

规避 AI 变脸的技术风险

□ 陈 磊

布虚假信息，制造舆论恐慌。

如何规避 AI 变脸技术所引起的风险

AI 变脸技术风险的规制涉及范围广、领域宽、内容多，蕴含活力与秩序、自由与安全、责任与信任等诸多方面，是一项复杂的系统工程。要强化高位统筹、顶层设计、系统谋划、整体推进，以数据算法、合作共治、信息保护、法律法规等维度的同向发力、良性互动，聚力推进 AI 变脸技术异化治理体系的构建创新，有效管控风险，提升整体治理效能。

以数据算法为基石，深化从源头治理。算法既是 AI 变脸的技术核心，又是其生成路径。智能算法具有的数据共享、自主学习、封装移植等特质，将普通民众彻底隔绝在信息内容的生产、过滤、推荐链条外，引致产生与 AI 变脸天然契合的“算法黑箱”。规制 AI 变脸技术的风险，要强化算法治理。强化主流价值引领算法，将其嵌入到 AI 变脸技术发展的全生命周期，引导算法设计者、信息分发者坚持正确价值取向，引导算法应用向上向善。构建算法数据库，收录 AI 变脸技术的生成算法、识别算法和检测算法等模型，以溯源防伪、反向破解辅助识别判定 AI 变脸作品，并将其同步收录数据库，通过样本的累积迭代，提升比对判定能力。优化传播媒介平台内部算法，构建事实核

查、分级检测和分级推荐机制，强化外部算法治理，协同优化算法的准入、监督、投诉和评估等配套制度体系。

以合作共治为架构，推进协同治理。AI 变脸技术异化的规制蕴含组织与制度等多元要素，应秉持审慎、理性、权衡和预防的原则，在网络化、包容性、无缝隙的协作框架下聚智共谋，推进协同治理。要秉持合作精神，培育共治价值，优化完善合作治理、共同治理机制，在规章制度框架内，通过资源整合、功能互补，寻求最大“公约数”，驱动 AI 变脸技术风险治理由点状分散向融合交互跃迁。要以主体跨域协同、路径集成协同、信息交互协同为路径，优化协同治理界面，贯通政府部门、传播平台、技术开发、产品制作、网络用户的责任传导链条，创新协同治理，提振治理质效。强化政府部门、社交平台、科研院所的良性互动，厘清权责关系，破除壁垒、疏解堵点、理顺机制、畅通渠道，共同开发检测审查 AI 变脸的技术算法，画好协同治理“同心圆”，共同应对 AI 变脸技术风险。

以信息保护为支撑，构筑立体屏障。生物识别信息具有人身表征性、不可替代性特点，既是 AI 变脸的原始材料，也是基础根基。AI 变脸技术异化的风险规制，应以“隐私控制”为原则，构建从收集到存储再到应用的多维立体保护体系。AI 变脸技术使用者采集面部识别信息时，应秉持知情同意、目的限制、正当诚信、必要安全的基本原则，

严格遵守个人信息保护法中基于“特定的目的和充分的必要性”“取得个人的单独同意”的条款规定。细化采集入库规则，建立媒介平台动态征询和信息披露机制，基于适度透明、隐私侵犯最小化原则收集信息。完善保管存储规则，媒介平台应采取加密技术，匿名化、匿名化存储用户生物识别信息，严格限定存储期限和敏感应用场景。丰富加工利用规则，制作者、上传者对原始视频采用语音标识、数字水印等方式进行信源标注，优化信息公开、转让，确保检测核查有据可循。

