智慧景观

人机协同视域下吐鲁番景观价值评价与智慧服务响应机制

——基于 AI 预测模型与公众实测数据的耦合校验

马 丁[1] 李又达[1] 李雨柯[2] 邵钰涵[1]

（1. 同济大学建筑与城市规划学院，上海 200092；2. 广安理工学院筹建处，四川广安 638000）

摘要：【目的】针对我国西部干旱-半干旱区景观评价面临的地域广阔、调研成本高昂、样本覆盖不足，以及服务响应滞后等痛点，构建一种低成本的“人机耦合”智慧服务生成新范式，以突破传统公众参与模式的时空局限，为西部欠发达地区精细化治理提供依据。【方法】构建“AI 预测—实测校验”双轨模型，利用 LLM 生成与实测人口学结构同构的虚拟公众代理形成价值基准，并以实地问卷作为地面真值开展耦合对比与差异诊断。【结果】AI 预测与公众实测之间存在显著（q<0.05）偏差：AI 更易高估科学、历史等知识性价值；公众评价更受在场体验及主客身份与受教育程度差异等因素影响，形成“在场沉浸增益区”高价值-低感知的“认知赤字区”与高期待-低体验“服务盲区”。【结论】基于偏差诊断提出 AR 遗产解译、微气候感知动态游线与社区数字记忆平台等智慧响应机制，验证 AI 可作为规划前期的理性参照与盲区发现工具。未来应引入多模态大模型（LMMs）结合微气象数据，建立“人在回路（Human-in-the-loop）”的反向校准机制，推动静态评估向动态自适应治理跨越。

关键词：人工智能；公众参与；智慧服务；耦合校验；吐鲁番

DOI：10.12169/zgcsly.2026.01.21.0001

Landscape Value Evaluation and Smart Service Response Mechanism in Turpan from the Perspective of Human-AI Collaboration: Coupling Verification Based on AI Prediction Models and Empirical Public Data

Ma Ding[1] Li Youda[1] Li Yuke[2] Shao Yuhan[1]

(1. School of Architecture and Urban Planning, Tongji University, Shanghai 200092, China; 2. Guang'an Institute of Technology, Guang'an 638000, Sichuan, China)

Abstract: 【Objective】 This study addresses key challenges in the evaluation of arid and semi-arid landscapes in western China, such as vast spatial extent, high fieldwork costs, insufficient sample coverage, and delayed service responses. It aims to construct a novel low-cost "human-AI coupling" paradigm for smart service generation in order to tackle the spatiotemporal constraints of conventional public participation and provide a basis for the target-centered governance of the underdeveloped west region. 【Method】 We develop a dual-track model of "AI prediction + ground-truth validation". Large language models (LLMs) are used to

收稿日期：2026-01-21

第一作者：马丁，男，1990 年生，博士后，研究方向为历史城市景观遗产和恢复性景观（文化恢复）。E-mail：mad@tongji.edu.cn

通信作者：邵钰涵，女，1984 年生，博士，副教授，研究方向为风景园林理论与技术、社会恢复性城市、景观规划设计、海岸带景观。E-mail：shaoyuhan@tongji.edu.cn

generate virtual public agents whose demographic structure is isomorphic to that of the surveyed population, thereby establishing an AI-based value baseline. Field questionnaires are collected as ground truth to enable coupling-based comparison, discrepancy quantification, and diagnostic analysis. 【Result】 Significant divergences ($q<0.05$) are identified between AI predictions and *in situ* public ratings. The AI tends to overestimate knowledge-oriented values (e. g., scientific and historical values), whereas public evaluations are more strongly shaped by factors such as on-site experience and differentiation in visitors/residents' identity and educational attainment. Discrepancy diagnostics further reveal three spatial-experiential patterns, i. e., "on-site immersion-for-gain zone", "cognitive deficit zone" of high-value but low-perception, and "service blind spot" of high-expectation but low-experience. 【Conclusion】 Based on the diagnostic biases, we propose smart response mechanisms including AR-enabled heritage interpretation, microclimate-sensing dynamic touring routes, and a community digital memory platform. The findings verify that AI can serve as a rational reference and an effective tool for detecting service blind spots in the early stages of planning. Furthermore, future research should introduce large multimodal models (LMMs) in connection with microclimate data to establish a "human-in-the-loop" reverse calibration mechanism, driving the transition from static evaluation to dynamic and adaptive governance.

