Habitat International 176 (2026) 103944

Contents lists available at ScienceDirect

Habitat International

journal homepage: www.elsevier.com/locate/habitatint

Measuring neighbourhood-level perceived residential environment quality (PREQ) from multi-platform user-generated text: An LLM-enhanced framework

Qiyuan Hong [a], Huimin Zhao [a], Yong Jiang [a], Cheng Cheng [b], Ying Long [c,*,1]

[a] School of Architecture, Tsinghua University, Beijing, 100084, China
[b] CSCEC JIUHE Development Group Co., Ltd., Beijing, 100160, China
[c] School of Architecture and Hang Lung Center for Real Estate, Key Laboratory of Eco Planning & Green Building, Ministry of Education, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Perceived residential environment quality (PREQ)
User-generated content (UGC)
Large language models
Geospatial analysis
Neighbourhood assessment
Urban renewal

ABSTRACT

Perceived residential environment quality (PREQ) is used to evaluate neighbourhood liveability, yet survey-based instruments are costly and difficult to update at scale. We propose an LLM-enhanced, PREQ-compatible pipeline to measure neighbourhood perceptions from multi-platform Chinese user-generated text. We analyse 2023-2024 texts from three complementary sources: online reviews, social media posts, and citizen complaints. GPT-4o is prompted to filter PREQ-relevant content, extract evaluation object-content pairs, and score sentiment intensity on a five-point ordinal scale from strongly negative to strongly positive. We compare zero-shot and few-shot prompting with supervised fine-tuning, and benchmark all configurations against two baselines: rule-based dictionary matching and BERT. The fine-tuned model performs best (90.0% object-content extraction accuracy; 92.5% sentiment accuracy), outperforming BERT by 66.6 and 4.5 percentage points, respectively. These outputs support a 45-indicator system mapped to the canonical four-aspect, eleven-scale PREQ framework, with data-driven weights combining mention frequency and sentiment intensity. Nationwide results show a negative-leaning sentiment distribution, with the most salient themes concentrated on environmental health, upkeep and care, and the organisation/accessibility of space, alongside platform-specific emphases. Comparing text salience with an urban-renewal literature corpus reveals attention gaps, especially for operational and management issues. A Beijing case study (2038 communities within the Fourth Ring Road) yields an average PREQ score of 3.40/5, lower in the historic core, and shows a statistically significant but weak positive association with survey-based PREQ, alongside weak correspondence with an objective residential environment index. The framework offers a scalable, resident-centred complement to surveys and objective metrics for neighbourhood diagnostics and renewal prioritisation.

1. Introduction

Perceived residential environment quality (PREQ) at the neighbourhood scale captures residents' overall judgments of built form and layout, social relations, service provision, and contextual ambience (Bonaiuto & Alves, 2012; Bonaiuto et al., 1999). These perceptions are closely linked to residential satisfaction, neighbourhood attachment, health, urban activities, and travel behaviour (Bonaiuto et al., 2004; Liu et al., 2022; Weden et al., 2008). Against the backdrop of rapid urban change, it is necessary to establish a PREQ-compatible indicator system with wide coverage, high-frequency updates, and a clear reflection of residents' salient concerns as an important support for urban governance. For fine-grained governance and urban renewal, stable, comparable, and scalable neighbourhood-level measures of perceptions can guide the identification of local deficits, the evaluation of policy performance, and the prioritisation of interventions such as redevelopment, maintenance, and service upgrading.

PREQ research has largely relied on surveys and interviews, yet inherent limits in spatiotemporal coverage, update frequency, and cost hinder large-scale, continuous monitoring (Fornara et al., 2010; Zhao

* Corresponding author.
E-mail address: ylong@tsinghua.edu.cn (Y. Long).
[1] Present/permanent address: Room 501, New Architecture Building, School of Architecture, Tsinghua University, Beijing 100084, China.

https://doi.org/10.1016/j.habitatint.2026.103944
Received 29 April 2026; Received in revised form 23 July 2026; Accepted 23 July 2026

et al., 2025). Being mostly cross-sectional, such data are susceptible to representativeness and response biases, and thus struggle to capture fine-grained neighbourhood heterogeneity and dynamic change. Fixed-item questionnaires can also constrain content validity, leaving salient, emerging, and locally specific concerns underrepresented; cross-context studies further indicate that scale structure and performance may vary outside Western linguistic and cultural settings, calling for localisation (Dębek & Janda-Dębek, 2015; Tu & Lin, 2008). Moreover, favourable environmental conditions do not necessarily translate into positive lived experience or well-being: the benefits of environmental features depend on accessibility, maintenance, safety, and person–environment fit (Lei et al., 2026), while perceived environmental quality may also diverge from residents' well-being (Zhu et al., 2026). Perceived residential environment indicators have therefore been proposed as a subjective complement to objective city-level quality-of-life indicators, yet a mature approach is still lacking to operationalise such subjective measures at fine spatial scales and update them routinely (Bonaiuto et al., 2015; Cummins et al., 2007).

User-generated content (UGC), by virtue of its high frequency and geo-referenced nature, offers valuable evidence for studying PREQ, helping to offset the limits of survey-based approaches in spatiotemporal coverage and update frequency. Among different forms of UGC, text records users' subjective evaluations of places more directly than uploaded images, and is therefore better aligned with the assessment of perceived quality (Ghermandi et al., 2026). Multi-source user-generated texts, including social media posts, government complaint portals, and service reviews, thus provide rich evidence of everyday experiences and neighbourhood problems. Prior studies have used platform reviews and listings (e.g., Airbnb, Zigbang) and social media streams (e.g., Twitter, Weibo) to extract semantics and sentiment and infer residents' perceptions of their living environments (Hollander et al., 2016; Kweon & Lee, 2023; Wang et al., 2023). In recent years, advances in natural language processing, exemplified by BERT and related transformer architectures, have enabled more reliable entity identification, content extraction, and sentiment intensity measurement from open text; with light human-in-the-loop verification, processing can be scaled and updated at high frequency while retaining interpretability. Accordingly, user-generated text provides a practical basis for producing comparable neighbourhood-level perceptual indicators, and can be benchmarked against objective spatial information and classical PREQ instruments.

Despite this promise, existing methods face several limitations. First, free-form UGC is semantically heterogeneous and noisy. Approaches that rely on predefined lexicons or large labelled sets often struggle with nuanced context, resulting in high false positives during filtering and shallow insights (Lore et al., 2024). Second, most studies analyse a single platform, overlooking cross-platform variation in topic salience and audience composition, which may bias assessments (Alipour et al., 2024). Third, many pipelines prespecify category labels, which hinders the detection of new or unexpected concerns when training data are sparse or absent (Hu et al., 2019; Wang et al., 2023). In addition, neighbourhood-level inference is constrained by spatial attribution uncertainty, because precise coordinates are rarely available and resolving fine-grained, informal place references in short texts remains challenging (Hu et al., 2024).

In response to these limitations, this study combines the canonical PREQ framework of four aspects and eleven scales with an LLM-enhanced pipeline for multi-platform UGC (Bonaiuto et al., 1999, 2003). It pursues three objectives. First, it develops and benchmarks an LLM-based workflow for identifying PREQ-relevant texts, extracting object-content evaluation pairs, scoring sentiment intensity, and assigning texts to neighbourhoods. Second, it constructs a PREQ-compatible indicator and weighting system by deriving resident-expressed issue salience from a nationwide multi-platform corpus, while using a comparison with urban renewal literature to contextualise the added value of resident-derived weights. Third, it applies the resulting framework to Beijing communities and evaluates its correspondence with an interview-based benchmark and objective residential environment indicators.

The remainder of this paper is organised as follows. Section 2 reviews the literature on PREQ, urban UGC, and text-analysis methods. Section 3 describes the data sources, LLM-based processing pipeline, model calibration, indicator construction, weighting strategy, and neighbourhood-scale scoring approach. Section 4 presents the model evaluation, nationwide resident-expressed concern patterns, UGC-derived indicator weights and their literature-side comparison, and the Beijing case results. Section 5 discusses the contribution, implications, and limitations of the study. Section 6 concludes the study.

2. Related work

2.1. PREQ indicator systems and survey-based approaches

Perceived residential environment quality was first systematically developed within the field of environmental and architectural psychology to capture residents' subjective evaluations of the overall quality of their everyday residential surroundings (Bonaiuto et al., 1999). Empirical evidence suggests that subjective assessments of neighbourhood conditions are more strongly associated with self-rated health than objective neighbourhood characteristics (Liu et al., 2022; Weden et al., 2008). Since 1999, Bonaiuto and colleagues have developed a relatively comprehensive set of survey-based Perceived Residential Environment Quality Indicators (PREQIs) (Bonaiuto et al., 2003).

