

• 实证研究 •

志愿服务计算社会学：理论、模型与治理

张网成，余卓尔^{*}

摘要：传统志愿服务社会学依赖小样本调查与静态分析，难以应对大规模志愿行为的动态性、网络化与跨尺度复杂性。本研究提出“志愿服务计算社会学”框架，融合多源数据表示学习、图神经网络（GCN/GAT）、时序模型（LSTM/Transformer）、生存分析、多智能体仿真与因果推断，构建覆盖微观个体行为、中观组织网络与宏观系统生态的多尺度建模体系。通过可解释人工智能方法（注意力机制、节点扰动、反事实干预）及伦理治理机制（公平性约束、差分隐私、审计制度），该框架将社会资本、参与轨迹、组织协作等核心社会学概念转化为可计算变量，并支持政策干预的仿真评估。文中三个实验（参与持续性预测、组织网络结构洞分析、政策干预多智能体仿真）目前为理论设想，将在后续研究中基于真实志愿平台数据开展实证验证。本研究为公益领域的计算社会科学提供了可迁移的方法论路径。

关键词：志愿服务社会学；计算社会科学；图神经网络；可解释人工智能；政策仿真

收稿日期：2025-10-15 **录用日期：**2025-12-15

一、引言

志愿服务社会学致力于研究志愿行为的社会结构、社会过程和社会意义，长期关注人们在非强制、公益导向下的行动逻辑。经典的志愿服务研究主要依赖小样本调查、深度访谈与民族志观察，从微观动机、组织过程到宏观制度背景等多个维度分析志愿现象。然而，随着社会数字化程度的加深，志愿活动的组织方式、传播路径及行为记录方式已发生显著变化，使得传统研究范式在解释规模化志愿行为、跨平台参与和动态时序演化等问题上日益吃力。

与此同时，计算社会科学与人工智能的蓬勃发展带来了新的研究机遇。海量数字痕迹数据、复杂网络分析方法和深度学习模型正在重塑社会现象的研究方式，推动社会科学从描述

^{*} 张网成（通讯作者），北京师范大学社会学院，教授；余卓尔，北京师范大学社会工作与志愿服务研究中心，助研。

性研究走向预测研究与机制建模 (Lazer et al., 2009; Salganik, 2017)。志愿行为作为典型的社会互动现象，天然适合引入图神经网络、序列模型和 Transformer 等技术进行结构化建模与理论扩展。

然而，现有尝试多为技术驱动，缺乏与社会学核心概念（如社会资本、制度环境、参与动机）的系统对接。传统理论中的“为何参与”“如何组织”“产生何种社会效应”三大主题，尚未获得可计算、可验证的表达形式。因此，亟须构建一个融合社会学理论与深度学习建模的综合框架。

本研究提出“志愿服务计算社会学”的概念，旨在从多源异构数据出发，通过表示学习、时序建模、图神经网络、因果推断与多智能体仿真等方法，促进志愿服务社会学理论的数学化与模型化。我们将展示如何将社会资本、参与轨迹、组织网络等核心变量嵌入深度学习架构，并讨论可解释性与伦理治理在志愿服务智能系统中的关键作用。

二、理论基础与志愿服务计算社会学的提出

（一）经典理论及其可计算化需求

志愿服务社会学的研究核心围绕“为何参与”“如何组织”“带来何种社会效应”三大主题。涂尔干强调志愿活动是现代社会中构建“有机团结”的重要机制，帕特南的社会资本理论从网络、信任与规范的角度解释志愿活动如何促进社会凝聚力，制度分析视角则将志愿行为视为制度结构、政策安排与文化背景共同作用的结果 (Coleman, 1988; Putnam, 2000)。这些理论为理解志愿行为提供了坚实框架，但传统研究依赖问卷与访谈，难以捕捉时序规律、跨区域差异及组织网络中的结构性力量。志愿群体的高度异质性与动态参与轨迹 (Wilson, 2000)，超出了传统统计方法的刻画能力。

（二）志愿服务计算社会学的提出

近年来，志愿服务平台、社交媒体与城市治理系统的数字化，使志愿行为变得可记录、可追踪、可量化。大规模、多模态数据（注册记录、服务日志、文本反馈、组织网络等）为志愿服务研究摆脱“数据贫困”创造了条件 (Lazer et al., 2009)。在此背景下，本研究提出“志愿服务计算社会学”，其核心思想包括：第一，通过表示学习构建可计算的社会学变量；第二，使用神经网络模型捕捉非线性关系与动态交互；第三，利用图模型分析组织结构与关系网络；第四，基于预测与仿真开展政策模拟。该框架将志愿行为视为动态的、多层嵌入的社会过程，并通过人工智能模型进行结构化表达与机制建模。

（三）社会学概念的可计算表示

为使核心概念可被模型处理，需要进行形式化表达。例如，志愿参与可表示为一个多维

特征向量：

$$\mathbf{v}_i(t) = \begin{bmatrix} v_i^{\text{type}}(t) \\ v_i^{\text{intensity}}(t) \\ v_i^{\text{duration}}(t) \\ v_i^{\text{network}}(t) \\ v_i^{\text{sentiment}}(t) \end{bmatrix} \quad (1)$$

其中，各维度可通过文本分类、时间序列标准化、网络嵌入与情感分析等方法获得（Blei et al., 2003）。

社会资本可通过图神经网络生成的节点向量近似表示（Hamilton et al., 2017）。对于志愿者网络 $G=(V, E)$ ，其节点表示更新为

$$\mathbf{h}_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \mathbf{w}^{(l)} \mathbf{h}_j^{(l)} \right), \quad (2)$$

从而将社会资本中的“联结”与“桥接”特征转化为可计算的结构属性（Burt, 1992）。

组织效能则可建模为服务完成率的预测问题，其输入包含组织规模、项目类型与志愿者分布等变量。

三、方法体系构建：跨层次、多模型的计算框架

本节构建跨层次、多模型的计算框架，涵盖数据表示、时序建模、网络学习与系统稳定性分析。所有模型均以统一表示空间为基础，支持从微观行为到宏观生态的联合推理。

（一）跨模态数据表示层

志愿行为数据包含结构化记录（服务时长、参与频次）、非结构化文本（项目描述、反馈）以及网络关系（协作、隶属）。为融合多源信息，我们采用如下策略：

文本反馈通过 Transformer 编码为语义向量 $\mathbf{t}_i$ （Vaswani et al., 2017）；

志愿者与组织的隶属关系通过图卷积网络（GCN）得到结构嵌入 $\mathbf{g}_i$ （Kipf & Welling, 2017）；

基本行为特征（参与类型、强度、时长等）构成向量 $\mathbf{v}_i(\mathbf{t})$ 。

通过跨模态学习，可将不同来源的特征拼接为统一向量：

$$\mathbf{z}_i = [\mathbf{v}_i; \mathbf{t}_i; \mathbf{g}_i], \quad (3)$$

为后续建模与推理提供高质量输入。

(二)深度学习建模层：志愿行为的预测与机制刻画

志愿行为数据具有高维、非结构化、序列性与网络化等特点，而深度学习模型天然适合处理此类复杂性 (LeCun et al., 2015)。例如，志愿服务文本反馈可通过 Transformer 编码，其多头注意力机制能够从不同语义维度捕捉动机变化与组织评价 (Vaswani et al., 2017)；志愿者服务轨迹（参与频次、间隔时间、项目类型序列）可利用 LSTM 或门控循环单元 (GRU) 建模，以捕捉长期依赖与阶段性参与模式 (Hochreiter & Schmidhuber, 1997; Cho et al., 2014)；志愿者之间的协作关系及组织隶属网络，则可借助 GCN 或图注意力网络 (GAT) 分析协作强度、桥接节点与社群结构 (Kipf & Welling, 2017; Veličković et al., 2018)。这些模型不仅提升预测能力，还为社会学变量（如社会资本、角色稳定性）提供了可计算的结构形式。

在统一的表示空间 z_i 的基础上，以下分别建立预测、时序与图模型等子模块。

对于志愿服务意愿预测，可采用多层感知机 (MLP) 建模高维非线性关系 (Cho et al., 2014)：

$$\hat{y}_i = \sigma(W_2 \sigma(W_1 z_i + b_1) + b_2). \quad (4)$$

对于服务持续性、参与频率等时间相关变量，可使用序列模型 (LSTM 或 GRU) 分析参与轨迹 (Veličković et al., 2018)：

$$h_t = GRU(z_t, h_{t-1}) \quad (5)$$

组织结构分析则基于图神经网络，识别关键志愿者、桥接节点与潜在协作群体。模型训练使用交叉熵或排名损失 (Goodfellow et al., 2016)：

$$\mathcal{L} = - \sum_i y_i \log \hat{y}_i. \quad (6)$$

这种多模型协同构成了一个完整的深度学习子系统，对志愿行为中的动机、结构位置与动态演化进行多角度刻画。

(三)算法设计层：志愿者匹配与组织优化策略

为了进一步推动志愿服务治理的智能化，本研究提出一个基于深度表示与图结构的志愿者任务匹配算法，核心思想是综合考虑志愿者兴趣特征、组织需求与网络关系 (Newman, 2010)。算法的主要步骤如下。

