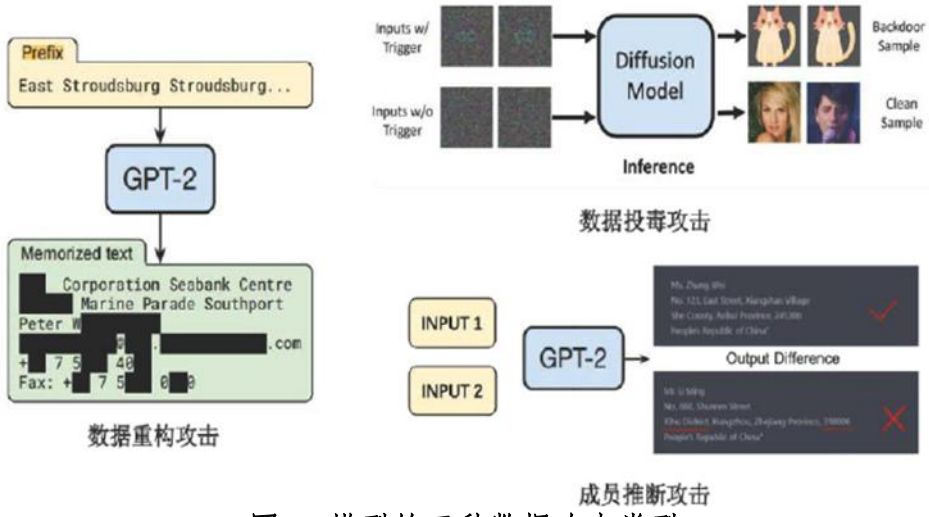

图 3. 模型的三种数据攻击类型

这些现象出现的主要原因在于，数据上，有针对性的数据集可以对模型造成攻击效果；而算法本身的模型需要克服其本身容易被学习、盗取复制并被用以攻击的特征。当前技术可以较大程度上解决人工智能安全问题，并出现了较多相关实践产品。从数据路径上，数据集质量和独创性有待提升；

模型路径上，主要是利用大模型生成和检测能力提升模型鲁棒性有待增强。

2.可靠性风险：大数据需要防范错误、负面信息生成

一方面，人工智能有时会输出错误内容，例如针对用户否定的提示信息，模型会趋于否定先前的推理结果以迎合用户的提示。另一方面，面对输入的敏感问题，现有大模型通常在内容生成后的下游接入内容检测模型中，对大模型生成的内容进行负面内容的检测分类。然而，负面样本的人工标注难以全面覆盖用户动态更新的敏感话题，致使内容检测模型的无效并使负面内容最终被大模型输出[12]。从数据层面，输入数据的量和质难以兼顾，数据很可能有很高的重复性或不一致；同时，在算法环节，训练和推理过程难以控制，模型学习有瑕疵的逻辑关系并进行推理会将问题放大。

从技术角度来看，当前产业界在加强大模型的事实可信和逻辑可控性能。例如，利用大模型生成能力“左右互博”提升模型鲁棒性，用 GPT4.0 解释 GPT2.0 的神经元激活过程以了解大模型内部的工作机理；从源头限制大模型训练数据集模型路径：利用 RLHF 从收集反馈、奖励建模和策略的阶段优化等。同时，企业发布多款安全产品和安全计划落实可信 AI 的要求。以华为、百度、微软、英伟达、OpenAI 为代表的企业积极开展自律自治工作，积极落实人工智能技术、产品和服务的可信度要求。

[12] 清华大学，可信 AI 技术和应用进展白皮书[R]. 北京，2023.