Keywords: artificial intelligence; public participation; smart service; coupling verification; Turpan

"数字中国"背景下，风景园林正经历从传统形态营造向智慧服务转型的关键期[1-2]。数字赋能要求景观系统具备感知与响应能力，实现从被动管理向主动预知的跨越[3-4]。

针对西部干旱-半干旱荒漠绿洲区等生态脆弱区的景观评价，现有研究主要沿循两条路径：一是基于遥感（RS）与GIS的宏观空间测度，侧重生态格局与风险评估[5-6]；二是基于专家赋权或网络大数据的微观评价，在量化视觉感知与风景价值[7]。然而，在面向"智慧普惠服务"转型时，传统范式在西部特殊语境下面临显著局限：宏观评价往往"见物不见人"；而网络大数据在欠发达区域存在显著的"数据稀疏性"。同时，若完全依赖线下调研，因地广人稀与极端气候（如新疆吐鲁番），不仅人力成本极高、数据滞后[8]，更易受数字鸿沟与环境阻隔影响产生"幸存者偏差"——由于排斥了老年群体及因恶劣环境缺席的潜在用户，导致弱势群体需求"隐形"，违背了普惠初衷[9]。

生成式人工智能（AIGC）的发展为上述痛点提供了"硅基样本"新工具[10]。经角色设定（Persona）的大语言模型（LLM）能高保真模拟群体行为，克服社会赞许性偏差并突破时空限制[11]；多模态模型更具备结合地理环境推理人类特定生理反应（如高温胁迫）的能力[12]。尽管已有AI在通用景观评价中的初步探索[13]，但其在复杂遗产与极端环境交织的西部地域，认知的准确性与感知差异仍待实证。鉴于此，本研究选取具"极端环境—地标遗产"二元特征的新疆吐鲁番为实证对象，构建"AI预测—实测校验"双轨机制，旨在利用实测数据校准AI模型，通过海量硅基模拟填补数据盲区，修正"幸存者偏差"；通过人机认知对比，精准识别景观服务中的"认知赤字"与"体验盲区"；期望建立一套低成本智慧服务生成范式，为西部欠发达地区精细化治理提供依据。

1 研究区概况

新疆吐鲁番盆地（88°51′—91°54′E，41°12′—43°40′N）属暖温带大陆性极端干旱荒漠气候，年均降水不足16 mm，夏季极端高温达49.6 ℃。研究区呈显著垂直地带性特征，由天山冰雪融水经坎儿井地下输送，穿越戈壁灌溉绿洲，最终汇入艾丁湖，构成"雪山-荒漠-绿洲-湖泊"生态系统。研究区内富集交河故城等世界文化遗产及生土乡土建筑。研究区极高的景观禀赋与极端气候带来的低游憩舒适度形成鲜明对比，是验证极端环境下人机感知差异的典型样本。

2 研究方法

针对传统景观评价中客观环境数据与主观感知数据难以融合的痛点，本研究提出一种基于人机耦合的智慧服务生成范式。研究通过构建实地

具身感知（Ground Truth）与 AI 预测（AI Prediction）的双轨数据集，利用差异化校验模型识别真实人群与大语言模型在同一问卷下的回答差异，剥离出环境压力对价值认知的影响，进一步识别吐鲁番景观服务中的“盲区”。

2.1　数据获取

2.1.1　实测数据：基于多源感知的问卷调查

采用问卷获取公众感知的“地面真值”。问卷含 6 大模块：1）人口学特征；2）景观特征识别；3）感官与环境感知；4）多维价值评价；5）环境感知与管理意愿；6）综合排序。采用分层与滚雪球抽样，在吐鲁番主要景区及社区线上线下分发。回收有效问卷 570 份（有效率 87.7%）。样本涵盖多元社会经济属性：游客占 65.4%，居民占 34.6%（区分汉族与少数民族）；年龄覆盖各生命周期；学历涵盖初中至硕博。多维社会属性为探析人群价值观分化奠定基础，样本符合正态检验要求（表 1）。