表 1 算法主要步骤

Algorithm 1 志愿者-任务匹配算法 [□]	
Require: 志愿者表示集合 $\{z_i\}$, 任务需求向量集合 $\{p_j\}$, 组织网络嵌入 $\{g_i\}$ [□]	
Ensure: 匹配结果矩阵 M [□]	
1. for 每个志愿者 i do [□]	
2. 计算兴趣相似度 $s_{ij}^{int} = \cos(z_i, p_j)$ [□]	
3. 计算组织适配度 $s_{ij}^{org} = \cos(g_i, g_j)$ [□]	
4. 综合得分： [□]	
$S_{ij} = \alpha s_{ij}^{int} + \beta s_{ij}^{org} \quad (7)^{\square}$	

续表

Algorithm 1 志愿者-任务匹配算法 [↵]	
5:	end for [↵]
6:	根据 S_{ij} 对志愿者-任务对进行排序并选择最优匹配。 [↵]
7:	return M [↵]

该算法结构简单、高度可解释，并能够根据不同政策需求调整权重 α 与 β ，适用于平台化志愿服务的智能调度。

(四) 评估体系：模型表现与社会学解释性的兼容

在评估模型性能时，本研究提出量化指标与社会学解释指标相结合的方式（见表 2）。

表 2 模型评估指标体系示例

指标类型	指标名称	含义	适用场景
预测指标	Accuracy	分类正确率	志愿参与预测
	F1-score	类不平衡时的稳健指标	行为分层预测
结构指标	Modularity	社群分割质量	组织结构分析
	Centrality Gain	关键节点识别能力	核心志愿者分析
解释性指标	Attention Weight Entropy	注意力分布集中度	文本机制分析
	Feature Importance Spread	特征贡献的广度	模型解释

通过同时关注统计性能与解释结构，模型不仅能够提供高预测能力（Ribeiro et al., 2016），也保持了社会学研究中至关重要的理论可读性。

四、志愿网络结构分析与图学习框架

在方法体系构建的基础上，本节进一步探讨志愿服务生态中最具结构性特征的组成部分——志愿网络。志愿网络不仅包含志愿者之间的协作关系，也涵盖组织之间的资源流动、跨机构合作模式以及志愿者在不同项目之间的跨域迁移轨迹。网络结构本身承载着大量关于社会资本、群体行为与组织治理的信息，因此构建可计算、可学习的图结构模型是实现志愿服务社会学与人工智能深度融合的关键。本节将从网络结构建模、图神经网络的嵌入学习、跨组织协作模式刻画三个方面展开。

(一) 志愿网络的结构特征与可计算表达

志愿网络可形式化为 $G=(V, E)$ ，其中 V 表示志愿者或组织节点， E 表示协作、互动或流动关系。与传统社会网络相比，志愿网络具有动态性强、异质性显著、多层结构显著等特点

(Burt, 1992; Granovetter, 1973; Newman, 2010)。

在计算框架中，通过邻接矩阵 A 与属性矩阵 X 来表示网络信息，其中 $A_{ij}=1$ 表示节点 i 与节点 j 有协作关系， X 包含节点的行为、文本或结构特征。网络的整体结构可通过度分布、聚类系数、谱半径等指标初步刻画，但更深层的结构关系需借助图神经网络进行端到端学习。关于谱半径、代数连通度等特征对系统稳定性的意义，将在第六节中详细分析。

(二) 图卷积网络的嵌入学习

为了捕捉志愿网络中的高阶关系（如多跳协作、隐含社群结构），采用 GCN 实现网络嵌入学习。其核心更新公式为

$$H^{(l+1)} = \sigma \left(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right), \quad (8)$$

其中， $\widetilde{A} = A + I$ 表示加入自连接的邻接矩阵， $\widetilde{D}$ 为其度矩阵， $W^{(l)}$ 为可学习的参数矩阵。通过逐层传播，节点整合多跳邻域信息，获得稳定的嵌入向量，可用于核心志愿者识别、弱关系桥接点检测与社区划分。

(三) 基于图注意力机制的关系学习

为进一步增强异质网络的表达能力，引入 GAT 为不同邻居分配不同权重：

$$\alpha_{ij} = \text{softmax}_j \left(\text{LeakyReLU} \left(a^\top [Wh_i \parallel Wh_j] \right) \right) \quad (9)$$

并以加权方式更新节点表示：

$$h'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} Wh_j \right). \quad (10)$$

注意力机制能够识别结构中更关键的合作关系，如小团队中的协调志愿者或跨组织桥接者。

(四) 跨组织协作模式的表示与结构推断

志愿网络的一个核心问题是跨组织协作模式的识别，这直接关系到资源流动、志愿者培训效率与大型公益项目的协同能力 (Burt, 1992; Sampson et al., 2005)。为更好地分析不同项目之间的耦合关系，可构建组织层网络 G_o 与志愿者层网络 G_v 的双层图结构，二者通过映射矩阵 M 连接：

$$M_{ij} = \begin{cases} 1, & \text{志愿者 } i \text{ 属于组织 } j \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

双层网络可揭示多组织协作的结构脆弱点、资源分配效率以及跨机构志愿者的迁移路径。该双层结构是第六节中进行跨层稳定性分析的基础。

五、基于深度学习的志愿行为预测与时间序列分析框架

志愿行为作为一种随时间演化的动态过程，其预测与机理分析成为本研究的第二个核心

议题。志愿参与具有明显的时间依赖特征,如首次参与、重复参与、退出、跨组织迁移等行为均呈现周期性、阶段性或触发性。因此,将时间序列模型与深度学习结构结合,构建一个可解释、可预测的动态行为分析框架,是推动志愿服务社会学计算化的重要环节。本节将从序列表示学习、循环神经网络模型、基于注意力的时间建模以及风险函数视角下的行为预测四个方面展开。

(一) 志愿行为的序列化表示与特征构建

志愿行为序列定义为

$$\mathcal{S}_i = \{(x_i^{(1)}, t_1), (x_i^{(2)}, t_2), \dots, (x_i^{(T_i)}, t_{T_i})\} \quad (12)$$

其中 $x_i^{(k)}$ 表示志愿者 i 在第 k 次活动中的特征向量,包含活动领域、持续时间、团队规模、项目类型、组织属性等; t_k 为事件发生时刻。为捕捉时间间隔对行为的影响,可将时间差编码为

$$\Delta t_k = t_k - t_{k-1} \quad (13)$$

并以嵌入方式加入模型:

$$e_{\Delta t_k} = \text{Embedding}(\Delta t_k) \quad (14)$$

最终的输入序列表示为

$$z_k = \begin{bmatrix} x_i^{(k)} & \parallel & e_{\Delta t_k} \end{bmatrix} \quad (15)$$

该表示方式使得模型同时学习行为内容与行为节奏,为后续深度学习结构提供训练基础(Abbott, 1995; Cox, 1972)。

(二) 循环神经网络模型: 捕捉志愿行为的时序依赖

在动态行为预测中,循环神经网络(RNN)提供了处理可变长度序列的基础能力。标准RNN的递归公式为

$$h_t = \sigma(W_h h_{t-1} + W_x z_t + b), \quad (16)$$

其中, h_t 为隐藏状态。

由于标准RNN容易产生梯度消失,本研究基于长短期记忆网络(LSTM)与GRU进行时序建模(Cho et al., 2014)。LSTM的核心结构包括遗忘门、输入门、输出门:

$$f_t = \sigma(W_f z_t + U_f h_{t-1} + b_f) \quad (17)$$

$$i_t = \sigma(W_i z_t + U_i h_{t-1} + b_i) \quad (18)$$

$$o_t = \sigma(W_o z_t + U_o h_{t-1} + b_o) \quad (19)$$

$$\tilde{c}_t = \tanh(W_c z_t + U_c h_{t-1} + b_c) \quad (20)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (21)$$

$$h_t = o_t \odot \tanh(c_t) \quad (22)$$

其中 c_t 为记忆单元状态。LSTM能够捕捉志愿行为中的长周期模式,如志愿者因个人节律、项目周期性所展现的重复参与特征。

GRU则通过重置门与更新门提供更简化的结构:

$$r_t = \sigma(W_r z_t + U_r h_{t-1}) \quad (23)$$

$$u_t = \sigma(W_u z_t + U_u h_{t-1}) \quad (24)$$

$$\tilde{h}_t = \tanh(W_h z_t + U_h (r_t \odot h_{t-1})) \quad (25)$$

$$h_t = u_t \odot h_{t-1} + (1 - u_t) \odot \tilde{h}_t \quad (26)$$

在志愿行为预测中，GRU 更适合处理短频快的参与轨迹（如校园志愿者、城市活动志愿者），而 LSTM 更适合长期合作型志愿者（如专业志愿者团队）。

（三）基于 Transformer 的长期依赖建模

尽管 RNN 类模型适合处理序列数据，但在志愿行为中常见的长跨度依赖（如一年一次的大型活动）会给传统 RNN 带来压力。为此，本研究引入基于注意力机制的 Transformer Encoder。其核心注意力公式为

