Leveraging Social Media Big Data and Large Language Models to Quantify the Temporal Dynamics of Cross-Border Spatial Identity: Evidence from Shenzhen-Hong Kong Region

Luyao Niu [a,1], Yurun Wang [a,1], Wenjia Zhang [b,*]

[a] School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, Guangdong, China

[b] College of Architecture and Urban Planning, Tongji University, China

Abstract

Cross-border spatial identity refers to the multifaceted, individually constructed perceptions that residents form through cross-border interactions and daily behaviors. It has emerged as a critical metric for evaluating the integration of cross-border regions, playing a pivotal role in fostering regional community cohesion and advancing integration initiatives. In the context of the Guangdong-Hong Kong-Macau Greater Bay Area, the Shenzhen-Hong Kong corridor exemplifies as a significant yet complex cross-border zone, where divergent political systems, administrative practices, and cultural norms impede seamless integration despite a shared national affiliation. Consequently, examining residents' subjective identity perceptions within cross-border spaces, as reflected in their daily activities, becomes imperative. However, the social, cultural, and linguistic heterogeneity in cross-border spaces complicates the clear definition of measurement dimensions, hampers the collection of long-term, high-quality data, and limits the ability to discern complex, multi-dimensional emotions. As a result, existing research has not adequately captured the temporal dynamics and addressed the multi-dimensional heterogeneity of cross-border spatial identity. To address these challenges, this study introduces an innovative measurement framework that integrates extensive long-term social media data with the semantic analysis capabilities of large language models. Focusing on the Shenzhen-Hong Kong cross-border space as a case study, we analyzed a comprehensive dataset of social media text from platforms such as Weibo and Xiaohongshu, covering the period from 2018 to 2023. By employing the chain of thinking prompt engineering of large language models (LLMs), we quantitatively assess the identity perceptions of Shenzhen and Hong Kong residents across three dimensions: spatial imagery, spatial satisfaction, and spatial sense of belonging. Our findings reveal that the spatial identity of Hong Kong residents experienced marked temporal fluctuations during critical events and policy shifts, such as the COVID-19 pandemic and the implementation of the Greater Bay Area development plan. Moreover, distinct variations in identity perception were observed across different types of cross-border spaces, with shopping and consumption areas displaying a notably positive spatial identity. This research provides a robust, efficient and scalable framework for the long-term tracking of cross-border spatial identity, offering valuable theoretical insights and quantitative evidence to better understand the temporal and spatial evolution of identity perceptions in cross-border populations. It offers important theoretical support and quantitative evidence for the evaluation of the cross-border region integration process.

Keyword: Cross-border spatial identity; Large language models; Social media data; Shenzhen-Hong Kong region.

1. Introduction

Globalization has blurred national borders, increasing resource sharing and economic-cultural interactions among border cities. Cross-border integration is a key to achieving spatial cohesion and regional equilibrium. Therefore, cross-border spatial identity is defined as the emotional attachment, psychological identification, and sense of place that residents develop through their cross-border activities. This identity, in turn, shapes the intensity of cross-border mobility, residents' willingness to participate in regional policies, social environment, and economic activities. Furthermore, it plays a pivotal role in the implementation of cross-border cooperation and cohesion policies, providing intrinsic motivation for resource sharing, social stability, and policy coordination within border regions. As such, cross-border spatial identity offers valuable insights into the mechanisms underpinning regional integration and serves as a crucial lens for evaluating the processes of cross-border spatial cohesion.

The Shenzhen–Hong Kong corridor exemplifies the highly complex dynamics of integration within the Guangdong–Hong Kong–Macau Greater Bay Area (GBA) Strategy framework, primarily shaped by deep-rooted structural differences and political disparities between the two systems. . Specifically, new cross-border transportation infrastructures, such as high-speed rail, metro, and maritime routes, have significantly boosted cross-border interactions, and strengthened cross-border spatial identity. However, complex social tensions in Hong Kong, exemplified in social movements such as the "Occupy Central" movement and the "Extradition Bill Protests", deep-seated cultural differences, mobility restrictions, and disparities in urban space complicate both regional

1 These authors contributed equally.

integration policy implementation and the assessment of residents' psychological connections to the integration.

The conceptualization of spatial identity is inherently multidimensional, encompassing cognitive, affective, and behavioral dimensions. Cognitively, it is shaped by individuals' mental representations of space, influenced by spatial perception, historical narratives, and collective memory (R Hay, 1998)such as aesthetics and a feeling of dwelling. Insider status and local ancestry are important toward the development of a more rooted sense of place. Three contexts are used to examine the development of sense of place: residential status in the place (superficial, partial, personal, ancestral, and cultural senses of place. Affectively, it is built through emotional attachment and place-based belonging from lived experiences and spatial interactions (Hernández et al., 2007). Behaviorally, it manifests in mobility patterns, spatial preferences, and engagement in place-based social practices (Dorfman et al., 2021)place, and partners, with emphasis on the impact of the social environment on spatial behavior and on how individual spatial representations converge to form collective spatial behavior - i.e., common places and routes. Recent studies suggest that in addition to their mental representation (cognitive map. Given this multidimensional nature, any attempt to measure spatial identity must account for these interrelated components, making its quantification particularly complex.

In cross-border regions, factors like intricate socio-political and cultural landscape further complicate the measurement of spatial identity. These regions are often characterized by linguistic diversity, institutional fragmentation, and historical legacies that shape distinct spatial perceptions among different groups (Laven and Baycroft, 2008). Socio-cultural tensions, policy discrepancies, and asymmetries in economic development add further complexity, shaping individuals' sense of belonging and attachment to space (Gardner and Richards, 2017)an indigenous people divided by the Guatemala-Mexico border, identify collectively. We further existing sociological literature on collective identity "boundary work" by demonstrating how it is shaped by spatial, and not just symbolic, boundaries. Mam individuals and organizations define symbolic boundaries that sustain political-administrative borders (such as municipal divisions within Guatemala and Guatemala's border with Mexico. Moreover, the dynamic nature of cross-border interactions further underscore the fluidity and heterogeneity of cross-border spatial identity. As a result, traditional approaches of measuring spatial identity (e.g., surveys and questionnaires) are challenged in fully capturing the nuanced and evolving aspects of spatial belonging in cross-border contexts.

Existing studies on spatial identity primarily rely on qualitative data, such as surveys and interviews. However, these approaches are often limited by small sample sizes, high costs, and low temporal efficiency. Moreover, as they are typically based on cross-sectional data, they fail to capture the temporal dynamics of spatial identity. Additionally, such studies are geographically constrained, which limit their ability to address the complexities of cross-border contexts. Recently, the use of large-scale social media text data to capture residents' emotional perceptions has emerged as a promising method. Social media text data with geolocation information, provides a novel source of linguistic emotional feedback for measuring spatial identity (Gao et al., 2022). This approach offers significant advantages, including large scale data availability, low collection costs, extensive geographic coverage, and the ability to track long-term temporal dynamics.

However, collecting high-quality labeled social media data for measuring cross-border spatial identity remains challenging. First, divergent social media platform preferences on each side of the border lead to fragmented data access and varied data. Second, linguistic and expressive difference (e.g. Mandarin, Cantonese, English,...) further complicate data processing. Moreover, cultural diversity in cross-border regions creates a range of expression styles that can also introduce biases in emotion interpretation, making it harder to accurately capture spatial identity using computational models.

Based on social media data, machine learning algorithms have proven effective in extracting textual emotions. These algorithms automatically identify nuanced user opinions and detect emotional fluctuations by training models to classify sentiment polarity(positive/negative) (Mutanov, Karyukin and Mamykova, 2021) and quantify intensity levels using features like lexical patterns, contextual embeddings, and temporal dynamics in posts (Mandhasiya et al., 2022). However, these methods face several challenges when applied to cross-border spatial identity measurement. They struggle to capture perceptions of abstract concepts accurately, rely heavily on large amounts of labeled data, interpret the prediction process, and their performance is limited by the quality of training data. The multi-source and heterogeneous nature of social media data in cross-border regions increases the costs of data labeling, reducing the efficiency and transferability of model training. Additionally, linguistic diversity in cross-border settings presents further challenges, as a single machine learning model typically struggles to conduct cross-lingual analysis. Furthermore, the black-box nature of machine learning models undermines the interpretability and credibility of prediction outcomes. Therefore, a critical research challenge lies in developing a spatial identity measurement framework tailored to cross-border contexts—one that effectively leverages large-scale social media data, addresses the complexities of multilingual and multicultural environments, and enhances the interpretability and reliability of analytical processes.

LLMs have shown potential in large-scale text analysis, semantic understanding, and complex emotion recognition tasks. Compared to traditional sentiment analysis models based on lexicons or supervised learning, LLMs offers clear advantages in capturing complex and implicit semantics(Xing, 2024), performing zero-shot or few-shot learning without heavy reliance on labeled data(Gandhi et al., 2024), and improving interpretability through self-explanation capabilities (Randl et al., 2024). These features open new avenues for conducting complex sentiment analysis based on social media data and for quantifying cross-border spatial identity perceptions.

Therefore, this study proposes an innovative framework for measuring cross-border spatial identity by integrating social media data with LLMs. We utilize a three-dimensional theoretical model that conceptualize cross-border spatial identity through spatial imagery, spatial satisfaction, and spatial sense of belonging. The framework involves several key steps: collecting and processing social media data, recognizing regional identity, developing chain-of-thought (COT) prompt engineering, and applying LLMs to quantify spatial identity scores. This approach enables precise and scalable measurement of cross-border spatial identity. [Specifically, incorporating LLMs introduces a novel semantic reasoning layer. By leveraging COT prompting, the framework decomposes complex spatial narratives into structured elements (location, activity, sentiment). Consequently, this approach supports dynamic, cross-lingual, and context-aware spatial identity scoring that surpasses traditional keyword-based methods, while also facilitating an analysis of the drivers behind spatial identity beyond conventional interpretability limits.] Using the Shenzhen-Hong Kong cross-border region as a case study, we employed mobile phone signaling data to identify the primary cross- border activity spaces for Hong Kong residents in Shenzhen and collected corresponding social media text data. We then examined the temporal fluctuations of spatial identity in Shenzhen's cross-border areas, the heterogeneity of identity perceptions across different functional spaces, and the differences in cross-border spatial identity between Guangdong and Hong Kong residents.

By systematically exploring the dynamic spatiotemporal patterns and underlying mechanisms, the framework contributes to a deeper understanding of spatial identity dynamics in complex cross-border environments. Compared to traditional methods, this research not only tackles the challenges such as multilingual and multicultural complexities and the heavy reliance on labeled data, but also significantly enhances semantic understanding and emotional analysis capabilities in large, heterogeneous datasets through LLMs. The findings from Shenzhen-Hong Kong cross-border region provide robust empirical evidence for understanding the cross-border integration dynamics and offer valuable theoretical foundations and practical insights for optimizing cross-border spaces and informing regional integration policies.

The remainder of this paper is structured as follows: Section 2 reviews the existing literature; Section 3 introduces the construction of the cross-border social media dataset; Section 4 presents the cross-border spatial identity analysis method based on large models; Section 5 discusses the experimental and analytical results using the research method; Finally, Sections 6 and 7 discuss the significance and limitations of this study and summarize the research conclusions.

2. Literature Review

2.1 Cross-Border Spatial Identity Measurement

Spatial identity refers to the emotional attachment, psychological affiliation, and cognitive bond that an individual or group has with a specific place. When a space's social and material resources meet residents' needs and preferences, they form a positive identity and attachment, which in turn encourages behaviors that support positive interactions with their environment.

