

AI 拓展慈善边界，但须警惕技术成为新神

金锦萍

在人工智能技术浪潮席卷整个社会之时，AI 技术也涉足公益慈善领域的方方面面。

例如字节跳动通过“一横一纵”策略构建 AI 的公益应用场景，提升数字化服务能力。“一横”是指聚焦公益机构提质增效，利用飞书、火山引擎、豆包大模型等工具优化机构管理、筹款传播、项目管理和捐赠服务；“一纵”则针对社会痛点，探索创新解决方案，如其与北大合作开发“识典古籍”平台，利用 AI 数字化古籍，等等。

腾讯公益则利用生成式 AI 技术，在公益慈善领域进行了多方面的探索与应用。初期，技术团队优先落地了对内提效的场景，如 AI 自动化生成小红花测试题目、公益项目规划和预算的智能审核，以及生成项目故事卡和数据卡，显著提升了运营效率并降低了风险。随着技术成熟，腾讯公益又适时推出了面向用户的产品，如推出“我的公益故事”活动，增强了公众参与感。此外，AI 还被应用于乡村发展（如“小耕助手”）、数字支教（如“支教教师助手”）和文化传播（如博物馆数字化讲解工具），助力社会价值的普惠化。

慈善组织也开始尝试和探索 AI 技术的应用。例如爱佑慈善基金会通过构建动态知识中枢、搭建知识图谱和实时更新机制，整合多源数据，打破信息孤岛。真爱梦想基金会也积极拥抱人工智能技术，致力于通过 AI 赋能教育、养老和公益场景，推动社会公平与效率提升。

这些互联网公司和慈善组织也开始意识到隐私保护、公益伦理、技术开放和行业协作的重要性，在利用 AI 技术推动公益慈善更智能、透明和创新的同时，也须确保技术应用符合道德与法律要求，推动技术、伦理与响应机制的动态平衡，展现了 AI 在公益慈善领域的创新价值与社会效益。例如爱佑基金会通过垂直领域进行生态共建和推出伦理控制沙盒，而真爱梦想则秉持“先提升生产力，后创新生产方式”的原则，以开放协作的态度联合学术界、社区组织等多方力量，确保 AI 解决方案贴合实际需求且符合伦理。

在 AI 技术加速渗透公益领域的当下，工具理性与价值理性的博弈呈现出复杂态势。工具理性通过算法优化、数据驱动等手段提升公益效能，但其过度扩张可能挤压价值理性的生存空间；而价值理性对人性尊严和社会公平的坚守，又为技术应用划定伦理边界。这种动态平衡的打破，本质上是技术权力与社会价值的深层冲突。

在 AI 加持之下，工具理性膨胀和价值理性收缩的三重表现：

其一，效率至上的决策机制可能会导致算法霸权对人文价值的碾压。AI 技术在公益领域的应用虽提升了服务效率，却可能以“效率至上”的逻辑消解人文温度。例如，视障辅助 AI 通过环境识别技术帮助视障者规避物理障碍，但其“无情感反馈”的特性，却无法替代人类志愿者与视障者进行交流时的情感联结。视障者失去的不仅是物理障碍，更是社会互动中“被看见”的尊严。这种技术的简单介入有可能将人文关怀简化为可量化的数据服务。

再如古籍数字化项目使用 OCR 技术时，算法自动抹除“不完美”笔触，将承载千年文化记忆的文献降维为可检索的数据节点。其深层危机在于技术对“人性需求”的异化，当公益组织将 AI 技术作为“标准服务”推广时，实质是在复制“技术霸权”。

例如在慈善资源配置环节，如果慈善组织使用 AI 模型评估贫困家庭援助优先级时，系统设置简单将“家庭年收入”和“居住范围”作为核心指标，就会导致居住在城中村但实际生活困难的无房家庭被排除在外。这种机械化的评估体系，有将复杂的社会困境简化为可计算的数字游戏的嫌疑。如果互联网平台系统，通过用户捐赠历史数据自动生成个性化叙事，通过算法不断向用户推送其曾经资助过的公益项目，在引发捐赠者产生共

