Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03645-7 OPEN

Place identity: a generative AI's perspective

Kee Moon Jang[1], Junda Chen[2], Yuhao Kang[1,3,4]✉, Junghwan Kim[5], Jinhyung Lee[6], Fabio Duarte[1] & Carlo Ratti[1]

Do cities have a collective identity? The latest advancements in generative artificial intelligence (AI) models have enabled the creation of realistic representations learned from vast amounts of data. In this study, we test the potential of generative AI as the source of textual and visual information in capturing the place identity of cities assessed by filtered descriptions and images. We asked questions on the place identity of 64 global cities to two generative AI models, ChatGPT and DALL·E2. Furthermore, given the ethical concerns surrounding the trustworthiness of generative AI, we examined whether the results were consistent with real urban settings. In particular, we measured similarity between text and image outputs with Wikipedia data and images searched from Google, respectively, and compared across cases to identify how unique the generated outputs were for each city. Our results indicate that generative models have the potential to capture the salient characteristics of cities that make them distinguishable. This study is among the first attempts to explore the capabilities of generative AI in simulating the built environment in regard to place-specific meanings. It contributes to urban design and geography literature by fostering research opportunities with generative AI and discussing potential limitations for future studies.

[1] Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA. [2] Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, USA. [3] GISense Lab, Department of Geography and the Environment, The University of Texas at Austin, Austin, TX, USA. [4] Department of Geography, University of South Carolina, Columbia, SC, USA. [5] Department of Geography, Virginia Tech, Blacksburg, VA, USA. [6] Department of Geography and Environment, Western University, London, ON, Canada. ✉email: yuhao.kang@austin.utexas.edu

Introduction

Place identity, as introduced by Proshansky et al. (1983: p.59), refers to the "sub-structure of the self-identity of the person consisting of broadly conceived cognitions about the physical world in which the individual lives". Emerging from an environmental psychology standpoint, such a traditional definition emphasizes an individual's socialization with the physical environment through a complex interaction of cognition, perception, and behavior to form an identity within their surroundings. Since its introduction, the notion of place identity has expanded to describe the people-place relationship, resulting in parallel terms such as place attachment, place uniqueness or sense of place. In particular, an important distinction has been made between people's identity *with* place and identity *of* place, which refers to properties that distinguish a place from others (Peng et al. 2020; Relph 1976). Further, a shifted focus toward the latter has offered insights into what features construct distinctive place identities in fields of urban design, geography and tourism (Larsen 2004; Lewicka 2008; Paasi 2003; Wang & Chen 2015). Despite the inherent vagueness in formalizing these concepts, prior studies have pointed out that physical settings, events that take in space, and associated individual (or group) meanings are key elements that shape distinctive place identities (Relph 1976; Seamon & Sowers 2008).

As an attempt to establish the theoretical foundation of place research, the notion of *place* has been discussed in contrast to *space*. Tuan defined the distinction between the two important concepts in human geography; *space* is an abstract physical environment that lacks substantial meaning, whereas *place* is a "center of felt value" (Tuan 1977) that is given meaning through human experience. Consequently, recognizing such place characteristics has been crucial to link individual behaviors to their surrounding environment and offered indicators for measuring urban form, function, emotion, and quality of life in cities (Gao et al. 2022; Nasar 1990). Prior studies have highlighted the benefits of understanding place identity in facilitating planning processes to create livable and legible places. By designing such places, individuals may develop a sense of attachment to their urban communities and cultivate environmentally friendly attitudes that are conducive to sustainability (Hernandez et al. 2010; Manzo & Perkins 2006). Thus, an important challenge in placemaking is to build physical as well as visual features that can trigger stronger subjective attachments to a place.

Despite its significance, measuring place identity has been a difficult task due to its intrinsically obscure and subjective nature (Goodchild 2010; Peng et al. 2020). Conventional studies attempted to capture built environment characteristics and human perceptions through qualitative research techniques. For instance, Hull et al. (1994) conducted a phone interview on the damaged place identity of Charleston, South Carolina after Hurricane Hugo, and Stewart et al. (2004) employed photo-elicitation, participant-employed photography followed by interviews to understand how residents' representation of their community identity can help shape visions for landscape change. Another stream of research explored the role of identity markers, such as towers, street signs, region names and (non)commercial establishments, in reflecting the unique identities of a place (Peng et al. 2020). However, such qualitative approaches pose limitations in terms of time and cost efficiency, where limited sample sizes may lead to biased results.