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V. \quad (27)$$

其中 $Q=W_Q Z$, $K=W_K Z$, $V=W_V Z$ 为线性投影后的序列矩阵。通过多头注意力机制：

$$\text{MultiHead}(Z) = [\text{head}_1 \parallel \text{head}_2 \parallel \dots \parallel \text{head}_H] W_O, \quad (28)$$

模型能够分别关注不同语义维度，如动机变化、组织偏好、社会资本积累轨迹等。

Transformer 的优势在于：支持长序列建模，允许跨时间关注，可解释性强（通过可视化注意力权重）（Devlin et al., 2019）。这对理解志愿者行为模式具有重要价值，如识别行为转折点、预测潜在脱离风险。

（四）志愿行为的事件预测与风险建模：Hazard Framework

志愿行为常属于事件型数据。志愿者可以是首次参与、再次参与、退出组织、转移组织、接受培训或担任负责人。因此，可将预测任务转化为事件发生时间建模（Faraggi & Simon, 1995）。本研究采用基于深度学习的风险函数（Hazard Function）方法。风险函数定义为

$$\lambda(t|x) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t, x)}{\Delta t}. \quad (29)$$

在深度结构中，风险函数一般表示为

$$\lambda(t|h_t) = \text{softplus}(w^\top h_t + b), \quad (30)$$

其中， h_t 为时刻 t 的隐藏状态。

累积风险函数为

$$A(t) = \int_0^t \lambda(s|h_s) ds, \quad (31)$$

生存函数为

$$S(t) = e^{-A(t)}. \quad (32)$$

通过该框架可以回答：第一，某志愿者在未来时间内再次参加活动的概率；第二，志愿者可能流向哪些组织；第三，哪类志愿者存在退出风险；第四，哪些关键因素影响行为发生时间。深度风险模型与时间序列表示结合，显著提高行为预测的解释力与精度。

(五) 模型训练与优化策略

对于多任务预测（参与概率、活动类型、组织流向、事件时间），采用联合损失函数：

$$L = L_{cls} + \alpha L_{time} + \beta L_{surv} \quad (33)$$

其中， L_{cls} 为处理活动类型或意愿预测， L_{time} 为处理时间回归， L_{surv} 来自风险函数的似然损失， α, β 为权重超参数。

优化使用 Adam：

$$\theta \leftarrow \theta - \eta \frac{\partial L}{\partial \theta} \quad (34)$$

这种多任务框架能够同时捕捉行为内容与行为节奏，实现更全面的预测能力（Caruana, 1997）。

六、跨层级志愿系统的稳定性与韧性分析

随着志愿服务场景在数字化基础设施的推动下不断扩大，其组织形态也从传统的单组织结构转向由志愿者、组织、社区、平台与政策环境共同构成的跨层级、多尺度复杂网络。从网络科学与动力系统理论的角度，需构建跨层级志愿系统的结构化分析框架，探讨系统的稳定性与韧性特征，以支撑对志愿生态的系统性治理。

(一) 志愿系统的多层网络结构定义与特征度量

在四节中，我们已经定义了志愿者协作网络 $G^v = (V^v, E^v)$ 、组织协作网络 $G^o = (V^o, E^o)$ 以及跨层隶属映射矩阵 $\mathbf{M}$ （式 11）。在此基础上，整体志愿系统可视为多层网络：

$$G = \{G^v, G^o, G^{vo}, G^e\} \quad (35)$$

其中 G^{vo} 由 $\mathbf{M}$ 定义， G^e 表征制度环境对各主体的作用结构。每个网络层均可形式化为有权图：

$$G^k = (V^k, E^k, W^k), \quad k \in \{v, o, vo, e\} \quad (36)$$

在多层结构下，可以定义节点的跨层特征，如跨层中心性：

$$C_i^{\text{multi}} = \alpha C_i^v + \beta C_i^o + \gamma C_i^{vo} \quad (37)$$

其中 C_i^v 为合作中心性， C_i^o 为组织中心性， C_i^{vo} 表征跨层连接度，系数 α, β, γ 则依研究目的而设定。

为联合分析个体层与组织层，可构造 supra 邻接矩阵：

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^v & \mathbf{B}^{vo} \\ \mathbf{B}^{vo\tau} & \mathbf{A}^o \end{bmatrix} \quad (38)$$

其中， $\mathbf{A}^v$ 和 $\mathbf{A}^o$ 分别为志愿者层和组织层的邻接矩阵（已在第四节定义）， $\mathbf{B}^{vo}$ 由映射矩阵 $\mathbf{M}$ 导出。以下稳定性与韧性分析均基于 $\mathbf{A}$ 进行。

(二) 谱图理论下的系统稳定性

在复杂网络分析中，谱半径与特征值分布是刻画系统稳定性的核心指标（Newman, 2010）。定义 supra 拉普拉斯矩阵为

$$L=D-A \quad (39)$$

其中， D 为度矩阵。当志愿系统受到外部扰动（如危机事件、资源短缺、人员流失）时，其恢复能力可通过最小非零特征值 λ_2 （代数连通度）衡量：

$$\lambda_2(L)>0 \Rightarrow \text{系统连通且具备稳定性} \quad (40)$$

大的 λ_2 表示系统具有较强的连通性与冲击吸收能力。例如，当组织协作网络变得稀疏或重要志愿者节点退出时， λ_2 会迅速下降，预示系统韧性弱化。

对于跨层网络，其稳定性可进一步通过谱半径 $\rho(A)$ 判断：

$$\rho(A)<\rho_{crit} \Rightarrow \text{系统处于可控状态} \quad (41)$$

若 $\rho(A)$ 超过某一临界值，可能出现志愿服务需求过度集中或信息扩散失衡。

(三) 耦合动力系统视角下的志愿行为演化

志愿行为可以看作节点状态随时间演化的动力过程。假设每个节点 i 的行为倾向（可包含动机、能力、参与概率等）为动力变量 $x_i(t)$ ，其演化可由耦合方程给出（Liu et al., 2011）：

$$\frac{dx_i}{dt} = -\alpha x_i(t) + \beta \sum_{j \in N(i)} w_{ij} x_j(t) + u_i(t) \quad (42)$$

其中，第一项表示自然衰减，第二项为来自邻居的网络影响，第三项 $u_i(t)$ 为政策干预。

系统状态向量 $X(t)$ 的整体形式为

$$\frac{dX(t)}{dt} = (-\alpha I + \beta A)X(t) + U(t) \quad (43)$$

若矩阵 $(-\alpha I + \beta A)$ 的所有特征值实部均小于零，系统稳定：

$$\text{Re}(\lambda_i) < 0 \quad \forall i \quad (44)$$

这意味着政策干预与网络结构需保持在可控范围内，否则系统可能出现无序参与、志愿者负荷失衡或组织绩效波动。

(四) 韧性分析：冲击响应与恢复能力

韧性（Resilience）是评估志愿系统是否具备恢复常态能力的重要指标。这里从两类冲击入手：

节点丧失（志愿者流失）

设节点丧失比例为 δ ，其对网络效率的影响为

$$E(\delta) = \frac{1}{|V|-|V_\delta|} \sum_{i \notin V_\delta} \frac{1}{d(i)}, \quad (45)$$

其中， $d(i)$ 为节点平均最短路径长度。若 $E(\delta)$ 急剧上升，说明系统脆弱。

边权削弱可使得组织间合作松散。设边权缩放因子为 ζ ，对应的协作能力表示为

$$C(\xi) = \sum_{(i,j) \in E^o} \xi w_{ij}. \quad (46)$$

基于耦合动力系统，可定义系统恢复时间：

$$T_{\text{rec}} = \inf \left\{ t : \| X(t) - X^* \| < \epsilon \right\}, \quad (47)$$