以法律规制为红线，厘清技术边界。秉持“技术—经济—社会”的模式，兼顾必要的社会面向，平衡调和公共利益和个人权益之间的张力。持续完善行政与民事相结合的法律规制体系构造，健全法律规制体系的层次性和梯度性，回应 AI 变脸技术异化导致权益侵害复合化，为技术发展提供方向指引。在 2023 年 1 月 10 日施行的《互联网信息服务深度合成管理规定》的基础之上，进一步厘清 AI 变脸源头制作者的法律责任，细化 AI 变脸可能带来的侵权行为，做好追踪溯源。构建要素上贯通生产、使用、维护、监管等节点，内容上统合审查、商议、互动机制的法律规范框架体系。加大平台处罚力度，增设传播媒介平台的技术检测义务和过滤删除的“守门人”责任，以疏堵结合的策略，细化不同场景应用的法律规制，以法律为 AI 变脸划定红线。

算法（Algorithm），原本只是解决特定问题的数理路径，但现在作为人工智能的“底层代码”，已经嵌入社会资源的调配过程。毫不夸张地说，基于算法的智能应用，正逐渐改变着人们的“存在方式”——社交媒体塑造电子身份、出行软件记录活动轨迹、电子黄页等革新生活方式、网络商城改变消费理念、搜索引擎划定知识边界、生成式人工智能主导知识传播。就此而论，“算法社会”已经悄然降临。

算法可能引发的风险及成因

算法虽具有便利生活、改善决策、提高效率、优化产能等优势，但同时也因缺乏有效的规制而存在诸多难以根除的风险。第一，过度依赖算法决策。即便是经验丰富的司机，完全离开手机导航，也会有如坐针毡之感。随着辅助性算法日趋频繁地介入人们的日常决策，算法同人类的关系好比“象与骑象人”，很难确定最终决定行走方向的究竟是作为潜意识的大象，还是作为显意识的骑象人。第二，“信息茧房”充斥信息获取过程。大型语言模型打底的生成式人工智能可轻易实现海量内容的海量生产，若辅之以基于推荐算法的“精准推送”，人们将被困于由人工智能主导的“信息茧房”之中，甚至将不得不迎合由算法所引导的潮流。第三，技术黑箱所隐匿的决策偏误。数据集偏差、编程人员的自我价值负载都可能导致非中立、不公正的算法结果。资信审查算法对于不同性别的利率差等，反映出算法对既有社会性偏见的强化，AI 聊天软件屡屡发表的有喜好偏见的言论，更是对这一问题的直观体现。

算法运行原理及步骤

无论是乘客在网约车平台选择不同档次的出租车、用户在手机短视频应用中的互动，抑或是商业银行通过评分系统考察客户资信等，这些日常行为所包含的算法活动可大致划分为三个典型步骤：第一步，平台端通过（主动或被动的）数据挖掘，获取用户、消费者及同类群体的原始数据，例如所在位置、健康信息、消费记录、浏览历史、行为言论、运动轨迹等；第二步，云端将规整后的数据输入机器学习模型，输出概率、排名、风险、倾向、估值等具有明确指向性的分数和结果；第三步，应用终端生成可供决策的评价和预测报告，例如，搜索引擎生成排序界面、自动驾驶系统规划行车路径、问诊机器人给出医疗建议、出行帮手设计旅行方案、电子商务网站识别消费偏好并推送广告、平台管理与控制的混合增强智能框架进行虚拟现实等。

人工智能的“智能”水平，事实上同“人工”在上述步骤中的含量成反比，但这并不能否认适当的“人工”干预对于风险治理的效果。工具的好坏取决于使用它的人，促进算法向善的关键在于算法背后的人之向善。从前述步骤拆解来看，有必要将“以人为本”的精神贯彻运用至算法活动的全流程。算法应用具有显著溢出效应，正确使用可推动经济社会各领域从数字化、网络化向智能化加速跃升，实现对社会、经济和生活的全面推动。尤其是先进的法律制度可以为确保算法活动不至于具有胁迫性而建立防火墙。

算法治理及对策

对算法规制的目标在于：一方面，防范本来应当是中立的人工智能应用，成为少数人的集权；另一方面，为技术优势方和算法受众提供商谈场域，避免法律沦为“技术寡头”的帮凶。回顾算法活动的典型步骤，规制算法技术风险可从以下三方面展开。