With the emergence of various user-generated contents, researchers have been leveraging these new data sources to understand the meaningful collective place identity of cities (Jang & Kim 2017). In particular, text and images have been the two most widely used data formats to advance our knowledge of place identity. Previous studies have employed natural language processing (NLP) methods such as sentiment analysis and topic modeling to process text-based datasets and understand individuals' opinions and emotions of places from online text corpora (Gao et al. 2017; Hu et al. 2019). In parallel, computer vision (CV) approaches have been effectively used to extract visual information about places from street-level images and geotagged photos (Kang et al. 2019; Liu et al. 2017; Zhang et al. 2018, 2019), which offer valuable insights to advance our understanding of place.

Recently, advancements in generative artificial intelligence (GenAI) have received significant attention due to their capabilities to generate realistic text and image outputs supported by large language models (LLM). Built on billions of inputs and parameters, researchers have noted that users can overcome the language barriers through GenAI by obtaining results that can be applied across diverse populations and settings (Gottlieb et al. 2023; Sajjad & Saleem 2023). The current advancements of GenAI have enabled people to communicate and interact with ChatGPT (OpenAI 2023) naturally and can generate vivid images given certain prompts with DALL·E2 (Mishkin et al. 2022). These GenAI models have been highlighted as powerful tools with potential for a wide variety of applications in different domains, including, transportation (Kim & Lee 2023), education (Latif et al. 2023), climate literacy (Atkins et al. 2024) and geospatial artificial intelligence (Mai et al. 2023).

In the meantime, researchers are wary of the inattentive use of GenAI tools despite its potential benefits and versatility across fields. As Shen et al. (2023) describes, LLMs may become a double-edged sword that produces plausible but logically incorrect results. For such misinformation being produced, Van Dis et al. (2023) pointed out the absence of relevant data in the training set of LLMs. The output quality in terms of accuracy and bias may heavily rely on the information that was included for training. Therefore, it is essential to acknowledge and address the ethical and societal concerns of these models that stem from the lack of transparency (Dwivedi et al. 2023; Kang et al. 2023).

While creative jobs were considered safe from technological innovations until now, compared to those of routine and repetitive tasks (Ford 2015), the emergence of GenAI is turning things around. Although concerns remain about the ethics and disruptive impact of their usage, generative models would inevitably replace or, at least, assist content generation in creative industries (Anantrasirichai & Bull 2022; Lee 2022; Turchi et al. 2023). Design fields are not an exception—architectural firms are nowadays utilizing AI-assisted tools to generate 100,000 designs per day for their building projects (see Supplementary Note). Researchers have also investigated the capability of various text-to-image generators to assist the initial process of architectural design (Paananen et al. 2023). Additionally, recent urban studies have explored the potential of GenAI in evaluating design qualities of the built environment scenes and obtaining optimal land-use configuration through automated urban planning process (Seneviratne et al. 2022; Sun & Dogan 2023; Wang et al. 2023).

Creating design alternatives, however, has been a *space*-making, rather than a *place*-making, approach; it has leaned towards the simulation of physical forms of the built environment with less consideration of the surrounding contexts. Paananen et al. (2023) argued that generative systems have mostly been used to represent the geometry of architecture, such as façade, form, and layout, while its conceptual creativity remains to be studied. DALLE-URBAN has demonstrated the potential of GenAI for effectively creating urban scenes, but fell short in depicting composition and locales for specific conditions (Seneviratne et al. 2022). Furthermore, Bolojan et al. (2022) called for the need to consider how human perception works in the computational

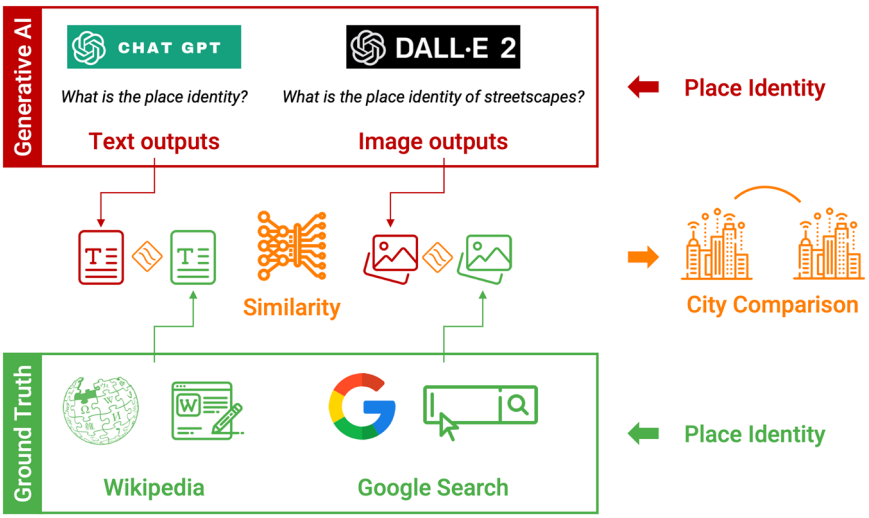


Fig. 1 The computational framework of this paper.

