

融合大语言模型的城市指标体检 评估智能诊断范式构建与探索*

Building an Intelligent Diagnostic Paradigm for Urban Health Assessment by Integrating Large Language Models

王康旭 张姗姗 陈成 甄峰 秦萧 张逸舒

WANG Kangxu, ZHANG Shanshan, CHEN Cheng, ZHEN Feng, QIN Xiao, ZHANG Yishu

关键词 城市体检评估; 国土空间治理; 大语言模型; 智能诊断

Keywords: city health examination and evaluation; territorial spatial governance; large language models; intelligent decision support

中图分类号 TU984 文献标志码 A
DOI 10.16361/j.upf.202602011
文章编号 1000-3363(2026)02-0082-07

作者简介

王康旭, 南京大学建筑与城市规划学院硕士研究生, 522024360118@smail.nju.edu.cn

张姗姗, 南京大学建筑与城市规划学院副教授、博士生导师, 通信作者, zhang-shanshan@nju.edu.cn

陈成, 江苏省测绘工程院高级工程师

甄峰, 南京大学建筑与城市规划学院教授、博士生导师, 城市AI与绿色人居环境营造省高校重点实验室主任

秦萧, 南京大学建筑与城市规划学院副教授、博士生导师, 城市AI与绿色人居环境营造省高校重点实验室副主任

张逸舒, 江苏省测绘工程院工程师

提 要 在国土空间治理数智化转型背景下, 城市体检评估正由指标监测转向综合诊断, 但指标体检中仍存在异常识别停留于表层描述、数据与专业知识衔接不足等问题。尝试将指标体检界定为面向复杂城市系统运行状态的系统性认知活动, 提出由“表征—解释—决策”构成的“S-E-D”诊断认知本体框架, 并构建基于知识增强大语言模型的智能分析方法。该方法通过结构化处理规划文本与技术规范, 在规划知识约束下支持指标分析、问题归因与策略推演。以南京市指标体检为案例, 结果表明, 该方法在多指标整合、因果解释一致性和策略生成可解释性方面表现较好。研究为指标体检由经验驱动转向可复现、可追溯的智能诊断提供了方法参考。

Abstract: In the context of digital and intelligent territorial spatial governance, city examination and evaluation is shifting from indicator monitoring toward comprehensive diagnosis. However, indicator-focused assessment still suffers from superficial anomaly descriptions and weak integration between data analytics and domain knowledge. This study conceptualizes indicator-focused assessment as a systematic cognitive process for understanding the operational state of complex urban systems and proposes an S-E-D diagnostic ontology framework comprising Symptom, Explanation, and Decision. Based on this framework, a knowledge-enhanced large language model approach is developed to support indicator analysis, causal explanation, and strategy generation even under constraints in planning knowledge. This is accomplished through structured processing of planning documents and technical standards. A case study of Nanjing indicates that the proposed method performs well in multi-indicator integration, reasoning consistency, and strategy interpretability. The study offers a methodological reference for transforming indicator-focused assessment from experience-driven practice toward reproducible and traceable intelligent diagnosis.

* 国家自然科学基金面上项目“虚实空间交互的城市生活服务设施优化配置机理与规划响应”(项目编号: 52478061); 国家自然科学基金重点项目“智能技术驱动下城市居民活动和建成环境交互作用机理研究”(项目编号: 42330510); 国家自然科学基金面上项目“基于多源要素流模拟的城镇低效用地智能识别与优化”(项目编号: 52478060); 自然资源部2023年度部省合作试点项目(项目编号: 2023ZRBSHZ061)

1 背景：城市体检评估的数智化转型

1.1 范式转型：从指标监测到智能认知诊断

城市体检评估是国土空间规划实施监督的重要机制^[1]。自2017年住房和城乡建设部提出“一年一体检、五年一评估”以来，城市体检评估逐步由试点走向全面推广^[2]。2021年，自然资源部发布《国土空间规划城市体检评估规程》，将其纳入“多规合一”的国土空间规划体系^[3]。在地方实践中，城市体检评估通常包括指标体检、任务体检和领域体检^[4]，其中，指标体检通过指标监测感知城市空间运行状态，成为体检评估的重要基础。

随着数字化基础设施和时空数据体系完善，指标体检正由事后统计转向动态感知。2025年，自然资源部提出建设国土空间规划实施监测网络（CSPON），系统推动评估分析动态化和辅助决策智能化^[5]，形成智慧规划新格局^[6-7]。这意味着指标体检需突破静态监测，转向问题识别、成因解释与策略生成相结合的智能诊断。

1.2 现实困境：经验依赖与知识整合的双重制约

尽管多源数据提升了城市运行监测能力，但指标体检的深度诊断仍受两方面制约。

一是经验依赖突出。城市系统具有高度耦合、多尺度和非线性特征，指标间存在复杂交互与滞后效应^[8]。规划人员难以在短时间内从碎片化、高频更新的数据中提炼稳定因果链条，评估结论往往停留于异常指标罗列，较大程度依赖个体经验，难以解释结构性成因^[9]。

