算法推荐在社区心理援助中的应用风险与治理路径

张力文
（山东建筑大学）
DOI：10.65285/RWYSHKX.V2I6.016

摘　要：居民的情绪画像正被社区平台经由算法逐一采集，干预内容也随之依个体差异定向推送，一套以数据为内核的社区心理援助范式由此成形。其运作原理并不复杂：自然语言处理与多模态情感识别先把散落于日常生活中的焦虑、孤独、睡眠障碍等零散信息转译为可供运算的风险指标，推荐引擎再据此匹配疏导资源。但技术理性的扩张未必价值中立。训练语料携带的社会偏见，会沿年龄、性别、户籍等维度造成服务供给的落差；情感数据自采集至存储的各环节均有暴露可能，隐私边界随之收紧；情绪体验一旦被反复标签化并分类归档，居民在“被诊断”的同时，其自我叙事的能力也可能悄然退化。本文立足基层社工实务，循“风险生成—深层机理—协同治理”的线索展开分析，尝试给出兼顾服务效能与人文关怀的治理思路，以期为智慧社区心理援助的规范化运行提供学理参考。
关键词：算法推荐；社区心理援助；数字鸿沟；伦理治理

Application Risks and Governance Paths of Algorithm Recommendation in Community Psychological Assistance

Liwen Zhang
Shandong Jianzhu University

Abstract: Community platforms collect residents' emotional profiles through algorithms and push targeted intervention content according to individual differences, forming a data driven paradigm of community psychological assistance. Its working mechanism is straightforward. Natural language processing and multimodal emotion recognition convert scattered information about anxiety, loneliness, sleep disorders and other emotional signals in daily life into computable risk indicators, and the recommendation engine matches counseling resources accordingly. However, the expansion of technological rationality is not necessarily value neutral. Social biases embedded in training corpora may lead to unequal service supply among groups divided by age, gender and household registration. Emotional data faces leakage risks in all links from collection to storage, which squeezes the scope of personal privacy. When emotional experiences are repeatedly labeled and filed, residents may gradually lose the ability of self narrative while being labeled. Based on grassroots social work practice, this paper follows the research thread of risk formation, internal mechanism and collaborative governance. It proposes governance strategies balancing service efficiency and humanistic care, so as to provide theoretical reference for the standardized operation of psychological assistance in smart communities.
Keywords: algorithm recommendation; community psychological assistance; digital divide; ethical governance

前言：

2022 年《互联网信息服务算法推荐管理规定》施行，算法治理自此有了制度依据。此后数年，大数据与人工智能技术持续向基层服务末梢渗透，多个城市的智慧社区平台陆续上线心理援助模块，试图借算法匹配打通“需求发现—资源对接—干预介入”的链条。供需之间的缺口是明确的：中国科学院心理研究所 2025 年发布的蓝皮书显示，全国约两亿人存在不同程度的心理困扰，其中获得专业干预者不足百分之五，基层社区被寄望于弥合这道断裂带。问题在于，当推荐算法从商业消费场景迁移至心理援助这类高敏感领域，其效率优先的技术逻辑便同服务伦理迎面相撞——算法求的是最优解，心理援助倚重的却是关系信任与共情陪伴，二者难以直接换算。本文聚焦这一张力，追问算法嵌入社区心理援助后产生了哪些非预期后果，制度设计又当如何回应。

一、算法推荐嵌入社区心理援助的运作图景

（一）数据采集与情绪画像的生成机制

智慧社区的心理援助模块，其原始数据池通常由三部分构成：居民健康档案、网格巡查记录与线上问卷反馈；语音情感分析、社交媒体公开文本挖掘等技术手段则作为补充。据此，平台可为每位居民生成多维度情绪画像。画像并非静态标签的堆积，而是随数据流入不断迭代的动态模型——社区 App 上的一次搜索、一通热线、网格员入户时的几句交谈，都可能成为模型更新的输入。画像颗粒度越细，后续推荐的“适配感”越强；可采集行为越是隐蔽，与居民知情同意之间的落差就越大。这道落差，构成第一

重伦理裂缝。

（二）干预内容的算法匹配与推送逻辑

画像生成之后，推荐引擎依据协同过滤、内容相似度计算与深度学习排序模型，从心理科普文章、自助练习音频、线上团体辅导课程等资源库中筛出所谓最优匹配。试点社区的平台日志里不乏这样的记录：一位老年居民连续数周在深夜浏览失眠相关内容，随后陆续收到助眠冥想音频与焦虑自评量表的推送。表面看，技术对个体需求的捕捉堪称敏锐；细究则会发现另一层问题——推荐系统的优化目标被设定为点击率与完播率，而非居民心理状态的实际改善。商业平台遗留的流量思维，正于不经意间改写心理援助的价值坐标。