design workflows with GenAI models. Motivated by their potential, we raise the question: Can GenAI contribute to our understanding of place-specific contexts in a trustworthy manner?

GenAI has the potential to revolutionize the way we perceive the world and offer a new paradigm for urban studies. In particular, we intend to suggest a more proper use of GenAI in urban studies for creating *place* by bringing the people and meanings intertwined with human experience to the fore. To this end, we aim to examine the potential of GenAI as new tools for understanding the place identity *of* different cities. In this study, we ask the following two research questions: (1) How does generative AI illustrate place identity? (2) To what extent can we trust generative models in terms of their place identity results when compared with fact-based descriptions? To address these questions, we propose a computational framework to collect place identity with GenAI and evaluate the quality and trustworthiness of the data. We first asked a mixture of questions about the place identity of 64 global cities using two GenAI models, namely, ChatGPT for texts, and DALL·E2 for images. The cities were selected across 6 continents and 49 countries that represent diverse spatial coverages and contexts in order to better evaluate the performance of GenAI models at a global scale. Then, we collected two fact-based datasets as ground-truth data, including Wikipedia texts and images retrieved from Google search for comparison. Finally, we comprehensively evaluated the similarity between the AI-generated results and their fact-based counterparts.

Methods

We present a computational framework of this study in Fig. 1. The framework primarily involves two steps: exploring place identity with GenAI and validating results by comparing with real-world settings. For each step, two types of datasets, namely, text-based, and image-based datasets were created to investigate the potential of GenAI models in capturing place identity. In particular, we employed ChatGPT to generate text descriptions of cities; and we leveraged DALL·E2 to generate images of representative streetscapes of different cities. We further collected two datasets including a text dataset from Wikipedia and an image dataset from Google search for validating the results produced by the generative models. After that, we performed cross-validation to compare similarities among these datasets, analyzed the characteristics of place identity produced by GenAI, and evaluated whether the results provided can be trusted.

Understanding place identity with generative AI

Place identity from ChatGPT. We first asked ChatGPT with a prompt "What is the meaning of place identity" to confirm that its understanding of place identity is consistent with the notion of identity *of* place that is to be explored from the generated outputs. Then, we created a text-based dataset by asking ChatGPT to generate descriptions of the place identity of various cities around the world. To accomplish this, we developed a set of prompts using the following format:

- "What is the place identity of {*city*}? Give me in ten bullet points."
- "What is the urban identity of {*city*}? Give me in ten bullet points."
- "What is the place identity of streetscapes in {*city*}? Give me in ten bullet points."

The {*city*} includes a list of 64 global cities around the world. A full list of cities is in Table 1. The prompts we used allowed us to retrieve the specific place identity information we sought to generate from the AI model for each city. It should be noted that responses generated by ChatGPT may vary in length and style, despite using the same prompt format. To ensure consistency and comparability across different cities included in our dataset, we limited the responses to ten bullet points. By doing so, the generated outputs are concise and well-structured and can be easily analyzed and compared.

Place identity from DALL·E2. Similar to the text-based datasets, we created an image-based dataset using DALL·E2 to understand place identity. We aim to capture the visual representations of the built environment and streetscapes of each city, which are essential components of its place identity. To achieve this, we input the following prompt into DALL·E2 to generate representative streetscapes for each city:

- "What is the place identity of streetscapes of {*city*}?"

We generated 20 images for each city, where each image has a size of 256*256 pixels. By combining the image-based dataset with the text-based dataset, we aim to provide a comprehensive and multi-modal understanding of the place identity of each city captured by GenAI models.

Table 1 The full list of cities included in this study.