Spatial identity develops through sustained interaction and comprehensive perception of both the physical environment and socio-cultural context (Robert Hay, 1998). Its abstract nature makes it challenging to represent concretely or intuitively. Researchers have constructed theoretical frameworks to measure spatial elements across multiple dimensions. For instance, Proshansky and other environmental psychologists proposed a three-dimensional model of spatial identity, which includes cognitive understanding, emotional experience, and behavioral intention (Proshansky, Fabian and Kaminoff, 1983). Breakwell, focusing on individual psychological needs, categorized spatial identity into distinctiveness, continuity, self-esteem, and self-efficacy (Breakwell, 1993). Knez further refined the concept of continuity into place-oriented coherence and place-consistent continuity, proposing a five-dimensional model of place identity (Knez, 2005)climate, place attachment and place identity using Breakwell's four processes model of place identity (e.g. Twigger-Ross, Bonaiuto, and Breakwell. (2003. In contrast to purely theoretical approaches, Lalli introduced the urban identity scale, dividing place identity into five dimensions: external evaluation, general attachment, commitment, continuity with personal past, and familiarity perception (Lalli, 1992)the following problems are seen to characterize the psychological research on 'place identity': heterogeneity of terms and their spatial extension, differing theoretical foundations and fragmented formulations, lack of adequate measuring instruments, and a scarcity of empirical work. This paper aims firstly, to present a systematic analysis of the theoretical traditions of the work on 'place identity'. Secondly, it uses constructs of medium range in order to systematize theory and research. An example of the latter is presented with respect to urban-related identity. A framework for conceptualizing urban-related identity and identification is developed on the basis of social psychological work on self-concept. The second part introduces a measuring instrument (the 'Urban Identity Scale'. This work marked the first quantification of spatial identity, providing an operational tool for theoretical inquiry.

Cross-border spatial identity emerges when residents traverse geographic, cultural, and institutional boundaries through ongoing engagement with cross-border activity spaces, developing emotional attachment and cultural affinity toward the border region (Newman and Paasi, 1998). Distinct from spatial identity confined within a single country, cross-border spatial identity is shaped by three fundamental asymmetries: political system disparities, uneven economic development, and cultural and linguistic divides (Pedersen, 2018), resulting in inherently multi-dimensional complexity. It demonstrates dynamic temporal evolution: individual level development occurs through cultural adaption, social integration, and life course progression (Rishbeth and Powell, 2013), whereas regional level transformations respond to historical trajectories and socio-political dynamics(Laven and Baycroft, 2008). Furthermore, exogenous shocks (e.g. pandemics) and planned interventions (e.g. economic zone development) affect both the intensity and structure of spatial identity. These factors make cross-border spatial identity a complex, dynamic phenomenon and position it as a critical interdisciplinary research focus, bridging geographical inquiry with political studies, psychological analysis, and urban planning (Robert Hay, 1998).

Existing research delineates cross-border spatial identity along several dimensions, reflecting its formation and connotations. For example, thereby building on Scannell and Giddord's three-dimensional model, Rishbeth et al. incorporated memory triggers, comparisons, and symbolic markers of local culture. (Rishbeth and Powell, 2013). Similarly, Pedersen et al. introduced the social interaction networks to examine changes in cross-border spatial identity (Pedersen, 2018), while Kajić et al. emphasized the significance of residents' perceptions of regional boundaries, emotional attachment, and socio-cultural and political systems in the construction (Kajić, Bogdanić and Fuerst-Bjeliš, 2022).

However, quantifying cross-border spatial identity presents considerable challenges. These difficulties stem from the interplay of emotional, behavioral, and cognitive factors, including subjective conceptual boundaries (Peng, Strijker and Wu, 2020), intricate cognitive mechanisms (Arbab, Azizi and Zebardast, 2018), multi-dimensional interactions(Raymond, Brown and Weber, 2010) and diverse socio-cultural conflicts (Fikfak, 2009). While current theoretical frameworks offer valuable insights, they often neglect the key spatial elements, fail to capture dynamic evolution of spatial identity, and lack verified its cross-border applicability. Consequently, measuring cross-border spatial identity remains problematic, hindering further research.

Traditional methods for measuring spatial identity primarily rely on surveys, in-depth interviews, and participatory observation. For example, researchers have used random sampling with the place identity scale to quantify spatial identity in specific areas (Williams and Vaske, 2003), employed Likert scales to explore characteristics during its formation(Lalli, 1992)the following problems are seen to characterize the psychological research on 'place identity': heterogeneity of terms and their spatial extension, differing theoretical foundations and fragmented formulations, lack of adequate measuring instruments, and a scarcity of empirical work. This paper aims firstly, to present a systematic analysis of the theoretical traditions of the work on 'place identity'. Secondly, it uses constructs of medium range in order to systematize theory and research. An example of the latter is presented with respect to urban-related identity. A framework for conceptualizing urban-related identity and identification is developed on the basis of social psychological work on self-concept. The second part introduces a measuring instrument (the 'Urban Identity Scale', and conducted in-depth interviews to examine shifts in individuals' identity perceptions (Chow and Healey, 2008). However, these methods are often limited by small sample sizes, restricted coverage, high costs, lack of timeliness, and potential biases from self-reporting and researcher interpretation. They struggle to capture the intricate features of cross-border spatial identity. As surveys and interviews typically yield cross-sectional data, they inadequately reflect the temporal dynamics essential for deeper research.

With the advent of the "humans as sensors" concept (Resch et al., 2016), social media data, as a sensor for residents' emotions, offering new perspectives for urban emotional perception research (Kong et al., 2022) and serving as a critical tool for measuring urban sentiment. As an essential component of urban emotional perception, social media data holds significant potential for measuring spatial identity due to its unique advantages. Specifically, it provides extensive geotagged, long-term time series information that transcends geographical and cultural boundaries, captures cognitive patterns across spatial scales, and offering long-term, continuous, large-scale data coverage. Public emotional responses and cognitive characteristics toward spaces can be effectively captured, enabling more refined spatial identity measurement. For instance, Kong et al. quantified the positive emotions of park visitors using long-term Sina Weibo check-in data(Kong et al., 2022); Felasari et al. analyzed Twitter text related to tourism sites to explore the emotional connections between visitors and attractions across dimensions like understanding, psychological involvement, and self-identity fusion (Felasari et al., 2017); Jang & Kim employed Instagram image hashtags and geotags for text semantic analysis to reveal the public's collective cognitive and emotional connections to specific urban spaces (Jang and Kim, 2019).

To efficiently analyze large volumes of social media data, machine learning and deep learning methods have been increasingly introduced. Machine learning models, such as support vector machines, random forests, and logistic regression, have been applied to identify spatial emotions, place attachment, and spatial identity perceptions within social media data. For instance, Zhou et al. used Biterm topic modeling and other NLP algorithms to analyze residents' emotions expressed in social media texts about urban green spaces (Zhou et al., 2023); Jang & Kim used self-organizing map (SOM) algorithms to cluster social media text keywords, revealing latent patterns in spatial identity, and providing graphical representations of semantic relationships that enhance both analysis efficiency and interpretability (Jang and Kim, 2019), offering a comprehensive understanding of place attachment when integrated with text-based analysis.

Despite their capacity to parse complex semantics in large-scale data, current machine learning models remain facing limitations in the context of cross-border spatial identity research. Cross-border spatial identity is complicated by socio-economic disparities, diverse historical cultures, varying political systems, and cross-border political dynamics. Its conceptual definition and measurement are inherently dynamic and multi-dimensional (Peng, Strijker and Wu, 2020). Moreover, these models typically require abundant high-quality labeled data for training, yet data collection in cross-border contexts is hindered by dispersed sources and heterogeneous data structures. They often exhibits poor generalization across different scenarios, performing optimally only within specific contexts. The complex socio-cultural environment of cross-border regions-characterized by multilingual and multicultural challenges, further complicates semantic understanding and emotional perception; a single machine learning model struggle to adapt effectively to cross-language text analysis. Finally, the "black-box" nature of machine learning models precludes transparent explanations for their outputs, thereby hindering insights into the underlying sources and mechanisms shaping cross-border spatial identity.

2.2 Utilizing LLMs for Measuring Spatial Identity

LLMs demonstrate advanced capabilities in reasoning, analysis, and generation, excelling particularly in tasks involving noisy data, complex contextual scenarios, and implicit semantic analysis. Their capability to capture bi-directional contextual dependencies facilitates the efficient analysis of large text corpora (Wan et al., 2024). Their extensive model parameters enable few-shot and zero-shot learning, reducing the dependence on labeled data and supporting transfer across diverse downstream tasks. LLMs trained on multilingual datasets can process and generate text in various languages, enhancing performance in cross-lingual tasks and enabling effective scenario transfer(Lai et al., 2023)large language models (LLMs. Furthermore, emerging approaches leveraging LLMs aim to provide more transparent reasoning process, potentially enhancing the interpretability of text-based quantification. Consequently, LLMs represent a promising tool for measuring spatial identity, particularly for capturing and quantifying it within cross-border regions characterized by cross-lingual and multicultural contexts.

The robust semantic understanding capabilities of LLMs provide a distinct advantage in interpreting implicit emotions like sarcasm, irony, and subjectivity, leading to strong performance in sentiment analysis tasks. Specifically, LLMs provide effective analysis pathways for tasks involving subjective and abstract emotional concepts, such as accurately distinguishing between moral foundations like care and fairness (Rathje et al., 2024). These capabilities allow LLMs to effectively address complex and ambiguous concepts to spatial identity, including place attachment (Zhou et al., 2024), social ideology (Liu and Shi, 2024), and personal values (Brigola, 2024), providing effective tools for its analysis.

LLMs inherently possess zero-shot and few-shot learning capabilities, which are among their most essential features. They can perform downstream tasks with little or no annotated data, bypassing the traditional reliance on large labeled datasets. This advantage reduces annotation costs and increases efficiency. For example, Kojima et al. introduced the zero-shot reasoning framework (Zero-shot-COT), demonstrating the effectiveness of zero-shot conditions in complex tasks (Kojima et al., 2022). Under zero-shot conditions, LLMs exhibit strong generalization and transfer capabilities across diverse tasks and scenarios. These zero-shot and few-shot capabilities are widely applied in sentiment analysis tasks. Studies indicate that LLMs effectively address sentiment reasoning tasks, including emotional inference, outcome reasoning and target identification, even with zero-shot prompts (Gandhi et al., 2024).

Training on multilingual data endows LLMs with robust cross-linguistic analytical capabilities, leading to superior performance in multilingual tasks. Lai et al. evaluated ChatGPT on 37 tasks, observing strong performance in high-resource languages like English, Chinese, and Spanish (Lai et al., 2023)large language models (LLMs, while also demonstrating consistent performance in cross-linguistic sentiment analysis. Consequently, the cross-linguistic capabilities of LLMs offer effective solutions for addressing the cross-cultural and cross-linguistic challenges inherent in measuring cross-border spatial identity.