二是知识整合不足。政策文件、技术规程和专家经验构成评估的专业逻辑，但现有数字化方法多聚焦指标计算和可视化，难以将非结构化知识嵌入分析过程，导致监测数据与专业知识缺乏稳定映射^[10]。而传统数据驱动方法难以刻画规划机理，规则推理又受限于多源信息整合能力^[11]，因此亟须连接数据关联与知识约束的新型方法框架。

1.3 技术机遇：知识增强型智能推理

大语言模型为指标体检智能化提供了新路径^[12]。通过语义检索、知识注入和推理链构建，模型可与规划文本、政策规范等外部知识结合，在专业约束下开展推理^[13]。已有研究表明，检索增强和领域知识注入能提升大模型的知识整合、跨文本推理和多步推演能力，降低事实偏差并增强推理一致性^[14]，相关方法已在金融监管^[15-17]、医疗临床决策^[18-21]和教育评估^[21-23]等高知识密度场景中得到验证。

然而，现有指标体检研究多停留于政策解读或指标释义等局部环节^[22-23]，尚未形成覆盖“指标理解—解释分析—综合判断”的智能诊断范式。基于此，本文将指标体检界定为面向复杂城市系统运行状态的认知诊断活动，提出“表征（symptom）—解释（explanation）—决策”（decision）（S-E-D）认知本体框架，并构建知识增强型智能分析方法，以探索其在指标体检中的应用路径。

2 认知本体构建与知识增强机制

2.1 诊断本体重构：“S-E-D”认知本体

从体检本义看，指标体检不仅是一项技术性监测活动，还是一种面向城市复杂适应系统开展综合诊断^[24]，其重点不在于单项指标的静态比对，而在于识别“空间—社会—环境”多子系统的耦合状态、运行偏离及潜在风险^[25-26]。随着城市系统呈现出显著的非线性、多尺度耦合与反馈特征，传统指标对比或线性加权的评价方式已难以有效解释运行偏离的成因与潜在风险，指标体检急需一种能够贯通现象识别、机理解释与决策干预的系统性认知框架。

基于此，本文尝试将指标体检的认知过程表述为一个由“表征—解释—决策”构成的“S-E-D”认知本体框架，用以刻画从状态感知到因果理解、再到策略形成的分析路径。

2.1.1 表征层：状态感知与偏离识别

表征层侧重于将城市运行状态转化为可识别、可比较的问题信号。该层通过多源指标体系识别偏离规划目标或合理区间的异常状态，并关注指标的动态变化及联动关系。相较于静态指标对比，

表征层更强调异常信号的语义化识别，为后续问题解释提供诊断入口。

2.1.2 解释层：情境归因与路径识别

解释层的核心在于揭示指标偏离背后的形成机理。该层将异常信号置于具体空间结构、制度环境与社会经济过程之中，识别多因素交互作用下的成因路径。由于相似指标异常在不同空间与制度条件下可能对应不同治理含义，解释层本质上是一种以规划知识为支撑的情境化归因过程。

2.1.3 决策层：策略生成与决策输出

决策层进一步将解释阶段形成的诊断认知转化为治理策略。该过程要求策略建议能够回应问题成因及其作用路径，并与规划目标、制度约束和实施条件相衔接。由此，决策层以“症状识别—机理诊断—策略干预”的逻辑组织输出，为规划决策与实施评估提供认知支撑。

2.2 知识约束下的智能推理路径

本文将知识约束机制划分为知识建模与结构化组织、语义检索与动态调度、融合生成与路径控制等三个层级，分别对应知识组织、知识调用与推理约束。

首先，知识组织将规划知识转化为可计算的语义单元。一方面，对政策文本、技术规范和专家经验中的知识进行主题切分与语义规整，形成相对独立的知识片段^[27]；另一方面，借助S-BERT、BGE等文本嵌入模型，将其映射为稠密向量表示，构建“文档—向量”语义嵌入空间^[28]，为模型推理提供专业语义基础^[29-30]。

其次，知识调用通过相似度计算，实现任务语境与知识单元的动态匹配^[31]。当模型接收具体分析指令时，检索高相关知识片段，并将其注入上下文。为兼顾召回范围与匹配精度，可结合密集向量检索与稀疏关键词检索，支持模型在不同语义层级上调用知识资源^[32]，从而限定其推理边界，形成“语境感知—知识调用—上下文扩展”的增强结构^[33]。

最后，推理约束通过检索增强生成（RAG）与结构化提示机制，对模型输出过程施加约束。具体包括3个方面：一是将高相关知识片段与任务提示联合输入，引导模型显式参考外部知识，提高结论的事实一致性与专业适配性^[34-35]；