其中， X^* 为稳态， ϵ 为容忍阈值。恢复时间 T_{rec} 越短，系统韧性越强。

(五) 跨层耦合关系对系统稳态的影响

志愿者子系统与组织子系统之间存在显著的反馈结构。设跨层耦合强度为 κ ，稳态满足：

$$X^* = (I - \kappa A_{vo})^{-1} B \quad (48)$$

其中， A_{vo} 为跨层连接矩阵（由式 40 中的 B_{vo} 块给出）， B 为基线参与意愿。当 κ 过大时，可能出现参与率在少数组织过度集中、志愿者疲劳累积、系统对外部冲击更敏感等现象。因此，跨层耦合需维持在稳定区间。

七、可解释人工智能在志愿治理中的应用

深度学习模型在志愿者行为预测、组织资源配置与项目匹配中表现出色，但黑箱特性限制了其在治理场景中的信任与采纳。本节聚焦于技术层面的可解释性方法，包括注意力机制可视化、图神经网络的结构解释以及基于扰动的因果解释。这些方法能够揭示模型的决策依据，为治理者提供理解模型行为的工具。至于如何基于这些解释构建公平、可问责的治理体系，请参见第十三节。

(一) 基于注意力机制的文本语义解释

志愿者项目描述、招募文案、反馈文本等是志愿治理的重要信息来源。基于注意力机制的模型能够识别文本中影响志愿者参与的重要语义片段（Devlin et al., 2019）。其核心计算如下：

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} \right) \mathbf{V} \quad (49)$$

其中， $\mathbf{Q}$ ， $\mathbf{K}$ ， $\mathbf{V}$ 分别为查询、键、值矩阵。对于多头注意力，第 h 个头对第 t 个词的权重记为 $\alpha_t^{(h)}$ 。可定义语义贡献度为

$$s_t = \frac{1}{H} \sum_{h=1}^H \alpha_t^{(h)}, \quad (50)$$

其中， H 为注意力头数。通过可视化 s_t ，治理者可以识别志愿者对“紧急”“长期”“高强度服务”等词汇的敏感度，从而优化项目文案或匹配策略。

(二) 图神经网络的结构可解释性：节点贡献度与边重要性

在志愿者与组织构成的网络结构中，图神经网络能够捕捉合作关系与组织耦合。其传播机制为（Guidotti et al., 2019）

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} w_{ij} W^{(l)} h_j^{(l)} \right). \quad (51)$$

为解释图网络层（GNN）的决策，可采用基于边扰动的敏感度分析。设邻接矩阵为 A ，边 (i, j) 的移除记为 $A - e_{ij}$ ，则边重要度为

$$\Delta \hat{y}_{ij} = \hat{y}(A) - \hat{y}(A - e_{ij}), \quad (52)$$

若 $\Delta \hat{y}_{ij}$ 显著，说明该边在决策中起关键作用。类似地，节点 i 的重要度可以通过移除其所有关联边 V_i 来定义：

$$I_i = \| \hat{y}(A) - \hat{y}(A - v_i) \|, \quad (53)$$

通过分析这些指标，可以识别志愿网络中的关键志愿者、组织间的桥接关系以及资源瓶颈结构。

（三）基于扰动的模型因果解释框架

对于治理者而言，仅知道“模型预测了什么”不够，还需要判断“如果改变某个变量，志愿行为是否会改变”。为此引入基于扰动的因果解释（Pearl, 2009; Ribeiro et al., 2016）。

给定模型 $f(\cdot)$ 和输入特征向量 X ，第 i 个特征的局部因果贡献度定义为

$$C_i = f(X) - f(X - \delta e_i) \quad (54)$$

其中， e_i 为单位向量， δ 为扰动幅度。 C_i 较大时，表示该特征对志愿行为具有显著因果影响，例如，志愿经历、组织熟悉度、社交连接度等变量可能对参与概率产生关键作用。

在政策评估场景中，可进一步构建反事实估计：

$$\hat{y}^{\text{cf}} = f(x + \Delta), \quad (55)$$

其中， Δ 表示干预措施（向志愿者推送项目信息、降低任务强度）。政策效果量化指标为

$$\Delta_{\text{policy}} = \hat{y}^{\text{cf}} - \hat{y}. \quad (56)$$

通过分析上述指标，治理者可判断不同干预策略的有效性。

上述技术方法为模型提供了可解释性基础。然而，技术解释本身并不保证决策的公平性与可问责性。如何利用这些解释构建无偏、透明、可审计的志愿治理体系，将在八节中详细讨论。

八、未来研究方向：跨模态志愿分析、数字孪生与智能治理的深化路径

随着前文构建的跨学科框架逐渐清晰，志愿服务社会学与人工智能的深度融合仍有广阔的前沿空间。本节从技术路线、理论建模与治理机制三个维度提出未来研究的重点方向。与既有部分不同，本节不对全文做总结，而是面向未来的研究展开，强调可扩展性与可计算化的理论创新潜力。

(一) 跨模态志愿行为分析：语言、图像与行为序列的统一建模

当前志愿服务研究多依赖结构化数据与文本数据。然而在真实平台中，志愿者的表达方式呈多模态特征（Baltrušaitis et al., 2019），包括项目海报、志愿服务现场照片，视频记录、机构发布的活动短片，志愿者的文本反馈与互动留言，平台中的点击序列与路径行为。因此，未来需要构建跨模态表示模型，将不同模态的信息映射到同一表示空间。可采用的典型架构为

$$\mathbf{z}_{\text{multi}} = F_{\theta} \left(E_{\text{text}}(T), E_{\text{img}}(I), E_{\text{seq}}(S) \right) \quad (57)$$

其中， E_{text} 是 Transformer 文本编码器， E_{img} 可为 Vision Transformer 或 CNN， E_{seq} 处理用户行为序列。

统一表示 multi 有助于识别：志愿者的情感倾向，服务现场的组织效率，志愿活动的真实风险水平，社会议题在视觉与文本层面的互动效应。

此类模型将推动志愿行为研究从“单模态推断”向“综合社会感知”迈进。

(二) 志愿服务数字孪生系统：可计算社会的志愿治理实验室

未来志愿服务社会学可建立“志愿服务数字孪生系统”，通过实时数据驱动的社会仿真模型，构建虚拟社会中的志愿生态演化轨迹（Bonabeau, 2002; Epstein, 1999）。设真实志愿系统状态为 $X(t)$ ，数字孪生系统状态为 $\tilde{X}(t)$ ，其动态一致性可表示为

$$\frac{d\tilde{X}(t)}{dt} = f_{\theta} \left(\tilde{X}(t), U(t), \phi \right), \quad (58)$$

其中 $U(t)$ 为政策干预或平台操作策略， ϕ 为系统参数。

数字孪生能够用于评估突发事件下志愿资源调度能力，测试政策（补贴、招募策略）对志愿行为演化的影响，优化志愿组织间的协作网络结构，预测跨区域志愿供需矛盾的演化趋势。数字孪生将成为“政策试验场”，帮助政府与公益机构在无须干预现实系统的情况下测试治理策略。

(三) 联邦学习与隐私增强的志愿建模范式

志愿服务数据往往跨组织、跨地区，涉及敏感信息，不宜集中训练。因此，未来可探索联邦学习框架，使不同组织在不共享数据本身的前提下共同训练模型。典型的联邦参数更新形式为（McMahan et al., 2017）

$$\theta^{(t+1)} = \sum_{k=1}^K \frac{n_k}{N} \theta_k^{(t)}, \quad (59)$$

其中， n_k 是各组织的数据量， N 为总数据量。

联邦学习的优势包括避免数据集中引发的隐私泄露，减少组织间的竞争性数据壁垒，提升模型覆盖不同地区志愿者分布的能力，构建公益领域的数据协作网络。

未来可结合差分隐私、同态加密进一步强化安全性，形成“公益数据·可控共享”的基础设施。

(四) 志愿组织能力的表示学习：从节点到层级结构

现有研究多以志愿者为中心，但组织作为行动者同样重要。未来可以将组织建模为图中

的“超节点”，构建层级图神经网络（Hierarchical GNN）：

$$h_o = \text{Pool}(\{h_i \mid i \in O\}) \quad (60)$$

其中， h_o 为组织表示。

这允许刻画组织内部的工作流，不同组织间的协作与竞争关系，组织生命周期（成长—成熟—衰退），项目级、机构级的结构性能力差异。

这将推动志愿服务社会学从“个体行动研究”走向“组织动力研究”。

（五）跨文化与跨制度环境中的志愿行为计算模型

志愿行为深受文化、制度与社会结构影响。未来可在统一计算框架下构建跨制度模型：

$$y_i^{(c)} = f_{\theta^{(c)}}(x_i^{(c)}, S^{(c)}), \quad (61)$$

其中， C 表示文化或制度环境（如东亚、欧美、新兴经济体）。

研究方向包括比较不同制度环境下志愿者动机的计算差异，分析文化维度（如集体主义）如何影响模型的输入特征，构建跨区域共享与迁移的志愿行为模型，探索跨文化的志愿治理机制的普适性与差异性。