LLMs can generate self-explanations based on instructions, thereby enhancing transparency and interpretability. For instance, Randl et al. distinguished between extractive explanations (e.g., identifying "the most important phrases") and counterfactual explanations (e.g., "modifying input text to change the prediction outcome"), verifying the effectiveness and consistency (Randl et al., 2024). Similarly, Fragkathoulas & Chlapanis (2024) demonstrated that instructing LLMs to generate self-reasoning chains improves contextual understanding and accuracy (Fragkathoulas and Chlapanis, 2024). In addition, incorporating self-explanation prompts reduces information loss and contextual misunderstanding, thereby enhancing comprehension. Research has explored the deep interpretability of LLMs, particularly their attribution capabilities. In attribution analysis, LLMs support their response by generating attribution paragraphs or identifying source texts, thus verifying authenticity and reliability. For instance, Bohnet et al. validated attribution-enabled question answering (Bohnet et al., 2023), while Hu et al. combined knowledge graphs to automatically generate attribution categories, underscoring the advantages of LLMs in complex attribution tasks(Hu et al., 2024). Therefore, integrating the self-explanation and attribution capabilities of LLMs into the measurement of cross-border spatial identity can reveal more reliable emotional associations and attribution mechanisms. This provides urban planners and policymakers with precise sentiment feedback tools and critical insights for optimizing regional integration policies.

In summary, LLMs offer a groundbreaking approach for measuring cross-border spatial identity, leveraging their efficient text parsing, advanced semantic understanding, zero-shot and few-shot learning, cross-linguistic proficiency, and self-explanatory features.

3. Data

This study investigates cross-border activities between Shenzhen and Hong Kong, focus on the primary locations in Shenzhen frequently visited by Hong Kong residents . As a highly mobile groups within the Guangdong-Hong Kong-Macau Greater Bay Area, Hong Kong residents predominantly use Shenzhen as their main entry point to the mainland China. Their activities usually encompass shopping, dining, cultural engagement, leisure, and transit through border checkpoint (Wang, 2004). Furthermore, the diversity and density of these activity spaces provide rich contexts for the expression of spatial identity.

In order to...., this study constructs two key datasets to measure Shenzhen-Hong Kong cross-border spatial identity: (1) A spatial dataset derived from China Unicom's mobile signaling data for Shenzhen, utilized to identify hotspots of cross-border activities, and (2) A social media text dataset comprising posts from Weibo and Xiaohongshu related to cross-border spaces, employed to assess spatial identity scores.

3.1 Shenzhen-Hong Kong Cross-Border Spatial Data

The study focuses on the Shenzhen region (Figure. 1), specifically targeting areas characterized by frequent and representative visits by Hong Kong residents. To identify hotspots of activities for Hong Kong residents, we utilized mobile signaling data sourced from China Unicom's Smart Footprint platform, covering the month of May in each year from 2018 to 2023.It contains user identifiers, card registration locations, dates, and activity locations with geographic coordinates, comprising approximately 3.27 million Hong Kong-registered users and about 32 million mobile signaling records representative of Hong Kong cross-border population.

The process for defining the cross-border activity space is illustrated in Figure.2. First, based on the primary cross-border activities of Hong Kong residents, five types of cross-border spaces were identified: shopping and consumption, leisure, socio-cultural, border checkpoint, and policy spaces. Specifically, shopping and consumption spaces encompass retail, dining, and service areas in major Shenzhen commercial districts, catering to the consumption needs of Hong Kong residents. Leisure spaces include outdoor areas such as parks and tourist attractions, offering relaxation and social opportunities. Socio-cultural spaces comprise venues like museums and cultural centers, fostering cultural exchange and social interaction. Border checkpoint spaces refer to key transit points-including train stations, ports, and land crossings-serving as primary conduits between the Hong Kong and Shenzhen. Finally, policy spaces are delineated by regional cooperation policies, exemplified by the Shenzhen-Hong Kong Cooperation Zone in the Qianhai and Lok Ma Chau Loop.

Spatial boundaries for shopping and consumption, leisure, socio-cultural, and border checkpoint spaces in Shenzhen were extracted using the 2023 Area of Interest (AOI) data from Baidu Maps, whilst policy spaces were delineated based on planning documents pertaining to the Qianhai and Lok Ma Chau Loop innovation cooperation zones. These boundaries were subsequently mapped onto a 250m×250m grid covering Shenzhen, allowing for the possibility of multiple space types co-occurring within a single grid cell.

Mobile signaling data was then used to calculate the number of Hong Kong cross-border residents in each grid cell for May of each year. For shopping and consumption, leisure, and socio-cultural, grids cell were ranked by resident count, with the top 20% designated as hotspots. All grid cells corresponding to border checkpoint spaces identified via AOI data were classified as activity hotspots. Policy space hotspots were defined by the geographical extent of the Qianhai and Lok Ma Chau Loop cooperation zones.

Finally, an expert panel refined the hotspots selection through discussion, considering criteria such as representativeness, geographic proximity, and uniqueness. Specifically, for shopping and consumption, leisure, and socio-cultural spaces, three representative areas were selected from the top-ranked grids. Core locations within these areas were subsequently identified using AOI and Points of Interest (POI) data. For policy and border checkpoint spaces, the expert panel confirmed all previously identified hotspot grids. Overall, 15 key areas and 100 core locations were identified (locations shown in Figure. 1b.)

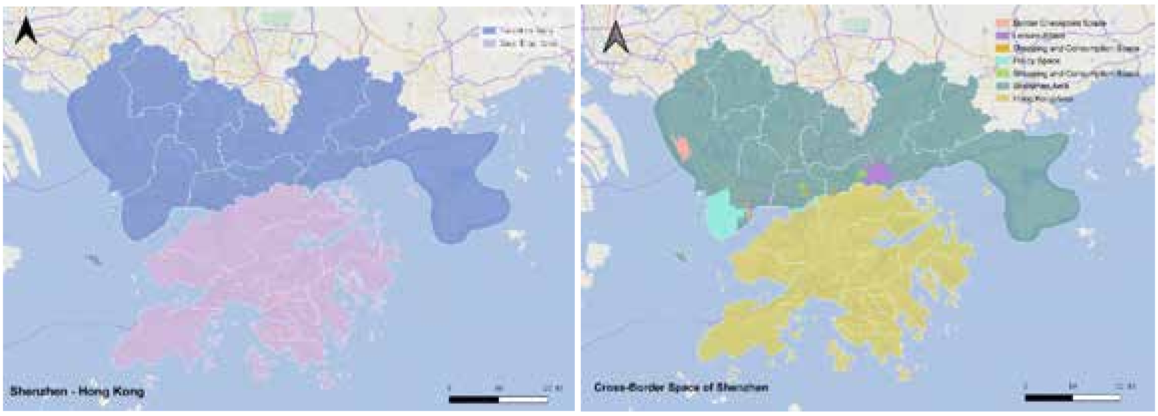


Figure. 1 a. Shenzhen-Hong Kong Map; b. Different types of cross-border space.

3.2 Cross-Border Spatial Social Media Text Data

Building upon the cross-border spatial areas defined above, we collected social media text data that either mentioned or was geo-located in those regions to construct a dataset for semantic analysis of cross-border spatial identity. Data was sourced from Weibo (www.weibo.com) and Xiaohongshu (www.xiaohongshu.com), selected based on platform activity level and data accessibility. As China's largest social media platform, Weibo reported 500 million daily active users in May 2020 (Kong et al., 2022). Xiaohongshu, an emerging platform integrating social networking, lifestyle sharing, and e-commerce, reached 200 million monthly active users in 2020, including a significant proportion of overseas users (He, 2024). Notably, with increasing engagement from Hong Kong users on mainland platforms, Xiaohongshu serves as a primary medium for accessing and sharing information related to mainland China (Digital Report, 2020). The dataset construction involved three steps: expansion of core location keywords, data collection, and filtering to isolate the Hong Kong cross-border population, as illustrated in Figure. 2.

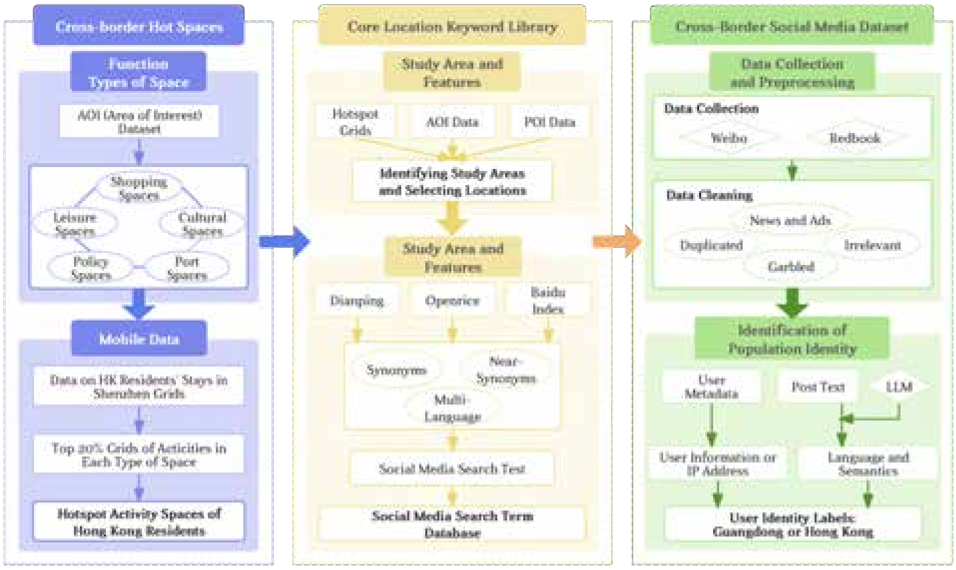


Figure. 2 Construction Process of the Cross-border Spatial Social Media Text Dataset

To ensure comprehensive data retrieval, the initially extracted core location keywords were expanded. Specifically, synonyms and related terms for each core location were generated by translating names into Simplified Chinese, Traditional Chinese, and English, accommodating Hong Kong residents' linguistic habits. Subsequently, the search functions of Weibo and Xiaohongshu were utilized to test and refine these keywords, identifying high-frequency terms and discarding low-quality ones. Tools such as Dianping, OpenRice, and Baidu Index were used to identify further high-frequency synonyms and phrases. Following an expert review to eliminate ineffective terms, a cross-border spatial location keyword bank comprising 405 search terms for the 100 core locations was established.

Using the keyword bank, we screened social media posts and location tags that mentioned relevant locations and collected 2.28 million original posts from January 1, 2018, to December 31, 2023. These posts include content, publication date, user ID, user type, follower count, IP location (since April-May 2022?), and location keywords. Since Weibo and Xiaohongshu began publicly sharing IP data in April-May 2022, we retained only content with IP addresses located in Guangdong and Hong Kong, while posts without IP data were reserved for later processing.

To enhance data quality, a systematic data cleaning process combining rule-based filtering and semantic analysis via LLMs was implemented. First, duplicate posts were removed by retaining only the earliest submission from each user in case of similar or identical post content. Second, special characters, irrelevant data fragments, and hyperlinks were removed to standardized text formatting. Third, a text length filter was applied, retaining posts containing between 20 and 1000 characters to ensure sufficient content relevance. Posts originating from official media or high-influence accounts were excluded to minimize promotional content bias. Finally, posts identified as advertising or news reports were removed using the ChatGLM, preserving only those reflecting genuine user opinions. This process resulted in 1.01 million valid posts, forming the analytical basis for the subsequent spatial identity assessments.

Based on this dataset, posts originating from the Hong Kong cross-border population were identified and extracted, considering metadata and semantic content, including Hong Kong-specific dialects, topics, and cultural factors. While existing research often relies on metadata (e.g., IP address, time zone, user profile data, social network data) and text analysis for population identification, this study employed a multi-level method combining post metadata, user profile data, and LLM-assisted content analysis to accurately capture posts from the target demographic.