In terms of indicator composition, classic PREQIs distinguish several major aspects and multiple subscales. A common structure is to differentiate four broad aspects: architectural and town-planning space, functional and service provision, social relations and interpersonal environment, and contextual qualities such as environmental health, safety and pace of life (Bonaiuto & Alves, 2012; Canter, 1983). These aspects are further decomposed into eleven subscales, for example building and spatial form, accessibility and road organisation, green areas and open spaces, welfare and recreational services, and commercial and transport services, each typically measured by 3-8 items. As the PREQ framework has diffused, subsequent studies have extended, shortened and cross-validated these instruments in different countries and cities. On the one hand, to address the large number of original items, researchers have used factor-analytic techniques to develop abbreviated or more aggregated versions in order to reduce respondent burden (Fornara et al., 2010; Mao et al., 2022). On the other hand, studies have documented differences in factor structure and indicator salience in countries that are more distant from Western linguistic and cultural contexts (Dębek & Janda-Dębek, 2015; Mao et al., 2015). Some work has focused on translation and adaptation of PREQ dimensions to specific languages and settings (Bonaiuto et al., 2015; Zhao et al., 2025), or on more substantive localisation. For instance, in high-density, mixed-use East Asian cities, researchers have combined interviews and factor analysis to derive new combinations of subscales (Tu & Lin, 2008; Zhang et al., 2020). Recent research further indicates that the relevance of residential-environment features varies with the population group, behavioural outcome, and geographic context under consideration; for intergenerational interaction, walkability, parks and open spaces, safety, and sense of community were rated as particularly important (Zhong et al., 2025). Overall, PREQ provides a widely used and relatively stable conceptual framework, while the concrete indicators are often adjusted to specific contexts and research aims.

At the same time, PREQ studies remain predominantly grounded in one-off surveys or interview-based data collection. Respondents are usually asked to rate 4- to 7-point Likert-type items such as "There is enough greenery in my neighbourhood" or "Public spaces in the neighbourhood are well maintained", and researchers then aggregate these ratings into subscale scores and composite indices. Both full and abbreviated versions of PREQIs are mostly based on cross-sectional data collected from specific samples at specific points in time, with sample

size and spatial coverage constrained by survey costs and limited opportunities for regular updating. The pre-defined questionnaire items inevitably delimit the problem space, and may struggle to capture newly emerging or highly localised residential issues (Song et al., 2026). Moreover, survey-based methods are susceptible to common method bias and social desirability bias (Sampson & Raudenbush, 2004; Sinha et al., 2017), which may lead to biased or distorted conclusions. These limitations indicate that there remains room for improvement and extension at the level of data sources and measurement approaches.

2.2. Urban text and UGC in neighbourhood perception research

In recent years, urban studies have increasingly used user-generated content (UGC) or crowdsourced data from social media, online reviews and government complaint platforms to capture residents' subjective perceptions of urban environments. Existing work shows that social media text, reviews and mobility traces are gradually becoming an important data layer for representing urban perceptions and behaviours (Gao et al., 2021). Compared with traditional surveys, such data are large-scale, temporally continuous and georeferenced, and can therefore be used to track spatial variations in emotions and everyday experiences at city-wide or even national scales (He et al., 2024; Nguyen et al., 2016; Plunz et al., 2019).

At the neighbourhood and residential-environment scale, studies based on urban text can be broadly grouped into three categories. The first category directly uses housing- or community-related texts to describe overall residential satisfaction and neighbourhood perceived quality. For example, Hu et al. (2019) analyse the semantics and sentiment of online neighbourhood reviews to reveal residents' overall satisfaction with their living environment and its determinants; Tan and Guan (2021) overlay Twitter sentiment with housing prices and activity frequency to examine residents' subjective perceptions of high-price locations. The second category focuses on specific perceived dimensions of the neighbourhood environment revealed by UGC. For instance, studies use emotions and activity traces derived from Twitter or Weibo to assess the impacts of urban parks and green spaces on subjective well-being and physical activity (Plunz et al., 2019; Resch et al., 2016); citizen complaint data from 311 or 12345 hotlines are employed to uncover the spatiotemporal patterns of urban public safety and crime issues (He et al., 2024; Zhao et al., 2026); and online reviews from Dianping are mined to extract passengers' perceptions for evaluating urban metro service quality (Dou et al., 2024). The third category uses UGC to identify and interpret neighbourhood change and community social relations. For example, Comber et al. (2024) construct a gentrification lexicon-based index from Twitter content to detect and track neighbourhood-level spatial change; Zahnow et al. (2024) analyse posts and interactions in locality-based Facebook groups to measure the spatial structure of neighbourhood social media networks.

Taken together, these studies demonstrate that multi-platform and multi-type UGC has been widely applied to capture subjective perceptions of environmental quality, safety, mobility convenience and social relations at the neighbourhood scale, providing fine-grained perceptual evidence for subsequent spatial analysis and governance interventions. Nevertheless, most existing work still concentrates on a single topic or platform—for example, only green space, public safety, or physical activity—and typically reports a limited set of stand-alone indicators or sentiment indices. Recent studies have also begun to incorporate LLMs into broader multimodal workflows. Du et al. (2026), for example, combined RAG-LLM analysis of professional documents with text, emotion, and image analysis of multi-platform social media data to compare architectural intention and public perception in industrial heritage regeneration, while Song et al. (2026) used LLMs to extract perceived accessibility, usability, and attractiveness of urban parks. However, these applications remain confined to specific domains and case settings. Compared with multidimensional instruments such as PREQ, the majority of current research has yet to organise different dimensions of neighbourhood perception within a unified conceptual framework or to systematically compare them.

2.3. Text analysis and AI methods for perception extraction

Existing AI-based approaches to urban text analytics can be broadly grouped into three methodological families: (1) keyword- and sentiment-lexicon pipelines that rely on manually specified vocabularies and rule-based pattern matching; (2) supervised machine-learning and deep-learning models for text classification and named entity recognition (NER); and (3) large language model (LLM) workflows that use prompting to perform information extraction, geocoding and sentiment analysis within a unified interface.

Early perception-related studies relied mainly on keyword dictionaries, affective lexicons and simple rules. For online neighbourhood reviews, researchers construct domain-specific lexicons for safety, noise, accessibility and amenities and aggregate sentiment scores at the neighbourhood level to approximate residential satisfaction and perceived residential environment quality (Ho et al., 2024; Hu et al., 2019). Similar strategies have been adopted for public open spaces, for example estimating the “happiness benefit” of urban parks from differences between in-park and out-of-park tweets (Schwartz et al., 2022), or building indices of gentrification pressure from geotagged tweets using lexicons related to renovation, displacement and affordability (Chapple et al., 2022). Lexicon- and rule-based methods are transparent and easy to implement, but they struggle with metaphors, sarcasm, multi-topic sentences and domain shifts across platforms and languages, and often perform poorly outside the context for which they were designed.

Supervised learning and deep neural networks were introduced to overcome the rigidity of hand-crafted lexicons. One line of work formulates perception-related tasks as sequence-labelling problems, using NER models to identify place names, facility types and event descriptors in text, followed by geoparsing pipelines that attach coordinates to extracted entities for spatial analysis (e.g. Hu et al., 2023). Another line applies convolutional, recurrent and transformer-based architectures to classify urban text corpora, such as public service requests and online rental or housing listings, in order to detect emerging urban problems, latent housing-market themes and their spatial patterns (Jiao et al., 2024; Lore et al., 2024). These supervised and neural models substantially improve predictive performance relative to lexicon baselines, but they typically require sizeable labelled datasets and task-specific model design and are usually optimised for a single platform and a single downstream task.

The rapid progress of LLMs has opened up more complex multi-step urban text analysis. First, in information extraction, geo-knowledge-guided GPT models have been used to extract fine-grained location descriptions from disaster-related social media, outperforming standard NER tools in recognising complex place expressions (Hu et al., 2023), and LLMs have been benchmarked against traditional NLP and qualitative coding when decoding public sentiment and themes in upzoning debates, where they improve efficiency and descriptive richness but still require human validation to address bias and reliability concerns (Rong et al., 2025). Second, in structuring unstructured texts, LLM-driven workflows have been developed to parse court judgements into crime records with standardised fields for time, location and case attributes, yielding nationwide spatiotemporal crime datasets that enable fine-grained spatial analysis (Zhang et al., 2025). Third, in sentiment analysis, recent studies apply multilingual transformers and GPT-4 to mobility-related tweets and citizen feedback from virtual urban living labs, showing that LLM-based sentiment analysis can capture nuanced framings of transport and street interventions yet remains sensitive to prompt design (Scerri & Attard, 2025; Tori et al., 2024). Compared with lexicon- and CNN/RNN-based approaches, LLMs provide a unified interface for perception extraction, event detection, summarisation and reasoning, but they also raise new issues around prompt design,

reproducibility and evaluation. Existing applications mostly call general-purpose commercial models (e.g. GPT-3.5, GPT-4) through APIs, focus on a single text source and task, and rarely engage with neighbourhood-scale residential environment quality. Using LLMs for multi-platform UGC and for systematically measuring neighbourhood perceived residential environment quality therefore remains underexplored.