保障用户和受众享有不受算法“支配”的权利。算法因异化而膨胀出的权力越是绝对，人类的尊严就越受威胁。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



□ 孙小雨 唐林荪

和延续在知识、经验和道德上的卓越。为此，有三个层次的权利，效果从低到高依次为允许数据主体修订或替换不合理数据的“更正权”、允许数据主体要求数据控制者擦除关于其个人数据的“被遗忘权”以及允许个体不受自动化决策限制的“脱离自动化决策权”。

加强以预期效果为导向的信息披露。“算法透明”一直是算法规制的重点方向，但各国政策制定者就如何实现“算法透明”观点不一。2021 年我国国家网信信息办公室和工信部等部门联合发布了《互联网信息服务算法推荐管理规定》，要求算法推荐服务提供者制定并公开算法推荐服务相关规则。落实算法透明的核心在于有效披露算法决策的逻辑、运算和分析过程。为此，技术方必须主动解释清楚智能应用的软硬件和外部环境之间的相互作用将如何影响算法的行为模式，并清晰直观地就以下事项作出真实、准确、完整的信息披露：第一，供机器学习模型使用的数据的来源、典型特征和分类方式；第二，终端程序的运作原理、代码逻辑和预期效果；第三，可能存在的系统偏差、运行故障和矫正机制；第四，智能应用的即期效果和远期影响。对于算法程序的重大修改，技术方有必要及时进行补充信息披露。

确保责任主体链条的追踪识别。目前，随着机器学习的不断进步以及数字技术的相互开源，很多算法程序已经自我改良到可能完全偏离其创作者预设但尚未获得独立法律地位的地步，这种复杂性进一步加大了损害赔偿诉讼确定责任主体的难度。在人工智能主体尚未获得明确的法律地位之前，确保其算法程序能够被追踪至可识别的造物者是当前责任体系构建的基础。可资借鉴的制度设计是无人机牌照试点，如，强化算法运用的事后监管和效果追踪，倒逼技术方事先排除某些会产生不公现象功能的置入。对于集成不同来源、多个算法程序的人工智能应用，跨境数据传输中的常见业务关联交易，提供了可类推适用的解决方案，可考虑在封闭应用和开放应用中都置入硬核编码的审计日志，以便为将来可能的责任划分留下记录。



数字技术如何激发基层党建新动能

□ 郭忠瑞

随着技术的发展进步，数字技术与“智慧党建”成为近年来各级党组织在创新开展党的建设工作中持续关注的热点问题。积极探索数字化赋能、智慧化管理、精细化服务的党建工作高效机制，是创新党建工作模式的重要选择。数字技术能够更加焕发基层党建工作强劲动能，更加精细化其靶向服务职能，在关照党员群众、党务工作者和党委三个层面发挥积极作用。

激发基层党员主动性。数字党建落脚于通过数字技术赋能提升广大党员群众在参与组织活动、开展理论学习等方面的积极性，不断激发基层党员的主观性、积极性。数字党建所包含的多便捷应用、多变场景空间、多维学习资源，为满足党员个性化需求提供了更多可选择的方

案。一方面，在保证组织生活严肃性的前提下，凭借大数据优势，精准分析、预测和把握党员基本需求，有针对性前置管理和提供人性化解决方案，尤其是突破了传统党建模式下时间和空间方面固有的局限性；另一方面数字党建致力于实现教育学习立体化。基于数字技术的新建党建系统涵盖丰富网络学习资源、网络名师课堂等多元知识库，使党员教育学习可不受授课地点、特定时长、限定专业类别的束缚，通过自由选择培训课程，实现自主开启全时段多领域立体式学习。

实现基层工作高效性。基层党务工作在落实中央各项部署中发挥着重要作用，直接关系到基层党组织的正常运转。以智能化方式赋能基层党务工作其优势体现在处理大量数据、精简办事流程时，既能保障党务工作严谨规范、高质高效，又能优化人、财、物各方配置。数字技术提升党务工作效率，一方面反映在通过数字革命扬弃传统模式下人工誊写、收储导致的非标准化、误差