Initially, posts were preliminarily identified based on metadata. Posts were classified as originating from the Hong Kong cross-border population if their associated IP address or user profile explicitly indicated a Hong Kong location. For posts lacking such definitive metadata, LLM was employed to analyze language style, writing patterns, and semantic content to infer the users' origin (see Figure. 3).

Step 1: Identifying Cantonese versus Mandarin Usage: This step focused on distinguishing between Cantonese and Mandarin text in the posts. First, an LLM prompt was developed, leveraging the distinct vocabulary and grammatical characteristics of Cantonese. The prompt design explicitly outlined the key differences between Cantonese and Mandarin and highlighted specific features indicative of Cantonese. Subsequently, few-shot learning examples representing both languages were provided. The LLM was then tasked with assigning a probability score indicating the likelihood that each post was written in either language (the probabilities sum to 1). Posts assigned a Cantonese probability exceeding 50% proceeded to Step2 for further refinement, while the remaining posts (classified as primarily Mandarin) were analyzed in Step3.

Step 2: Distinguishing Cantonese Variants (Guangdong and Hong Kong): Given that Cantonese spoken and written in Hong Kong exhibits differences from that used in Guangdong, particularly regarding character usage, vocabulary choices, frequency of English loanwords/transliterations, code-switching patterns, and grammatical structures. Posts identified as Cantonese in Step 1 underwent further examination. Recognition rules specific to Hong Kong Cantonese were established, focusing on the mentioned features. Manually identified social media examples illustrating these distinctions, along with discriminative criteria were provided to the LLM. Posts assessed as having a Hong Kong Cantonese probability exceeding 50% were classified as originating from Hong Kong users; otherwise, they are classified as originating from Guangdong users.

Step 3: Identifying Hong Kong Cross-border Semantics in Mandarin Posts: For posts identified as primarily Mandarin, semantic prompts were used to detect content indicative of a Hong Kong perspective. Indicators included phrases suggesting movement patterns typical of cross-border residents, such as "going to Shenzhen" (implying departure from Hong Kong) or "returning to Hong Kong" / "back to Kowloon" (implying residence in Hong Kong). Accordingly, the prompt design first introduced key features characterizing a Hong Kong perspective—including references to departure location, habitual residence, return destination, and expressions of belonging—along with culturally specific Hong Kong terms (e.g., "high tea," "OT," "taking the lift"). Next, manually curated examples demonstrating these features were provided. Explicit judgment rules were defined, instructing the LLM to assign a binary value (1 for Hong Kong perspective present, 0 otherwise). Consequently, posts assigned a value of 1 were classified as originating from Hong Kong residents, while the remainder were classified as likely originating from Guangdong residents.

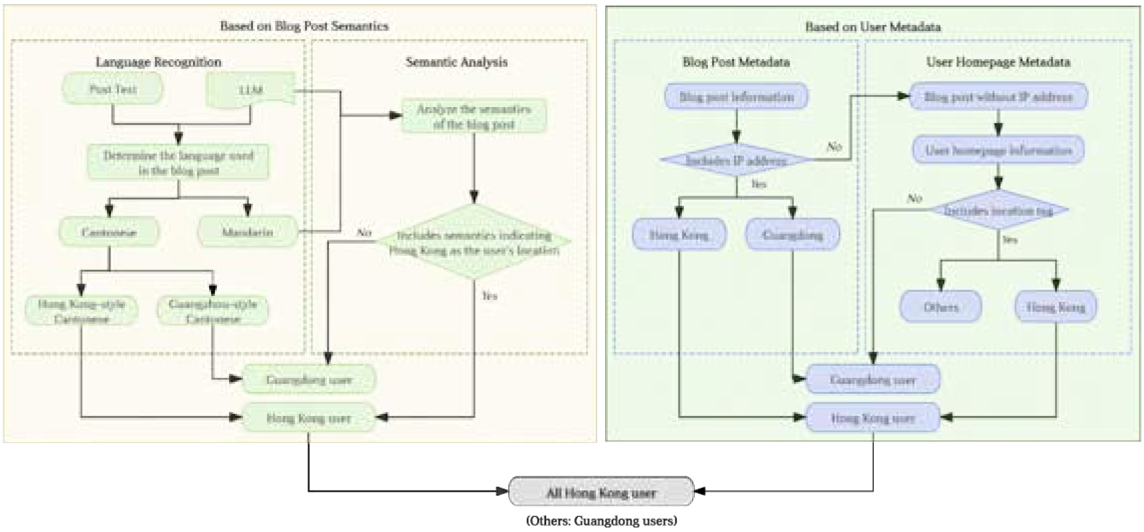


Figure 3: Crowd Identity Recognition Framework

To evaluate the performance of the developed LLM-based framework in differentiating user origins, we assessed its accuracy on distinct tasks. First, for differentiating Cantonese varieties (Steps 1 and 2), we compiled a test set comprising 100 samples each from the "Guangzhou Spoken Corpus" (representing Guangdong Cantonese) and the "Hong Kong MapTask Corpus" (representing Hong Kong Cantonese), supplemented by 200 Mandarin samples. The model (ChatGLM3) achieved 88% accuracy in distinguishing Cantonese from Mandarin, 100% accuracy in identifying Hong Kong Cantonese, and 74% accuracy for Guangdong Cantonese. Second, to evaluate the model's ability to identify a Hong Kong perspective in Mandarin text (Step 3), we addressed the absence of a suitable labeled corpus by generating synthetic data. Specifically, ChatGPT 4.0 was prompted to create 300 text samples exhibiting distinct Hong Kong-related semantic features; these samples were subsequently manually reviewed for quality and relevance. Evaluating the ChatGLM3 model against this curated, labeled dataset yielded a recognition accuracy of 81% for identifying the Hong Kong perspective. Overall, these results indicate that the proposed population identity recognition framework possesses robust discriminative capabilities. Following this validation, the framework was applied to classify the social media posts that initially lacked identifying metadata, assigning either a Hong Kong or Guangdong origin label to each. Finally, these newly classified posts were merged with the posts previously identified via metadata (IP address or user profile) to construct the final cross-border spatial social media text dataset. This comprehensive dataset serves as the foundation for the subsequent measurement of cross-border spatial identity.

4. Method

This study employed LLMs to assess cross-border spatial identity along three dimensions: spatial satisfaction, spatial imagery, and spatial sense of belonging. The methodology involved defining and labeling these dimensions, designing corresponding prompts for the LLM, computing identity scores for text data, and conducting attribution analysis (Figure. 4). After comparing models based on Chinese text analysis capabilities, response speed, and cost effectiveness, ChatGLM-3 was selected for implementation.

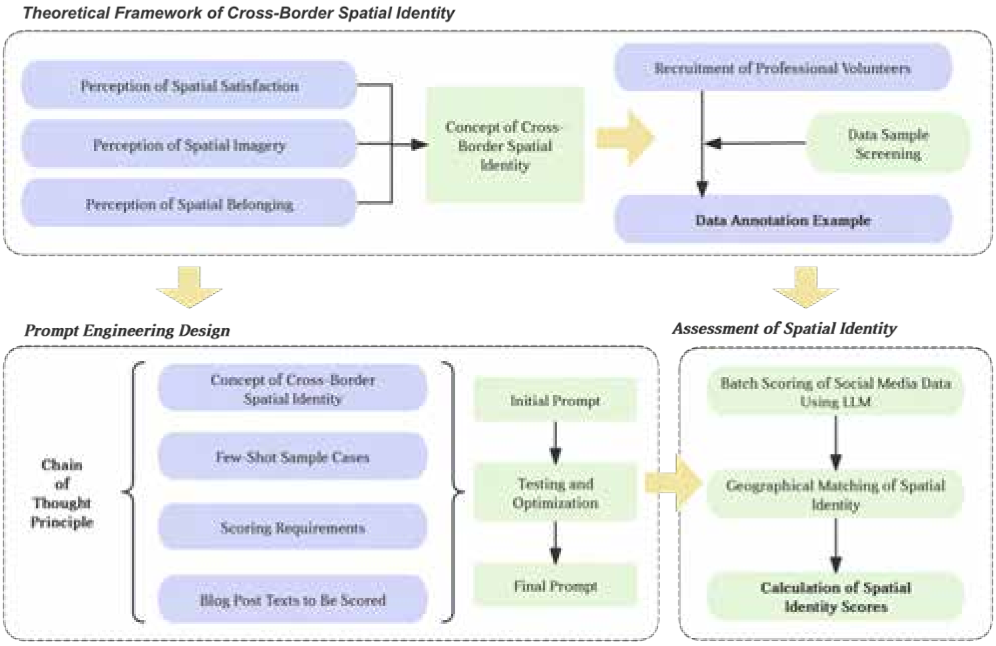


Figure. 4: Cross-Border Spatial Identity Recognition Framework

4.1 Definition and Labeling of Cross-Border Spatial Identity

Drawing on dynamic and hierarchical perspectives, this study proposes a framework comprising three dimensions to measure the cross-border spatial identity of Hong Kong residents concerning their activity spaces in Shenzhen (Casakin, Hernández and Ruiz, 2015).

1. Spatial Satisfaction: This dimension captures residents' evaluations of the quality and adequacy of the environment and facilities within the cross-border space, reflecting perceived convenience and suitability (Li et al., 2022). It assesses whether spatial functions meet individual needs, highlighting aspects significant for regional planning, such as public service provision, living convenience, and accessibility.

2. Spatial Imagery: This dimension relates to the residents' direct perceptual experiences of the cross-border space. It involves awareness of position and orientation, familiarity with spatial components and distinctive feature and the recalled events or memories associated with the space (Blajenkova, Kozhevnikov and Motes, 2006). Spatial imagery thus reflects the subjective impressions that shape the perceived character and image of a place.

3. Spatial Sense of Belonging: This dimension signifies the deep emotional attachment and loyalty residents develop towards the cross-border space. It manifests in behaviors such as willingness to engage in cross-border activities, protect the local environment, contribute to development, or make personal sacrifices for the place (Casakin, Hernández and Ruiz, 2015). This dimension underscores the planning objective of fostering inclusive and shared spatial environments.

Collectively, these dimensions represent a potential developmental trajectory where a strong sense of belonging emerges from cumulative positive spatial imagery and satisfaction. To quantify the perceived intensity for each dimension, this study employed an integer scale ranging from -3 (strong negative perception) to +3 (strong positive perception), with 0 indicating no discernible association. This approach frames the assessment of cross-border spatial identity as a three-dimensional multi-classification task, where each social media post is classified according to the connotations expressed for each of the three dimensions.

To establish a benchmark for evaluating the LLM's performance, 10 volunteers with backgrounds in urban studies were recruited to manually annotate a sample of the social media data. These volunteers were provided with the definitions of the three cross-border spatial identity dimensions and asked to rate 500 randomly selected posts on the -3 to +3 scale for each dimension based on the text content. The resulting annotations were reviewed and refined by experts, establishing a reference standard for the abstract concept of cross-border spatial identity. This standard subsequently informed the design of few-shot learning prompts for the LLM and served as the basis for evaluating its measurement accuracy.

4.2 Prompt Engineering Design

Following the COT principle, the study designed a prompt engineering framework for scoring spatial identity using LLMs. The method guides the model's reasoning process step-by-step, thereby enhancing its ability to interpret complex semantic meanings in texts. Specifically, the framework comprises three steps: defining cross-border spatial identity, providing scoring examples, and iteratively refining the prompts.