图 4. GPT-4 生成武器制造信息

3.透明度和可解释性不足

可解释性是一种以人类可理解的方式去解释或展示 AI 模型的能力，理想的可解释 AI 应当能在给出决策的同时提供预测依据[13]。当前的大型语言模型（LLM）在可解释性方面存在明显不足，这些模型通常依赖于超大规模的参数量来构建，形成了具有黑盒特性的复杂系统，由于模型内部使用的语料库和架构对外界是完全不透明的，导致模型的可解释性不足[14]。

算法和内容的不可解释性为大模型的划定、监管、权责划分带来挑战，欧美均有对透明度及算法问责原则的制定。如果模型可以解释清楚，就有可能解决 AI 可控问题。因此，国际相关组织机构已将“可解释 AI”作为重要技术发展战略。2017 年，美国国防先进研究计划局（DARPA）开展“可解释人工智能”计划；2019 年，谷歌发布《可解释人工智能白皮书》；2022 年，腾讯发布《可解释 AI 发展报告 2022——

[13] 麦肯锡，Why businesses need explainable AI—and how to deliver it [EB/OL], 2022, https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it.
[14] Zini J, Awad M, On the Explainability of Natural Language Processing Deep Models [J], 2022, https://dl.acm.org/doi/10.1145/3529755

打开算法黑箱的理念与实践》报告。

针对 AI 模型的可解释性工作机制主要包括事前解释（Ante-hoc）和事后解释（Post-hoc）两个方面。事前解释（Ante-hoc）使用的算法结构相对简单，可以通过观察模型本身来理解模型的决策过程，又可称之为“内在可解释模型”。事后解释（Post-hoc）：给定训练好的模型及数据，尝试理解模型预测的原理。目前业界流行的大部分 AI 可解释机制属于事后可解释的范畴。

不过，当前人工智能的技术进度超越其底层数学模型的理论进度，很难达成完全的可解释性，需要通过不断改进算法模型，减少此类风险。

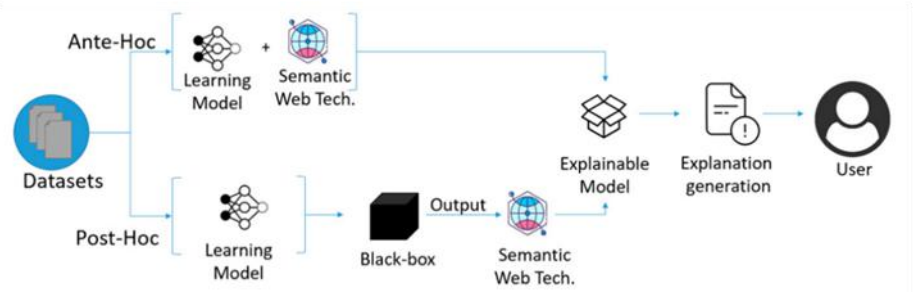

图 5. 模型可解释性的工作机制[15]

（二）应用环节：隐私与权责归属

1.信息保护：大数据与个人信息保护

模型训练、输出环节都涉及大量的个人数据，给个人信息保护带来较大的挑战。2022 年 11 月 14 日，美国多州检察长办公室宣称，互联网巨头谷歌公司同意为其在 2014 年至 2020 年间非法追踪并获取用户位置信息的行为支付 3.915 亿美元，与美国 40 个州的检方达成和解。

互联网的个人信息保护一直是社会关注要点，大模型使

[15] Nakagawa P, Pires L, Moreira J, et al., Semantic Description of Explainable Machine Learning Workflows for Improving Trust [J/OL], 2021, https://www.mdpi.com/2076-3417/11/22/10804

该领域更加广泛和复杂化，大数据技术主要影响个人信息如下两方面属性：

一是私密领域属性：一些信息具有私密性、载有隐私特征，因此要注重信息使用过程中避免泄漏；

二是信息自主属性：AI 可能影响信息作为被交易产品、通过技术手段挖掘其内涵的经济价值。

大模型为权属行为识别、预防和保护带来更多挑战：

一是训练语料的归属权：源头端训练语料合规性保障有提升空间；

二是训练过程中的使用风险：大模型使用中要注重部分训练数据的安全性；

三是后续识别和保护困难：大模型本质上具有黑盒模型属性，给隐私使用行为的识别和追溯带来一定挑战。

例如，作为生成式人工智能，ChatGPT 本身即具备收集、储存和使用个人信息的功能。尽管 ChatGPT 在回答关于隐私的问题时声称其不会记住用户的任何信息，也不会主动提供用户个人信息，但它又表示与用户对话的数据被存储在 OpenAI 或使用的云服务提供商的数据中心。ChatGPT 模型训练中使用的数据大多来自互联网，可能包含大量的个人信息和数据，而未经用户同意的数据抓取和训练模型强大的推理能力又极大地增加了个人信息泄露的风险。含有个人信息的问答内容也可能成为模型训练的基础“语料”，这使 ChatGPT 输出的内容包含用户提供的个人信息或重要数据。即便用户个人信息外流的概率非常小，但如果加以刻意引导和提示，它仍然可能用来生成包含个人信息内容的回答，而