	City Name	Country	Population	Latitude	Longitude
1	Abu Dhabi	United Arab Emirates	1,362,000	24.45	54.38
2	Almaty	Kazakhstan	2,447,000	43.24	76.88
3	Amsterdam	Netherlands	1,736,000	52.37	4.90
4	Athens	Greece	3,309,000	37.98	23.73
5	Auckland	New Zealand	1,537,000	−36.85	174.76
6	Bangkok	Thailand	18,884,000	13.76	100.50
7	Barcelona	Spain	5,317,000	41.39	2.17
8	Beijing	China	18,883,000	39.90	116.41
9	Berlin	Germany	4,286,000	52.52	13.40
10	Bogota	Colombia	10,252,000	4.71	−74.07
11	Brussels	Belgium	2,238,000	50.85	4.36
12	Bucharest	Romania	2,097,000	44.43	26.10
13	Budapest	Hungary	2,407,000	47.50	19.04
14	Buenos Aires	Argentina	15,748,000	−34.60	−58.38
15	Busan	South Korea	3,843,000	35.21	129.07
16	Cairo	Egypt	22,679,000	30.04	31.24
17	Caracas	Venezuela	2,521,000	10.48	−66.90
18	Chicago	United States	8,954,000	41.88	−87.63
19	Copenhagen	Denmark	1,650,000	55.68	12.57
20	Delhi	India	31,190,000	28.70	77.10
21	Dubai	United Arab Emirates	4,945,000	25.20	55.27
22	Dublin	Ireland	1,386,000	53.35	−6.26
23	Helsinki	Finland	1,146,000	60.17	24.94
24	Ho Chi Minh City	Vietnam	14,953,000	10.82	106.63
25	Hong Kong	China	6,468,000	22.32	114.17
26	Istanbul	Turkey	14,441,000	41.01	28.98
27	Jakarta	Indonesia	35,386,000	−6.21	106.85
28	Johannesburg	South Africa	15,551,000	−26.20	28.05
29	Kobe	Japan	14,916,000	34.69	135.20
30	Kyoto	Japan	14,916,000	35.01	135.77
31	Lagos	Nigeria	14,540,000	6.52	3.38
32	Lisbon	Portugal	2,832,000	38.72	−9.14
33	London	United Kingdom	10,803,000	51.51	−0.13
34	Los Angeles	United States	15,587,000	34.05	−118.24
35	Madrid	Spain	6,798,000	40.42	−3.70
36	Manila	Philippines	24,156,000	14.60	120.98
37	Melbourne	Australia	4,709,000	−37.81	144.96
38	Mexico City	Mexico	21,905,000	19.43	−99.13
39	Milan	Italy	5,471,000	45.46	9.19
40	Montreal	Canada	3,750,000	45.50	−73.57
41	Moscow	Russia	17,878,000	55.76	37.62
42	Mumbai	India	25,189,000	19.08	72.88
43	Munich	Germany	2,112,000	48.14	11.58
44	New York	United States	21,396,000	40.71	−74.01
45	Osaka	Japan	14,916,000	34.69	135.50
46	Oslo	Norway	1,007,000	59.91	10.75
47	Paris	France	11,108,000	48.86	2.35
48	Prague	Czechia	1,240,000	50.08	14.44
49	Rio de Janeiro	Brazil	12,306,000	−22.91	−43.17
50	Rome	Italy	3,239,000	41.90	12.50
51	Santiago	Chile	7,099,000	−33.45	−70.67
52	Seoul	South Korea	23,225,000	37.55	126.99
53	Shanghai	China	24,042,000	31.23	121.47
54	Shenzhen	China	17,778,000	22.54	114.06
55	Singapore	Singapore	5,926,000	1.35	103.82
56	Stockholm	Sweden	2,200,000	59.33	18.07
57	Sydney	Australia	4,836,000	−33.87	151.21
58	São Paulo	Brazil	21,486,000	−23.56	−46.64
59	Tel Aviv	Israel	3,006,000	32.09	34.78
60	Tokyo	Japan	37,785,000	35.68	139.65
61	Toronto	Canada	6,837,000	43.65	−79.38
62	Vienna	Austria	2,030,000	48.21	16.37
63	Warsaw	Poland	2,028,000	52.23	21.01
64	Zurich	Switzerland	863,000	47.38	8.54

Collecting real-world settings

Text-based dataset from Wikipedia. Despite the high performance of ChatGPT in generating texts, researchers and the public have raised concerns regarding its reliability and trustworthiness (Shen et al. 2023). However, the subjective nature of place identity, which is intrinsically related to human experience and may vary across different individuals, poses a significant challenge in validating responses generated by ChatGPT. Moreover, the absence of a large-scale ground-truth place identity dataset further complicates the validation process. To address these challenges, we collected data from Wikipedia on the full list of cities as a source of textual introduction to each case. As Jenkins et al. describes, it is plausible to consider Wikipedia entries that are created through users' collaborative efforts as a collective perception of places with contents on main characteristics of different locations (Jenkins et al. 2016).