Step 1: Definition. Cross-border spatial identity is defined in three dimensions: spatial satisfaction, spatial imagery, and spatial sense of belonging. This step clarifies the evaluation criteria and establishes a scoring range from -3 to 3, where negative values indicate negative identity, positive values indicate positive identity, and 0 denotes irrelevance.

Step 2: Scoring Examples. To instruct the model on the required logic and output format, the study present a small set of sample posts with corresponding scores and detailed reasoning for each emotional dimension. This step aids the model in accurately scoring spatial identity.

Step 3: Iterative Refinement. The effectiveness of the prompt design is evaluated through multiple rounds of testing, with continuous optimization based on model performance. Retry and backtracking principles are applied, requiring the model to review and correct its output (Yang et al., 2024). A self-consistency evaluation mechanism is also integrated to ensure the logical coherence and enhance result reliability (Wang et al., 2023). Additionally, the model is required to output its reasoning process and score explanations, thereby enhancing transparency. This approach has been shown to significantly enhance the reasoning capabilities of LLMs (Kojima et al., 2022).

Subsequently, the study employs the ChatGLM API, configured with the ChatGLM-3-turbo model, to assess the cross-border spatial identity of social media posts. Each post is individually embedded into the prompt and process to mitigate contextual interference. Posts are sequentially input into the model, which returns structured data. After sorting and cleaning, scores are assigned values of -3, -2, -1, 0, 1, 2, and 3, corresponding respectively to "Extremely Dissatisfied", "Dissatisfied", "Slightly Dissatisfied", "Unrelated to Spatial Identity", "Slightly Satisfied", "Satisfied", and "Extremely Satisfied". Posts scoring 0 were excluded (indicating irrelevance). Consequently, the final dataset comprises nearly 20,000 posts from Hong Kong and 687,000 posts from Guangdong, covering 15 spatial regions within Shenzhen. Each post is scored across the three defined dimensions.

4.3 Cross-Border Spatial Identity Scoring Model Evaluation

After computing the cross-border spatial identity scores, the study evaluates the LLM's performance. Precision and F1 score are adopted as evaluation metrics, as they are widely used in machine learning and have proven effective in sentiment analysis tasks. Precision is defined as the proportion of true positive instances (according to human annotations) correctly classified by the model. For example, in a post with a spatial sentiment score of 2, precision is the ratio of posts with a human-assigned score of 2 within the model's predicted category. The F1 score, calculated as the harmonic mean of precision and recall, provides a balanced evaluation of overall performance. The formulas are as follows:

$$Precision = \frac{|Correctly\ recognized|}{|All\ recognized|} \quad (1)$$

$$F\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

In this study, 500 posts were randomly selected from the cross-border social media dataset. Sentiment analysis tools (e.g., SnowNLP) and mainstream machine learning models (e.g., Support Vector Machines and Random Forests) were employed to score the spatial identity of these posts. These supervised models were trained using manually annotated spatial identity scores from Section 4.1, with the data partitioned into training, testing, and validation sets at ratios of 0.4, 0.4, and 0.3, respectively. Furthermore, since SnowNLP outputs continuous values between 0 and 1, the three-dimensional scores (spatial imagery, spatial satisfaction, and spatial belonging) from both the LLM and machine learning models were averaged and standardized to a range between 0 and 1 for comparison. Finally, using manual spatial identity scores from volunteers as benchmark labels, the models' performance was evaluated by calculating precision and F1 scores, thereby assessing the LLM's applicability in measuring cross-border spatial identity.

4.4 Cross-Border Spatial Scoring Indicators

Based on the cleaned scoring data, regional multidimensional spatial identity indicators were calculated for each region using the following formulas to quantify spatial imagery, satisfaction, and sense of belonging:

$$M_i = \frac{\sum_{j=1}^{N_i} p_{ij}}{N_i} \quad (3)$$

$$S_i = \frac{\sum_{j=1}^{N_i} q_{ij}}{N_i} \quad (4)$$

$$O_i = \frac{\sum_{j=1}^{N_i} t_{ij}}{N_i} \quad (5)$$

Where:

p_{ij} is the spatial imagery score of the j-th post related to region i;
q_{ij} is the spatial satisfaction score of the j-th post related to region i;
t_{ij} is the spatial sense of belonging score of the j-th post related to region I;
N_i is the total number of posts related to region i;
M_i is the average spatial imagery score for region i;
S_i is the average spatial satisfaction score for region i;
O_i is the average spatial sense of belonging score for region i;

4.5 Attribution Analysis of Spatial Identity

LLMs can perform multi-dimensional evaluations of spatial identity on large-scale social media data and are capable of analyzing long texts and conducting inductive topic analysis (De Paoli, 2024). They can identify underlying meanings from sample annotations and qualitative data, thereby summarizing and generalizing vast amounts of qualitative text. In this study, the LLM analyzes social media data to uncover the driving factors behind spatial identity, akin to conducting semi-structured interviews (Dengel et al., 2023). The interview material comprises posts from platforms such as Xiaohongshu and Weibo, and the model's long-text analysis capabilities reveal core emotional features, enhancing the interpretability and credibility of the analysis. The process involves the following steps:

1. Interview Guide Design: A semi-structured interview guide was developed for the LLM based on temporal and spatial influencing factors, spatial functions, changes in emotional identity.

2. Data Organization and Input: After obtaining the spatial identity scores for each region, social media posts were categorized into 90 groups based on five spatial types, three emotional dimensions, and six scoring outcomes. Each group was then summarized to form a database serving as the LLM's prior knowledge during the interview.

3. Model Inference and Interview: A semi-structured interview method was employed, using prompt engineering to guide the LLM in analyzing social media data and identifying spatial identity drivers. The prompts, designed in a role-playing format, instruct the LLM to act as a senior urban planning analyst to enhance its performance. Iterative questioning was applied to further probe underlying attributions when ambiguous or unexpected response emerged (Dengel et al., 2023).

4. Emotional Cause Induction: The LLM conducted an in-depth analysis of each text group, extracting underlying causes related to spatial identity and identifying potential influencing factors from the perspective of spatial function, along with providing targeted optimization suggestions.

Due to the ChatGLM's limitations in processing long texts, the Kimi Chat model was selected for further exploration. Through conversational prompts, attribution results for both positive and negative spatial identity across different spatial types were obtained. This approach combines quantitative and qualitative analyses, leveraging LLMs' strengths to offer a more comprehensive understanding of cross-border spatial identity.

5. Results

5.1 Measurement Results

This study evaluated the model's performance in scoring multi-dimensional cross-border spatial identity, comparing it against sentiment analysis tools (e.g., SnowNLP) and machine learning models (e.g., SVM) (Table 1). The LLM demonstrated superior performance, achieving 99% precision and an F1 score of 0.86, significantly exceeding the other models while closely aligning with human ratings.

Table 1. Comparison of cross-border spatial identity model assessments

	Precision	F1
LLM	0.99	0.86
SnowNLP	0.92	0.85
Random Forest	0.71	0.77
SVM	0.71	0.77
Logistic Regression	0.71	0.77
LightGBM	0.71	0.77

Table 2 presents the spatial identity perceptions of Shenzhen and Hong Kong residents. Guangdong residents consistently reported higher spatial identity than Hong Kong residents, consistent with literature indicating that local familiarity foster stronger place attachment (Hernández et al., 2007). Specifically, among the three dimensions, Hong Kong residents scored highest in spatial satisfaction (mean = 0.51) and lowest in spatial sense of belonging (mean = 0.34), with spatial imagery in between, indicating a more functional than emotional spatial identity toward Shenzhen. Similarly, Guangdong residents showed the lowest spatial sense of belonging, suggesting that robust emotional attachment and regional identity are more challenging to establish through short-term cross-border interactions (Qian and Zhu, 2014).

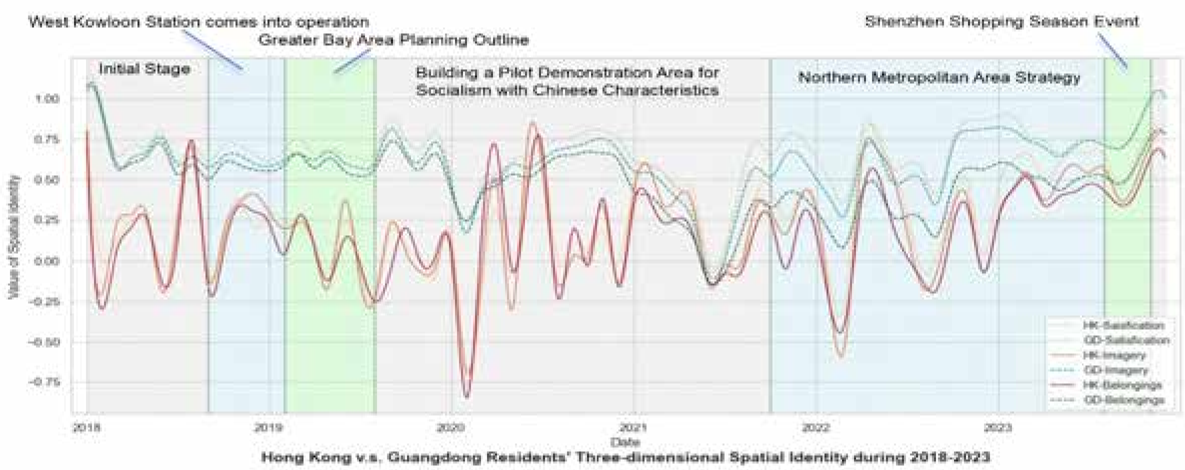
Despite a smaller sample size, Hong Kong residents exhibited greater variability in spatial identity perception than Guangdong residents. In contrast, Guangdong residents demonstrated more consistent spatial identity, likely reflecting more uniform spatial promotion and local culture in mainland China (Xie et al., 2020).