二是通过模板化提示与逻辑路径引导，约束模型按照任务路径展开推理，避免跳跃式判断^[36]；三是通过生成超参数调控采样策略，在不改变模型参数的前提下约束输出分布，降低随机性与发散性。相较于模型微调，该路径降低了重复训练成本与过拟合风险，使模型由知识记忆转向知识调用与组织。见表1。

3 基于“S-E-D”本体的智能推理方法

基于此，本文提出以大语言模型为核心的指标体检智能推理方法（图1），融合规划知识库、语境信息与约束推理机制，引导模型沿“表征—解释—决策”链条展开分析，为规划诊断提供可复现、可追溯的支撑。

3.1 表征分析：规划知识库构建

表征分析旨在对指标异常进行语义化识别。通过文本清洗、结构解析和术语规范化，将分散的政策法规、技术标准、规划文本和评估报告转化为可检索、可计算的知识单元（图2）。其中，指标数据按类别、年份、数值、标准和目标值标注，规划文本按“对象—状态/问题—成因—对策”归类，并构建“父段—子段”双层语义结构，以兼顾精准检索与语境恢复，最终通过向量化表征形成规划领域知识库。

3.2 解释归因：知识驱动检索策略

指标异常识别后，分析重点由“是否异常”转向“为何异常”。为避免通用模型脱离空间语境进行泛化解释，引入规划领域知识库，将指标变化置于区位条件、功能属性、发展阶段和制度约束中进行归因。检索机制采用符号约束与语义相似并行策略，前者限定对象范围，后者识别指标异常与规划知识的关联，经权重融合后为模型提供知识支撑。

3.3 决策生成：循证诊断推理范式

在完成异常识别与成因解释后，指标体检进入决策生成阶段，即将诊断结果转化为治理建议。通过结构化Prompt模板，将“指标识别—成因解析—对策生成”嵌入模型推理过程，约束其按照

表1 大语言模型生成超参数配置及其调控机制

Tab.1 Configuration and regulation mechanisms of large language model generation hyperparameters

参数名称	数值	范围	参数说明	取值说明
温度	0.3	0—2	控制生成随机性与多样性	采用较低取值，增强推理稳定性与逻辑一致性
Top P	0.1	0—1	按累积概率截断候选词集合	压缩候选空间，减少低概率表达，提高结果确定性
频率惩罚	0	0—1	降低高频词重复概率	保留术语复现，维持结构化文本表达一致性
存在惩罚	0	0—1	鼓励引入新内容	避免偏离检索知识，确保分析围绕既定信息展开
最大标记	16 384	0—16 384	控制单次生成最大长度	支持长文本输入与多步骤推理

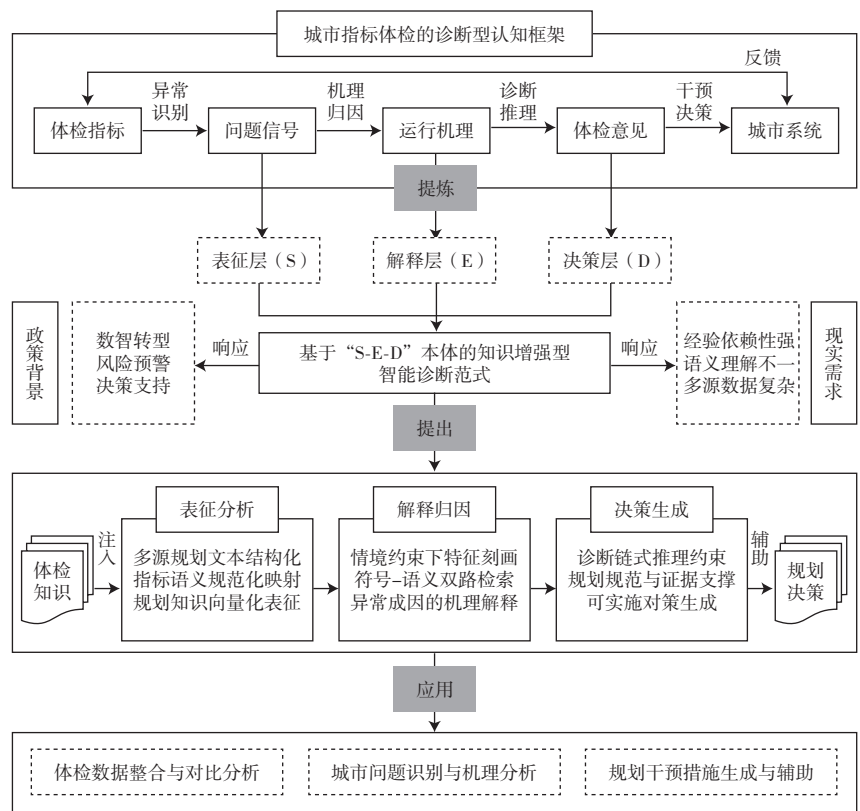


图1 基于“S-E-D”认知闭环的指标体检智能诊断技术路线

Fig.1 Technical framework of intelligent diagnosis for indicator-based urban health assessment based on the "s-e-d" cognitive closed loop