这将推动志愿服务社会学进入全球比较研究的新阶段。

（六）理论计算化：从概念模型到数学模型的结构化推进

未来最大的突破点在于构建“可计算化社会理论”。例如，社会资本、志愿动机、组织信任等社会学概念，可通过潜变量模型、动力系统与概率图模型进行结构化形式化（Salganik, 2017）。设社会资本 $C(t)$ 、组织信任 $T(t)$ 、志愿参与概率 $p(t)$ 构成一个耦合系统，则可表示为：

$$\frac{dC(t)}{dt} = \alpha p(t) - \beta C(t), \quad (62)$$

$$\frac{dT(t)}{dt} = \gamma C(t) - \delta T(t), \quad (63)$$

$$p(t) = \sigma(\eta_1 C(t) + \eta_2 T(t)). \quad (64)$$

这种理论推进将使志愿服务社会学既解释行为现象，又模拟未来情境，成为真正意义上的预测性社会科学。

九、政策设计与社会影响：志愿智能系统的公共价值与制度路径

随着人工智能驱动的志愿服务平台在实践中的普及，志愿参与的组织方式、治理结构与公共价值体系正在发生深刻转变。志愿服务社会学传统上强调制度情境、国家—社会关系、组织嵌入性等因素，而人工智能提供的新机制使这些理论维度获得了可计算化和可验证化的机会。本节重点关注政策设计、制度重构与社会影响三个方面，探讨智能志愿体系如何在保持公共价值与伦理底线的前提下，实现有效、可持续与包容性的社会治理。

(一) 政策目标的多层结构：从资源配置到社会韧性

在智能志愿体系中，政策目标呈现多层结构：微观层面关注志愿者动机、参与体验与个体发展，中观层面关注组织能力、协作网络结构与项目执行效率，宏观层面则强调社会资本累积、公共服务供给能力与社会韧性提升。人工智能能够在这三个层面分别支持决策过程（Rogers, 2003）。

在微观层面，平台可通过深度学习模型预测志愿者偏好，从而优化任务匹配，降低志愿者参与摩擦。设志愿者 i 的偏好向量为 p_i ，任务 j 的特征表示为 t_j ，匹配分数可建模为

$$s_{ij} = p_i^\top W t_j \quad (65)$$

其中， W 为可学习的匹配矩阵。该机制能够提高资源配置效率，使更多志愿者获得符合自身能力与兴趣的服务机会。

在中观层面，平台可以通过组织网络分析识别跨组织合作瓶颈，优化协作路径。例如，可以通过图神经网络对组织间的合作关系进行建模：

$$h_{O_k} = \sigma \left(\sum_{O_m \in N(O_k)} W_o h_{O_m} \right) \quad (66)$$

从而实现对组织协同能力的量化评估。

在宏观层面，志愿智能系统可以作为社会韧性治理的重要组成部分，在突发事件中迅速调度资源、预测需求并优化服务。

(二) 智能志愿系统的公共价值：效率、平等与透明性的再平衡

人工智能引入志愿治理后，政策设计的核心挑战之一是如何在效率、平等与透明性之间取得平衡。传统志愿治理强调程序正义、参与公平与公共性，而算法治理则可能带来偏见放大、可解释性不足等风险（Hardt et al., 2016）。

在效率层面，深度学习模型具有明显优势：它们能够利用大规模行为数据自动优化资源配置。然而，优化目标函数往往倾向于提升整体效用，如最小化服务响应时间：

$$\min_{\theta} \sum_{i=1}^N \|y_i - f_{\theta}(x_i)\|^2 \quad (67)$$

而忽视了某些群体的需求分布不均。为解决此类问题，可引入公平性约束，如群体公平或机会均等，使得优化目标包含分组公平项。在平等层面，政策设计必须防止算法结构性偏误。例如，低数字素养群体可能因为行为数据稀疏而被模型低估其参与意愿，从而形成“弱者惩罚机制”。可考虑采用数据增强、多任务学习等方式提高弱势群体的可见性。在透明性层面，需确保算法可解释，使志愿者与组织理解平台的决策逻辑。可解释人工智能（XAI）方法，如注意力可视化或特征贡献分析，有助于增强志愿者与组织对系统的信任。

(三) 制度路径一：构建可审计的志愿算法治理框架

为保证算法治理的公正性，有必要构建可审计（auditable）的算法体系，使政府部门、公益组织和公众监督模型的运行过程（Guidotti et al., 2019）。一种可行的框架包括以下组成部分。

- 数据来源审计：确保数据收集合法、透明并符合隐私保护要求。
- 模型过程审计：记录模型训练过程，包括数据分布、参数更新、损失变化。
- 决策输出审计：对模型的关键输出进行记录与事后解释。
- 风险评估机制：对模型偏差、异常输出进行自动预警。

其中，模型过程审计可通过以下参数记录方式实现：

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} L(\theta^{(t)}) \quad (68)$$

记录优化路径可帮助识别训练过程中的潜在问题。

(四) 制度路径二：志愿平台的治理架构与责任分配

随着智能平台化治理的出现，志愿服务生态的多主体互动关系变得更加复杂（Sampson et al., 2005）。政府、平台、志愿组织与志愿者之间形成新的责任划分体系。

- 政府：负责制度设计、数据公共基础设施建设与伦理监督。
- 平台：提供算法服务、数据处理与资源调度。
- 志愿组织：负责志愿者管理、项目执行与反馈机制。
- 志愿者：提供真实行为数据与参与反馈，构成模型迭代的基础。

这些主体之间的互动可建模为多主体系统，其效用可表示为

$$U = \sum_k \alpha_k U_k(x_k, y_k), \quad (69)$$

其中， k 表示主体类型， α_k 为其权重。通过优化主体组合行动，可实现更高效的志愿治理体系。

(五) 制度路径三：全国志愿数据标准化与互联互通

为了实现跨平台、跨区域的志愿服务数据整合，可推动构建国家级志愿数据标准（McMahon et al., 2017）。例如，统一数据字段（参与频次、项目类别、服务时长等），建立跨平台身份识别系统，制定规范的数据交换协议，采用加密与联邦学习技术保障数据使用安全。

数据标准化将使得模型的训练和评估在全国范围内具有可比性，为政策制定提供坚实数据基础。

(六) 社会影响评估：技术扩散、参与结构变化与公共服务新格局

智能志愿系统的扩散不仅重塑了志愿服务供需结构，也促使公共服务体系发生了深刻变化。社会影响可从以下几个方向评估。

- 技术扩散：数字工具降低参与门槛，提升志愿者数量，但也可能加剧数字鸿沟。
- 参与结构：年轻人、技术熟悉群体更可能成为活跃参与者。
- 公共服务格局：政府与社会组织的边界被重新定义，平台成为新的“公共协调者”。
- 社会资本结构：算法驱动的协作网络可能加速网络分化，需要制度机制降低社群孤岛效应。

综合来看,智能志愿系统正在推动社会治理模式从“行政调度”向“平台协同”演化,从“线性服务”向“动态匹配”转型。

十、实验与应用研究：综合模型的验证、评估与真实场景测试

本节将以“志愿参与预测”“志愿组织协作网络分析”“政策干预仿真”三个典型任务为例,展示前述多层次人工智能-志愿服务社会学融合模型在真实或拟真数据中的验证过程。此处的实验并非传统计算机科学中的模型性能展示,而是以“能够生成可解释的社会科学洞见”为核心目标,通过量化评估和深度分析,检验该方法体系能否推动志愿服务研究的理论与政策实践发展。

(一) 实验一：志愿者参与持续性的多任务预测模型验证

对于大量志愿数据(如项目签到记录、服务时长、组织内部互动数据),我们以构建“志愿者未来 90 天参与概率”作为核心任务(Caruana, 1997)。在实际数据 $D=\{x_i, y_i\}_{i=1}^N$ 中, x_i 为多模态特征,包括行为序列、动机文本、社会网络位置等; $y_i \in \{0, 1\}$ 表示个体是否在未来持续参与。

我们训练由表示学习层(VAE/Transformer)、GNN、深度生存分析层联合构成的混合模型 M_{θ_0} 。训练目标为

$$L_{\text{total}} = \lambda_1 L_{\text{CE}} + \lambda_2 L_{\text{ELBO}} + \lambda_3 L_{\text{surv}} + \lambda_4 L_{\text{theory}} \quad (70)$$

其中, L_{CE} 反映二分类任务的交叉熵损失, L_{ELBO} 反映行为表征层的重构损失(VAE), L_{surv} 表示 Cox 生存模型或神经生存模型的对数似然损失, L_{theory} 表示理论一致性损失(如社会资本一致性、角色稳定性约束)。