Table 2: Descriptive Statistics of Spatial Identity of Hong Kong and Guangdong Residents

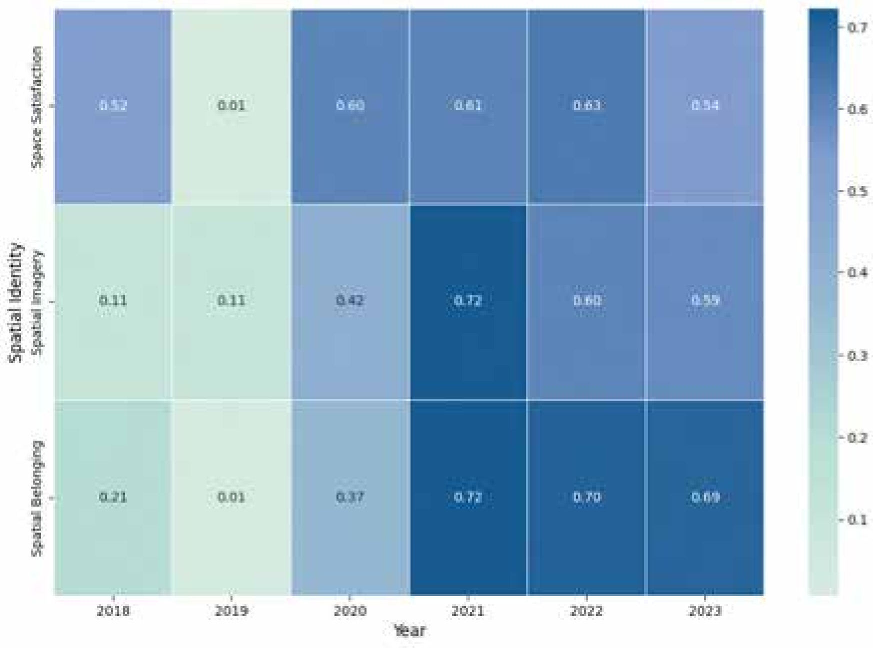
		Spatial Satisfaction	Spatial Imagery	Spatial sense of Belonging
	Mean	0.51	0.44	0.34
Hong Kong	SD	1.59	1.60	1.53
	observations	18374	17397	18421
	Mean	0.71	0.70	0.47
Guangdong	SD	1.55	1.57	1.44
	observations	664025	620594	669932

5.2 Temporal Heterogeneity

Figure. 5a illustrated the temporal evolution of cross-border spatial identity perceptions among Hong Kong and Guangdong residents, according to the timings of post publication. Cross-border spatial identity measured across three dimensions, fluctuated over time with alternating positive and negative values and consistent trends. For instance, in February 2020, Hong Kong residents' reached their lowest scores (satisfaction: -0.76; imagery: -0.69; belonging: -0.85), then experienced a brief peak in June-July 2020 (satisfaction: 0.68; imagery: 0.72; belongings: 0.70), followed by a rapid decline and recovery. Fluctuations intensified from 2021 to 2022, with the lowest scores in March 2022 (satisfaction: -0.24; imagery: -0.50; belongings: -0.37), and the highest in May 2022 (satisfaction: 0.76; imagery: 0.70; belongings: 0.55). After 2023, all dimensions gradually increased and stabilized. In contrast, Guangdong residents showed a more stable and consistently positive trend, with only minor fluctuations before 2020 and a slight decline in 2021 that recovered rapidly. Although the emotional disparity between the regions narrowed after 2021, the convergence reflects distinct regional dynamics in cross-border spatial identity.



a. Hong Kong v.s. Guangdong Residents' Three-dimensional Spatial Identity during 2018-2023



b. Correlation Matrix of Multidimensional Spatial Identity between Hong Kong and Guangdong Residents

Figure. 5a. Monthly Fluctuation Chart (Smoothed) b. Correlation Matrix of Multidimensional Spatial Identity between Hong Kong and Guangdong Residents

Fluctuations in Hong Kong residents' spatial identity were closely associated with key events in Hong Kong and Shenzhen, as well as regional integration policies. Significant shifts occurred during critical periods, such as the COVID-19 pandemic outbreak, the social unrest surrounding the extradition bill, the opening of West Kowloon Station, and the launch of the Greater Bay Area development plan. Specifically, between 2020 and 2023, the COVID-19 pandemic triggered marked declines in cross-border spatial identity, consistent with previous findings (Luo and Lam, 2020). Before the pandemic, gradual policy implementation maintained emotional connections; however, 2020's cross-border restrictions disrupted mobility and altered perceptions of Shenzhen. By 2023, as cross-border travel resumed, Hong Kong residents' spatial identity rebounded, with emotional intensity even surpassing pre-pandemic levels.

Similarly, sudden events such as the extradition bill protests might also affect Hong Kong residents' cross-border emotions. The protests in June 2019 led to a marked decline in cross-border spatial identity due to increased social tension and reduced mobility, while the 2020 National Security Law helped stabilize these fluctuations by boosting confidence in regional integration. (He et al., 2023). Although regional integration policies influence spatial identity, their net impact is moderated by key events and the pandemic, warranting further evaluation using econometric models.

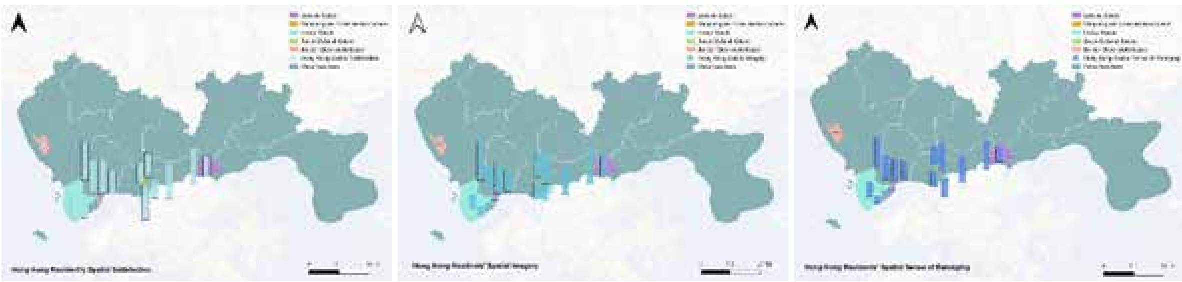
Figure. 5a reveals significant differences in the trends and intensity of spatial identity dimensions between Hong Kong and Shenzhen residents. As shown in Figure. 5b, early in the study, the correlation between the regions' spatial identities was weak and divergent, with 2019 marking a clear trough when differences peaked. However, in 2020 and 2021, the correlation increased markedly, indicating reduced heterogeneity in spatial identity between the regions. This convergence likely reflects the progressive implementation of the Guangdong-Hong Kong-Macau Greater Bay Area development. The spatial satisfaction dimension consistently showed higher correlation, suggesting minimal heterogeneity, while spatial imagery and spatial sense of belonging initially showed more pronounced differences that diminished over time, particularly during 2020-2021. Overall, although some differences remain, the spatial identities of Hong Kong and Shenzhen residents are gradually converging under the influence of policy initiatives and evolving cross-border dynamics.

5.3 Spatial Heterogeneity

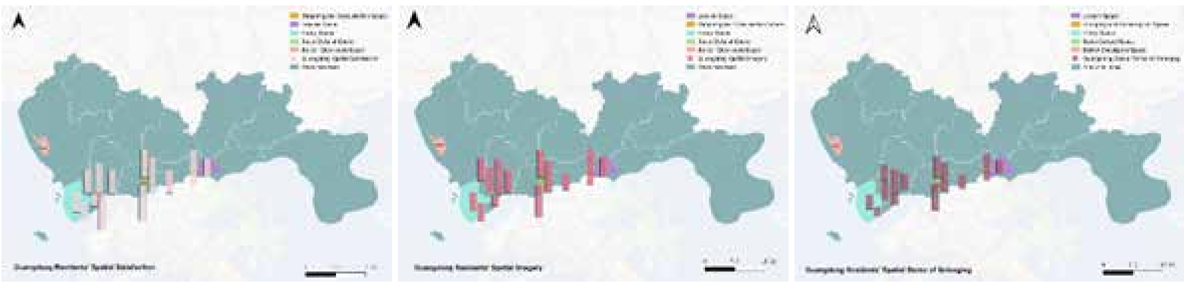
The study categorized cross-border activity spaces into five types based on spatial functions: shopping and consumption, socio-cultural, leisure, border checkpoint, and policy spaces. Figure. 6a illustrates the geographical distribution of Hong Kong residents' cross-border spatial identity perceptions in Shenzhen, with colors denoting spatial types and bar heights representing identity strength. Evidently, Hong Kong residents' spatial identity varies with location types.

The study summarizes the spatial identity of Guangdong and Hong Kong residents across different spatial types. Hong Kong residents display distinct identity characteristics by function. In border checkpoint spaces, their spatial identity scores were low across all dimensions (0.1, 0.21, 0.05), significantly lower than in other spaces. This suggests that, although border checkpoint spaces are essential for cross-border movement, their limited facilities and services fail to fully meet cross-border needs. However, Hong Kong residents' spatial sense of belonging in these spaces was slightly higher than that of Guangdong residents, reflecting a stronger functional dependence. In contrast, shopping and consumption as well as socio-cultural spaces elicited the highest spatial identity scores among Hong Kong residents. In shopping and consumption spaces, scores peaked for spatial imagery (0.99), spatial satisfaction (1.33), and spatial sense of belonging (0.92), suggesting that these spaces best meet Hong Kong residents' needs for convenience and diverse facilities (Wang, 2004; Sharma, Chen and Luk, 2018). Shenzhen's advanced shopping services foster positive spatial imagery and, through routine activities, enhance emotional attachment and spatial belonging. Conversely, policy spaces received relatively lower identity scores, likely because their services primarily cater to mainland China.

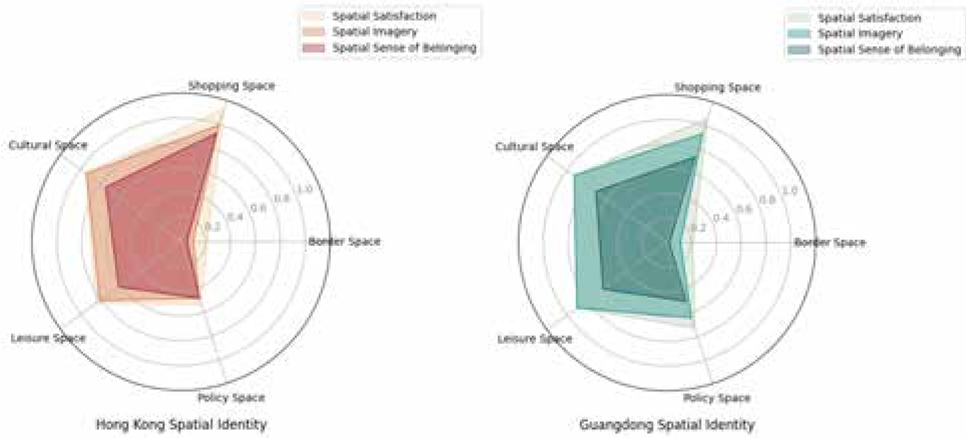
Overall, Hong Kong residents demonstrate the strongest emotional attachment to shopping and cultural spaces, while their attachment to border and policy spaces is weaker. Guangdong residents demonstrate a similar trend. These findings marked significant spatial heterogeneity in cross-border spatial identity, suggesting service and facility imbalances that offer a comprehensive geographic perspective for policy development.



a. Geographic distribution of Hong Kong spatial identities in various types of cross-border spaces



b. Geographic distribution of Guangdong spatial identities in various types of cross-border spaces



c. Hong Kong v.s. Guangdong Residents' Three-dimensional Spatial Identity during 2018-2023

Figure. 6 a. Geographic distribution of spatial identities in various types of cross-border spaces; b. Radar map of three-dimensional spatial identity perceptions of Hong Kong and Guangdong residents.

5.4 Qualitative Attribution Analysis

The study employed LLM's long-text mining and summarization capabilities to conduct an attribution analysis on blog posts with varying identity scores among Hong Kong residents. A semi-structured prompt process was established to investigate the drivers of spatial identity. The results indicate that although there some common factors significant differences emerge concerning spatial functions.

Negative spatial identity among Hong Kong residents was chiefly driven by functional deficiencies and inadequate management services. Specifically, issues such as traffic congestion, overcrowding, and poor spatial layouts were prevalent in border checkpoint, shopping and consumption, and leisure spaces (e.g., "Futian checkpoint is too crowded," "Shenzhen Bay Park... the long queues and traffic jams are exhausting," "The mall's layout is poor, I always get lost," "During National Day, I couldn't even get on the subway"). These factors diminished spatial identity perceptions. Moreover, each space type presented specific problems. For example, negative reviews of border checkpoint spaces cited problematic customs policies and poor transportation hub management, whereas critiques of shopping and leisure spaces focused on product quality and service attitudes (e.g., "The crabs bought in Shenzhen are half the price of those in Hong Kong").