规划诊断逻辑生成内容。同时，结合RAG机制引入规划规范和政策文本，并通过参数控制降低推理跳跃与证据缺失风险，提升策略生成的可追溯性和适用性。

4 实证研究：以南京市为例的实验探索

4.1 研究区概况与指标体系

南京市位于长江下游与江淮之间，是江苏省省会和长三角核心枢纽城市。截至2024年末，常住人口达到957.7万

人，城镇化率为87.3%。自2019年起，南京持续推进城市体检评估工作，建立了连续、结构化的指标体系与体检评估成果（表2）。

在此基础上，本文筛选与国土空间规划密切相关且数据连续可比的基本指标，剔除2项数据不连续指标，形成29项实验指标。同时，整理2019—2023年南京市城市体检评估报告，并整合59项核心政策文件，构建结构化知识库与标准政策知识库（表3）。

4.2 基于“S-E-D”本体的实验验证

依托 DeepSeek 671B 模型开展知识增强型模型实验。结果表明，模型能够结合外部知识库完成指标关联分析，形成“指标识别—原因解释—策略生成”的递进式诊断路径。

4.2.1 指标识别与信息调用

以南京市生态保护红线面积为例，模型首先调用规划知识库，明确其评价基准为“不得减少，保持稳定或适度增加”。随后，模型对 2020—2024 年数据进行比较，识别出该指标由 2020 年的 505.74 km² 降至 2022 年的 496.64 km²，2024 年回升至 498 km²，呈现“先扩张—后回落—小幅修复”的变化特征。结合城乡建设用地面积总体趋稳并小幅回落的背景，判断该指标虽存在空间压力，但仍处于政策底线范围内，实现了对指标状态、变化趋势与规范符合性的综合判读。

4.2.2 问题解释与因果推理

以南京市城市体检问题诊断为例，模型整合安全、生态、土地利用、创新与空间发展等指标，将异常现象置于资源禀赋、发展阶段和制度约束中解释。通过耕地保护、地下水水位、建设用地扩张等指标，识别出自然条件受限和建设用地分配不足对土地与生态安全的影响；通过城乡建设用地增长、工业用地占比下降和闲置土地处置效率偏低等指标，揭示产业结构调整滞后与存量更新机制不足对创新动能的制约。同时，结合人口密度、建筑密度和公共服务指标，指出新城空间扩张快于人口与功能导入，反映出多规协同不足和区级自主决策增强带来的治理约束，从而将指标异常转化为可追溯的因果解释。

4.2.3 策略整合与决策输出

以南京市创新类指标对策设计为例，模型在跨年份、跨维度信息整合基础上生成策略。针对土地与创新发展的关系，模型提出优化闲置土地处置、划定工业用地保障线、完善创新空间布局等措施；针对生态与安全领域，提出加强生态红线保护、湿地修复和应急设施布局等建议。上述策略能够对应指标异常、解释其作用路径，兼顾结构性制约与实施可行性，体现了“发现问题—分析成因—制定策略”的诊断逻辑。