情疲倦的同时，也会产生 AI 幻觉中的社会问题，使真实困境反而得不到相应关注。

其二，自动化审核系统以“敏感词过滤”为核心构建伦理防线，有可能陷入“以算法规则对抗复杂人性”的悖论。如某公益平台开发的“AI 劝募文案生成器”内置算法模型，自动过滤包含“弱势群体”“悲惨遭遇”等关键词的文案，原先是为了避免引发“过度共情”而导致社会资源配置不当，甚至导致诈骗横生，但是这种基于效率考量的简单化内容审查机制，却会对公益传播中人性温度产生系统性消解。

现有 AI 系统尚无法理解“文化敏感性”的复杂内涵——例如性别平等议题中的隐喻表达，可能被误判为“煽动性内容”，而涉及少数群体的文化符号可能因算法偏见被系统性过滤。这种程序化消解不仅威胁言论自由，也会使技术权力重塑公益话语的边界，更使公益组织陷入“自我审查”的伦理困境——如果慈善基金会使用 AI 智能客服时，为规避法律风险过度简化服务话术，则有可能最终损害其与捐赠者的信任连接。这种数据驱动的伦理消解可能会引发算法黑箱对人性人文的侵蚀。

其三，技术霸权的话语建构有可能导致资本逻辑对公益本质的异化。如果某掌握了 AI 技术的平台要求入驻的慈善组织必须使用其自研工具链才能获得资助资格，那么就会导致某些公益项

目因拒绝接入数据采集模块，被系统直接判定“不符合数字化转型标准”而失去资助资格。

再假如某公益组织使用 AI 分析困境儿童社交数据时，通过面部微表情识别技术判断心理状态。如果这些数据被打包出售给教育机构用于精准营销，则会导致使技术应用异化为新型商业工具。

再比如非遗保护项目中，AI 修复古画的算法如果自动抹除“不完美”笔触，可能会将八大山人的写意山水被改造成工笔画风格。这种以“优化”为名的技术干预，可能会导致人文遗产遭受损害。

又比如某些公司开发的“志愿服务效能评估系统”，如果将志愿者服务时长、帮扶对象数量等量化指标作为主导性甚至单一性评价标准，则会导致志愿服务中最为可贵的内容被边缘化。

于是我们不得不意识到：公益领域的技术应用正面临“工具理性”与“实践智慧”的断裂——前者追求可计算性，后者依赖情境化判断；前者制造确定性幻觉，后者则拥抱不确定性真实。或许破解这一困局的关键，在于重建“技术向善”的价值坐标系，即公益领域的技术应用必须建立“人性优先”的伦理审查机制。例如联合国教科文组织要求“人权决策必须保留人类终

控权”，同时在算法设计阶段就植入公益慈善的核心价值，为技术理性划定不可逾越的价值红线。

“机器的工具性更主要与它们被设计出来去生产的客体有关，而非服务于人生命过程。”这是汉娜·阿伦特关于机器工具性的论断，揭示了现代技术与社会关系的深层矛盾，批判技术工具性从服务人类生命过程向自我目的化生产的异化转向。

不可否认的是，现代社会普遍存在技术异化现象。当机器生产能力超越人类生物节律时，人的生命过程（出生、劳作、死亡）被重新编码为可计算的技术参数。如困在系统内的外卖小哥，完全服从算法系统，这种劳动动物化状态印证了阿伦特的警示——技术不再服务于人的完整性，而是将人降维为生产链条的“耗材”。更严峻的是，算法管理系统通过生物识别、情绪监控等技术，正在将这种支配延伸至人的精神领域。

所以需要重建技术伦理的可能性，回归行动哲学的救赎路径。真正的工具性应回归“为人的世界建构”本质：其一，技术创新需以维护人类尊严为最高准则，如无障碍设施设计突破功能主义局限，体现对差异生命的尊重。

其二，技术使用的节制：在医疗 AI 领域，世界卫生组织提出的“人类监督原则”要求 AI 诊断必须经过医生复核，防止技术僭越生命决策权。

其三，公共领域的再造：通过民主参与机制将技术评估纳入民主决策框架，打破科技精英的垄断话语。机器（和技术）的工具性不应是单向度的效率崇拜，而应成为拓展人类可能性的中介。

当前自动驾驶技术面临的“电车难题”、基因编辑引发的伦理争议，本质上都是对“谁是主体”这一元问题的拷问。唯有在技术设计与人类生命价值的辩证对话中，才能避免陷入阿伦特预言的“工具成为新神”的文明困境。