Image-based dataset from Google search. We further employed a Python web scraper to collect images of each city via Google Images (https://images.google.com/) search engine. Performing content analysis on images sampled from Google has been approved as an effective method to retrieve visual information on various places and thus represent place-specific meanings (Choi et al. 2007; Coghlan et al. 2017). In this study, this was accomplished by entering a search query in the format of "{*city*}", such as "*Singapore*". The top returned images appear based on their relevance to the search query, which we assume to reflect the place identity of that city. We then collected the top 30 image search results among all returned images for each query. By doing so, we were able to collect a representative set of images for each city, allowing us to compare with the outputs generated by DALL·E2.

Validation of place identity by generative AI

Measuring text similarity. To validate the place identity results generated by ChatGPT, we utilized a cross-validation approach after collecting the two text-based datasets from ChatGPT and Wikipedia. More specifically, we assessed the similarity between sentences generated by ChatGPT and the sentences in Wikipedia to determine whether the AI-generated results may capture and reflect place identity.

To achieve this, we first conducted data cleaning of the Wikipedia data to ensure that the text was in a clean format and could be processed further. We utilized the *tokenizer* function in the Natural Language Toolkit (NLTK) Python library to segment the corpus into individual sentences. To analyze the two text-based datasets and extract their semantics, we leveraged a sentence transformer BERT model (Devlin et al. 2018) based on a modified version of MiniLM (Wang et al. 2020). Such a model has been widely used in prior studies to convert each sentence in the Wikipedia corpus and each bullet point in ChatGPT responses into word embeddings to capture their underlying semantics. This model has been distilled for efficiency and fine-tuned on 1 trillion triples of annotated data, making it highly accurate in measuring short sentence topic similarity. By

inputting each sentence from ChatGPT responses and the Wikipedia corpus into such a sentence transformer BERT model, we transformed them into embedding vectors. Then, we measured cosine similarity for sentence embeddings from ChatGPT responses and Wikipedia corpuses to assess the relevance between the two datasets. This similarity score indicates the degree of relatedness between sentences through values ranging from 0 to 1. Specifically, for a sentence pair (one from ChatGPT and one from the Wikipedia corpus), higher similarity scores indicate that the ChatGPT response is highly relevant to a particular topic within the Wikipedia corpus, while lower scores denote that the response is not closely aligned with any topic in the corpus. To quantify the similarity, we iterated each bullet point in the ChatGPT responses and compared it to every sentence in the Wikipedia corpus. We identified the sentence in the Wikipedia corpus that had the highest similarity score in response to each bullet point. This allowed us to further quantify the overall similarity between the ChatGPT responses and the Wikipedia corpus.

In addition to the text similarity measurements, we also created word cloud images of each city based on ChatGPT-generated responses and the introduction from Wikipedia. A word cloud image offers a vivid graphical representation of text data, where the size of each word corresponds to its frequency in the given text. These word cloud images serve as visual representations of the topics covered in the texts of place identity, allowing for a comparison between outputs generated by ChatGPT and their corresponding Wikipedia introductions of each city.

Measuring image similarity. Similar to the comparison between ChatGPT-generated sentences with Wikipedia corpus, we also compared images generated by DALL·E2 and those collected from Google image search. We aim to evaluate the reliability and generative capability of the text-to-image model in producing realistic representations of place-specific scenes of cities. For this purpose, we adopt the Learned Perceptual Image Patch Similarity (LPIPS) to assess the perceptual similarity between AI-generated and real-world images (Zhang et al. 2018). This metric was evaluated against a large-scale dataset of human judgments on image pair similarity and found to outperform other perceptual similarity metrics. LPIPS computes the Euclidean distance between feature vectors of images extracted from a pretrained deep convolutional network for image classification. We employed AlexNet as the feature extractor for LPIPS calculation, which was tested to output the best performance. Noting that a lower LPIPS score indicates greater similarity, and vice versa, we defined the image similarity score ($S_{i,j}$) between any two images i and j as follows:

$$S_{i,j} = 1 - LPIPS_{i,j}$$

Subsequently, we compare each image generated by DALL·E2 with all images from the Google image search in the same city, and identify the three most similar images based on the image similarity scores. This allows us to quantitatively compare and determine whether the results generated by the text-to-image model are consistent with the real-world urban settings of each city.