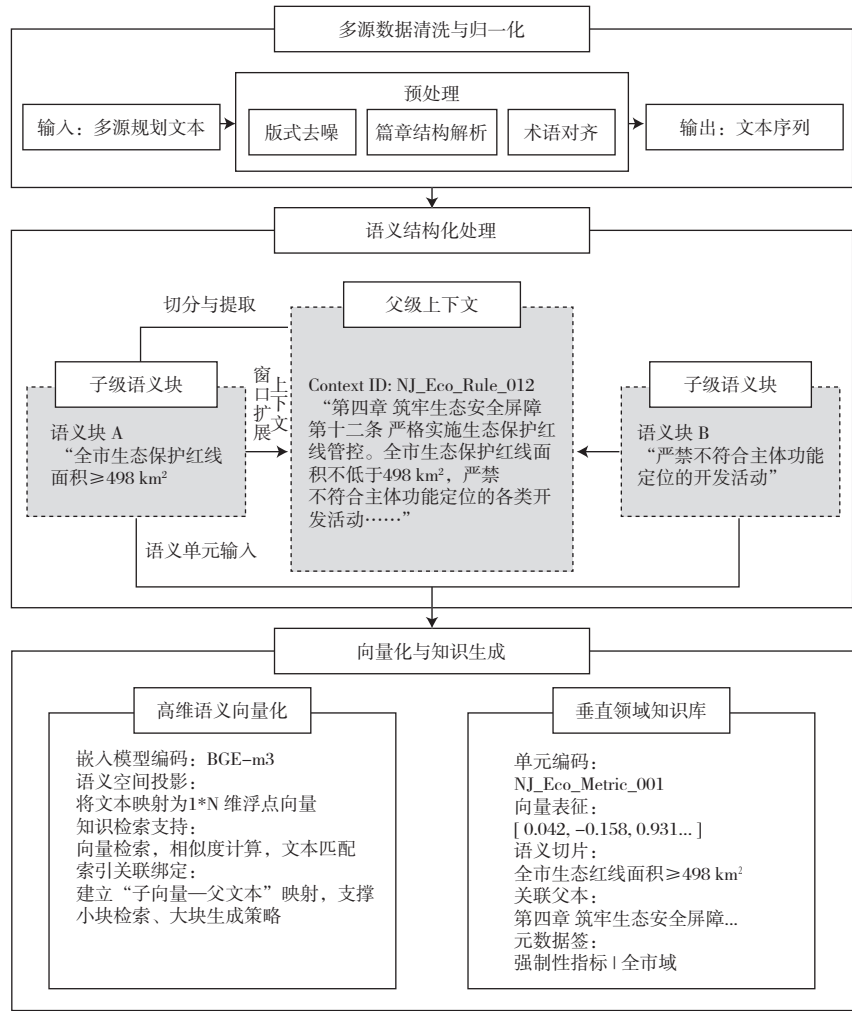


图2 面向指标诊断的规划知识库语义解构与建库流程

Fig.2 Semantic deconstruction and database construction processes of planning knowledge base for indicator-based diagnosis

表2 南京市 2023 年城市体检评估指标体系构成

Tab.2 Composition of the 2023 Nanjing city health examination and evaluation indicator system

一级分类	类别构成			总计 / 个
	二级分类 / 种	基本指标 / 个	推荐指标 / 个	
安全	6	11	12	23
创新	3	6	8	14
协调	3	5	13	18
绿色	3	2	12	14
开放	3	1	11	12
共享	4	6	18	24
总计	22	31	74	105

4.3 模型能力对比分析

4.3.1 实验设计与评测框架

为比较预训练大模型与知识增强模型的能力差异,构建分层评测框架。测试集基于南京市2019—2023年城市体检评估报告及相关规范,包含50道客观题和5道论述题,考察多指标识别、问题归因及综合决策能力,见表4。

实验中,两类模型在相同指标数据条件下作答。主观题采用成对比较问卷,由18名城乡规划专业研究生盲评,评价内容包括回应有效性、逻辑清晰度、错误情况、采纳意愿、评价把握度和总体偏好。

4.3.2 客观题测评结果与分析

客观题结果显示,两类模型差异显著。50道题中,知识增强模型仅错1题,预训练模型错11题,错误主要集中在本地语境理解、规范匹配、数据锚定和跨期推理等方面,反映出二者知识调用机制的差异。

预训练模型在事实锚定类题目(Q1—Q10)中错5题,表现最弱。例如Q6中,模型将南京“研发飞地”误解为制造业向低成本地区转移,未能识别其推动产业链与创新链融合、支撑市级重大科技专项协同落地的本地政策含义,说明其易以通用产业逻辑替代具体制度语境。知识增强模型在该维度全部正确,表明知识库提供了有效语义锚点。

在数据推演与机制诊断维度(Q11—Q20)中,预训练模型错2题,主要表现为归因偏移。如Q18中,其将内涝问题归因于生态调蓄能力下降,而忽略管网不联通这一关键短板,说明缺乏数据约束时模型易生成一般性解释。知识增强模型在该维度保持正确,体现出外部数据对推理方向的约束作用。

在本地约束推演维度(Q21—Q30)中,两类模型整体表现较稳定。预训练模型未出错,说明在约束条件明确时,其通用推理能力仍可发挥;知识增强模型在Q30出现唯一错误,将广州误匹配为杭州/苏州,反映出多源语境调用中的语义对齐风险。

在规范映射与跨期推理维度(Q31—Q50)中,预训练模型错4题,主要表现为制度适用性判断模糊和关键证据识别不足。例如:Q32中,模型未能

表3 指标—知识多层映射关系示例

Tab.3 Examples of multi-level correspondences between indicators and knowledge

典型指标	体检指标类知识	规划现状类知识	政策规范类知识	对策建议类知识
永久基本农田保护面积	表征耕地保护底线与农业空间安全(目标阈值约1240 km ²)	存在阶段性下降、补划压力、后备资源不足及非粮化等问题	依据《基本农田保护条例》,强化特殊保护与占补平衡	推进高标准农田和耕地质量提升,稳定粮食空间与产能
城市内涝积水点数量	表征排水防涝韧性与基础设施承载能力	管网老化,雨洪调蓄和智慧监测能力不足	依据《海绵城市建设评价标准》,完善系统防涝框架	推进海绵城市建设,易涝点整治、管网更新和智慧监测
土地出让收入占预算比例	反映土地财政依赖及结构风险	土地出让收入波动加剧,存量低效用地消化压力较大	依据国发[2017]11号等制度,规范土地出让收支管理	推动城市更新和存量再开发,降低土地财政依赖