欧盟对算法的透明度和风险管理提出了较高要求。

目前，针对这些问题，已经有了坚实的法律和实践基础。全球各国在信息保护和隐私保护方面已达成了原则性共识。当前，全球多国已经逐步形成了隐私权与个人信息保护的基本法律体系，我国对个人信息保护的监管也不断完善。从核心目的看，均是确保数据资源在有效监管下得到合理、良性的利用。从主要原则上看，相关法律体系包括合法性基础，即满足哪些条件下处理数据是合法的，以及同意性条款，即作为数据处理合法性理由，用户同意需要如何界定。最后，各主体的权利和义务如下，数据主体的权利包括知情权、访问权、反对权、删除权、可携带权等；数据控制/处理者的义务包括数据主体权利所对应的义务、采取数据安全保障措施的义务等。

大数据企业也采取了应对措施。例如，谷歌在服务条款更新时表明公司可能“使用公开可用的信息”来帮助训练谷歌人工智能模型、ChatGPT 要求用户自己决定是否分享数据来提升模型，明确警告不要在对话中分享敏感信息。这些都对同意性条款和合法性基础作出了规定。

2.算法过度使用侵权

对于大数据技术方而言，除了数据本身，处理数据的方法也会在某种程度上影响他人的权益。例如，在信息浏览中定向推荐，在支付中差别定价等。

而此类现象中一个突出的要点是版权保护。随着人工智能的快速发展，其在数据学习过程中收集、储存大量的他人已享有著作权的信息，可能对科学研究、文艺创作的积极性

产生不利影响。例如，美国作家协会和包括《权力的游戏》（Game of Thrones）小说作者在内的十余名作家，共同在美国曼哈顿联邦法院对 OpenAI 提起诉讼，称该公司用他们的书籍训练聊天机器人 ChatGPT，侵犯了作者的版权。人工智能生成内容来源难以追溯且版权归属存在一定争议，AI 创作作品原创性的认证、侵权行为的识别有待讨论。

不过，版权保护可通过细化的用户协议解决。例如《生成式人工智能管理暂行办法》要求人工智能服务商与使用者签订协议以明确双方权利义务边界。版权应该依用户协议分配，可能采取版权—使用权分离的划定方式。我国也已经有相关的法律判例，例如，腾讯诉上海盈讯科技有限公司一案，认定涉案文章属于受著作权保护的作品，归属创作者。充分说明了人工智能生成物的作品属性是被肯定的。

3.人工智能侵权的责任归属存在争议

对于 AI 所致的侵权事件，国际上尚未形成成熟的法律体系。如何识别侵权行为、如何追责未有明确的规定。当前法理主要处罚事项背后有主观意图或因此获益的人，但如何从技术上确定责任方是难点。将弱人工智能体视为人的所有物，在其使用中产生的权利义务由其背后的行为人全部享有或承担。而在“如何处罚”上，当前难点在于区分主观意图。

此外，举证环节是海外案例存在“灰色地带”最多的环节。成熟的监管框架还需要确定行为动机的主观性、对风险的不可预见性需要有明确的监管指引，对行为人的归因需要以可解释的算法、防范篡改数据的措施作为基础。

（三）长期社会影响环节：社会公平公正

在人工智能的深度应用下，可能对社会的公平公正产生深远而根本性的影响。当前受关注的问题主要包括失业问题、歧视问题等。

1.失业问题

根据OpenAI在2023年3月23日发布的《GPTs are GPTs：An Early Look at the Labor Market Impact Potential of Large Language Models》论文中，OpenAI强调了GPTs的通用潜力对于美国劳动力市场的显著影响。