二、算法推荐在社区心理援助中催生的三重风险

（一）情感茧房效应：负面信息的螺旋式强化

推荐算法的底层激励倾向于投其所好。一位情绪低落的居民点开一条抑郁症状科普视频后，系统往往在随后数日密集推送同类内容。针对国内某头部短视频平台的计算实验显示，其新闻推荐算法对用户消极情绪偏好的捕捉与强化格外敏感。社区调研中也观察到类似轨迹：中年居民从浏览“职场焦虑十大征兆”“抑郁症前兆自查”一类推送起步，逐步滑向持续性失眠与自我怀疑。负面信息被螺旋式投喂，个体遂被同质化的痛苦叙事围困，这便是“情感茧房”。主观世界困在狭窄的病理化框架之内，心理弹性反倒在“被关心”的过程中遭到削弱。

（二）数字排斥：算法服务对弱势群体的结构性遮蔽

社区心理援助的对象谱系里，独居老人、残障人士、低学历流动务工者占比可观。这一群体或欠缺操作智能终端的技能，或难以准确理解线上问卷，甚至从未注册过社区 App 账号。平台若以“活跃用户数据”训练推荐模型，沉默群体的心理需求便在数据层面被自然蒸发——算法看不见他们，也就不会向他们推送任何资源。技术能力落差由此转化为服务排斥，并与传统数字鸿沟叠加，使本应发挥兜底功能的社区心理援助，反而加剧了资源分配的不公。

（三）隐私侵蚀与信任瓦解：数据流动的暗面

心理援助得以成立的前提，是一段足够安全的信任关系；来访者袒露内心，靠的是对隐私保护的确信。算法驱动的服务链条却把这份脆弱暴露无遗：情感数据须流经采集终端、云端服务器、模型训练管线、推荐决策引擎等节点，每一处都可能是泄露敞口。更深的隐患在于数据的越界使用。部分社区平台将心理画像与物业管理、信用评价等系统打通，曾有被标记为“中度焦虑”的居民在申请社区志愿者岗位时遭到隐性筛选——心理数据一旦脱离援助语境而被挪作他用，居民对社区服务体系的基本信任便随之崩解。

三、风险生成的深层机理剖析

（一）技术理性的越位：效率逻辑对关怀伦理的挤压

算法推荐的实质，是把复杂的人类情感体验压缩为可运算的特征向量，再以效率最大化为目标输出决策。心理援助的专业伦理却要求从业者悬置预判，以“在场”的姿态倾听那些未被言说的部分。两种逻辑的冲突并非单纯的技术缺陷，其根须扎在工具理性与价值理性的哲学张力之中。社区管理者若以覆盖人数、响应时长、匹配精准度作为绩效考核指标，算法的效率优势便被进一步放大；关怀伦理所要求的耐心、留白与对不确定性的容忍，则被一步步挤到边缘。

（二）训练语料的偏见内嵌：社会结构的数字化复刻

模型从历史数据中学习，而历史数据本身已凝结着社会偏见。训练样本若显示女性焦虑障碍诊断比例偏高，模型便倾向将女性的情绪表达更多归入“焦虑倾向”；某社区的流动人口心理健康数据若长期缺失，模型对该群体的推荐精度就会系统性偏低。算法并未创造新的歧视，它只是以数学语言把既有不平等重新编码、加以固化——再披上一层“客观计算”的外衣，令其更难被察觉，也更难被质疑。

四、协同治理路径的构建

（一）“人在回路”的服务架构：让算法退回辅助位

北京海淀区东升镇八家社区已在居民微信群部署 AI 社工，实现全天候秒级应答；社区党委书记涂桂敏坦言，机器只能处理标准化诉求，真正涉及情绪疏导的对话仍需人工接手。这一经验指向一个明确定位：与其说算法是关系的缔造者，不如说它更像信息的分拣员。可行的框架是把干预流程拆为三段。推荐引擎先从健康档案、网格巡查记录中提取情绪波动信号，生成候选干预方案清单；驻点社工逐一审阅，结合入户走访得来的非结构化信息（独居老人家中堆积的过期食品、青少年房间角落散落的空酒瓶，皆属此类）作出专业研判；干预决策最终由社工与居民面对面协商敲定。灾后心理援助的实务复盘也反复印证一点：AI 预筛出的高危个案，须经持证咨询师二次确认，方可启动危机干预，毕竟误判造成的二次伤害往往比漏判更难收拾。人的判断力与算法的计算力之间，需要一条清晰的权限边界。这条边界牵涉的并非技术参数，而是一个伦理命题：谁有权定义一个人的痛苦。

（二）包容性设计：跨越数字鸿沟的服务冗余机制

葫芦岛市 2026 年启动的“心邻相伴情暖社区”项目覆盖全市 24 个试点社区，其策略之一，是专业社工与志愿者搭档入户巡访，面对面采集独居老人、残障人士的心理状态——绕开智能终端的操作门槛，以最朴素的人际接触填补数据盲区。平台层面的包容性设计可沿两个方向展开。线下：铺设纸质心理自评量表投放点，开通固定电话热线，安排网格员定期敲门探访，确保从未下载社区 App 的居民不被挡在服务体系之外。算法内部：植入“沉默用户发现”模块，居民若连续六十天未留下任何线上行为痕迹，系统不再将其静默归档，而是自动生成入户核查工单，推送至对应网格的社工。这种设计的立足点，是承认技术覆盖天然存在结构性盲区，并以制度化的冗余安排加以填补——冗余不是效率的对立面，它是社会公平的具象化表达。既