3. Materials and methods

The methodology developed in this study integrates the theoretical dimensions of the PREQ framework with an LLM-enhanced pipeline to measure neighbourhood perceptions from multi-platform UGC. The workflow comprises four stages: (i) data acquisition and preprocessing from social media, service reviews, and complaint portals; (ii) LLM-based processing of the text corpus, including topic relevance judgement, object-content extraction, sentiment scoring, and spatial attribution; (iii) benchmarking of fine-tuned GPT-4o against rule-based and supervised NLP baselines; and (iv) construction of a PREQ system through data-driven weighting and neighbourhood-level aggregation. Fig. 1 illustrates the overall methodological framework.

3.1. Data sources and preprocessing

Textual data were sourced from three major Chinese platforms: the Message Board for Leaders (https://liuyan.people.com.cn/), Weibo (https://weibo.com/), and Dianping (https://www.dianping.com/). The corpora were collected by the authors from publicly accessible web pages. Our dataset covers a two-year time window (2023–2024) and includes 48,091 complaint records from the Message Board for Leaders, 409,103 geotagged posts from Weibo, and 210,226 reviews from Dianping. The three platforms correspond to three major UGC sources commonly used in urban studies: social media, online reviews, and citizen complaint or feedback portals. Within the Chinese context, Weibo, Dianping, and the Message Board for Leaders represent widely used and institutionally relevant platforms in these three categories, respectively. This multi-platform design broadens the observable range of resident-expressed concerns and reduces dependence on the user composition and communication norms of any single platform.

Before LLM-based analysis, we conducted a pre-filtering and cleaning step to reduce obvious noise while retaining potentially relevant texts. We first compiled an inclusion lexicon for residential and neighbourhood contexts to support platform queries and local filtering (e.g., “community/compound”, “housing/apartment”, “building”, “property management”, “parking”, “greenery”), with synonym expansion and fuzzy matching. We then applied exclusion rules to remove texts clearly unrelated to the residential environment, including (i) housing advertisements (price-area templates, rent/sale keywords, and explicit promotional language), (ii) non-residential senses of key terms (e.g., “digital/online community”, “residence permit”, “community hospital”), and (iii) pandemic-related public health or policy discussions not concerning neighbourhood quality. Texts were further normalised by removing emojis and irregular white space and collapsing repeated punctuation, and were deduplicated using the concatenated “title + body” field. Records shorter than 10 Chinese characters or 20 characters in total were discarded. This stage prioritised high recall, while final relevance decisions were made by the downstream LLM.

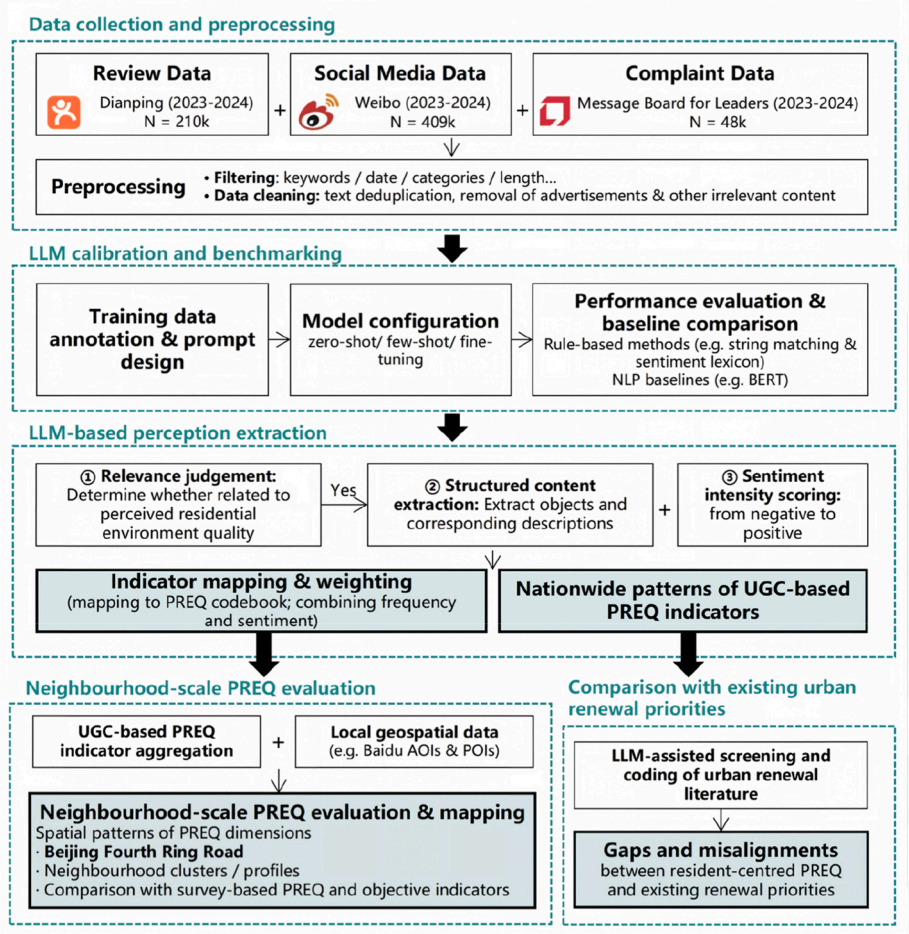

Fig. 1. Research framework.

3.2. LLM-based processing and spatial attribution

After data filtering and preprocessing, we used GPT-4o to process the text corpus through a single-call hierarchical prompt that integrated three nested tasks: topic relevance judgement, object-content extraction, and sentiment scoring. For each input text, the model first judged whether it contained a substantive evaluation of the residential environment. Only when the text was judged as relevant did the model proceed to extract the community name, object-content evaluation pairs, and sentiment orientation. We then assigned the relevant texts in the study area to neighbourhoods for subsequent aggregation and mapping. To improve reproducibility, the system prompt, user prompt template, output rules, and few-shot examples are provided in Supplementary Note S1.

First, topic relevance judgement determined whether a text contained substantive evaluations of the residential environment. This step was essential because the previous rule-based filtering stage prioritised recall and still retained many texts that mentioned residential terms without providing meaningful evaluations of neighbourhood quality. Texts identified as irrelevant were excluded from subsequent object-content extraction, sentiment interpretation, PREQ indicator mapping, and neighbourhood-level aggregation.

Second, for texts judged as relevant, the same prompt extracted object-content evaluation pairs. The evaluated object refers to the residential-environment element being assessed, such as parking, property management, greenery, sanitation, noise, accessibility, or public space. The evaluation content records the corresponding description or judgement expressed in the text. When a text contained multiple evaluated objects, multiple object-content pairs were returned in the same output.

Third, each extracted object-content pair was assigned a sentiment orientation. Sentiment was classified into five categories: strongly negative, negative, neutral, positive, and strongly positive. For subsequent scoring and aggregation, the five categories were encoded as −2, −1, 0, 1, and 2, respectively. Sentiment was judged at the level of each evaluation pair rather than the whole text, allowing mixed evaluations within the same text to be retained.

After semantic processing, we assigned the relevant texts to neighbourhoods. Because the three platforms differed in their spatial anchoring mechanisms, different matching strategies were adopted. For Weibo, posts with valid geographic coordinates were assigned according to spatial location. For Dianping, reviews were linked to neighbourhoods based on the location and name information of their associated POIs. For the Message Board for Leaders, where explicit coordinates were generally unavailable, candidate neighbourhood names and other place cues were identified from the text and matched to a reference list of neighbourhoods. The matching procedure combined exact matching, alias matching, fuzzy matching, and LLM-assisted disambiguation for ambiguous cases. Records that could not be matched with sufficient confidence were excluded from neighbourhood-level aggregation.