表 1　调查问卷核心维度及涵盖问题范围

核心维度	二级分类	具体指标/问卷题项涵盖内容	设计意图与理论支撑	题目编号参考
人口学特征	基础画像	年龄段、最高学历	构建受访者基础画像，用于分析不同生命周期与认知水平群体的评价差异	Q1，Q2
	社会身份	身份类型（游客、当地少数民族居民、当地汉族居民、科研人员）	区分“内部视角”与“外部视角”，识别文化背景对景观认知的调节作用	Q3
景观特征识别	物理设施与遗产	坎儿井系统（竖井/暗渠）、葡萄晾房（生土建筑）、交河/高昌故城遗址、苏公塔、柏孜克里克千佛洞、烽燧/古驿站	评估物质实体作为“地标”的代表性强度，通过识别率判断遗产的社会知名度	Q4，Q5，Q8，Q9，Q10，Q14，Q18
	自然与生态景观	连片葡萄园（绿洲）、火焰山（丹霞地貌）、库木塔格沙漠、艾丁湖（洼地）、农田防护林网	考察公众对吐鲁番“极端干旱与人工绿洲”二元生态结构的认知程度	Q7，Q11，Q12，Q13，Q15
	生活与公共空间	传统巴扎（集市）、清真寺（社区中心）	关注日常生活空间在文化景观体系中的地位，强调“活态遗产”属性	Q16，Q17
感官与环境感知	具身/生理感知	极度干热气候感（“火洲”热浪）、极度的甜味（味觉记忆）、干燥的泥土气息（嗅觉）	引入“具身认知”维度，测量极端环境压力（如高温）如何介入并重塑景观体验。这是本研究区别于传统视觉评价的关键创新	Q19，Q22，Q24
	视觉与光影	强烈的色彩对比（黄沙/红山/绿洲）、光影效果（葡萄架下斑驳）、夜间星空通透感、艾德莱斯绸（视觉符号）	评估景观的视觉冲击力与美学氛围，尤其是干旱区特有的光环境体验	Q20，Q21，Q29，Q32
	听觉与空间感	流水的声响（坎儿井听觉）、封闭与围合感（盆地地形）、幽深的街巷（迷宫感）、历史沧桑感（废墟美学）	捕捉多感官通道的景观叙事能力，考察空间形态对心理感受的塑造	Q23，Q25，Q26，Q31
	文化氛围体验	多民族融合氛围、木卡姆/麦西热甫（歌舞律动）、馕坑与烤肉（烟火气）、“丝路驿站”商贸联想	测量非物质文化要素的场景化感知，评估文化景观的“场所精神”	Q27，Q28，Q30，Q33
多维价值评价	历史与科学价值	丝路文明见证（历史）、坎儿井适应智慧（科学）、晾房风资源利用（建筑学）、遗址原真性、垂直地带性地貌（地学）	采用李克特 5 级量表，量化公众对景观深层认知价值的认同度。旨在发现“高认知门槛”的价值点（如科学原理）	Q34，Q35，Q36，Q37，Q39
	美学与艺术价值	荒漠绿洲色彩美学、废墟悲剧美（艺术感染力）、建筑装饰纹样	评估景观的艺术审美层级，从形式美到意境美的转化	Q41，Q42，Q43
	社会与生态价值	景观完整性、农业物种基因（生态）、归属感与自豪感（麻扎/寺庙）、社区邻里关系、爱国主义教育、民族交流融合	考察景观在维系社会结构、传承集体记忆及生态服务方面的功能	Q38，Q40，Q44，Q45，Q46，Q47
	经济与品牌价值	旅游直接收入、农产品品牌溢价、影视艺术衍生价值	评估景观资源的资本化潜力及其对地方经济的带动作用	Q48，Q49，Q50
环境感知与管理意愿	风险与保护态度	破坏风险感知、生态文物优先性、风貌协调性要求、商业开发限制、政府（责任）VS 个人责任	测量公众的环保意识与对过度商业化的警惕性，为管理政策制定提供民意基础	Q51，Q52，Q53，Q54，Q55
	服务与支付意愿	数字化展示需求（VR/AR）、支付/牺牲利益意愿、非遗抢救性保护	探索公众对“智慧服务”及数字化介入的接受度，评估保护资金的社会筹措潜力	Q56，Q57，Q58
综合排序	价值维度排序	人文景观、自然景观、旅游开发、居民感知、游客体验的重要性排序	通过直观排序，确立公众心目中吐鲁番景观价值的层级结构，辅助决策优先序	Q59，Q60，Q61，Q62，Q63

2.1.2　AI 预测数据：基于角色扮演的大语言模型模拟

基于 GPT-4o 构建“虚拟公众代理”，通过 3 阶段提示词工程（Prompt Engineering）生成对照数据。1）角色映射：依据实测样本人口学比例，重构 90 个等比虚拟画像（如“50 岁本地果农”）；2）情境注入：植入吐鲁番夏季极端气象参数（气温 42 ℃、相对湿度 10%），强制模型在

环境压力下响应；3）思维链推演（CoT）：设定“角色内省”指令，要求 AI 评分前基于角色处境逻辑推演，还原复杂情境下的真实决策过程。最终生成 90 份对照数据集。

2.2 人机耦合校验的统计方法与差异分区规则

以清洗后的公众问卷样本（n = 570）和 GPT-4o 生成的预测数据集（n = 90）作为两类独立样本开展人机耦合校验。鉴于李克特量表属于定序数据、分布形态可能偏态且两组样本量不等，本文采用以非参数检验为主要统计框架[14]。