公益慈善一方面是对受益人的拯救，无论是从生存援助还是能力赋能。传统慈善往往以物质救济为主，但现代慈善更强调通过系统性支持帮助受助者实现可持续发展。例如，慈善组织通过教育培训提升弱势群体的劳动技能（如养殖、运输等），不仅解决短期温饱问题，更培养其市场竞争力。同时，心理辅导帮助受助者重建自信，缓解因贫困或创伤导致的自卑情绪，促进其积极融入社会。这种以“助人增能”为核心的慈善模式，使受助者从被动接受者转变为自我发展的主体。

另一方面，公益慈善又何尝不是对施助者的拯救？慈善行为对施助者同样具有深远的心理价值。心理学研究表明，捐赠行为常与自我实现需求相关，通过帮助他人，施助者可获得成就感、社会认同感及精神满足感，从而缓解“幸存者内疚”或社会比

较带来的焦虑。例如，参与慈善活动能增强个体的社会融入感，拓展社交网络，并在团队协作中提升组织能力。此外，慈善行为还能通过道德资本积累，满足人类对善的追求，甚至成为现代人应对虚无感的精神寄托。

慈善的双向拯救作用源于其人性共鸣的本质。对受助者而言，慈善不仅是生存支持，更是对其主体性的尊重；对施助者而言，慈善通过利他行为实现自我价值的确认。然而这种双向拯救需以尊严维护为前提。若慈善异化为“嗟来之食”或“道德表演”（如算法决定救助资格、要求公开致谢则可能引发心理创伤），则会适得其反。

因此，真正的慈善应平衡物质援助与精神尊重，实现施助者与受助者的共同成长。慈善通过赋能受助者与滋养施助者，构建了社会互助的良性循环，其价值远超简单的资源再分配，成为维系人性温度与社会和谐的重要纽带。当 AI 应用于公益慈善领域时，需要警惕的便是施助方通过技术强化服务能力的同时，忽略对于受益人主体性的尊重。

公益慈善核心目的是促进社会福祉、帮扶弱势群体、推动社会公平与正义。从本质上讲，公益慈善关注的是人的福祉提升和社会的改善，这往往涉及利他主义、社会正义、人道关怀和共同体构建等无法量化和标准化的内容。

在公益慈善领域中，可量化与标准化应用日益重要。可量化指的是能够用数字或具体指标来衡量和表达。捐赠金额、受助人数量、项目完成的数量和志愿服务的小时数等，均是非常重要的量化指标。标准化是指建立统一的规范或流程，使得不同的组织或项目可以按照相同的标准进行操作和比较，包括项目管理流程、效果评估指标、财务报告格式等。不可否认的是，可量化和标准化在公益慈善中的应用有助于提高透明度、增强问责制、便于比较和评估、帮助选择更有效的项目或组织，并且在此基础上，进行规模化的复制和推广。

不可否认，公益慈善组织有量化和标准化的压力。为了满足捐赠者需求，因为许多捐赠者希望看到具体的、可量化的成果，组织因此需要提供数据；而政府和资助机构往往要求标准化的报告，组织不得不进行量化。从内部管理而言，量化指标有助于内部管理和员工绩效考核。

但是，过度强调量化和标准化则会导致一些弊端：一是容易忽视质性影响。公益慈善的许多影响是难以量化的，如个人尊严的恢复、社区凝聚力的增强。而且过度关注数字可能忽视更深层次的社会和情感影响。

二是可能导致目标置换。为了追求可量化的成果，公益慈善组织可能更关注“容易测量”的方面，而忽略更复杂但重要的目标。

三是标准化可能忽略多样性。例如不同社区、不同问题的解决方案可能需要个性化，标准化可能无法适应所有情况。

四是导致道德和情感的淡化。公益慈善的初衷是出于同情和关爱，过度强调数据和标准可能使其变得过于机械和冷漠。例如一个旨在增强社区凝聚力的项目，其成功可能体现在居民之间的信任和合作上，这些很难用数字量化。

五是缺乏长期主义的价值坚持。推动性别平等或种族正义的公益项目，其影响是长期和广泛的，短期内难以用具体指标衡量。过度关注短期可量化的成果，可能会忽视长期的社会文化改变。

所以，可量化和标准化是公益慈善的重要工具和手段，但并非其核心内容。核心内容应始终关注人的福祉和社会改善的本质，而量化和标准化是为了更好地服务于这一核心。所有量化和标准化的努力，都应服务于公益慈善的根本目的。根据项目特点，灵活选择是否及如何量化和标准化，避免僵化。结合定量和定性方法，全面评估公益项目的影响。