表4 分层评测框架

Tab.4 Hierarchical evaluation framework

测评类型	核心考察机制	典型测试情境与样例
事实锚定	本地指标、规划数值和专有名词的识别能力	提取永久基本农田面积、15分钟生活圈覆盖率等具体数值
语境辨析	特定时空与政策语境下的概念区分能力	区分历史街区“微更新”与大拆大建式开发
因果链条	多重约束下的问题归因与路径分析能力	分析潮汐交通、内涝等问题的结构性成因
约束推演	现实情景条件下的本地化决策推演能力	分析建设用地受限与财政压力下的园区扩张路径
政策研判	假设情景下的合规判断与风险识别能力	评估单位GDP地耗上升风险与预警

区分《自然资源督察办法》与《关于加强国土空间规划监督管理的通知》的适用范围;Q45中,未能抓住“新增指标仅占7.6%”这一直接证据。知识增强模型在该维度均正确,说明结构化知识和数据有助于提升规范匹配、证据识别与因果约束能力。

总体而言:预训练模型短板在于本地语境、规范知识和数据锚点不足,易以通用知识替代具体诊断;知识增强模型提升了准确性和推理稳定性,但风险也由知识缺失转向多源信息调用中的语义对齐问题。

4.3.3 主观题测评表现与分析

在客观题已识别出两类模型在知识调用机制上的结构性差异基础上,主观题进一步从回答质量与决策支持角度,对知识增强路径的实际作用进行检验。基于问卷评分结果与偏好选择数据,从整体表现、决策支持价值、群体差异及议题适配性等四个方面展开分析(图3)。

从整体偏好看,知识增强模型获得更高认可。54.5%的受访者倾向于其回答,高于预训练模型的28.9%;16.7%认为二者接近。在显著偏好层面,知识增强模型占比为27.8%,明显高于预训练

模型的7.8%,显示出较稳定的质量优势。

在具体维度上,知识增强模型在“有效回应程度”和“逻辑清晰度”上得分更高,分别为4.43和4.31,高于预训练模型的3.79和3.73,说明本地化数据与语境约束提升了回答的针对性和结构性。在决策支持方面,知识增强模型的采纳意愿得分为4.23,高于预训练模型的3.54;评价把握度得分为4.21,高于预训练模型的3.87,表明其输出更接近可修订的规划分析初稿,也降低了评价者的信息甄别成本。但在“无明显错误”指标上,两类模型得分均低于3分,且差距有限,说明知识增强对复杂推理错误的约束仍不足。

进一步看,知识增强优势具有明显情境依赖。在副城职住比等定量问题中,知识增强模型凭借数据支撑和本地语境锚定获得72.2%的偏好率;而在应急避难盲区识别等偏重空间推演和定性判断的问题中,预训练模型仍有44.4%的偏好率,说明其通用推理与语言组织能力仍具适用性。由此可见,知识增强模型更适合量化、本地化任务,在抽象性较强的问题中优势相对有限。从群体差异看,领域知识越熟悉,受访者越偏好