缺漏等弊端，达到对党员数据的全面录入、量化积分和实时更新，实现党务数据分析精确化、科学化；另一方面反映在以互联网为依托实现无纸化管理与无缝衔接，通过数字驱动系统实现业务制度化、制度流程化、流程表单化、表单规范化，方便党务工作者对党务相关事项进行高效审批、全程跟踪和及时复盘，实现党务工作智能化、高效化。

辅助基层决策科学性。党委科学决策是推进党建与业务工作深度融合的坚强保障。在用人决策方面，秉承政治标准始终是选人用人第一标准的原则，借助数字技术能够对党员的创新理论学习情况、组织生活表率情况、急难险重任务攻坚等情况等进行提醒、统计、监督、反馈，纪实呈现党员干部是否一以贯之坚持党的领导、是否始终致力于为党和国家建设锤炼过硬工作技能、是否愿服务社会民生，据此生成量化考评结

果，辅助党委选人用人决策；在经营决策前瞻化方面，通过整合生产、销售、财务、人资等内部链接模块，打造内部共享大数据中台，数字党建着力实现内部党建数据和企业营运数据的自动采集、存储、计算，综合生成直观立体的图表信息，及时为党委推送科学、全面的分析报告，有利于党委作用的发挥，为科学制定规划提供有效支撑。

数字技术驱动下的党建系统涵盖维度多元、涉及领域全面，既是密切党群联系的连心桥，也是传播党的方针政策的扬声器，更是促进企业更好发展的助推剂。数字技术赋以党建工作互联网时代下的智慧模式，成为基层党建工作的重要手段。



（上接 1 版）以科技创新推动产业创新，特别是以颠覆性技术和前沿技术催生新产业、新模式、新动能，发展新质生产力。完善新型举国体制，实施制造业重点产业链高质量发展行动，加强质量支撑和标准引领，提升产业链供应链韧性和安全水平。大力推进新型工业化，发展数字经济，加快推动人工智能发展。打造生物制造、商业航天、低空经济等若干战略性新兴产业，开辟量子、生命科学等未来产业新赛道，广泛应用数智技术、绿色技术，加快传统产业转型升级。深化科技体制改革，深化科技评价改革，加大多元化科技投入，加强知识产权法治保障，形成支持全面创新的基础制度。提高维护重要产业链供应链安全的能力。

按照党的二十大关于确保重要产业链供应链安全的部署，加快实施创新驱动发展战略，加大科研开发工作的投入与力度，不断减轻重要产业链供应链对国外的依赖程度，不断提升维护重要产业链供应链安全能力。要从高水平科技安全的长远发展战略着眼，加强基础研究和前沿研究，强化企业科技创新主体地位，真正实现科技创新自立自强。针对我国目前重大原创性成果缺乏、顶尖基础研究人才匮乏、基础研究的社会支持环境亟待优化等问题，应进一步加大对基础前沿研究的支持与投入，提高基础研究投入总量，优化基础研究投入的整体结构，加大中央财政对基础研究的稳定支持力度，引导鼓励地方、企业和社会力量增加基础研究投入，构建基础研究多元化投入机制。