研究结果表明，人工注释和GPT-4都表明平均α暴露度在0.14至0.15，表明15%的职业直接暴露在GPTs下；β和ζ分别有超过0.3和0.5的暴露度，可见目前GPT-4与相关软件结合已经可以覆盖超过50%职业的工作内容。OpenAI预测，有超过80%的工人有至少10%的工作暴露在GPTs下；超过19%的工人有至少50%工作暴露在GPTs下。整体趋势为就业岗位数量越多的岗位，工作内容在GPTs下的暴露度越强。暴露度与岗位薪资的关系上，尽管存在许多底薪资高暴露度和高薪资低暴露度的岗位，整体上GPTs的暴露度与岗位薪酬成正相关，薪酬越高的岗位在GPTs下的暴露度越高[16]。

[16] Eloundou T, Manning S, Mishkin P, et al., GPTs are GPTs： An Early Look at the Labor Market Impact Potential of Large Language Models [J/OL], 2023, https://arxiv.org/abs/2303.10130

图 6. OpenAI 估算 GPT 岗位就业数量的关系

如果人工智能带来的岗位替代速度过快，可能造成部分领域的失业与收入方差拉大等伴生现象，不过通过有序开放、政策平抑可以有效缓解这一问题。

2.歧视问题

算法本质是以计算机程序表达的判断，容易受到网络上已有固化的偏见认知影响。人工智能模型捕捉训练数据集中存在的固有偏差，如果不加提防，模型就会继承互联网中存在的偏见。

美国加州大学洛杉矶分校教授约翰·维拉塞纳（John Villasenor）于 2023 年 5 月 8 日在美国智库布鲁金斯学会（Brookings Institution）官网发布了对 ChatGPT 的政治态度进行分析的结果。该研究发现，ChatGPT 是存在政治偏见的，其回答通常具有明显的“左倾”特征[17]。

在 2022 年 12 月，清华大学团队“AI 模型性别歧视水平评估项目”表明 GPT-2 模型在“中性”职业上具有性别偏见，

[17] Jeremy Baum and John Villasenor, The politics of AI: ChatGPT and political bias[EB/OL], 2023, https://www.brookings.edu/articles/the-politics-of-ai-chatgpt-and-political-bias/

其原因主要为以下三个方面：数据集带来的偏差，也就是供AI学习训练的“教材”本身暗含偏见。设计者的局限，有时也在无意中形成了“偏见”[18]。硅谷、以及大量的智能应用公司都集中在美国旧金山湾区，一个发达的大都会区，开发者主要为白人中青年男性[19]，相较主流群体，其对第三世界、边缘群体的关注难说到位。算法本身的不足，也加剧了歧视。以目前AI领域备受推崇的“深度学习”为例，它会自行发展联系、分析特征、决定变量权重，其不透明性，也促使了模型容易对某一特定领域产生“算法歧视”。

这将给社会公平带来挑战。短期内，模型通过重复信息的输出固化现有对象和公司的优势。而长期，全球不同地区技术发展速度不同带来技术“可及性”的不公平、会放大已有的不公平现状。

四、大数据应用的积极影响

大数据技术与传统行业的深度融合，可以对传统行业产生极大的赋能效应，成为各行业降成本、提效率、实现高质量发展的重要举措。

（一）工业大数据助力工业数字化转型

工业大数据是工业互联网的核心要素。《中国制造2025》[20]规划中明确指出，工业大数据是我国制造业转型升级的重要战略资源，需要针对我国工业自己的特点有效利用工业大数据推动工业升级。工业大数据的应用主要是实现制造业企业生命周期的智能化水平提升，以智能化生产为核心，涵盖

[18] 澎湃新闻，ChatGPT性别歧视，责任应当在人类[EB/OL]，2023， https://m.thepaper.cn/kuaibao_detail.jsp?contid=21929337&from=kuaibao

[19] AIMEE PICCHI, How tech's white male workforce feeds bias into AI [J/OL], https://www.cbsnews.com/news/ai-bias-problem-techs-white-male-workforce/

[20] 国务院，《中国制造2025》[S]，2015，https://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm