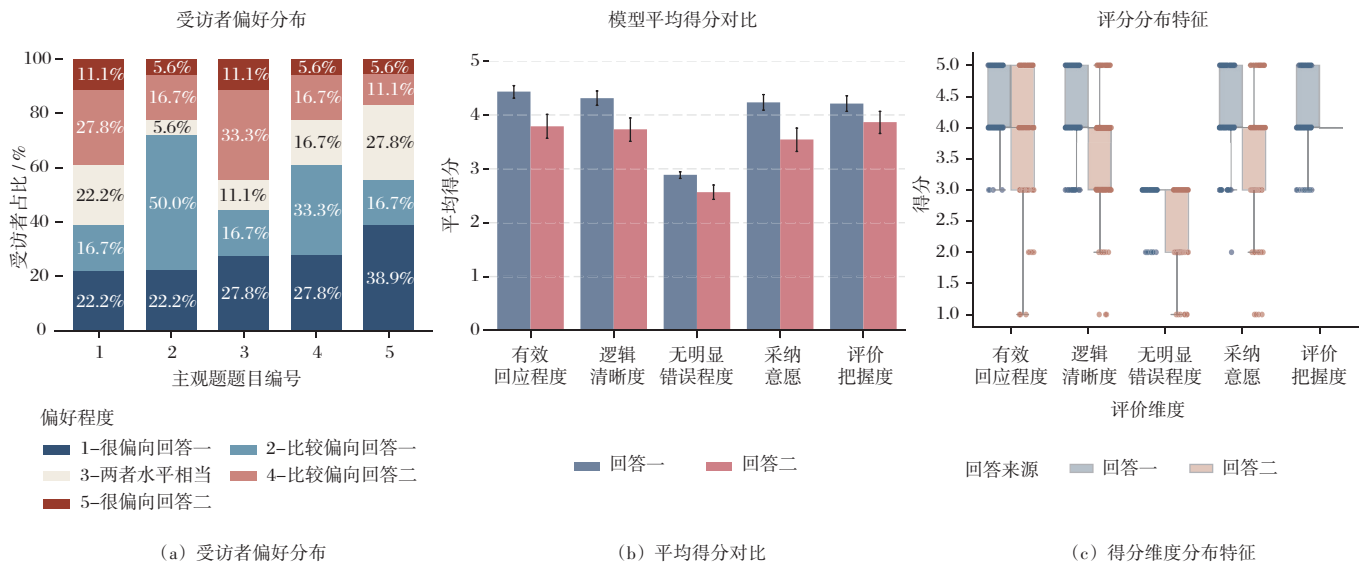


图3 主观题表现对比分析

Fig.3 Comparative analysis of subjective question performance

知识增强模型，高熟悉度群体中的偏好率达到 86.7%。专业评价者更重视指标依据和数据支撑，也更能识别预训练模型中的泛化表达；低熟悉度群体则更易受语言流畅性和文本完整性影响。

总体而言，知识增强模型提升了输出的信息密度、逻辑完整性和决策辅助潜力，但其效果受议题类型与评价者专业水平影响，在弱数据约束或面向泛用户的场景中仍需专家主导的人机协同。

5 结论与展望

城市体检评估中，指标体检的关键不在于数据获取，而在于指标监测、专业知识与规划判断之间的推理衔接。本文将指标体检界定为贯通“表征识别—解释归因—决策生成”的诊断过程，提出“S-E-D”认知本体，并构建知识增强型大语言模型分析方法。该方法通过结构化规划知识与情境约束，将模型生成限定在规划知识、制度逻辑与专业语境之内，支撑跨指标整合、问题归因与策略推演。

南京市指标体检实验表明，知识增强机制能够为模型提供语义锚点和推理约束，使其在本地语境理解、规范知识匹配和跨期逻辑推演中表现出更高稳定

性与准确性。相比预训练模型，知识增强模型生成结果更具数据支撑和分析深度，在回应有效性、逻辑清晰度与决策支持潜力方面获得更优评价，但其作用并非均质呈现。在数据约束明确、地方语境清晰的定量任务中，知识增强优势更明显；而在依赖空间推演或抽象判断的情境下，模型仍较多依赖通用推理，提升有限。同时，专业知识水平较高的评价者更容易识别知识增强带来的信息增益与结构优势，说明其效果具有情境依赖性和认知依赖性。机制上，知识增强并非简单叠加模型能力，而是以外部知识重构推理过程。其价值在于缓解知识缺失、降低信息甄别成本并提升分析稳定性；但在多源信息语义对齐、复杂推理可靠性和表达优化方面仍有局限，尚不能替代专业判断。

本文仍存在不足：一是知识库构建与更新依赖人工或半自动方式，动态适配能力有待提升；二是实证研究基于单一城市案例，跨区域适用性仍需验证；三是测评场景较真实规划决策过程有所简化。未来可从跨模态知识融合、多城市验证和证据驱动生成等方面深化研究。

参考文献

[1] 吴志强, 郭仁忠, 张兵, 等. “国家空间规划

- 系统化建构”学术笔谈[J]. 城市规划学刊, 2024(5): 1-11.
- [2] 赵民, 张翔晨. 城市体检评估的发展历程与高效运作的若干探讨: 基于公共政策过程视角[J]. 城市规划, 2022, 46(8): 65-74.
- [3] 黄伊婧, 张姗姗, 林昀, 等. 城市级国土空间规划实施监测体系的构建思路与实践探索: 以宁波市为例[J]. 自然资源学报, 2024, 39(4): 823-841.
- [4] 杨明, 王吉力, 谷月昆. 改革背景下城市体检评估的运行机制、体系和方法[J]. 上海城市规划, 2022(1): 16-24.
- [5] 中华人民共和国自然资源部. 自然资源部办公厅关于征求《国土空间规划实施监测网络建设技术指南(征求意见稿)》意见的函[EB/OL]. [2025-12-11]. <https://policy.mofcom.gov.cn/claw/clawContent.shtml?id=104237>.
- [6] 蔡玉梅. CSPON 建设的基本思路和总体框架[J]. 中国土地, 2024(5): 9-13.
- [7] 吴志强, 沈清基, 王雅娟, 等. 新规划范式: 概念提出、方法体系与实践路径[J]. 城市规划学刊, 2025(6): 11-17.
- [8] BATTY M. The new science of cities[J]. Building Research & Information, 2010, 38: 123-126.
- [9] 刘晓畅, 吴志强. 人工智能辅助国土空间诊断的理论范式建构与思行方法变革[J]. 城市规划学刊, 2026(1): 12-20.
- [10] CALDARELLI G, ARCAUTE E, BARTHELEMY M, et al. The role of complex-