为表征 AI 预测与公众实测之间的方向与幅度差异，并将数据差异转化为服务策略，研究构建“认知偏差指数”——对于景观评价指标 i，认知偏差指数 G_i 定义为实测组均值（μ_H）与 AI 预测组（μ_{AI}）的差值：

$$G_i = \mu_H(i) - \mu_{AI}(i) \quad (1)$$

式（1）中：μ_H 与 μ_{AI} 分别为公众组与 AI 组的指标均值；$G_i>0$ 表示公众实测高于 AI 预测，可解释为现场体验或具身感知强于知识性预期，$G_i<0$ 表示公众实测低于 AI 预测，可解释为价值表述向可感知体验的转化存在不足或服务支撑欠缺。

使用 Mann-Whitney U 验证两组数据差异的显著性[14]。并对 P 值实施 Benjamini-Hochberg 假发现率（FDR）校正，以 $q<0.05$ 作为显著性判据[15]。

为评估差异的实质性影响，在显著性之外计算 Cliff's δ 效应量[16]，并依据 Romano 标准分为：可忽略（$|\delta|<0.147$）、小效应（$0.147\leqslant|\delta|<0.33$）、中效应（$0.33\leqslant|\delta|<0.474$）、大效应（$|\delta|\geqslant0.474$）[17]。据此，满足 qFDR<0.05 且 $0.147<|\delta|$ 指标被定义为“有效差异指标”，用于后续差异分区，其中 Gi 不作为差异成立的判据，仅用于偏差方向与幅度的描述。此外，采用 Fligner-Killeen 离散度检验评估两组在离散程度上的差异，并结合 IQR 与标准差提供描述性解释[18]。

最后，为检验 AI 代理在指标均值层面对公众评价结构的整体一致性，采用 Lin 一致性相关系数（concordance correlation coefficient，CCC）进行评价。CCC 能同时反映相关程度与系统偏差，适用于比较两组连续或近似连续指标在数值层面的可比性，用于整体层面的模型有效性验证，而不作为单个指标项差异判定的依据[19]。上述统计过程在 Python 环境中实现。

3 结果与分析

3.1 总体评价特征与模型有效性验证

Lin 一致性相关系数（CCC = 0.565）显示，AI 代理与公众实测在指标均值层面呈现中等程度的一致性，表明模型能够捕捉到景观价值评价的总体趋势，验证了将其作为基础参考系的有效性。方向稳定性检验进一步证实，在具有统计学意义的差异指标项中，偏差方向具有高度稳健性（符号一致率中位数达 1.000）；说明观测到的差异并非样本随机性带来的扰动，而是反映了人机之间客观存在且稳固的认知结构分歧，具有明确的统计学推断价值。

经 BH-FDR 校正后，Mann-Whitney U 检验显示 60 个（Q1—Q3 为人口学变量，不纳入差异检验；统计检验对象为 Q4—Q63 共 60 项评价指标。）指标项中有 39 个指标项存在中心位置差异（$q<0.05$），表明人机评价在相当比例的指标项上存在稳定的方向性偏离。Cliff's δ 效应量统计进一步指出，差异强度呈现出明显的“长尾分布”特征：尽管半数指标的差异强度处于“可忽略”或“小效应”区间，但仍有 7 个关键指标达到中、大效应水平。这表明人机认知的核心冲突并非均匀弥散，而是高度集中于特定的景观类型与感知维度，构成了关键的偏差来源。

此外，Fligner-Killeen 离散度检验显示 60 个指标项中有 54 个显著的离散度差异（$q<0.05$），这一结果提示，人机差异不仅仅体现为均值的高估或低估，更深层地体现为数据分布形态的异质性：AI 代理倾向于形成算法化的趋同评价（集中度高），而真实公众评价则因个体偏好与体验情境的不同而表现出显著的离散特征。这种分布形态的错位，客观反映了“标准化的 AI 理性”与“情境化的公众感性”在群体共识度上的结构性张力，表明单一的均值预测可能掩盖真实人群内部复杂的意见分歧（图 1）。