有社区心理服务研究亦指出，线下人际渠道对高龄独居群体的服务可及性具有补位作用。

（三）数据治理的全链条伦理规制

已有社区 AI 救助系统因算法漏洞致居民健康档案遭第三方商业机构抓取，此类事件暴露出心理数据在流转链条上的脆弱。情绪画像自采集终端出发，经云端服务器、模型训练管线、推荐决策引擎层层传递，每一环都须设防。治理的着力点应贯穿数据全生命周期。采集环节，以图形化界面逐条呈现数据用途，居民在手机上滑动勾选、逐项授权，而非以一纸冗长格式条款裹挟其一揽子同意。存储环节，引入联邦学习架构：各社区原始心理数据留存于本地服务器，仅将加密后的模型梯度上传至中央聚合节点，辅以差分隐私算法为统计结果注入可控噪声——“数据可用不可见”由此落地。使用环节，借权限隔离机制将心理画像的调用范围锁定于援助服务场景，物业管理、信用评估、商业推送等无关模块一律不得越界读取。社区层面还可组建由持证社工、居民代表、数据安全技术专家三方构成的伦理审查小组，按季度对算法模型作公平性与合规性的独立审计。

（四）算法透明与可解释性的制度化保障

推荐系统判定某位居民属于“中度抑郁风险”并推送危机干预信息时，驻点社工需要知道这一判断从何而来。权重给了深夜时段的 App 活跃度，还是语音情感分析捕捉到了通话中的低频颤音？答案无从知晓。算法决策若始终是打不开的黑箱，社工便只能被动接受机器给出的标签，丧失独立复核能力；居民同样无从理解自己为何“被诊断”。《互联网信息服务算法推荐管理规定》已要求平台以适当方式公示推荐服务的基本原理，落到社区心理援助场景，还需更具操作性的安排。其一，平台向驻点社工开放算法审计接口，允许其查看每条推荐决策背后的关键特征变量及贡献权重。其二，居民端的每条推送下方设“为什么推荐这条内容”的简明入口，以非技术语言说明触发逻辑。其三，运营方每年发布算法影响评估报告，披露模型在不同年龄、性别、户籍群体间的推荐准确率差异，接受外部学术团队与社区监督委员会的交叉审查。公开决策依据本身不削弱技术效能；恰恰相反，信任正是在这种可查证的状态下才有可能积累起来。

结语：

算法推荐进入社区心理援助场域，是数字技术向社会治理纵深推进的产物，同时也照出了技术乐观主义的盲区。本文沿技术嵌入的实际运作逻辑展开分析，揭示了情感茧房的负面强化、数字弱势群体遭到的结构性遮蔽、隐私信任链条的脆弱化这三重风险，并将其根源追溯至效率逻辑对关怀伦理的挤压，以及训练语料中社会偏见的数字化复刻。治理这些风险，没有单一的技术补丁可用，回应必须分层次展开。服务架构上，人的判断须优先于算法输出，这一条居于首位；工程设计需要为弱势群体预留非数字化的服务通道；数据管理则以全链条伦理规制封堵泄露与滥用的敞口；至于决策黑箱，最终要靠透明度方面的制度安排来打破。技术工具的用处，在于拓展人类行动的边界；对另一个生命的理解与尊重，却始终只能由人来完成。社区心理援助借算法之力走得越远，这一点越不该被遗忘：信任是在人与人面对面坐下、目光交汇的片刻之中，缓缓铺展开来的。

参考文献：

[1] 冯子轩 . 生成式人工智能应用的伦理立场与治理之道：以 ChatGPT 为例 [J]. 华东政法大学学报 ,2024,27(1):61-71.

[2] 张恩典 . 算法的法律治理研究 [M]. 中国政法大学出版社 :2024.

[3] 李静辉 . 算法推荐意识形态属性的生成逻辑、风险及治理 [J]. 理论导刊 ,2022,(2):70-76.

[4] 蔡丽丽 , 廖琴 , 岳晓红 . 社区心理服务建设路径研究 [J]. 黄冈师范学院学报 ,2025,45(4):50-54.

[5] 匡文波 . 智能算法推荐技术的逻辑理路、伦理问题及规制方略 [J]. 深圳大学学报 (人文社会科学版),2021,38(1):144-151.

[6] 林春婷 , 蒋柯 . 智慧社区与社区心理服务的智慧机制 [J]. 社区心理学研究 ,2020,9(1):3-16.

作者简介：张力文（2007 年 3 月 -）、女、汉族、山东省烟台市福山区、大学本科、研究方向：数字化赋能社区社会工作的机遇、伦理风险与对策。