3.3. LLM calibration and benchmarking

To calibrate the LLM-based pipeline and assess the contribution of supervised fine-tuning, we designed a controlled benchmark centred on a fine-tuned GPT-4o model, with prompting-only variants and conventional NLP baselines as comparators. As shown in Fig. 2, a total of 1000 texts were manually annotated through stratified quota sampling across the three platforms, including 500 from Weibo, 300 from Dianping, and 200 from the Message Board for Leaders. The annotation was conducted by two urban researchers familiar with residential environment studies and the PREQ framework, covering topic relevance, community name, object–content evaluation pairs, and sentiment orientation. Uncertain or inconsistent cases were reviewed in a second round and resolved through discussion to produce the final consensus labels. The annotated dataset was split into a development set (n = 800) and a held-out test set (n = 200). The development set was used for prompt refinement, in-context example selection, supervised fine-tuning, dictionary and rule design, and baseline training, while the held-out test set was reserved exclusively for final evaluation.

We fine-tuned the base model (gpt-4o-2024-08-06) using the OpenAI fine-tuning API to improve the stability and consistency of structured outputs. Training data were formatted as chat-style JSONL records, each containing a fixed system instruction, the original user text, and the target structured output. To isolate the gain attributable to fine-tuning beyond prompt engineering, we also evaluated two prompting-only configurations of the same base model. In the zero-shot setting, the model was queried without in-context examples. In the few-shot setting, 10 manually labelled examples, selected exclusively from the development set, were embedded in the prompt and held fixed across runs.

For comparison, we implemented two conventional baselines: a rule-based pipeline combining object and content keyword dictionaries with sentiment scoring based on the HowNet sentiment lexicon, and a supervised BERT model trained on the same 800 development entries. All configurations were evaluated on the same 200-entry held-out test set for object-content extraction and sentiment intensity prediction.

3.4. Indicator construction and weighting

Based on the structured outputs generated by the calibrated LLM pipeline, we constructed a UGC-based indicator system aligned with the PREQ framework. Each relevant text was decomposed into one or more structured units consisting of an evaluation object and a content description. Semantically similar descriptions were consolidated into 45 indicators through GPT-assisted coding with manual review and assigned to the corresponding PREQ scales; the resulting indicator codes

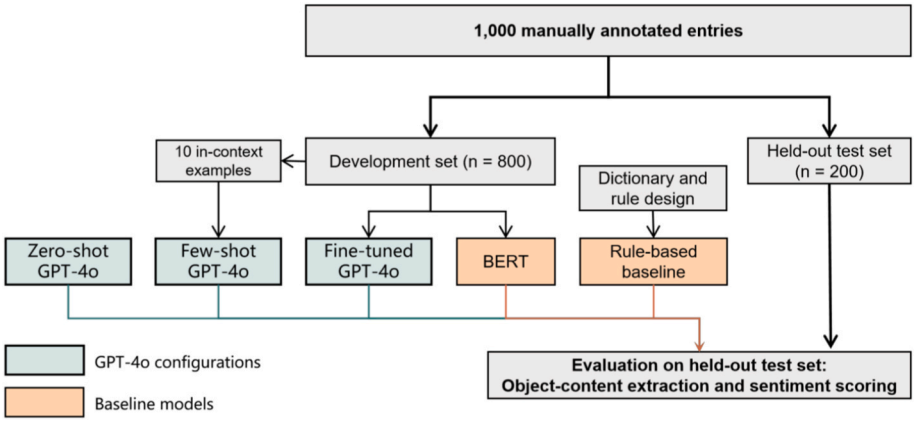

Fig. 2. Design of the calibration and benchmarking experiment.

and scale assignments are shown in Fig. 3.

Weights were derived from the nationwide multi-platform corpus to capture the relative salience of factor items. Specifically, we used two components: item frequency and sentiment intensity. Frequency was log-transformed to reduce the dominance of extremely frequent items, while sentiment intensity was defined as the absolute value of the mean sentiment score for each item. Final item weights were obtained by multiplying the two components and normalising them across all items.

$$F_i' = \log(F_i + 1) \tag{1}$$

$$I_i = |\overline{S}_i| \tag{2}$$

$$W_i = \frac{I_i \cdot F_i'}{\sum_{i=1}^{n}(I_i \cdot F_i')} \tag{3}$$

Where:

F_i: Frequency of item i.

F_i': Logarithmic frequency

$\overline{S}_i$: mean sentiment score of item i

I_i: sentiment intensity (absolute mean sentiment)

W_i: Weight of item i

n: Total number of factor items.

To examine whether the weighting scheme was sensitive to platform composition, we recalculated indicator weights separately within each platform using the same log-frequency × sentiment-intensity formula, normalised the platform-specific weights, and then assigned equal weight to the three platforms. To contextualise the resident-derived weights against the priorities of existing urban renewal research, we also constructed a literature-side salience profile using the same 45 PREQ indicators. Following the screening and reporting logic of PRISMA (Moher et al., 2009), 832 CNKI records on problems in existing residential communities were retrieved, and 81 studies identifying specific and mappable community-quality problems were ultimately included after LLM-assisted manual title-and-abstract and full-text screening. Problems identified in these studies were mapped to the 45 PREQ indicators. For each indicator i, L_i denotes the number of included studies mentioning that indicator, whereas U_i denotes the number of UGC mentions mapped to it in the nationwide corpus. Because the total volumes of the two sources differed, both were normalised into shares, and the indicator-level misalignment index was calculated as follows. The complete literature coding dataset is provided in Supplementary Table S1.

$$L_i^{share} = \frac{L_i}{\sum_{j=1}^{45} L_j}, U_i^{share} = \frac{U_i}{\sum_{j=1}^{45} U_j} \tag{4}$$

$$M_i = U_i^{share} - L_i^{share} \tag{5}$$

3.5. Neighbourhood-scale PREQ scoring

At the neighbourhood level, we first transformed evaluation-pair-level sentiment scores into indicator-level scores. For each PREQ indicator i in neighbourhood j, we averaged the sentiment of all text snippets mapped to that indicator and linearly shifted the result from the original [−2, 2] scale to a 1-5 quality scale, consistent with common Likert-type PREQ ratings. Formally:

$$x_{ij} = \frac{1}{n_{ij}} \sum_{t=1}^{n_{ij}} S_{ijt} + 3, S_{ijt} \in [-2, 2], x_{ij} \in [1, 5] \tag{6}$$

When no text in neighbourhood j was mapped to indicator i ($n_{ij} = 0$), we assigned the scale midpoint $x_{ij} = 3$ as a non-directional imputation value to retain a complete indicator matrix. This assignment does not assume that the absence of UGC represents genuinely neutral perceived quality. We therefore tested two alternative treatments: excluding missing indicators from neighbourhood-level aggregation and renormalising the weights of the observed indicators, and replacing missing values with indicator-specific national mean scores.

For each PREQ scale k, we then computed a neighbourhood-level scale score as a weighted aggregation of its constituent indicators.

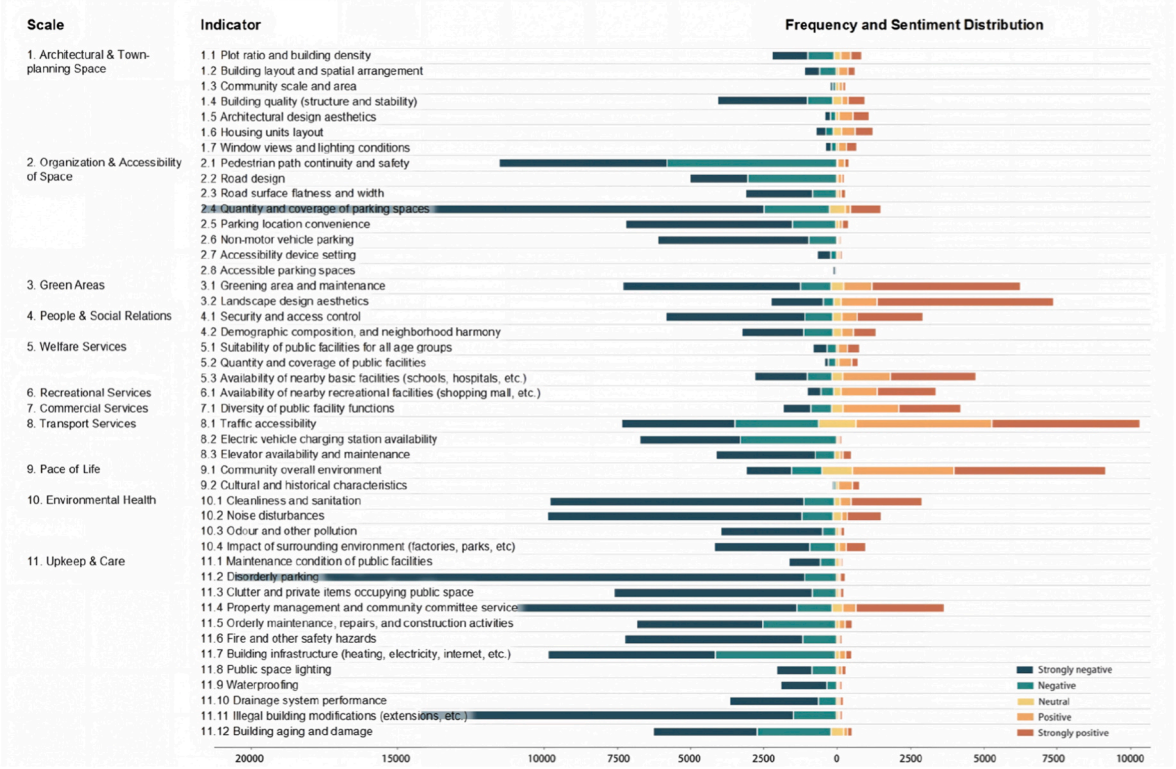

Fig. 3. Nationwide frequency and sentiment distribution of UGC-based PREQ indicators.