3.2 “AI-公众”偏差识别与分类

依据“显著性+效应量”的双重判断，筛选出 35 个有效差异指标（表 2），并在“认知偏差指数”（G_i）上呈现出显著的“双向非对称”结构特征。

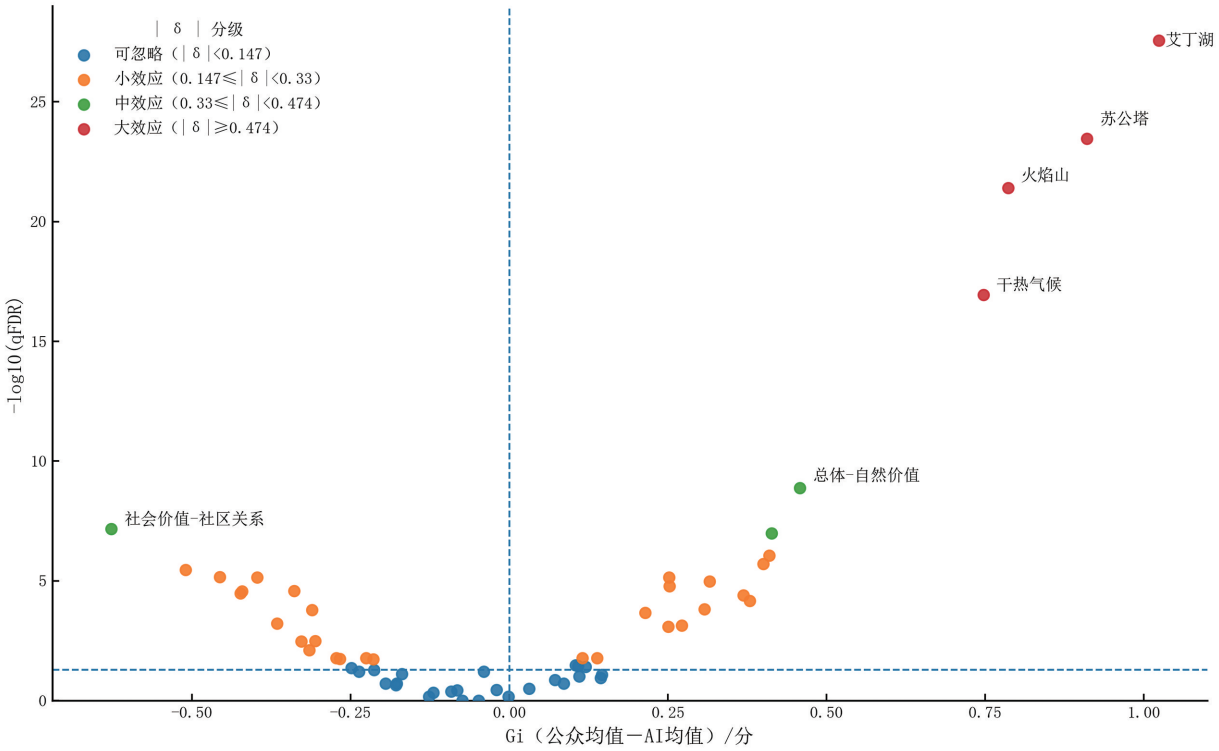


图 1　人机差异的显著性与方向

表 2　有效差异指标

题号	模块	指标名称	G_i	q_{FDR}	*Cliff's δ*
Q45	价值评价	社会价值：社区关系	−0. 627	6. 689E−08	−0. 359
Q40		生态价值：物种	0. 413	1. 030E−07	0. 336
Q36		建筑价值：晾房	0. 401	1. 977E−06	0. 306
Q34		历史价值：丝路	−0. 398	7. 031E−06	−0. 29
Q44		社会价值：归属感	−0. 421	2. 688E−05	−0. 272
Q35		科学价值：坎儿井	−0. 339	2. 614E−05	−0. 261
Q39		地学价值	0. 308	1. 529E−04	0. 241
Q42		美学价值：废墟	0. 272	7. 285E−04	0. 215
Q48		经济价值：旅游	0. 115	0. 017	0. 149
Q19	体验表征识别	干热气候	0. 747	1. 150E−17	0. 54
Q30		饮食烟火	−0. 424	3. 314E−05	−0. 273
Q29		传统服饰	0. 379	6. 834E−05	0. 263
Q31		幽深街巷	−0. 306	0. 003	−0. 198
Q27		民族融合	−0. 315	0. 008	−0. 177
Q24		泥土气息	−0. 273	0. 017	−0. 161
Q32		星空通透	−0. 267	0. 018	−0. 159
Q13	要素代表性识别	艾丁湖	1. 024	2. 790E−28	0. 705
Q10		苏公塔	0. 911	3. 434E−24	0. 649
Q11		火焰山	0. 786	3. 992E−22	0. 598
Q12		库木塔格沙漠	0. 41	8. 763E−07	0. 325
Q16		巴扎	−0. 51	3. 493E−06	−0. 31
Q08		交河故城	0. 252	7. 031E−06	0. 284
Q09		高昌故城	0. 253	1. 624E−05	0. 274
Q18		烽燧驿站	0. 369	4. 004E−05	0. 271
Q06		生土民居	−0. 328	0. 003	−0. 187
Q17		清真寺	−0. 215	0. 018	−0. 159
Q54	治理偏好与响应	限制开发	−0. 456	6. 770E−06	−0. 292
Q58		非遗保护	−0. 311	1. 643E−04	−0. 234
Q51		破坏风险感知	−0. 366	5. 948E−04	−0. 229
Q52		保护优先	0. 25	8. 100E−04	0. 216
Q57		支付意愿	−0. 226	0. 017	−0. 153
Q60	总体重要性权重	总体：自然价值	0. 458	1. 297E−09	0. 385
Q61		总体：旅游价值	0. 316	1. 064E−05	0. 278
Q63		总体：体验价值	0. 214	2. 104E−04	0. 232
Q62		总体：居民价值	0. 139	0. 016	0. 154