In contrast, positive spatial identity was more consistent, driven by modern infrastructure, comprehensive services, diverse consumer experiences, and rich cultural activities. Improved facilities at some border checkpoint spaces enhanced spatial satisfaction ("Luo Hu station's free dedicated passage always has enthusiastic and polite staff"), and shopping and consumption spaces were praised for competitive pricing and products variety (e.g., "Prices are half of what they are in Hong Kong"). Similarly, leisure spaces were appreciated for natural scenery and outdoor activities, boosting both spatial imagery and satisfaction (e.g., "The beautiful view at Wutong Mountain helped me relax"). Socio-cultural spaces fostered a sense of participation and belonging (e.g., "The cultural atmosphere here makes me feel at home"), and policy spaces reinforced confidence in regional integration (e.g., "The future of the Lok Ma Chau Loop area will be prosperous!").

Additionally, the spatial drivers of three emotional evaluations reflect not only functional attributes but also subjective perceptual differences in cross-border activities. For example, the impact of border-crossing efficiency on spatial identity varies with time and crowd density, while evaluations of shopping spaces involved interactions among price, layout, and service quality. Cultural activities and natural landscapes uniquely contributed to positive identity formation.

Based on these findings, the study further validates the accuracy of the LLM's attribution extraction by analyzing original negative posts for the presence of identified factors. Frequency data in Table 3 supports the reliability of the text analysis; common factors such as traffic congestion, overcrowding, and disordered spatial layouts were frequently noted. Notably, in border checkpoint spaces, complex customs procedures (32.26%) and poor transportation hubs service (4.74%), were prominent. Similarly, issues of poor product quality and inadequate service in shopping and leisure spaces were substantiated. Overall, these results confirm the effectiveness of the LLM in long-text analysis and support the credibility of the spatial attribution analysis.

Table 3. Frequency of occurrence of spatial identity-driven elements in the original blog post

Space Type	Posts No.	Traffic congestion	Pedestrian congestion	Confused spatial layout	Poor quality of goods	Poor services
Political	1359	0.37%	2.87%	0.66%	1.25%	0.59%
Cultural	2061	0.92%	8.05%	1.36%	0.87%	0.63%
Leisure	2110	0.95%	6.59%	0.57%	1.09%	0.90%
Consumption	2191	0.27%	10.45%	1.14%	2.83%	1.51%
	No. Posts	Traffic congestion	Pedestrian congestion	Confused spatial layout	Cumbersome clearance policies	Poor services at transport hubs
Border	10078	4.20%	15.24%	1.88%	32.26%	4.74%

In conclusion, LLM-based long-text mining revealed the key spatial drivers of diverse perceptions of Shenzhen's cross-border spaces. Specifically, functional factors (border-crossing efficiency, transportation convenience, and facility layout) and experiential factors (cultural activities, consumer experiences, and natural landscapes) emerged as critical. Negative emotions were primarily linked to functional deficiencies, while positive emotions were correlated with high-quality spatial attributes and experiential benefits. Analyzing original blog posts further improved the interpretability and reliability of cross-border spatial identity, offering valuable insights for spatial planning and management.

6. Conclusion & Discussion

As a critical psychological mechanism underlying cross-border activities, cross-border spatial identity not only shapes residents' cross-border behaviors but also exerts a profound impact on the planning and construction of cross-border spaces and the process of regional integration. However, measuring it is challenging due to conceptual ambiguity, contextual complexity, data scarcity, and methodological limitations. The study develops a systematic, scalable, and data-driven approach to quantifying cross-border spatial identity, integrating social media data to capture the perspectives and emotions of cross-border residents while leveraging large language models (LLMs) to interpret spatial identity perceptions. We utilize the framework to examine its temporal evolution, spatial and demographic heterogeneity in Shenzhen-Hong Kong cross-border areas. The findings contribute to a deeper understanding of intra-regional connectivity and residents' sense of belonging, providing a theoretical foundation for cross-border spatial planning and development.

This study introduces a precise, efficient, and scalable framework for the long-term tracking and recognition of cross-border spatial identity, validating the effectiveness of LLMs in integrating quantitative and qualitative approaches for measuring the perception. Traditional methods like surveys are limited by small samples, high costs, low temporal resolution, and inability to capture dynamic evolution. Traditional machine learning models also struggle with the abstract nature of spatial identity, hindering precise identification and quantification. Moreover, they often neglect socio-cultural nuances, require language-specific retraining, and demand extensive labeled data, limiting scalability. In contrast, our LLM-based approach leverages zero-shot learning and advanced text comprehension to analyze complex semantics. First, by embedding a detailed definition of cross-border spatial identity via prompt engineering, and using a multilingual LLM supporting both Chinese and English, we achieve unified text recognition across different languages. Subsequently, validation against expert annotations confirmed the LLM's effectiveness in analyzing complex spatial identity-related text and performing fine-grained sentiment classification. The findings underscore the substantial advantages of LLMs in analyzing complex emotions and multidimensional perceptual constructs, capturing nuanced emotional fluctuations and variations in identity. Particularly in cross-cultural and cross-regional sentiment analysis, LLMs have demonstrated unique potential, significantly enhancing the depth and breadth of cross-border spatial identity measurement.

Also, the study uses long-time-series social media data to reveal the dynamic evolution of cross-border spatial identity in the Shenzhen-Hong Kong cross-border areas. By analyzing residents' perceptions over the period from 2018 to 2023, the study captures shifts in cross-border social relations and regional integration. This longitudinal perspective enables identification of key regional events (e.g. the COVID-19 pandemic, the extradition bill protests, and major policy changes) influencing cross-border spatial identity, aligning closely with prior research (Guinjoan and Rodon, 2014; Borz, Brandenburg and Mendez, 2022). Unlike static cross-sectional studies, our longitudinal approach identifies cyclical patterns, sudden disruptions, and gradual transformations in cross-border spatial identity over time. By capturing the interplay between structural factors (e.g., policy shifts, infrastructure development) and individual experiences, this approach facilitates a deeper exploration of the causal mechanisms underlying changes in cross-border spatial identity.

Additionally, the study explores the differing cross-border spatial identity between Guangdong and Hong Kong residents. The findings indicate that Guangdong residents exhibit significantly higher spatial identity than Hong Kong counterparts. This dual-perspective analysis reveals distinct identity preferences and underscores the need to accommodate diverse demands in cross-border spatial planning. Addressing these variations requires a nuanced understanding of the unique socio-cultural characteristics of border regions for targeted and equitable policy interventions.

Furthermore, the study identifies significant heterogeneity in cross-border spatial identity across different functional spaces. For Hong Kong residents, border checkpoint areas exhibit the lowest levels, whereas shopping and consumption spaces generate the highest. These variations across spatial dimensions provide a deeper understanding of identity differentiation in Shenzhen's cross-border spaces and support empirically grounded optimization of spatial design and policy formulation. For example, areas with low spatial identity, such as border checkpoint spaces may require transportation infrastructure and efficiency improvements. At the same time, for specific groups (such as people with disabilities and business commuters), fast-track channels and convenient services should be added to improve travel experience. While areas with high spatial identity like socio-cultural and leisure spaces could benefit from optimizing spatial design and environmental quality, introducing more cultural integration elements, and diversifying activity forms. For shopping and consumption spaces, where spatial identity is stronger, it is important to maintain advantages in pricing, product quality, and services, while further exploring the diversity of spatial functions. For example, integrating dining, entertainment, and other functions to expand the consumption experience can enhance the overall attractiveness and impact of cross-border spaces. By tailoring interventions to specific functional space characteristics, overall spatial identity can be enhanced, promoting the integrated development of Shenzhen-Hong Kong cross-border spaces.

This study also has certain limitations that warrant further investigation. First, the proposed framework for cross-border spatial identity recognition may be affected by LLM hallucinations in text reasoning and generation, stemming from inherent stochastic processes in the LLM architecture. This variability could lead to inconsistencies in sentiment evaluation, so future studies should conduct comprehensive validation and performance assessments of LLMs to ensure their reliability, applicability, and robustness in spatial identity measurement. Secondly, the reliance on social media data, which predominantly represents younger demographics, may introduce bias due to limited participation from children and older adults (Marchi et al., 2022). To mitigate this, future studies should incorporate complementary data sources—such as surveys, interviews, and other additional textual datasets—to achieve a more comprehensive understanding of..... Finally, the linguistic similarities between Cantonese and Mandarin mean that language distinctions have not been fully explored in this study. Consequently, the extent to which LLMs can effectively handle complex cross-linguistic variations remains insufficiently tested. Future research should expand the geographical scope to include multilingual and multicultural border regions, enabling a more rigorous evaluation of the applicability and robustness of LLM-based spatial identity measurement across diverse linguistic contexts.

References

Arbab, P., Azizi, M.M. and Zebardast, E. (2018) 'Place-identity formation in new urban developments of Tehran metropolis', Journal of Urban Planning and Development, 144(2), p. 5018010. Available at: https://doi.org/10.1061/(ASCE)UP.1943-5444.0000446.

Blajenkova, O., Kozhevnikov, M. and Motes, M.A. (2006) 'Object-spatial imagery: A new self-report imagery questionnaire', Applied Cognitive Psychology, 20(2), pp. 239–263. Available at: https://doi.org/10.1002/acp.1182.

Bohnet, B. et al. (2023) 'Attributed question answering: evaluation and modeling for attributed large language models'. arXiv. Available at: https://doi.org/10.48550/arXiv.2212.08037.

Borz, G., Brandenburg, H. and Mendez, C. (2022) 'The impact of EU cohesion policy on european identity: A comparative analysis of EU regions', European Union Politics, 23(2), pp. 259–281. Available at: https://doi.org/10.1177/14651165221076444.

Breakwell, G. (1993) 'Social representations and social identity', Papers on social representations, 2, pp. 198–217.
Brigola, R. (2024) 'VIVE: an LLM-based approach to identifying and extracting context-specific personal values from text'.

Casakin, H., Hernández, B. and Ruiz, C. (2015) 'Place attachment and place identity in israeli cities: The influence of city size', Cities, 42, pp. 224–230. Available at: https://doi.org/10.1016/j.cities.2014.07.007.

Chow, K. and Healey, M. (2008) 'Place attachment and place identity: first-year undergraduates making the transition from home to university', Journal of Environmental Psychology, 28(4), pp. 362–372.

De Paoli, S. (2024) 'Performing an inductive thematic analysis of semi-structured interviews with a large language model: an exploration and provocation on the limits of the approach', Social Science Computer Review, 42(4), pp. 997–1019. Available at: https://doi.org/10.1177/08944393231220483.

Dengel, A. et al. (2023) 'Qualitative research methods for large language models: conducting semi-structured interviews with ChatGPT and BARD on computer science education', in Informatics. MDPI, p. 78. Available at: https://www.mdpi.com/2227-9709/10/4/78 (Accessed: 11 December 2024).

Digital Report (2020) Digital 2020 hong kong (january 2020) v01, SlideShare. Available at: https://www.slideshare.net/slideshow/digital-2020-hong-kong-january-2020-v01/227793853 (Accessed: 10 December 2024).

Dorfman, A. et al. (2021) 'Social spatial cognition', Neuroscience & Biobehavioral Reviews, 121, pp. 277–290. Available at: https://doi.org/10.1016/j.neubiorev.2020.12.023.