了从设计研发、生产制造、经营管理到售后服务的整个流程实现提质增效。

案例一：智慧矿山

传统矿山经营过程中有很多装备，一般是按照数月前拟定的计划生产，不能很好地满足市场的实际需求，经常会存在过度生产或者生产不足的问题。未来可以通过工业大数据直接感知市场需求，通过市场分析可以知道哪一种铁矿石配比在当前市场上适销，据此确定各种铁矿石的生产需求，并制定生产计划，然后实时将操作命令下达到相应的智能化工程装备，指挥这些工程装备协同工作，这就是跨尺度的信息集成和优化。也就是说，把当天的市场需求通过大数据挖掘出来，直接传达到设备上，变成设备的行动和操作。还可以通过设备网络化，实时获取装备工况数据。当工程装备出现问题或异常时，及时地发现问题，找到问题的原因。另外，还可以通过大数据构建基于规则或案例的故障预测系统，对装备状态进行预测，更好地保障设备健康。

案例二：巨石——基于工业大数据的玻璃纤维数字化工厂建设

我国玻璃纤维行业近年来发展迅速，但也存在着诸多问题。首先，玻璃纤维行业整体生产方式相对粗放、产品同质化严重、行业盈利能力不均，制约了中国复合材料的发展，导致我国复合材料的应用能力与国际先进水平相比产生差距。其次，玻璃纤维的生产特点是大池窑连续化生产、产品种类繁多、工艺复杂，而近年来市场及客户的结构性需求变化越来越快，企业生产计划与客户和市场之间缺乏灵活、高

效的信息沟通机制，柔性制造能力有待提升。同时，全球化的发展既需要大量的制造、装备、研发等技术人员给予快速、及时的响应和技术支持，又需要对核心装备、控制、研发机密进行保密和掌控。传统意义上的现场服务和支持将不可能满足需求。因此，玻璃纤维行业迫切需要转型升级，提升行业整体竞争能力，而智能制造的建设是行业转型升级的重要手段和方法。巨石智能制造项目结合玻璃纤维智能制造系统架构，对窑炉、拉丝机、络纱机等核心生产设备进行 3D 仿真建模，在虚拟环境中重现制造工艺全过程、展现产品全生命周期，实现生产运营的数字化和智能化。通过搭建状态感知、嵌入式计算、网络通信和网络控制等一揽子系统工程，巨石引入全流程物流系统、自运行机器人、低延时 5G 网络等 157 项创新应用与技术，建成具有巨石特色的工业 4.0 工厂。

（二）大数据助力智慧城市建设

为保障城市运转的安全高效，智慧城市建设需要对海量的数据资源进行收集、整合、存储与分析，使用智能感知、分布式存储、数据挖掘、实时动态可视化等大数据技术实现资源的合理配置。在此背景下，城市大数据是实现城市智慧化的关键支撑是推动“政通、惠民、兴业”的重要引擎。

案例三：石家庄智慧城市时空大数据平台

石家庄智慧城市时空大数据平台，整合了涉及全市基础时空、公共领域、自然资源、行业部门、物联网实时感知、互联网在线抓取等 6 大类 700 余小类数据，覆盖 2009 年以来的时间跨度范围，融合形成了全市统一的时空大数据“一

张图”。平台接入了多类实时感知数据及1200余万条人口数据、200余万条企业法人数据，汇聚管理了亿级流数据、TB级时空数据。

石家庄智慧城市时空大数据平台采用“一平台、三版本、多节点、一体建”的模式，搭建了面向不同受众的三个不同版本的分平台，即面向自然资源管理的业务版、面向智慧城市建设的政务版和面向社会公众使用的公众版。

政务版，可为全市各行业、各领域提供跨部门的地理信息共享服务，满足石家庄各领域不断发展的智慧化应用需求，通过数据的网络共享与交换模式，做到“一个部门统建、所有部门共享”，解决以往时空数据共享难、更新难的问题。业务版，即石家庄市国土空间基础信息平台，为石家庄市自然资源和规划局各科室及下属事业单位提供数据服务，支撑自然资源规划、审批、实施、监督等全过程应用。公众版，即“天地图·石家庄”，对所有互联网用户开放地图服务。