Using the UGC-based weights W_i derived in Section 3.4, the score for scale k in neighbourhood j was defined as :

$$Score_{kj} = \frac{\sum_{i \in I_k} W_i x_{ij}}{\sum_{i \in I_k} W_i} \quad (7)$$

Where:

S_{ijt}: sentiment score of the t-th extracted evaluation pair in neighbourhood j mapped to indicator i (S_{ijt} in [−2, 2]).

n_{ij}: number of extracted evaluation pairs in neighbourhood j mapped to indicator i.

x_{ij}: indicator-level PREQ score for indicator i in neighbourhood j on a 1-5 scale.

I_k: set of indicators that belong to PREQ scale k.

W_i: nationwide UGC-based weight of indicator i.

$Score_{kj}$: neighbourhood-level score of PREQ scale k in neighbourhood j on a 1-5 scale.

4. Results

4.1. Model performance

Table 1 summarises the performance of all model configurations on the 200-sample test set for both object-content extraction and sentiment scoring. For object-content extraction, the fine-tuned GPT-4o clearly performs best, achieving 90.0% accuracy, followed by the few-shot setting (84.2%), while the zero-shot setting is substantially lower. Both conventional baselines perform poorly on this task. For sentiment scoring, the fine-tuned GPT-4o again achieves the highest accuracy (92.5%), with BERT (88.0%) and the few-shot setting (86.9%) also performing well. Overall, the fine-tuned GPT-4o yields the strongest results on both tasks, with its advantage being especially pronounced for object-content extraction.

4.2. Nationwide frequency and sentiment distribution

Within the nationwide three-platform corpus, negative and strongly negative evaluations account for the largest share of the extracted sentiment distribution, whereas neutral and positive evaluations account for smaller shares (Fig. 3).

At the scale level, Environmental health (Scale 10), Upkeep and care (Scale 11), and Organisation and accessibility of space (Scale 2) receive the highest numbers of mentions in the nationwide corpus. By contrast, Architectural and town-planning space (Scale 1) and Green areas (Scale 3) show lower salience in the three-platform UGC corpus, while service-related scales occupy an intermediate position.

At the indicator level, the most frequently mentioned items are concentrated in parking, sanitation, noise, odour, and property management, and these indicators are predominantly associated with negative sentiment. Indicators related to structural safety and technical infrastructure are mentioned less often, while those concerning green areas and pace of life show more mixed sentiment profiles.

Additionally, the study identified platform-specific differences in PREQ categories. On Dianping, discussions most frequently concerned Transportation and Roads (19%) and the Community Overall Environment indicator (15%); on Weibo, texts were dominated by issues in Neighbourhood and Management (21%); whereas on Message Board for Leaders, Parking-related indicators accounted for 22% of all mentions. The sentiment distribution also varies by platform, with review data containing more positive evaluations, social media posts showing a more negative skew, and complaint records being dominated by negative reports.

Table 1
Performance on object-content extraction and sentiment scoring.

Model/Method	Zero-shot GPT-4o	Few-shot GPT-4o	Fine-tuned GPT-4o	String matching	BERT
Accuracy of object-content extraction	52.10%	84.20%	90.00%	15.90%	23.40%
Accuracy of sentiment scoring	71.70%	86.90%	92.50%	25.00%	88.00%

4.3. UGC-derived indicator weights and comparison with literature-side salience

4.3.1. Estimated indicator weights

Using the weighting scheme described in Section 3.4, the estimated indicator weights were normalised to sum to 100 and used as fixed parameters in subsequent aggregation. As shown in Table 2, the highest weights are concentrated in everyday management and environmental nuisance issues, particularly disorderly parking, parking supply and convenience, property management services, cleanliness and sanitation, noise, and the occupation of public space by private items. Indicators related to greening maintenance and landscape aesthetics also receive relatively high weights. Overall, operational/maintenance and environmental nuisance items are more prominent than built-form or aesthetic items.

The platform-balanced sensitivity check further supports the overall robustness of the weighting scheme. The original and platform-balanced weights remained strongly correlated, with a Pearson correlation of 0.850 and a Spearman rank correlation of 0.786, indicating substantial consistency in the overall weighting structure, although several indicator-level rank shifts remained. Detailed results are provided in Supplementary Table S2.

4.3.2. Comparison with urban renewal literature

Using the literature-side comparison described in Section 3.4, the Spearman correlation between literature-side and UGC-side indicator shares is weak ($\rho = 0.289$, $p = 0.0539$, $n = 45$), indicating limited alignment in issue salience between the two sources. As shown in Fig. 4, several indicators lie near the diagonal, but substantial deviations

Table 2
Top-15 UGC-derived indicator weights.

PREQ dimension	PREQ scale	Indicator	Weight (%)
Contextual	Upkeep & Care	Disorderly parking	3.11
Spatial	Organisation & Accessibility of Space	Quantity and coverage of parking spaces	3.00
Contextual	Upkeep & Care	Illegal building modifications (extensions, etc.)	2.92
Contextual	Upkeep & Care	Property management and community committee service	2.85
Contextual	Environmental Health	Cleanliness and sanitation	2.84
Contextual	Environmental Health	Noise disturbances	2.78
Spatial	Green Areas	Greening area and maintenance	2.75
Contextual	Upkeep & Care	Clutter and private items occupying public space	2.71
Spatial	Green Areas	Landscape design aesthetics	2.65
Contextual	Upkeep & Care	Fire and other safety hazards	2.63
Spatial	Organisation & Accessibility of Space	Accessibility device setting	2.61
Spatial	Organisation & Accessibility of Space	Non-motor vehicle parking	2.59
Social	People & Social Relations	Security and access control	2.59
Spatial	Organisation & Accessibility of Space	Parking location convenience	2.55
Contextual	Environmental Health	Odour and other pollution	2.49

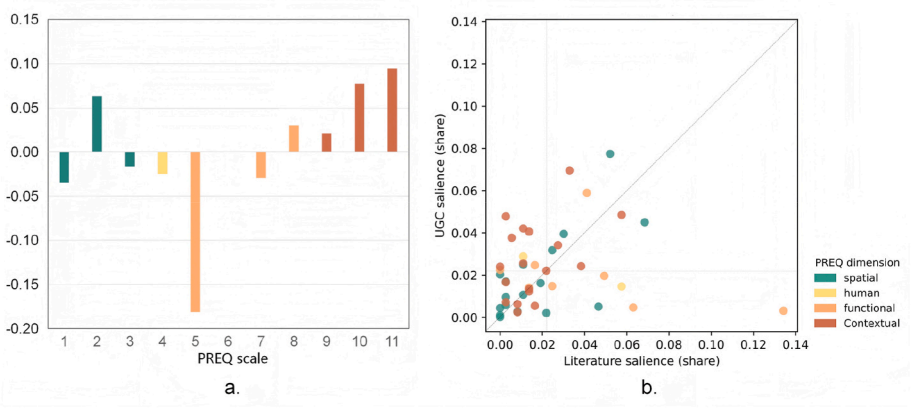

Fig. 4. Misalignment between UGC-based and literature-based salience of PREQ indicators: (a) Scale-level aggregation of the misalignment index M_i; (b) indicator-level comparison of literature and UGC salience shares.

remain across much of the indicator set.

The main mismatches are substantive rather than random. Indicators related to everyday management and environmental disturbances, especially those under Environmental Health and Upkeep & Care, tend to be more salient in residents' UGC. By contrast, indicators associated with spatial form, structural quality, public facilities, and infrastructure are generally more salient in the literature sample. At the scale level, this pattern appears as greater resident-side salience for operation- and environment-related issues, versus greater literature-side salience for spatial and facility-oriented concerns. Typical resident-salient items include noise disturbances, disorderly parking, and maintenance-related problems, whereas building quality, public facilities, and transport infrastructure are relatively more literature-salient.