正向偏差（G_i>0）主要集中于地标性自然景观与物理遗产维度，其中艾丁湖、苏公塔、火焰山等指标均呈现显著（q<0.05，$|\delta| \geqslant 0.474$）的大效应正向偏差，可定义为“在场沉浸增益区”——对于具有强烈视觉冲击力、巨大空间尺度与显著场景氛围的景观要素，公众在实地情境中获得的在场沉浸感使其综合评价显著高于 AI 基于文本知识与规范的理性预估（q<0.05）。自然价值与生态价值维度的整体评分上扬（图 2），也印证了实地场景的沉浸效应对景观价值认知具有显著的放大作用。

负向偏差（Gi<0）主要集中于社会文化感知与服务支撑维度，社区关系、巴扎体验、归属感及饮食烟火等指标均显示出实测显著低于 AI 预测（q<0.05）。根据指标类型可进一步分解为两类分区：其一是“高价值-低感知”的认知赤字区，典型表现为坎儿井地下工程原理等具有较高认知门槛的条目——AI 往往依据文本知识对其“应然价值”给出较高评价，但现场缺少有效解译路径、学习支撑或可视化解释时，公众难以将抽象价值转化为可感知理解，从而出现显著低估。其二是“高期待-低体验”的服务盲区，以景区热舒适度及相关服务支撑条目为典型——公众的实际评价往往受到热舒适度不足、设施运维短板、拥挤与商业化干扰等现实条件的影响而显著低于 AI 预测（q<0.05），揭示了景观价值从“意义与叙事”向“实际游憩体验”转译过程中的不足。

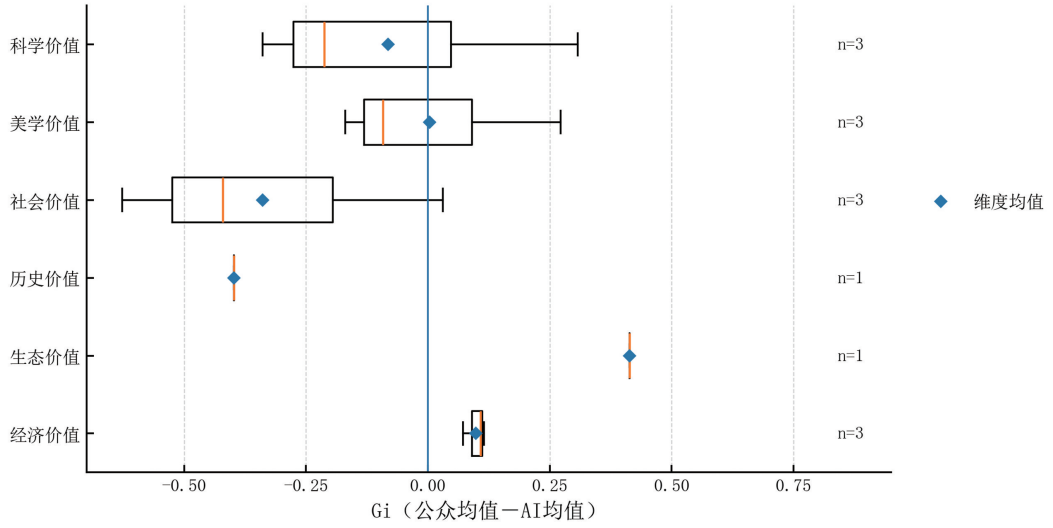


图 2 6 类价值维度下人机认知偏差的分布特征

3.3 群体差异视角：AI 视角的局限

除均值差异外，离散度指标揭示了人机评价在“群体共识度”上的本质区别（表 3）。进一步结合社会经济属性分析发现，公众评价的显著离散性（IQR>0）本质上源于人群认知的结构性分化。例如，在身份属性上，代表“外部视角”的游客群体对视觉奇观（如火焰山、交河故城）的评分普遍高于本地居民；而代表“内部视角”的本地居民，则在“社区关系（Q45）”与“饮食烟火（Q30）”等社会价值指标上的评分显著更高（q<0.05），反映出对活态生活空间的强烈情感依恋。此外，在“坎儿井科学价值（Q35）”的认知上，不同学历群体呈现显著的阶梯状分布，低学历群体受限于知识门槛出现了明显的“认知赤字”。总体来说，AI 代理在 63.3%的指标（38/60）上表现出极高的趋同性（IQR = 0），呈现出标准化的“算法平均意见”特征。相比之下，公众评价在绝大多数指标上表现出显著的离散性（IQR>0），反映了真实人群因社会阶层、文化背景及情境不同而产生的感知分化。