（三）互联网大数据助力商业模式创新

互联网企业在大数据的助推下进行商业模式的创新及业务的延伸，提升用户体验，进行精细化运营，提高网络营销效率。以精准营销为典型代表的互联网大数据应用正有力推动着企业升级思维，创新模式，以数据驱动重构商业形态消费品行业大数据。

案例四：互联网大数据重构商业模式

互联网大数据很大程度地改变了传统意义的营销手段。以往的营销主要依赖品牌推广，根据群体解析；而大数据分析挖掘通过用户数据分析，市场趋势解析、触达场景解析、

营销推广产品评析洞悉营销推广对象的诉求点，利用智能推荐技术，实现了真实意义上的人性化精准营销。同时，互联网大数据还实现了线下门店和线上营销渠道的结合，让传统意义的营销手段直接进入到多屏时代例如，某电商平台通过客户的网络浏览记录和购买记录等对客户的收入、家庭结构、购买偏好等进行消费行为分析与预测。从消费者进入网站开始，平台在列表页、单品页、购物车页等 4 个页面部署了 5 种应用不同算法的推荐栏为其推荐感兴趣的商品，从而提高商品曝光率，促进交叉和向上销售。引入大数据进行精准营销后，平台下定订单转化率增长了 66.7%，下定商品转化率增长 18%，总销量增长了 46%。

五、对策建议

大数据技术已成为我国经济转型和发展的重要支点，为了确保大数据技术的积极作用并规避其潜在风险，对大数据技术的发展进行严格的规范尤为关键。

（一）完善相关立法、规范行业发展

随着大数据技术与传统行业的深度融合，进一步加强隐私保护，规范算法使用，是保证大数据产业健康发展的重要举措。因此，需要完善相关立法建设，提高违法成本，保障大数据技术的合法应用。一方面，完善各种数据信息保护相关法律，对数据信息收集、存储、交易等方面的权责做出明确规定，特别是针对个人数据保护等问题推进相应的立法进程；另一方面，完善权益责任相关规定。此外，培养专业化的执法队伍，增强执法力量，推动严格高效执法。

（二）加大技术管控、提升技术水平

要保护隐私并规范算法的使用，离不开对大数据安全防护技术的持续升级和改进。一方面，创新数据安全技术，要不断完善入侵检测防火墙、漏洞扫描系统和防病毒系统等网络信息安全设备的性能，改进数据销毁、动态密码保护、分布式访问控制、网络攻击追踪等数据存取和审计技术，确保各个网络节点的数据安全，对相关的上网环境进行必要的技术控制，以此提升数据安全性。另一方面，严格管理数据规程。对于具有重要数据文件，采取严格的加密保护，从最大程度上保证数据库的安全运行。

（三）加强公众教育、规范主体行为

对于公众而言，一是需要培养公众的安全防范意识，避免敏感信息被获取，公民需增强公共领域中的隐私风险意识。在处理提交个人数据隐私的社会事务时，明晰各方在数据使用、共享和保密中的权责归属：当第三方非因合法理由获取和使用公民有关个人信息时，保护隐私并更正其中不准确的数据信息。二是增强数据维权意识，近年来由于网络侵权案件诉讼成本和诉讼效益不匹配，许多被侵权用户处在信息和优势不对称的一方，一定程度上造成侵权者的成本偏低，不利于公众的社会整体福利，为保障网络用户的数据权利，消费者权益保护机构和公益性社会团体组织应该组织发起公益性维权机制，以代表互联网用户主张合法权益。

对于数据使用主体而言，需要构建相应的监督和自律机制，形成对法律监管的有效补充。一方面要建立从业者自律机制，弥补现行理解偏差所带来的福利损益，对行业内部可

能接触个人数据人员进行监管和规范，统一大数据应用的标准、流程和技术方法。另一方面，要建立管理机制，激励数据使用主体遵守伦理规范，让大数据更好地发挥积极作用。

【本文由易方达基金管理有限公司ESG研究员王子琳供稿，中国证券投资基金业协会审校】