4.4. Evaluation results for Beijing

Within the Fourth Ring Road of Beijing, 2038 of 6198 residential communities could be matched with sufficient textual signals to derive useable UGC evidence, representing approximately one third of the communities in the study area (Fig. 5). A comparison between matched and unmatched communities showed close balance across the available objective residential environment indicators: the mean absolute standardised mean difference was 0.036, and the maximum absolute SMD was 0.074. Spatial coverage nevertheless varied across the study area, with match rates of 14.8% within the Second Ring Road and 44.7% between the Second and Fourth Ring Roads.

Among the matched communities, the average PREQ score is 3.40 on the 1-5 scale, indicating an overall perceived quality slightly above the midpoint. Spatial heterogeneity is evident: communities inside the Second Ring Road score lower on average than those outside (3.368 vs. 3.419). Across PREQ scales, Recreational Services and Commercial Services perform best, whereas Upkeep & Care scores lowest, with parking-related complaints remaining a prominent source of negative evaluations.

To assess consistency with a classical subjective measure, we compared the community-level UGC-based PREQ with the interview-based residential environment quality scores reported by Zhang et al. (2020). For comparability, both datasets were aggregated to 2 km × 2 km grid cells. The two measures show a statistically significant but weak positive association ($n = 440$; Pearson $r = 0.209$, 95% CI: 0.12–0.30, $p < 0.0001$; Spearman $\rho = 0.267$, 95% CI: 0.17–0.37, $p < 0.0001$). They also reveal similar broad spatial gradients, including relatively higher scores between the Second and Fourth Ring Roads and lower scores in the historic core and peripheral fringe.

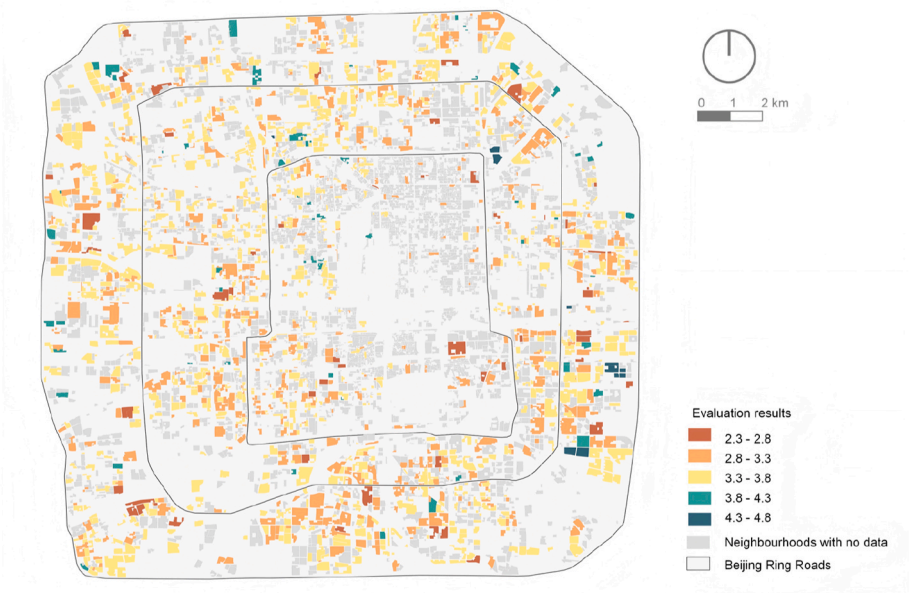

Fig. 5. Beijing Evaluation results.

We further compared the community-level PREQ with the objective index reported by Zhao and Long (2022). The correlation is positive but weak ($\rho = 0.15$), indicating limited correspondence between UGC-based PREQ and objective morphological and administrative measures.

The neighbourhood PREQ scores were highly consistent with the baseline results when missing indicators were excluded and the weights of observed indicators were renormalised (Pearson $r = 0.963$), and remained substantially correlated when missing values were replaced with indicator-specific national mean scores ($r = 0.749$).

5. Discussion

5.1. LLM-enabled PREQ extraction from multi-platform UGC

This study examined whether LLM-assisted analysis of multi-platform UGC can support scalable, PREQ-compatible evaluation of residential environments in Chinese cities. The results show that a fine-tuned GPT-4o pipeline can transform noisy and heterogeneous urban texts into structured object-content pairs, sentiment scores, PREQ indicators, and neighbourhood-level measures. The fine-tuned model outperformed zero-shot prompting, few-shot prompting, rule-based matching, and BERT, with the largest gain observed in object-content extraction. This indicates that supervised fine-tuning is especially valuable for converting open-ended residential narratives into standardised evaluative units, extending recent LLM-based urban text applications from location extraction, public sentiment coding, and structured event datasets to PREQ-compatible residential environment evaluation (Hu et al., 2023; Rong et al., 2025; Zhang et al., 2025).

The methodological contribution lies in embedding PREQ categories into the full processing chain, from relevance judgement and semantic extraction to scoring, weighting, and spatial aggregation. Existing urban text studies often focus on single platforms, single topics, or stand-alone sentiment indicators. By contrast, the present framework links platform-mediated residential narratives to a multidimensional PREQ structure, allowing resident-expressed concerns to be compared across platforms, indicators, and neighbourhoods. In addition, as the workflow is not tied to a specific model architecture, future applications could evaluate locally deployed open-source LLMs as lower-cost alternatives to commercial APIs.

5.2. Resident-expressed salience and PREQ-compatible weighting

At the national scale, the generated PREQ profiles reveal that residents' online narratives are predominantly problem-oriented, focusing on immediate operational and contextual disruptions rather than static physical attributes. Negative sentiments prevail across the three platforms, particularly within the scales of Environmental Health, Upkeep & Care, and Organisation & Accessibility of Space, where parking congestion, noise disturbances, sanitation problems, and property management failures form the most visible sources of dissatisfaction. This distribution differs from classical survey-based PREQIs evidence, in which dimensions such as building aesthetics and green areas have been reported as important predictors of neighbourhood attachment and residential satisfaction (Bonaiuto et al., 1999, 2015). The contrast suggests that predictive importance in structured surveys and salience in spontaneous platform-mediated expression capture different aspects of perceived residential environment quality. Features such as greenery and architectural appearance may shape residential satisfaction and attachment, while everyday management failures and environmental nuisances are more likely to trigger explicit online expression. This interpretation is also consistent with comparative PREQ research showing that residents' evaluative priorities vary across urban contexts. In high-density and mixed-use environments, for example, ground-floor access, transportation and commercial services, environmental health, and facility management emerged as salient components of residents' evaluative structure (Tu & Lin, 2008). The present results similarly show that, in large-scale Chinese UGC, daily operational and management issues become especially visible. These “low-status” topics, often treated as routine management details in formal urban renewal agendas, occupy a central place in residents' expressed experience. The systematic misalignment between these expressed concerns and the focus of existing urban renewal literature suggests that current evaluation systems may under-represent the factors that shape everyday comfort.

The weighting scheme operationalises this resident-expressed salience by combining mention frequency and sentiment intensity. The platform-balanced sensitivity check supports the overall stability of this scheme: the original and platform-balanced weights remained strongly correlated, although several indicator-level rank shifts were observed. The comparison with urban renewal literature further shows that resident-expressed concerns and academic research priorities do not fully align: management and environmental nuisance issues are more prominent in UGC, whereas spatial form, facilities, and infrastructure are more prominent in the literature sample. This misalignment may reflect the different production logics of the two sources: UGC records immediate and repeatedly experienced disruptions, whereas urban renewal research is more often organised around planning, design, infrastructure, and project-based interventions. As a result, evaluation systems oriented toward renewal prioritisation may under-represent factors that shape everyday comfort.

5.3. Neighbourhood-scale PREQ evaluation and external comparison

The Beijing case study demonstrates how the proposed framework can be applied to neighbourhood-scale PREQ evaluation. Among the 2038 matched communities, lower PREQ scores were concentrated in the historic core, while relatively higher scores appeared between the Second and Fourth Ring Roads. Scale-level results similarly indicate better performance in Recreational Services and Commercial Services and weaker performance in Upkeep & Care, with parking-related complaints continuing to shape negative evaluations. Although UGC coverage was spatially uneven, matched and unmatched communities were highly comparable in terms of the available objective residential environment indicators. Thus, incomplete coverage did not translate into an observable imbalance in objective environmental quality, although the lower coverage of the historic core should still be considered when interpreting the spatial gradient.