坚持社会高质量发展与高水平公共安全的良性互动

新时代新征程，我国发展进入战略机遇和风险挑战并存、不确定难预料因素增多的时期，各种“黑天鹅”“灰犀牛”事件随时可能发生。我们必须增强忧患意识，坚持底线思维，做到居安思危、未雨绸缪。党的二十大报告对提高公共安全治理水平作出了重要部署，明确指出要坚持安全第一、预防为主，建立大安全大应急框架，完善公共安全体系，推动公共安全治理模式向事前预防转型。推进安全生产风险专项整治，提高防灾减灾救灾和重大突发公共事件处置保障能力，加强国家区域应急力量建设。公共安全主要涵盖自然灾害、安全生产、公共卫生、社会安全四个领域，实现高水平公共安全、增强维护公共安全的能力需要从这四个领域入手。在自然灾害领域，全球气候变暖导致极端天气事件易发多发，我国自然灾害风险治理面临严峻复杂的形势。坚持“以防为主、防抗救相结合”的自然灾害治理方针，建立健全自然灾害风险预警与处置机制，完善自然灾害风险应急处置体系，不断提高自然灾害的防灾、减灾、救灾能力。在安全生产领域，目前我国安全生产事故总量仍然过大，重特大事故尚未得到有效遏制，安全生产形势依然严峻。进一步加强安全生产风险治理能力，推动安全生产风险治理模式向事前预防转型，推进安全生产风险专项整治，加强重点行业、重点领域安全监

管。实施城市更新行动，打造宜居、韧性、智慧城市，有效化解城市发展高风险。在公共安全领域，加强重大疫情防控救治体系和应急能力建设，有效遏制重大传染病疾病传播。随着自然因素和社会因素的不断变化，我国公共卫生面临鼠疫霍乱等法定报告传染病、突发急性传染病、境外输入传染病以及生物技术滥用、滥用、滥用的风险，及公共卫生事件防控难度加大，食品、药品安全基础依然薄弱等诸多风险与问题。应从事前预防、事中处置、事后恢复等多个环节，全方位、全流程提升社会公共卫生治理水平与治理能力。在社会安全领域，当前我国改革进入深水区，经济发展进入新常态，各种矛盾叠加，风险隐患积聚，不同类型的社会安全事件相互交织。全面做好就业、教育、社会保障、社会治安等各方面工作，坚决维护社会稳定，不断增加人民群众获得感、幸福感、安全感。坚持尽力而为、量力而行，兜住、兜准、兜牢民生底线。更加突出就业优先导向，确保重点群体就业稳定。织密扎牢社会保障网，健全分层分类的社会救助体系。坚持和发展新时代“枫桥经验”，健全共建共治共享的社会治理制度，完善社会治理体系，提升社会治理效能，坚决维护社会安全稳定，努力实现高水平的社会安全。

坚持高水平对外开放与海外利益安全的良性互动

随着全球化趋势不断加强，我国与世界各国的交流日益增多。我国海外利

益主要涉及对外贸易、对外投资、对外工程承包、驻外机构、海外中国公民等主要方面。我国境内投资者在国（境）外直接投资的企业越来越多，境外企业从业员工、出入境人员、留学生均在快速增加。近年来，中国机构、企业、公民等广泛走向世界，海外利益安全的迫切性、重要性日益增加。特别是海外安全事件频发，使得我国海外利益安全面临严峻考验，必须高度重视我国海外重大项目、驻外机构和人员安全，确保国家利益、企业利益和个人利益安全。

习近平总书记指出，当前，世界大变局加速深刻演变，全球动荡源和风险点增多，我国外部环境复杂严峻。国家安全法第 33 条明确规定：国家依法采取必要措施，保护海外中国公民、组织和机构的安全和正当权益，保护国家的海外利益不受威胁和侵害。党的二十大报告指出，加强海外安全保障能力建设，维护我国公民、法人在海外合法权益。加强海外利益保护，确保海外重大项目 and 人员机构安全。探索建立境外项目风险的全天候预警评估综合服务平台，对于国家海外利益和国家海外安全风险状况做到及时预警、定期评估。从国家层面上加强海外利益保护、国际反恐、安全保障等机制的协同协作，形成维护海外利益的合力。进一步加强涉外法治建设，积极参与国际法治，依法应对国际形势。积极推进国际传播能力建设，讲好中国故事。加强对社会组织走出去的指导，推动社会组织参与海外利益安全维护。

〔作者系中央党校（国家行政学院）应急管理培训中心主任、教授〕