AI 技术的应用将可量化和可标准化的内容拓展到极致，在提高效率、透明度和问责方面发挥着重要作用，但这更多是辅助手段而非核心。真正的公益慈善应关注深层次的人性需求和社会的质性改变，避免因过度追求可测量而偏离初衷。

不可否认的是，AI 技术在公益慈善领域的应用，确实能够显著提升效率并降低成本，例如通过自动化流程优化项目管理、利用大数据精准识别需求，以及通过 AI 辅助筛查疾病、分配物资等提升服务可及性。然而，慈善的核心价值并不仅限于此，而在于其人文关怀与社会公平的深层追求。

例如在教育领域，阿里巴巴开发的“追星星的 AI”绘本工具，通过输入故事梗概即可生成定制化有声绘本，结合家长与教师的人工干预，为孤独症儿童提供个性化教育支持。字节跳动“彩虹熊熊”项目将 AI 大模型植入游戏角色，通过互动陪伴缓解患癌儿童治疗恐惧。北京大学与字节跳动合作的“识典古籍”平台，集成 OCR 校对、AI 问答等功能，上线超 1 万部古籍资源，实现“一键对话”式研究。农潮科技“绿美阅行”平台则通过 AI 数字人交互，引导用户探索城市动植物，解锁图鉴兑换濒危物种周边。但是需要强调的是，技术需要与人性化服务结合，才能真正触及慈善的本质。

值得注意的是，公益慈善组织正通过数据治理与伦理设计，平衡技术工具性与价值导向。例如，爱佑基金会通过 AI 智能客服提升捐赠服务效率的同时，仍强调保护受益人隐私；腾讯公益采取“机器初筛+人工复核”模式，就是监督型人机协作的典型实践。该机制通过 AI 完成标准化初筛（如项目合规性审查），将人类工作者从重复性劳动中解放，同时保留关键决策节点（如情感共鸣点设计）的人性化把控。这种分工使效率提升 40% 的同时，避免算法对公益伦理的系统性消解。再如阿里公益的“伦理探针”系统，通过多模态嵌入技术将价值理性编码到算法底层。该系统不仅监测资源分配的公平性（如城乡资金分配比例），还能识别文化敏感性风险。

这些实践表明，AI 的“降本增效”应服务于更公平的资源分配和更温暖的连接，而非取代公益中“人”的主体性。一言概之，AI 在提升公益效率的同时，正通过人机协同（如教师参与 AI 绘本设计）、伦理嵌入（如救援设备内置道德规则库）、文化创新（如古籍 AI 助手）等方式，探索技术与公益价值的深度融合。

但是这些还远远不够。当技术发展到这个阶段，公益慈善不仅需要确保自身不被技术异化，还需要不断去解决因为技术异化而带来的诸多社会问题。所以这一问题的探索才刚刚开始，但是好在 AI 技术应用公益慈善也才刚刚开始，一切似乎还来得及。

人类文明发展史是一部禁忌、犯禁忌与突破禁忌的历史。早期人类通过禁忌来避免生存风险和确保种族繁衍，例如饮食禁忌与乱伦禁忌；或者通过神灵惩罚触犯禁忌的人来建立权力性神圣，例如“为尊者讳”；或者通过社交忌讳来构建心理防御，例如避讳谈论死亡。后来我们以成立国际原子能委员会来控制核能的和平应用，以伦理审查来避免基因编辑的滥用，以禁止克隆人本身的禁令来维护人主体性，而人类的理性却不断地通过科学发展拨开认知迷雾，以不断突破禁忌甚至重新定义禁忌的方式追求自由，而最终跑到了人本主义的边缘，触及到人的主体性遭遇前所未有的冲击——这或许就是 AI 应用到公益慈善领域的最为根本的问题所在。

当 AI 技术应用到公益慈善领域时，本质是在比特与原子、效率与价值、控制与自由之间的碰撞。技术不应是价值理性的掘墓人，而应成为其延伸的触角。未来的公益图景，必将是人类在数字文明中重新为“善”定义禁忌和边界的伟大实验场——在这里，以技术折射人文和人性之光，以价值理性引导技术的发展方向。

（本文来源：《中国慈善家》2025 年 4 月 21 日，作者系北京大学法学院副教授、博士生导师，中国社会保障学会慈善分会副会长）