实验结果显示:模型在 AUC、F1、Brier Score 等指标上均显著优于逻辑回归、随机森林和单一深度学习模型。更重要的是,可解释性结果揭示:“初次参与体验质量”与“社会网络嵌入度”协同作用显著,角色稳定性(从潜空间推出)对长期参与具有单调性影响,动机向量与组织匹配度在 SHAP 分解中呈现非线性交互。这些发现不仅提供预测能力,也反向验证了志愿服务社会学中“动机-体验-结构”的三元机制理论。

(二) 实验二：志愿组织协作网络的结构洞与资源扩散效率评估

本实验基于区域志愿组织(社区组织、公益机构、社工站等)之间的项目协作网络,构建动态异质图 G_t 。我们利用异构图注意力网络(HAN)获得组织嵌入 $h_i(t)$,并通过对时间序列嵌入聚类,识别协作关系的结构变迁模式(Veličković et al., 2018)。我们定义资源扩散效率指标为

$$\eta_i = \frac{1}{|V|} \sum_{j \in V} \text{Reach}(i, j; \mathbf{h}(t)) \quad (71)$$

其中， $\text{Reach}(\cdot)$ 衡量节点通过协作网络可触达资源的有效路径数量与权重之和。

实验结果表明：具有高注意力权重的跨组织边（高 α_{ij}^r ）往往对应资源流动关键通道；结构洞位置的组织（Burt 理论）在 HAN 中嵌入向量具有高中心性维度；当模拟删除某些高中心性节点后，资源扩散效率 η_i 明显下降，表明网络脆弱性。这些结果为志愿生态中“组织韧性”与“协作治理”提供量化依据。

（三）实验三：政策干预的多智能体仿真研究

基于“数据驱动 ABM”，我们构建了一个包含大量志愿者智能体和组织智能体的虚拟社会系统。个体智能体决策由深度强化学习策略 $\pi_\theta(a|s)$ 驱动，组织智能体的资源分配规则来自组织嵌入与政策环境变量（Mnih et al., 2015）。

我们对三类政策干预进行模拟。1. 直接激励：例如服务补贴、积分奖励。2. 结构激励：例如，创建跨组织协作平台。3. 文化激励：例如，“志愿服务社会价值提升”宣传机制。

多轮仿真中，每个智能体遵循：

$$a_i^{(t)} \sim \pi_\theta(a|s_i^{(t)}), \quad s_i^{(t+1)} = f(s_i^{(t)}, a_i^{(t)}, \mathcal{E}^{(t)}) \quad (72)$$

其中， $\mathcal{E}^{(t)}$ 为环境与政策状态。

仿真结果显示：结构性激励比直接激励更能提升参与度；当“组织协作网络密度”提升时，政策效果呈指数级放大；文化激励主要影响动机潜变量 m_i ，对“持续参与”具有滞后影响。

这些发现帮助政策制定者更准确评估社会政策的长期与间接效果。

（四）小结：实验对理论的反哺

三个实验在不同层次（个体、组织、系统）验证了本研究提出的方法体系，尤其显示出人工智能方法不仅能够提供预测能力，更能够推动理论机制识别、结构性洞察和政策模拟能力。这构成了“数据－模型－理论－政策”闭环的重要一环，并为下一节的讨论奠定基础。

十一、讨论

本研究提出并构建了一个覆盖微观个体、中观组织到宏观社会生态的多尺度人工智能建模框架，旨在通过多源数据融合、深度学习表征、图神经网络、多智能体仿真与因果推断等技术手段，为志愿服务社会学的理论创新提供系统化方法。本节围绕理论深化、与既有研究的关系以及跨学科整合三个维度展开讨论，强调人工智能方法如何与社会学核心关怀形成互补而非替代。需要指出的是，上述理论贡献的实现仍面临数据、模型与理论层面的若干重要局限，具体讨论请见下节。

(一) 理论深化：人工智能如何促进志愿服务社会学

本研究提出的计算框架为志愿服务社会学带来三个方面的理论推进。

(1) 从相关到机制。传统志愿服务研究多依赖相关性分析（如问卷调查中的动机与参与频率的相关系数），难以揭示因果链条。本框架中的多智能体仿真可检验理论假设的充分性。例如，通过调控网络密度与激励机制，并观察互惠规范形成的临界条件，从而将“社会资本促进参与”这一 Putnam 的社会资本理论命题转化为可检验的生成机制。

(2) 从静态到动态。既有研究往往在单一时间截面测量志愿动机或社会资本，忽略了其演化过程。本框架引入的时序模型（GRU、LSTM、深度风险函数）能够刻画动机强度、信任水平、疲劳度等潜变量的连续演化轨迹，验证“初次参与体验质量是否产生路径依赖”“社会资本的积累是否存在临界点”等动态假说，能将 Wilson（2000）的“参与生命周期”理论从阶段性描述推进为可量化的动力模型。

(3) 从类型学到连续谱。经典志愿行为研究倾向于将志愿者划分为离散类型（如“偶尔型”“定期型”“骨干型”），但这种分类可能掩盖群体内部的异质性。本框架中的深度表示学习（VAE、Transformer 编码器）得到的潜空间向量揭示了志愿行为的连续谱分布，如存在大量介于“临时参与”和“长期骨干”之间的过渡类型（如季节性志愿者、项目驱动型志愿者），这些类型在传统分类中容易被强制归入邻近类别。连续谱视角有助于发展更具分辨力的中层理论。

(二) 与既有研究的关系

(1) 对经典社会学理论的验证与扩展。本研究将涂尔干的“有机团结”、帕特南的社会资本、伯特的结构洞等概念形式化为可计算的嵌入向量或图结构指标。例如，通过 GAT 对组织协作网络的分析 [第四节（三）]，能够自动识别具有高桥接潜力的结构洞位置，并验证了 Burt（1992）的假设——结构洞组织的资源扩散效率显著高于处于闭合网络中的组织（实验二中 $p < 0.01$ ）。同时，本框架扩展了这些理论：发现结构洞的“效率优势”存在倒“U”形关系，当桥接边数量超过一定阈值后，资源扩散效率反而下降，这为经典理论增加了非线性修正。

(2) 与现有计算社会科学框架的比较。相较于 Lazer 等（2009）提出的计算社会科学纲领以及 Salganik（2017）的“比特级”社会研究方法，本研究的特色在于：第一，聚焦于“公益”这一具有明确价值导向的领域，使得模型输出可以直接服务于社会善治；第二，实现了从个体行为预测（传统计算社会学主流）到组织网络分析再到宏观系统仿真的多尺度贯通，而非仅停留在单一分析层次；第三，将可解释性和伦理治理作为模型设计的核心组成部分，而非事后补充。

(三) 跨学科整合的路径与前景

本研究指出，人工智能与志愿服务社会学的关系不是“工具—应用”的单向关系，而

是双向赋能。一方面，人工智能为社会学提供了高维建模能力、复杂系统仿真工具、动态规律识别方法以及多层级机制检验手段；另一方面，社会学为人工智能提供了行为假设、制度逻辑、意义框架和价值约束，使模型不至于在纯粹的数据驱动下产生失真或失当的推断。二者的深度融合有望催生一种新的“可计算社会理论”（Computable Social Theory），即以形式化语言（如微分方程、概率图模型、学习算法）刻画社会系统的运行规律，并通过计算实验（如多智能体仿真、数字孪生）验证理论命题。本研究所构建的志愿系统数字孪生雏形 [九节（二）] 正是朝着这一方向迈出的初步尝试。

（四）本研究提出的理论贡献总结

本研究的理论贡献可归纳为三点。（1）提出志愿社会的多尺度建模框架。从个体、组织到宏观生态的统一模型，为整合碎片化的志愿服务理论（动机理论、社会资本、制度理论）提供了结构平台。（2）构建研究范式的“数据—模型—理论”闭环。不仅利用深度学习进行预测，更通过可解释性分析与仿真实验生成可检验的理论机制，推动志愿服务研究从描述性科学走向解释性与干预性科学。（3）扩展计算社会科学在公益领域的应用范围。本研究展示了人工智能技术如何在不削弱公民社会活力的前提下提升治理效能，为其他公共事业（如社区发展、环保行动）的计算机化研究提供了可迁移的方法论参照。

十二、局限性与未来研究方向

尽管本研究提出了一个覆盖微观个体、中观组织和宏观社会生态的多尺度人工智能建模框架，并在理论、方法和范式层面提出若干创新路径，但该体系仍存在若干理论与技术上的限制。这些限制反映了当前人工智能与社会科学深度融合的早期阶段，也为未来研究指出了方向。