研究识别出 21 个“隐性异质指标”（如歌舞艺术、风貌协调、光影效果、历史沧桑感等），这些指标虽然在均值层面的差异未达统计显著水平（$qFDR \geqslant 0.05$），但其离散度差异极显著（$qFDR$<0.05）。这表明在审美判断、感官细节及文化共情等高阶认知维度，AI 无法模拟真实公众

内部由不同社会经济属性（如学历鸿沟、主客身份差异）交织形成的复杂的偏好差异。均值拟合但分布存在显著差异的现象（$q<0.05$），表明单一的 AI 预测值虽然在平均水平上接近公众，但掩盖了景观体验中基于不同受众画像的多元利益诉求与个性化、情境化差异。

表 3　离散度差异指标

题号	模块	指标名称	qFDR(均值差异)	qFDR(F-K 离散度差异)	IQR_公众	IQR_AI	F-K 离散度
Q38	价值评价	完整性	0.368	7.38E-14	1	0	TRUE
Q41		美学价值-色彩	0.41	1.15E-04	1	0	TRUE
Q43		艺术价值-纹样	0.075	1.62E-05	1	0	TRUE
Q47		社会价值-融合	0.309	1.26E-09	1	0	TRUE
Q50		经济价值-衍生	0.138	1.40E-08	1	0	TRUE
Q46		教育价值	0.053	3.12E-04	1	1	TRUE
Q23	体验表征识别	流水声响	0.668	1.27E-11	2	0	TRUE
Q22		味觉甜味	0.35	1.27E-12	1	0	TRUE
Q25		历史沧桑	0.468	8.33E-13	1	0	TRUE
Q28		歌舞艺术	0.061	4.59E-15	1	0	TRUE
Q21		光影效果	0.195	8.28E-07	1	1	TRUE
Q26		封闭围合	0.082	1.76E-06	1	1	TRUE
Q33		丝路联想	0.111	3.07E-04	1	1	TRUE
Q05	要素代表性识别	葡萄晾房	0.985	8.94E-05	1	0	TRUE
Q04		坎儿井	0.668	4.27E-04	1	1	TRUE
Q07		葡萄园	0.195	4.26E-05	1	1	TRUE
Q15		防护林	0.193	3.03E-07	1	1	TRUE
Q53	治理偏好与响应	风貌协调	0.22	6.12E-15	1	0	TRUE
Q55		保护责任-反向	0.061	6.38E-05	1	1	TRUE
Q56		数字化展示	0.096	3.42E-05	0	0	TRUE
Q59	总体重要性权重	总体-人文价值	0.985	3.12E-04	1	1	TRUE

4　讨论

研究证实，在样本受限时，LLM 可作为低成本的“价值先验结构”辅助前期研判[13]。但 AI 对高阶体验存在“算法平滑”效应[11]，其认知源于文本“语义共识”，难以重现微气候与空间尺度带来的具身感知。因此，AI 应定位于剥离情感的“理性参照系”，用于反衬真实公众感知的痛点。偏差诊断揭示了两大关键机制：文本模型易系统性低估大尺度地标的现场震撼力；基于应然叙事的乐观预判常忽视现实热胁迫与设施短板。更具治理价值的是评价的分布异质性：AI 趋同的标准化输出极易掩盖在地群体的利益分化；智慧服务绝不能被算法的“平均意见”主导，必须基于人群画像实施精细化供给。

为此，本文针对 3 大痛点提出具体智慧响应路径：1）针对认知赤字区，开发 AR 遗产解译系统。依托轻量化 WebAR 技术，将地下暗渠三维可视化叠加于地表，并调用 LLM 按受众学历生成降维讲解，低成本跨越知识门槛。2）针对服务盲区，构建微气候感知导引机制。在暴晒景区部署低功耗气象节点，实时计算热舒适指标（UTCI）并驱动寻路算法，向游客动态推送“低热暴露、高遮阴”避暑游线，以数字“软干预”替代实体工程。3）针对诉求异质性，搭建社区数字记忆平台。引入公众参与式 GIS（PPGIS），鼓励以去键盘化的语音交互上传口述记忆（UGC），动态生成文化热力图，从技术底层保障老弱及少数民族群体的话语权。