The comparison with the interview-based benchmark from Beijing (Zhang et al., 2020) shows a weak positive association between the two measures. This association indicates that UGC-based PREQ captures some broad spatial gradients also reflected in interview-based evidence, while retaining sensitivity to platform-visible management failures, environmental nuisances, and operational issues. The divergence between the two measures may partly reflect differences in survey and spatial measurement. Survey-based PREQ is derived from structured self-report, where social desirability and evaluative framing can affect responses in environmental psychology research (Vesely & Klöckner, 2020). In addition, the two datasets are not fully aligned in time, and their spatial units and matching procedures differ; such differences can reduce correspondence when neighbourhood-level indicators are compared across data sources (Guo & Bhat, 2004). The weak correlation with the objective residential environment index developed by Zhao and Long (2022) further indicates that subjective residential experience cannot be fully inferred from morphology or administrative statistics alone. This is consistent with recent evidence that perceived environmental quality, objective spatial conditions, and residents' well-being may not move in parallel (Zhu et al., 2026).

The sensitivity analysis further indicates that the principal neighbourhood scoring pattern is not dependent on midpoint imputation. The high consistency obtained when missing indicators were explicitly excluded and the observed weights were renormalised supports the robustness of the baseline scoring approach. The lower, although still substantial, correspondence under national-mean imputation may

partly reflect cross-regional differences in residential conditions and UGC expression patterns embedded in the nationwide reference values.

These findings clarify the role of UGC-based PREQ in neighbourhood diagnostics. It is most useful as a scalable, frequently updatable, resident-centred layer that complements survey instruments and objective spatial indicators. For planning practice, this layer can help identify specific management deficits, locate communities with concentrated negative experience, and track changes in resident-expressed concerns during neighbourhood renewal and city health check processes.

5.4. Limitations

Despite these contributions, our study has several important limitations that should be acknowledged. First, UGC coverage is inherently selective and uneven across users, neighbourhoods, and indicators, which may result in incomplete community coverage and attenuated between-neighbourhood variation where textual evidence is sparse (Ghermandi et al., 2026; Hollander et al., 2016). These coverage constraints highlight the complementary role of UGC-based PREQ alongside survey-based and objective indicators. Second, the analysis is essentially cross-sectional and confined to a limited time window, which constrains our ability to identify causal effects of specific interventions or to track longer-term trajectories in residential environment quality. Third, prompt design and manual coding involve researcher judgement and may affect extraction outputs, particularly for borderline cases. Transparent reporting and full prompt disclosure are therefore important for reproducibility (Rong et al., 2025). Fourth, the PREQ framework and indicator set are grounded in Chinese-language UGC and the institutional context of Chinese cities, so their meanings and relative weights cannot be assumed to transfer directly to other linguistic and urban settings.

Future work should therefore combine UGC with probabilistic surveys and administrative indicators, extend the framework to longitudinal and comparative designs, and systematically test and recalibrate the PREQ schema across different cities and countries.

6. Conclusions

This study develops a scalable pathway for assessing residential environment quality in Chinese cities by integrating multi-platform user-generated content (UGC) with large language models (LLMs) within the PREQ framework. The approach combines topic relevance judgement, object-content extraction, sentiment scoring, indicator construction, UGC-based weighting, and neighbourhood-level aggregation, extending PREQ measurement from conventional survey settings to large-scale digital text data. Its main contribution lies in providing an operational framework through which resident perceptions can be systematically extracted, structured, and translated into comparable neighbourhood-quality measures.

Empirically, the proposed framework yields stable and interpretable patterns across scales. At the national level, it identifies clear concentrations of resident concern in environmental health, upkeep and care, and the organisation and accessibility of space. In Beijing, the resulting PREQ scores show a weak positive association with an interview-based benchmark and weak correspondence with objective indicators, suggesting that the framework captures a meaningful layer of perceived residential environment quality that is not fully reflected in morphology or administrative statistics. Overall, this study shows that LLM-enabled UGC analysis can extend existing PREQ research with a scalable and resident-centred measurement pathway.

Ethical approval

Our institution does not require ethical approval for reporting individual cases or case series.

CRediT authorship contribution statement

Qiyuan Hong: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Huimin Zhao:** Data curation, Formal analysis, Investigation, Validation. **Yong Jiang:** Resources, Supervision. **Cheng Cheng:** Data curation, Investigation. **Ying Long:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China [Grant No. 2023YFC3805400] and Program of Tsinghua University (School of Architecture) - CSCEC JIUHE Development Group Co. Ltd. Joint Research Center (2025A1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.habitatint.2026.103944.

References

Alipour, S., Galeazzi, A., Sangiorgio, E., Avalle, M., Bojic, L., Cinelli, M., & Quattrociocchi, W. (2024). Cross-platform social dynamics: An analysis of ChatGPT and COVID-19 vaccine conversations. *Scientific Reports, 14*(1), Article 2789.

Bonaiuto, M., Aiello, A., Perugini, M., Bonnes, M., & Ercolani, A. P. (1999). Multidimensional perception of residential environment quality and neighborhood attachment in the urban environment. *Journal of Environmental Psychology, 19*(4), 331–352.

Bonaiuto, M., Bonnes, M., & Continisio, M. (2004). Neighborhood evaluation within a multi-place perspective on urban activities. *Environment and Behavior, 36*, 41–69. https://doi.org/10.1177/0013916503251444

Bonaiuto, M., Fornara, F., Ariccio, S., Ganucci Cancellieri, U., & Rahimi, L. (2015). Perceived residential environment quality indicators (PREQIs) relevance for UN-HABITAT City Prosperity Index (CPI). *Habitat International, 45*, 53–63. https://doi.org/10.1016/j.habitatint.2014.06.015

Bonaiuto, M., Fornara, F., & Bonnes, M. (2003). Indexes of perceived residential environment quality and neighbourhood attachment in urban environments: A confirmation study on the city of Rome. *Landscape and Urban Planning, 65*(1–2), 41–52.

Bonaiuto, M., & Alves, S. (2012). Residential places and neighbourhoods: Toward healthy life, social integration, and reputable residence. In S. Clayton (Ed.), *The Oxford handbook of environmental and conservation psychology* (pp. 221–247). New York: Oxford University Press.

Canter, D. (1983). The purposive evaluation of places: A facet approach. *Environment and Behavior, 15*(6), 659–698. https://doi.org/10.1177/0013916583156001

Chapple, K., Poorthuis, A., Zook, M., & Phillips, E. (2022). Monitoring streets through tweets: Using user-generated geographic information to predict gentrification and displacement. *Environment and Planning B: Urban Analytics and City Science, 49*(2), 704–721. https://doi.org/10.1177/23998083211025309

Comber, A., Kieu, M., Bui, Q.-T., & Malleson, N. (2024). Using social media data to identify neighbourhood change. *AGILE: GIScience Series, 5*, Article 20. https://doi.org/10.5194/agile-giss-5-20-2024

Cummins, S., Curtis, S., Diez Roux, A. V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine, 65*(9), 1825–1838. https://doi.org/10.1016/j.socscimed.2007.05.036

Dębek, M., & Janda-Dębek, B. (2015). Perceived residential environment quality and neighborhood attachment (PREQ & NA) indicators by Marino Bonaiuto, Ferdinando Fornara, and Mirilia Bonnes: Polish adaptation. *Polish Journal of Applied Psychology, 13*(2), 111–162. https://doi.org/10.1515/pjap-2015-0032

Dou, M., Gu, Y., & Gong, J. (2024). How do people perceive the quality of urban transport service? New insights from online reviews of Shanghai metro system. *Journal of Urban Management, 13*(4), 705–719.

Du, L., Yang, C., & Yan, J. (2026). Decoding the continuum of architectural intention and public perception in industrial heritage regeneration: A multimodal social media data analysis. *Habitat International, 170*, Article 103754. https://doi.org/10.1016/j.habitatint.2026.103754

Fornara, F., Bonaiuto, M., & Bonnes, M. (2010). Cross-validation of abbreviated perceived residential environment quality (PREQ) and neighborhood attachment

(NA) indicators. *Environment and Behavior, 42*(2), 171–196. https://doi.org/10.1177/0013916508330998

Gao, S., Liu, Y., Kang, Y., & Zhang, F. (2021). User-generated content: A promising data source for urban informatics. In W. Shi, M. F. Goodchild, M. Batty, M.-P. Kwan, & A. Zhang (Eds.), *Urban informatics (The urban Book Series* (pp. 503–522). Springer. https://doi.org/10.1007/978-981-15-8983-6_28.

Ghermandi, A., Depietri, Y., & Orenstein, D. E. (2026). Human-nature interactions through a digital prism: Heterogeneity in user socio-demographics and content across social media platforms. *Landscape and Urban Planning, 268*, Article 105568.