- ity for digital twins of cities[J]. Nature Computational Science, 2023, 3(5): 374-381.
- [11] 钮心毅, 刘思涵, 桑田, 等. 大模型的专业学习: 构建融入城市空间形态设计知识的图像生成模型[J]. 城市规划学刊, 2025(1): 55-63.
- [12] 吴志强, 严娟, 徐浩文, 等. 城乡规划学科发展年度十大关键议题(2024-2025)[J]. 城市规划学刊, 2024(6): 8-11.
- [13] HU L, LIU Z, ZHAO Z, et al. A survey of knowledge enhanced pre-trained language models[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(4): 1413-1430.
- [14] SASAKI M, WATANABE N, KOMANAKA T. Enhancing contextual understanding of Mistral LLM with external knowledge bases[J/OL]. Research Square, 2024. DOI:10.21203/rs.3.rs-4215447/v1.
- [15] HAN X, ZHANG S, ZHU X. Does ChatGPT provide better advice than robo-advisors?[J]. Finance Research Letters, 2024, 58: 104398.
- [16] BOMMARITO J, BOMMARITO M, KATZ D M, et al. GPT as knowledge worker: a zero-shot evaluation of (AI) CPA capabilities[J/OL]. arXiv preprint, 2023. <https://arxiv.org/abs/2301.04408>.
- [17] Lee J, Stevens N, Han S C. Large language models in finance (finllms) [J]. Neural Computing and Applications, 2025, 37(30): 24853-24867.
- [18] LEE P, BUBECK S, PETRO J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine[J]. New England Journal of Medicine, 2023,388(13): 1233-1239.
- [19] KUNG T H, CHEATHAM M, MEDENILLA A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models [J]. PLoS Digital Health, 2023, 2(2): e0000198.
- [20] GILSON A, SAFRANEK C W, HUANG T, et al. How does ChatGPT perform on the United States medical licensing examination (USMLE)? the implications of large language models for medical education and knowledge assessment[J]. JMIR medical education, 2023,9(1): e45312.
- [21] KASNECI E, SEBLER K, KÜCHEMANN S, et al. ChatGPT for good? on opportunities and challenges of large language models for education[J]. Learning and individual differences, 2023, 103: 102274.
- [22] 林勇军. 基于大语言模型的本地规划知识智能问答系统研究与应用实践[J]. 自然资源信息化, 2024(6): 63-70.
- [23] 王腾, 宋知达, 姜冬睿, 等. 知识图谱驱动的多类型城市体检协同方法研究[J]. 上海城市规划, 2022(1): 25-31.
- [24] 马璇, 郑德高, 孙娟, 等. “规-建-治”一体: 杭州“规划建设评估”的创新实践与方法[J]. 城市规划学刊, 2025(2): 81-88.
- [25] 伍江, 王信, 陈焯, 等. 超大城市城市体检的挑战与上海实践[J]. 城市规划学刊, 2022(4): 28-34.
- [26] 石晓冬, 徐勤政, 曹祺文, 等. 规划治理的新内核: “最初一公里”与“最后一公里”: 以首都北京为例[J]. 城市规划学刊, 2023(4): 18-24.
- [27] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: a survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [28] TANG J, QU M, MEI Q. PTE: predictive text embedding through large-scale heterogeneous text networks[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2015). Sydney, NSW, Australia: ACM, 2015: 1165-1174.
- [29] XIONG L, XIONG C, LI Y, et al. Approximate nearest neighbor negative contrastive learning for dense text retrieval[J/OL]. arXiv preprint, 2020. <https://arxiv.org/abs/2007.00808>.
- [30] XIAO S, LIU Z, ZHANG P, et al. C-Pack: packed resources for general Chinese embeddings[EB/OL]. (2023-09-14) [2026-06-01]. <https://arxiv.org/abs/2309.07597>.
- [31] CHENG Y, ZHANG C, ZHANG Z, et al. Exploring large language model based intelligent agents: definitions, methods, and prospects[J/OL]. arXiv preprint, 2024. <https://arxiv.org/abs/2401.03428>.
- [32] KARPUKHIN V, OGUZ B, MIN S, et al. Dense passage retrieval for open-domain question answering[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020: 6769-6781.
- [33] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.
- [34] ORRÙ G, PIARULLI A, CONVERSANO C, et al. Human-like problem-solving abilities in large language models using ChatGPT[J]. Frontiers in Artificial Intelligence, 2023, 6: 1199350.
- [35] ROSS S I, MARTINEZ F, HOUDE S, et al. The programmer's assistant: conversational interaction with a large language model for software development[C]//Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI 2023). Sydney, NSW, Australia: ACM, 2023: 491-514.
- [36] ZHAO W X, LIU J, REN R, et al. Dense text retrieval based on pretrained language models: a survey[J]. ACM Transactions on Information Systems, 2024, 42(4): 1-60.