本研究利用“人机差值”作为诊断工具，将 AI 从“模拟者”转为发现短板的“探索者”。但研究仍存局限：虚拟代理偏向缺乏身体经验的静态模型；评价量表跨区应用时需重校文化适配性。未来应引入多模态大模型（LMMs）结合微气象数据，建立“人在回路（Human-in-the-loop）”的反向校准机制，推动静态评估向动态自适应治理跨越。

5 结论

本研究构建了“人机耦合”的双轨评价与智慧服务生成新范式，揭示了空间感知中“机器知识理性”与“人类具身感性”的双向非对称偏差规律：一方面呈现出实测值高于预测值的“具身增量效应”；另一方面呈现出实测值低于预测值的“体验损耗效应”。AI代理能够生成具备参考价值的景观价值先验结构，适用于规划前期的快速研判，但因其脱离在地情境，存在系统性偏差，必须引入实测数据进行双轨校验，不可完全替代现场调查。

研究确立的“实测校准先验、差异诊断盲区”响应范式，为西部生态脆弱区实现从“被动管理”向包容且精细的“主动预知”智慧化治理转型提供了兼具理论深度与工程实践价值的可复制范式参考。未来应引入多模态大模型（LMMs）结合微气象数据，建立“人在回路（Human-in-the-loop）”的反向校准机制，推动静态评估向动态自适应治理跨越。

参考文献

[1]成实,张潇涵,成玉宁.数字景观技术在中国风景园林领域的运用前瞻[J].风景园林,2021,28(1):46-52.

[2]CHENG Y N. Digital landscape architecture: logic, structure, method and application[M]. Singapore: Springer Nature Singapore, 2023.

[3]张晓彤,刘晖,左翔,等.数据为媒:数字孪生技术在生态实践中的应用[J].中国城市林业,2025,23(6):171-177.

[4]吴志强.人工智能辅助城市规划[J].时代建筑,2018(1):6-11.

[5]张学渊,魏伟,周亮,等.西北干旱区生态脆弱性时空演变分析[J].生态学报,2021,41(12):4707-4719.

[6]张学斌,石培基,罗君,等.基于景观格局的干旱内陆河流域生态风险分析:以石羊河流域为例[J].自然资源学报,2014,29(3):410-419.

[7]毛志睿,高钰桔,陈笑葵.基于多源数据的昆明主城区城市公园景观视觉质量评价[J].中国城市林业,2022,20(4):96-102.

[8]BROWN G, KYTTÄ M. Key issues and research priorities for public participation GIS (PPGIS): a synthesis based on empirical research[J]. Applied Geography, 2014, 46: 122-136.

[9]ROBINSON L, COTTEN S R, ONO H, et al. Digital inequalities and why they matter[J]. Information, Communication & Society, 2015, 18(5): 569-582.

[10]ARGYLE L P, BUSBY E C, FULDA N, et al. Out of one, many: using language models to simulate human samples[J]. Political Analysis, 2023, 31(3): 337-351.

[11]PARK J S, O'BRIEN J, CAI C J, et al. Generative agents: interactive simulacra of human behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. ACM, 2023.

[12]GROSSMANN I, FEINBERG M, PARKER D C, et al. AI and the transformation of social science research[J]. Science, 2023, 380(6650): 1108-1109.

[13]LIU J X, ZHANG T C, MA Y C, et al. Generative artificial intelligence perspectives on typical landscape types: Can ChatGPT compete with human insight? [J]. Landscape and Urban Planning, 2025, 264: 105479.

[14]MANN H B, WHITNEY D R. On a test of whether one of two random variables is stochastically larger than the other[J]. The Annals of Mathematical Statistics, 1947, 18(1): 50-60.

[15]BENJAMINI Y, HOCHBERG Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1995, 57(1): 289-300.

[16]CLIFF N. Dominance statistics: Ordinal analyses to answer ordinal questions[J]. Psychological Bulletin, 1993, 114(3): 494-509.

[17]ROMANO J, KROMREY J D, CORAGGIO J, et al. Appropriate statistics for ordinal level data: should we really be using t-test and Cohen'sd for evaluating group differences on the NSSE and other surveys[C]//annual meeting of the Florida Association of Institutional Research. 2006.

[18]FLIGNER M A, KILLEEN T J. Distribution-free two-sample tests for scale[J]. Journal of the American Statistical Association, 1976, 71(353): 210-213.

[19]LIN L I. A concordance correlation coefficient to evaluate reproducibility[J]. Biometrics, 1989, 45(1): 255-268.