Guo, J. Y., & Bhat, C. R. (2004). Modifiable areal units: Problem or perception in modeling of residential location choice? *Transportation Research Record: Journal of the Transportation Research Board, 1898*(1), 138–147. https://doi.org/10.3141/1898-17

He, J., Zhang, W., & Yang, M. (2024). The spatial and temporal characteristics of urban public safety under the residents' complaints: Evidence from 12345 data in Beijing, China. *Journal of Urban Management, 13*(2), 217–231.

Ho, E., Schneider, M., Somanath, S., Yu, Y., & Thuvander, L. (2024). Sentiment and semantic analysis: Urban quality inference using machine learning algorithms. *iScience, 27*(7), Article 110192. https://doi.org/10.1016/j.isci.2024.110192

Hollander, J. B., Graves, E., Renski, H., Foster-Karim, C., Wiley, A., & Das, D. (2016). *Urban social listening: Potential and pitfalls for using microblogging data in studying cities*. Palgrave Pivot. https://doi.org/10.1057/978-1-137-59491-4

Hu, X., Elßner, T., Zheng, S., Serere, H. N., Kersten, J., Klan, F., & Qiu, Q. (2024). DLRGeoTweet: A comprehensive social media geocoding corpus featuring fine-grained places. *Information Processing & Management, 61*(4), Article 103742. https://doi.org/10.1016/j.ipm.2024.103742

Hu, Y., Deng, C., & Zhou, Z. (2019). A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers, 109*(4), 1052–1073.

Hu, Y., Mai, G., Cundy, C., Choi, K., Lao, N., Liu, W., Lakhanpal, G., Zhou, R. Z., & Joseph, K. (2023). Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science, 37*(11), 2289–2318.

Jiao, J., Jin, Y., & Yang, R. (2024). An approach to exploring the spatial distribution and influencing factors of urban problems based on land use types. *Sustainable Cities and Society, 104*, Article 105321.

Kweon, J., & Lee, S. (2023). Analysis of resident's satisfaction and its determining factors on residential environment: Using Zigbang's apartment review bigdata and deeplearning-based BERT model. *Journal of the Korean Regional Science Association, 39*(2), 47–61. https://doi.org/10.22669/krsa.2023.39.2.047

Lei, D., Liang, W., Chen, W., Jian, I. Y., & Mo, K. H. (2026). The impact of microscale environmental features on older adults' recreational activities in shared spaces of high-density residential neighbourhoods: A recreationist-environment fit perspective. *Habitat International, 170*, Article 103738. https://doi.org/10.1016/j.habitatint.2026.103738

Liu, J., Yang, L., Xiao, L., & Tao, Z. (2022). Perceived neighborhood environment impacts on health behavior, multi-dimensional health, and life satisfaction. *Frontiers in Public Health, 10*, Article 850923.

Lore, M., Harten, J. G., & Boeing, G. (2024). A hybrid deep learning method for identifying topics in large-scale urban text data: Benefits and trade-offs. *Computers, Environment and Urban Systems, 111*, Article 102131.

Mao, Y., Fornara, F., Manca, S., Bonnes, M., & Bonaiuto, M. (2015). Perceived Residential environment quality indicators and neighborhood attachment: A confirmation study on a Chinese sample in Chongqing. *PsyCh Journal, 4*(3), 123–137. https://doi.org/10.1002/pchj.90

Mao, Y., Luo, X., Guo, S., Xie, M., Zhou, J., Huang, R., & Zhang, Z. (2022). Validation of the abbreviated indicators of perceived residential environment quality and neighborhood attachment in China. *Frontiers in Public Health, 10*, Article 925651.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nguyen, Q. C., Li, D., Meng, H. W., Kath, S., Nsoesie, E., Li, F., & Wen, M. (2016). Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health and Surveillance, 2*(2), Article e158. https://doi.org/10.2196/publichealth.5869

Plunz, R. A., Zhou, Y., Vintimilla, M. I., McKeown, K., Yu, T., Uguccioni, L., & Sutto, M. P. (2019). Twitter sentiment in New York City parks as measure of well-being. *Landscape and Urban Planning, 189*, 235–246.

Resch, B., Summa, A., Zeile, P., & Strube, M. (2016). Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary space-time-linguistics algorithm. *Urban Planning, 1*(2), 114–127.

Rong, H. H., Davis, J., & Rada-Orellana, M. (2025). Benchmarking large language models against qualitative coding and natural language processing in decoding public sentiment on urban upzoning. *Urban Informatics, 4*(1), Article 17.

Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social Psychology Quarterly, 67*(4), 319–342. https://doi.org/10.1177/019027250406700401

Scerri, K., & Attard, M. (2025). Active travel street intervention ideas in Malta: Evaluating citizen feedback using sentiment analysis. *Journal of Transport Geography, 126*, Article 104241.

Schwartz, A. J., Dodds, P. S., O'Neil-Dunne, J. P. M., Ricketts, T. H., & Danforth, C. M. (2022). Gauging the happiness benefit of US urban parks through Twitter. *PLoS One, 17*(3), Article e0261056.

Sinha, R. C., Sarkar, S., & Mandal, N. R. (2017). An overview of key indicators and evaluation tools for assessing housing quality: A literature review. *Journal of The Institution of Engineers (India): Series A, 98*(3), 337–347. https://doi.org/10.1007/s40030-017-0225-z

Song, Q., Tian, S., Zheng, L., Zheng, Y., Qiu, L., Huang, B., & van Ameijde, J. (2026). Beyond sentiment: Using large language models to decode multidimensional urban park perceptions for enhanced equality. *Landscape and Urban Planning, 268*, Article 105571.

Tan, M. J., & Guan, C. (2021). Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using Twitter data. *Applied Geography, 132*, Article 102474.

Tori, F., Tori, S., Keseru, I., & Ginis, V. (2024). Performing sentiment analysis using natural language models for urban policymaking: An analysis of Twitter data in Brussels. *Data Science for Transportation, 6*(2), Article 5.

Tu, K.-J., & Lin, L.-T. (2008). Evaluative structure of perceived residential environment quality in high-density and mixed-use urban settings: An exploratory study on Taipei City. *Landscape and Urban Planning, 87*(3), 157–171.

Vesely, S., & Klöckner, C. A. (2020). Social desirability in environmental psychology research: Three meta-analyses. *Frontiers in Psychology, 11*, 1395. https://doi.org/10.3389/fpsyg.2020.01395

Wang, J., Chow, Y. S., & Biljecki, F. (2023). Insights in a city through the eyes of Airbnb reviews: Sensing urban characteristics from homestay guest experiences. *Cities, 140*, Article 104399.

Weden, M. M., Carpiano, R. M., & Robert, S. A. (2008). Subjective and objective neighborhood characteristics and adult health. *Social Science & Medicine, 66*(6), 1256–1270. https://doi.org/10.1016/j.socscimed.2007.11.041

Zahnow, R., Verrier, J., Hames, S., & Corcoran, J. (2024). Mapping and measuring neighbourhood social media groups: The case of Facebook. *Applied Geography, 172*, Article 103415.

Zhang, X., Du, S., Du, S., & Liu, B. (2020). How do land-use patterns influence residential environment quality? A multiscale geographic survey in Beijing. *Remote Sensing of Environment, 249*, Article 112014. https://doi.org/10.1016/j.rse.2020.112014

Zhang, Y., Kwan, M.-P., & Fang, L. (2025). An LLM-driven dataset on the spatiotemporal distributions of street and neighborhood crime in China. *Scientific Data, 12*(1), Article 467. https://doi.org/10.1038/s41597-025-04757-8

Zhao, H., Hu, J., Yang, M., & Lai, Y. (2026). Leveraging LLMs and explainable AI to decode citizen complaints for neighborhood respiratory health prediction. *Health & Place, 97*, Article 103606.

Zhao, Y., Zhang, Z., & Mao, Y. (2025). Perceived residential environment quality, emotional experience, and sense of community among the retired people in China. *American Journal of Community Psychology, 76*(1–2), 133–146. https://doi.org/10.1002/ajcp.12805

Zhao, H., & Long, Y. (2022). Research on the construction of index system for residential appearance space quality in China and its application in Pingdingshan. *Urban Design*, (2), 48–55. https://doi.org/10.16513/j.urbandesign.2022.02.006

Zhong, S., Hahm, Y., Chen, X., Wang, N., & Lee, C. (2025). Environmental factors influencing intergenerational interactions in residential communities: A US nationwide survey of built environment experts. *Habitat International, 156*, Article 103296. https://doi.org/10.1016/j.habitatint.2025.103296

Zhu, Y., Su, F., Wang, D., Fu, Q., Han, X., & Liu, J. (2026). Will the 'poor' perception of the environment in old urban areas lead to lower resident well-being? Exploring 'perception bias' in residential environment through spatial features and interpretable machine learning. *Habitat International, 171*, Article 103786. https://doi.org/10.1016/j.habitatint.2026.103786