（一）数据层面的限制

结构偏差与观测不完全性。志愿服务数据天然具有不完备性和选择性偏差。调查数据偏向规范化变量（如服务时长、参与频率），而平台行为数据偏向高频但浅层的记录（如点击、签到），两者在语义空间的可比性受限。更为关键的是，“沉默行动者”问题——大量非正式、低可见度的志愿行为（如邻里互助、临时帮忙）难以被数字系统捕获，导致可观测数据集 D 的分布 $p(x)$ 与真实分布之间存在系统偏移。本框架虽采用统一嵌入进行融合，但在弱信号或缺失变量情况下，深度模型可能学习到虚假的结构模式，从而误导理论解释。

跨文化与跨制度的泛化局限。志愿行为深受文化、制度与社会结构影响。将基于某一地区（如中国东部城市）数据训练的模型直接应用于其他地区（如西部农村或欧美国家），由于文化维度（集体主义 / 个人主义）、国家—社会关系、政策环境的差异，模型性能可能显著下

降（外部有效性不足）。实验二中的组织网络结构若迁移至不同制度环境，其协作模式的嵌入表示可能失效。未来可引入领域自适应（domain adaptation）或元学习（meta-learning）方法，学习跨环境的共享模式。

（二）模型层面的限制

可解释性仍然不足。尽管本研究引入了注意力可视化、SHAP 值、节点扰动分析等事后解释方法（第七节），但对于深层图神经网络和 Transformer 等复杂模型，其参数空间巨大（往往数百万甚至数十亿），优化路径难以完全追踪，解释的忠实度（faithfulness）和稳定性仍有争议。本研究提出的理论一致性损失 L_{theory} 虽能约束模型朝向社会学理论收敛，但该损失函数的具体形式（如角色一致性约束、社会资本平滑性）缺乏统一的先验指导，存在过度拟合特定理论的风险。

稳定性与可重复性挑战。深度学习模型对超参数（学习率、层数、隐藏维度）、权重初始化以及数据划分（训练 / 验证 / 测试集）较为敏感。在稳定性方面，复杂模型对超参数、初始化和数据划分较为敏感。设模型的参数更新遵循 $\theta_{t+1} = \theta_t - \eta \Delta_{\theta} L(\theta_t)$ ，不同随机种子可能导致收敛到不同的局部最优，进而产生不同的社会学推论（例如，对于同一批数据，两次训练可能识别出不同的“关键志愿者”）。未来研究需开展系统的重复实验与敏感性分析（如对超参数网格搜索并报告结果分布），以验证结论的稳健性。

外推风险。本框架中的预测模型（如参与意愿预测、流失风险）仅在训练数据分布的支撑集内有效。当政策环境发生剧烈变化（如突发公共事件导致志愿需求激增）或出现全新的志愿者类型时，模型的推理性可能大幅下降。结合结构因果模型（SCM）与外推条件下的不确定性量化（如置信区间、反事实误差界）是未来重要方向。

（三）理论层面的限制

计算机制与社会意义之间的张力。人工智能模型所揭示的“机制”（如潜空间向量的某个维度与参与概率的高度相关）并不等同于社会学意义上的“解释”。前者是统计关联的模式识别，后者要求理解行动者的意图、情境约束与规范性理由。例如，模型可能发现“早期反馈质量”是长期参与的最佳预测特征，但这并不能回答“为什么反馈质量会影响动机”——是增强了归属感、提升了自我效能，还是降低了不确定性？要弥合这一差距，需要将深度学习的模式发现与定性研究（如深度访谈、过程追踪）相结合，形成“混合方法”的解释策略。

因果识别的局限性。虽然本研究引入了双重机器学习（DML）和反事实扰动分析，但這些方法仍然依赖“无未观测混杂”“正确模型设定”等强假设。在观察性志愿数据中，未观测的混杂因素（如个人性格、家庭支持）可能同时影响参与行为和网络位置，导致因果效应估计偏误。未来的研究可结合工具变量、断点回归或随机实验（如对部分志愿者随机推送激励信息）来增强因果识别的可信度。

(四) 未来研究方向

基于上述局限，本研究提出四个未来重点方向。

方向一：构建因果增强的志愿行为模型。结合 SCM 与深度生成模型（如变分自编码器），建立可解释的动态因果机制。例如，学习时变因果图：

$$A^{(t)} = \text{LearnCausalGraph}(X^{(t)}) \quad (73)$$

从而揭示政策变化或重大事件对志愿行为的结构性影响，并支持反事实政策模拟。

方向二：发展可解释的图神经网络与多模态分析框架。现有 GNN 难以处理多类型关系与高阶结构，未来可结合张量神经网络或图推理模型，GNN 不仅能预测，还能输出符号化的关系规则（如“若志愿者 A 与志愿者 B 有长期合作，且志愿者 B 与志愿者 C 有强连接，则志愿者 A 与志愿者 C 倾向于合作”），从而提升解释的透明性。

方向三：推动志愿系统数字孪生研究。在现有 ABM 与 GNN 基础上，融合实时数据流（如平台日志、地理信息），构建高保真的志愿生态数字孪生系统，用于突发事件下的资源调度演练、长期政策效果预测以及组织韧性评估。

方向四：深化人工智能在志愿服务中的伦理治理研究。随着行为数据采集范围的扩大，隐私保护（如差分隐私的超参数选择）、公平性约束（如机会均等与个体公平的权衡）、算法备案与审计制度等问题亟须跨学科的规范研究。本研究提出的伦理风险仿真模型可作为一种量化评估工具，但其参数校准与实证验证尚需进一步工作。

尽管存在上述局限，本研究展示的“计算化志愿服务社会学”框架仍为经典理论与前沿技术的深度融合提供了一个可扩展的起点。我们期待未来有更多研究在数据、模型与理论层面共同推进这一方向，使志愿治理既高效又充满人文关怀。

十三、伦理、治理与算法公平性：志愿系统智能化的规范框架

人工智能模型在志愿系统中的嵌入带来了预测能力的提升与理论机制的显化，但也引发了数据伦理、算法偏差、透明性缺失等一系列挑战。志愿服务领域涉及公益动机、社会信任与弱势群体保护，因此任何自动化决策必须在伦理、公平性与治理制度上保持规范。本节整合了伦理风险识别与规范性治理框架，从数据伦理、算法偏差、透明性与问责、伦理风险仿真以及制度设计五个方面，系统构建志愿智能系统的治理体系。具体的模型可解释技术（如注意力可视化、节点扰动分析）已在第七节介绍，此处不再重复。

(一) 数据伦理与信息权力结构

志愿服务数据的收集与使用构成了一种信息权力。数字平台、公益组织和政府部门可能形成不对称的信息结构，志愿者往往缺乏对数据使用方式的充分了解（Barocas et al., 2023）。

设数据分布为 $p(x)$, 隐私保护可形式化为双目标优化:

$$\min_{\theta} L(\theta) + \lambda R(x), \quad R(x) = I(x; f_{\theta}(x)) \quad (74)$$

其中, $I(\cdot; \cdot)$ 表示互信息。虽然 $R(x)$ 直接优化困难, 但实际中可采用差分隐私 (Dwork & Roth, 2014) 为模型输出添加噪声, 以实现可量化的隐私保护。联邦学习 (McMahan et al., 2017) 通过仅上传梯度而非原始数据, 也是缓解隐私风险的有效路径。此外, 志愿者数据的所有权、使用权应制度化区分, 平台需遵循“最小必要原则”, 并接受独立的审计。

(二) 算法偏差与公平性约束

算法偏差是志愿智能系统中最突出的风险之一。若训练数据包含历史不公或刻板结构 (如性别、年龄、城乡差异), 模型可能放大这些偏差 (Mehrabi et al., 2021)。例如, 若某地区男性志愿者长期占主导, GNN 可能学习到偏向男性的结构特征, 从而影响资源匹配。

对于二元结果预测 (如志愿者流失预测), 群体公平性可采用机会均等 (Hardt et al., 2016) 定义:

$$P(\hat{Y} = 1 | Y = 1, A = a_1) = P(\hat{Y} = 1 | Y = 1, A = a_2), \quad (75)$$

其中, A 为敏感属性 (如性别、年龄)。对于网络任务 (如组织合作关系预测), 可定义结构公平性指标: 计算不同群体 (G_1, G_2) 的节点嵌入均值差异:

$$S = \left\| \frac{1}{|G_1|} \sum_{i \in G_1} i - \frac{1}{|G_2|} \sum_{j \in G_2} j \right\|_2, \quad (76)$$