Felasari, S. et al. (2017) 'Exploring collective memory and place attachment using social media data', in IOP Conference Series: Earth and Environmental Science. IOP Publishing, p. 12015. Available at: https://iopscience.iop.org/article/10.1088/1755-1315/99/1/012015/meta (Accessed: 21 November 2024).

Fikfak, J. (2009) 'Cultural and social representations on the border: From disagreement to coexistence', Human affairs, (4), pp. 350–362. Available at: https://doi.org/10.2478/v10023-009-0049-1.

Fragkathoulas, C. and Chlapanis, O.S. (2024) 'Local explanations and self-explanations for assessing faithfulness in black-box LLMs'. Available at: https://doi.org/10.1145/3688671.3688775.

Gandhi, K. et al. (2024) 'Human-like affective cognition in foundation models'. arXiv. Available at: http://arxiv.org/abs/2409.11733 (Accessed: 7 October 2024).

Gao, Y. et al. (2022) 'Measuring urban sentiments from social media data: a dual-polarity metric approach', Journal of Geographical Systems, 24(2), pp. 199–221. Available at: https://doi.org/10.1007/s10109-021-00369-z.

Gardner, J.A. and Richards, P. (2017) 'The spatiality of boundary work: Political-administrative borders and maya-mam collective identification', Social Problems, 64(3), pp. 439–455. Available at: https://doi.org/10.1093/socpro/spw047.

Guinjoan, M. and Rodon, T. (2014) 'Beyond identities: Political determinants of support for decentralization in contemporary spain', Regional & Federal Studies, 24(1), pp. 21–41. Available at: https://doi.org/10.1080/13597566.2013.818980.

Hay, Robert (1998) 'A rooted sense of place in cross-cultural perspective', Canadian Geographies / Géographies canadiennes, 42(3), pp. 245–266. Available at: https://doi.org/10.1111/j.1541-0064.1998.tb01894.x.

Hay, R (1998) 'Sense of place in developmental context', Journal of Environmental Psychology, 18(1), pp. 5–29. Available at: https://doi.org/10.1006/jevp.1997.0060.

He, S. et al. (2023) 'COVID-19 impacts on cross-border mobility of senior population between shenzhen and Hong Kong', Frontiers in Public Health, 11, p. 1285288. Available at: https://doi.org/10.3389/fpubh.2023.1285288.

He, X. (2024) 'Sentiment classification of social media user comments using SVM models', in 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE, pp. 1755–1759. Available at: https://doi.org/10.1109/AINIT61980.2024.10581547.

Hernández, B. et al. (2007) 'Place attachment and place identity in natives and non-natives', Journal of Environmental Psychology, 27(4), pp. 310–319. Available at: https://doi.org/10.1016/j.jenvp.2007.06.003.

Hu, N. et al. (2024) 'Benchmarking large language models in complex question answering attribution using knowledge graphs'. arXiv. Available at: https://doi.org/10.48550/arXiv.2401.14640.

Jang, K.M. and Kim, Y. (2019) 'Crowd-sourced cognitive mapping: a new way of displaying people's cognitive perception of urban space', PLOS One. Edited by R. Schifanella, 14(6), p. e0218590. Available at: https://doi.org/10.1371/journal.pone.0218590.

Kajić, S., Bogdanić, M. and Fuerst-Bjeliš, B. (2022) 'Syrmia – rethinking the regional identity of a (trans)border region: perception, self-identification, and place attachment', in O.-R. Ilovan and I. Markuszewska (eds) Preserving and Constructing Place Attachment in Europe. Cham: Springer International Publishing (GeoJournal Library), pp. 219–238. Available at: https://doi.org/10.1007/978-3-031-09775-1_13.

Knez, I. (2005) 'Attachment and identity as related to a place and its perceived climate', Journal of Environmental Psychology, 25(2), pp. 207–218. Available at: https://doi.org/10.1016/j.jenvp.2005.03.003.

Kojima, T. et al. (2022) 'Large language models are zero-shot reasoners', Advances in neural information processing systems, 35, pp. 22199–22213.

Kong, L. et al. (2022) 'How do different types and landscape attributes of urban parks affect visitors' positive emotions?', Landscape and Urban Planning, 226, p. 104482. Available at: https://doi.org/10.1016/j.landurbplan.2022.104482.

Lai, V.D. et al. (2023) 'ChatGPT beyond english: towards a comprehensive evaluation of large language models in multilingual learning'. arXiv. Available at: https://doi.org/10.48550/arXiv.2304.05613.

Lalli, M. (1992) 'Urban-related identity: theory, measurement, and empirical findings', Journal of Environmental Psychology, 12(4), pp. 285–303. Available at: https://doi.org/10.1016/S0272-4944(05)80078-7.

Laven, D. and Baycroft, T. (2008) 'Border regions and identity', European Review of History: Revue Européenne D'histoire, 15(3), pp. 255–275. Available at: https://doi.org/10.1080/13507480802082581.

Li, Q.-Q. et al. (2022) 'Towards a new paradigm for segregation measurement in an age of big data', Urban Informatics, 1(1), p. 5. Available at: https://doi.org/10.1007/s44212-022-00003-3.

Liu, M. and Shi, G. (2024) 'PoliPrompt: A High-Performance Cost-Effective LLM-Based Text Classification Framework for Political Science', arXiv preprint arXiv:2409.01466 [Preprint]. Available at: https://arxiv.org/abs/2409.01466 (Accessed: 8 October 2024).

Luo, J.M. and Lam, C.F. (2020) 'Travel anxiety, risk attitude and travel intentions towards "travel bubble" destinations in Hong Kong: effect of the fear of COVID-19', International Journal of Environmental Research and Public Health, 17(21), p. 7859. Available at: https://doi.org/10.3390/ijerph17217859.

Mandhasiya, D.G. et al. (2022) 'Evaluation of machine learning performance based on BERT data representation with LSTM model to conduct sentiment analysis in indonesian for predicting voices of social media users in the 2024 indonesia presidential election', 2022 5th International Conference on Information and Communications Technology (ICOIACT), pp. 441–446. Available at: https://doi.org/10.1109/ICOIACT55506.2022.9972206.

Marchi, V. et al. (2022) 'Attitudes towards urban green during the COVID-19 pandemic via twitter', Cities, 126, p. 103707. Available at: https://doi.org/10.1016/j.cities.2022.103707.

Mutanov, G., Karyukin, V. and Mamykova, Z. (2021) 'Multi-class sentiment analysis of social media data with machine learning algorithms', Computers, Materials & Continua, 69(1), pp. 913–930. Available at: https://doi.org/10.32604/cmc.2021.017827.

Newman, D. and Paasi, A. (1998) 'Fences and neighbours in the postmodern world: boundary narratives in political geography', Progress in Human Geography, 22(2), pp. 186–207. Available at: https://doi.org/10.1191/030913298666039113.

Pedersen, H.D. (2018) 'Touched by place: identity construction, place attachment and place attractiveness from a life course perspective in the context of migration'. Available at: https://vbn.aau.dk/en/publications/touched-by-place-identity-construction-place-attachment-and-place (Accessed: 21 November 2024).

Peng, J., Strijker, D. and Wu, Q. (2020) 'Place identity: how far have we come in exploring its meanings?', Frontiers in Psychology, 11(294), p. 294. Available at: https://doi.org/10.3389/fpsyg.2020.00294.

Proshansky, H.M., Fabian, A.K. and Kaminoff, R. (1983) 'Place-identity: physical world socialization of the self', Journal of Environmental Psychology, 3(1), pp. 57–83. Available at: https://doi.org/10.1016/S0272-4944(83)80021-8.

Qian, J. and Zhu, H. (2014) 'Chinese urban migrants' sense of place: Emotional attachment, identity formation, and place dependence in the city and community of guangzhou', Asia Pacific Viewpoint, 55(1), pp. 81–101. Available at: https://doi.org/10.1111/apv.12039.

Randl, K. et al. (2024) 'Evaluating the reliability of self-explanations in large language models'. arXiv. Available at: https://doi.org/10.48550/arXiv.2407.14487.

Rathje, S. et al. (2024) 'GPT is an effective tool for multilingual psychological text analysis', Proceedings of the National Academy of Sciences, 121(34), p. e2308950121. Available at: https://doi.org/10.1073/pnas.2308950121.

Raymond, C.M., Brown, G. and Weber, D. (2010) 'The measurement of place attachment: personal, community, and environmental connections', Journal of Environmental Psychology, 30(4), pp. 422–434. Available at: https://doi.org/10.1016/j.jenvp.2010.08.002.

Resch, B. et al. (2016) 'Citizen-centric urban planning through extracting emotion information from twitter in an interdisciplinary space-time-linguistics algorithm', Urban Planning, 1(2), pp. 114–127. Available at: https://doi.org/10.17645/up.v1i2.617.

Rishbeth, C. and Powell, M. (2013) 'Place attachment and memory: landscapes of belonging as experienced post-migration', Landscape Research, 38(2), pp. 160–178. Available at: https://doi.org/10.1080/01426397.2011.642344.

Sharma, P., Chen, I.S.N. and Luk, S.T.K. (2018) 'Tourist shoppers' evaluation of retail service: a study of cross-border versus international outshoppers', Journal of Hospitality & Tourism Research, 42(3), pp. 392–419. Available at: https://doi.org/10.1177/1096348015584439.

Wan, M. et al. (2024) 'TnT-LLM: text mining at scale with large language models', in Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Barcelona Spain: ACM, pp. 5836–5847. Available at: https://doi.org/10.1145/3637528.3671647.

Wang, D. (2004) 'Hong kongers' cross-border consumption and shopping in shenzhen: patterns and motivations', Journal of retailing and consumer services, 11(3), pp. 149–159. Available at: https://doi.org/10.1016/S0969-6989(03)00014-6.

Wang, X. et al. (2023) 'Self-consistency improves chain of thought reasoning in language models'. arXiv. Available at: https://doi.org/10.48550/arXiv.2203.11171.

Williams, D.R. and Vaske, J.J. (2003) 'The measurement of place attachment: validity and generalizability of a psychometric approach', Forest Science, 49(6), pp. 830–840. Available at: https://doi.org/10.1093/forestscience/49.6.830.

Xie, T. et al. (2020) 'Parallel evolution and response decision method for public sentiment based on system dynamics', European Journal of Operational Research, 287(3), pp. 1131–1148. Available at: https://doi.org/10.1016/j.ejor.2020.05.025.

Xing, F. (2024) 'Designing heterogeneous LLM agents for financial sentiment analysis', ACM Transactions on Management Information Systems, p. 3688399. Available at: https://doi.org/10.1145/3688399.

Yang, J. et al. (2024) 'Intercode: Standardizing and benchmarking interactive coding with execution feedback', Advances in Neural Information Processing Systems, 36. Available at: https://proceedings.neurips.cc/paper_files/paper/2023/hash/4b175d846fb008d540d233c188379ff9-Abstract-Datasets_and_Benchmarks.html (Accessed: 10 February 2025).

Zhou, C. et al. (2023) 'Investigating the dynamicity of sentiment predictors in urban green spaces: a machine learning-based approach', Urban Forestry & Urban Greening, 89, p. 128130. Available at: https://doi.org/10.1016/j.ufug.2023.128130.

Zhou, J. et al. (2024) 'Measurement and implications of "people-place" emotional association in urban blocks based on large language models: a case study of duolun road block in shanghai (In chinese)', Shanghai Urban Planning, (3), pp. 102–108.