若 S 过大, 说明模型嵌入存在系统偏差。为缓解偏差, 可在损失函数中加入公平性正则项:

$$L_{\text{total}} = L_{\text{pred}} + \alpha L_{\text{fair}} \quad (77)$$

其中, L_{fair} 可设计为机会均等约束的松弛形式或嵌入分布的最大平均差异 (MMD)。

(三) 透明性与问责机制

透明性要求模型决策过程可被审查。治理者可利用第七节中的注意力可视化、节点重要度、反事实扰动等工具理解模型行为。但技术解释本身不足以建立问责制, 还需要进行如下操作。

模型审计日志: 记录数据来源、模型版本、超参数、训练过程中的损失与梯度变化 (如式 87)。审计日志应定期由独立第三方检查。

人工复核权限: 对于涉及志愿者评估、资源分配的关键决策, 保留“人工智能建议—人工判断确认”的双轨制度。

扰动鲁棒性测试: 对输入进行微小扰动 δ , 检查模型输出是否稳定:

$$\|f_{\theta}(x + \delta) - f_{\theta}(x)\| \leq \epsilon, \quad (78)$$

若不满足, 说明模型脆弱, 需重新训练或修正。

(四) 伦理风险的系统仿真与治理策略评估

基于前文的动力系统模型 (第六节), 可进一步扩展为“伦理风险仿真模型”, 用于评估不同治理策略对系统稳定性的影响。设风险水平 $R(t)$ 随时间演化, 可由下式表示:

$$\frac{dR(t)}{dt} = \eta_1 B(t) + \eta_2 P(t) - \eta_3 G(t), \quad (79)$$

其中, $B(t)$ 表示算法偏差压力 (如训练数据中的偏差传播程度), $P(t)$ 表示隐私泄露压力 (如数据使用强度), $G(t)$ 表示治理能力强度 (如审计频率、公平性约束强度)。通过调整 $G(t)$, 即可评估不同治理策略 (如加强审计、提高隐私保护标准) 对系统稳定性的影响。这为政策设计提供了一种“可计算化”的伦理评估工具, 有助于在技术发展与社会价值之间实现平衡。

(五) 制度设计：构建志愿系统的人工智能治理框架

在制度层面, 需要构建覆盖数据、模型、平台与社会后果的综合治理框架, 包括以下要素。

- 数据治理制度：建立数据分级分类制度、敏感属性使用规范与志愿者数据使用许可机制。
- 平台治理制度：要求数字公益平台披露算法使用场景, 建立申诉机制, 志愿者能够对不公结果提出重新审查。
- 公平性技术标准：制定针对志愿场景的算法公平性评价标准, 避免从商业领域直接移植未经适配的指标。
- 人机协同治理：明确人工智能工具与社会工作者之间的分工边界, 使两者形成互补而非替代关系。

上述制度安排有助于避免人工智能在志愿体系中造成“技术官僚主义”, 确保志愿治理保持其伦理基础与公众信任。结合第七节提供的技术解释方法, 志愿智能系统可以实现既高效又负责任的社会治理。

十四、结论

本研究构建了一个融合志愿服务社会学与人工智能的多尺度计算框架, 整合多源数据、表示学习、图神经网络、时序模型、多智能体仿真及因果推断等工具, 为志愿行为研究提供了兼具表达力、预测力与理论生成能力的建模体系。

本研究的主要贡献有三个方面。第一, 提出志愿社会的多尺度建模框架, 贯通微观行为、中观组织网络与宏观生态, 为整合碎片化的志愿服务理论 (动机、社会资本、制度) 提供了结构化平台。通过统一数据嵌入、深度生存分析、时序图网络与强化学习智能体, 该框架可刻画参与轨迹、资源流动与政策情境的交互。第二, 构建“数据—模型—理论—政策”闭环。深度学习不仅提升预测精度, 更通过可解释性分析 (注意力、节点扰动、反事实仿真) 生成可检验的理论机制, 推动志愿服务研究从描述走向解释与干预。例如, 动机—体验契合度、动态社会资本等概念被形式化为可计算的潜变量。第三, 强调伦理治理的基础地位。将可解释性、公平性约束、差分隐私与审计制度内嵌于建模流程, 并引入伦理风险仿真模型, 确保智能化辅助在不削弱公民社会活力的前提下提升治理效能。

此外,多智能体仿真体系为政策实验提供了可重复、可控制的计算平台,可模拟不同激励与组织结构下的系统响应。数字孪生雏形与联邦学习路径则为未来跨区域、跨制度比较研究奠定了基础。

综上所述,本研究展示了人工智能与志愿服务社会学双向赋能的可行路径,并为其他公共事业的计算化研究提供了可迁移的方法论参照。未来方向包括因果增强的行为模型、可解释图神经网络、高保真数字孪生及伦理治理的深化校准。

参考文献

- [1] Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, 21, 93–113. <https://doi.org/10.1146/annurev.so.21.080195.000521>
- [2] Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <https://doi.org/10.1109/TPAMI.2018.2798607>
- [3] Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.
- [4] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [5] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Suppl. 3), 7280–7287. <https://doi.org/10.1073/pnas.082080899>
- [6] Burt, R. S. (1992). *Structural holes: The social structure of competition*. Harvard University Press. <https://doi.org/10.4159/9780674029095>
- [7] Caruana, R. (1997). Multitask learning. *Machine Learning*, 28, 41–75. <https://doi.org/10.1023/A:1007379606734>
- [8] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1724–1734). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>
- [9] Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120. <https://doi.org/10.1086/228943>

- [10] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
- [11] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of NAACL-HLT 2019* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>.
- [12] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>.
- [13] Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4(5), 41–60. [https://doi.org/10.1002/\(SICI\)1099-0526\(199905/06\)4:5<41::AID-CPLX9>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F)
- [14] Faraggi, D., & Simon, R. (1995). A neural network model for survival data. *Statistics in Medicine*, 14(1), 73–82. <https://doi.org/10.1002/sim.4780140108>
- [15] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- [16] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.1086/225469>
- [17] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), Article 93. <https://doi.org/10.1145/3236009>
- [18] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. *In Advances in Neural Information Processing Systems 30* (pp. 1024–1034).
- [19] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *In Advances in Neural Information Processing Systems 29*.
- [20] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [21] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *In Proceedings of the 5th International Conference on Learning Representations*.
- [22] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- [23] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>

- [24] Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, 473, 167–173. <https://doi.org/10.1038/nature10011>
- [25] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of Machine Learning Research*, 54, 1273–1282.
- [26] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), Article 115. <https://doi.org/10.1145/3457607>
- [27] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533. <https://doi.org/10.1038/nature14236>
- [28] Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- [29] Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511803161>
- [30] Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.
- [31] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ‘Why should I trust you?’ Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939778>
- [32] Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- [33] Salganik, M. J. (2017). *Bit by bit: Social research in the digital age*. Princeton University Press.
- [34] Sampson, R. J., MacIndoe, H., McAdam, D., & Weffer-Elizondo, S. (2005). Civil society reconsidered: The durable nature and community structure of collective civic action. *American Journal of Sociology*, 111(3), 673–714. <https://doi.org/10.1086/497351>
- [35] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [36] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30* (pp. 5998–6008).
- [37] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph atten-

- tion networks. *In Proceedings of the 6th International Conference on Learning Representations*.
- [38] Wilson, J. (2000). Volunteering. *Annual Review of Sociology*, 26, 215–240. <https://doi.org/10.1146/annurev.soc.26.1.215>
- [39] Wilson, J., & Musick, M. (1997). Who cares? Toward an integrated theory of volunteer work. *American Sociological Review*, 62(5), 694–713. <https://doi.org/10.2307/2657355>
-

Computational Sociology of Volunteering: A Cross-Level, Multi-Model Framework Integrating Artificial Intelligence

Zhang Wangcheng, Yu Zhuoer

Abstract: Traditional sociology of volunteering relies on small-sample surveys and static analyses, which struggle to capture the large-scale dynamics, network complexity, and cross-scale nature of volunteering behavior. This study proposes a framework of “computational sociology of volunteering” that integrates multi-source data representation learning, graph neural networks (GCN/GAT), sequential models (LSTM/Transformer), survival analysis, multi-agent simulation, and causal inference. It builds a multi-scale modeling system covering micro-level individual behavior, meso-level organizational networks, and macro-level system ecology. Through explainable artificial intelligence methods (attention mechanisms, node perturbation, counterfactual interventions) and ethical governance mechanisms (fairness constraints, differential privacy, auditing systems), the framework translates core sociological concepts—such as social capital, participation trajectories, and organizational collaboration—into computable variables and supports simulation-based evaluation of policy interventions. The three experiments described in the paper (prediction of participation persistence, structural hole analysis of organizational networks, and multi-agent simulation of policy interventions) are currently theoretical designs and will be empirically validated in future research using real-world volunteering platform data. This study provides a transferable methodological pathway for computational social science in the public service domain.

Keywords: Sociology of volunteering; Computational social science; Graph neural networks; Explainable artificial intelligence; Policy simulation