In addition, considering the subjectivity of place identity, it is necessary to keep human-in-the-loop and involve human evaluations. Therefore, we conduct a survey specifically designed to collect human ratings on the similarity between DALL·E2-generated images and Google images. We aim to invite humans to evaluate whether the two images are similar or not. An image pair that is nearest to the mean of $S_{i,j}$ image similarity scores for each city is selected as the representative case to be included in the survey. Hence, respondents were provided with a total of 64 questions that asked about the similarity of a given pair rated using a 7-point Likert Scale. Then we ordered the 64 cities based on the mean values of human-rated similarity to see whether the GenAI-based images might be similar to those representative images.

Last, we measure city-by-city similarity to test whether GenAI can identify cities that are visually distinctive or similar. In order to perform this experiment, we calculate the normalized Chamfer distance (CD) between DALL·E2 generated outputs of two cities. CD is a similarity metric that measures the distance between point clouds of latent representations of images. The normalized CD value ranges between [0,1], and is subtracted from 1 so that higher value indicates higher similarity, and vice versa. Recollecting that place identity refers to properties that distinguish one place from another, we attempt to examine two hypotheses from the similarity of each city pair:

- Similarity between generated outputs of the same city is greater than that of different cities;
- Similarity between generated outputs is greater in cities that are geographically and culturally close than in cities that are geographically and culturally distant.

This pairwise comparison enables us to explore whether the represented traits captured by GenAI are place identity-related or generic characteristics of cities, providing valuable insights for urban design research and practice.

Results

In the following section, we will first showcase several example results created by GenAI to illustrate its potential in capturing place identity in urban studies. Although GenAI models have demonstrated outstanding capabilities in several tasks such as communication with people and painting creation, their trustworthiness has been questioned by researchers. The nature of GenAI models is that they do not produce outputs based on explicit logic but fit statistical distributions present in the training dataset. Consequently, there is a need to evaluate the trustworthiness of GenAI outputs, particularly with respect to place identity in this study, as they may contain 'fake' information and lead to erroneous decisions. However, the absence of a ground-truth place identity dataset at a global level poses a challenge.

To address this issue, we employed two datasets including Wikipedia introductions and Google images of cities, under the assumption that they reflect place identity as found from previous studies (Choi et al. 2007; Coghlan et al. 2017; Jenkins et al. 2016). We further tested whether outputs generated by GenAI models align with information retrieved from these datasets, which could be considered as a reliable representation of a city's place identity.

Results of place identity generated by ChatGPT. First, we explored whether ChatGPT understands place identity as the identity *of* place, as suggested by Relph (1976), which provides the conceptual foundation of this study to explore the place-specific meanings from generated outputs of different cities. In response to the prompt, "What is the meaning of place identity?", we obtained an answer (see Supplementary Fig. S2):

"*Place identity refers to the unique characteristics and meanings that individuals or groups associate with a particular location. It encompasses the social, cultural, historical, and environmental elements that shape people's perceptions and emotions about a place. Place identity is formed through experiences, interactions, and attachments to a specific geographical area, such as a neighborhood, city, or region*.

Table 2 Examples of {*place*}, {*urban*}, and {*street*} identity generated by ChatGPT.

City	Place identity	Urban Identity	Street Identity
Beijing	1. Capital of the People's Republic of China	1. Rich culture and history: Beijing has a long and storied history and is home to many important cultural sites, such as the Forbidden City and the Great Wall of China.	The streetscapes of Beijing are characterized by their traditional Chinese architecture, including historic hutongs, courtyards, temples, and other landmarks. The city is known for its rich cultural heritage, and its streetscapes are a reflection of this. The streetscapes of Beijing provide an insight into the city's past, and its unique cultural identity.
	2. Political, cultural, and educational center of the country	2. Bustling metropolis: Beijing is a bustling metropolis, with a population of over 21 million people and a thriving economy.	
	3. Home to many iconic landmarks, including the Forbidden City, Tiananmen Square, and the Great Wall	3. Food: Beijing is known for its delicious cuisine, with a variety of regional dishes and street food.	
	4. Host of the 2008 Summer Olympics	4. Shopping: Beijing is a great place to shop, with a wide range of international stores, malls, and markets.	
	5. Largest city in China, with over 20 million inhabitants	5. Nightlife: Beijing has a vibrant nightlife, with a variety of bars, clubs, and restaurants.	
	6. Home to some of the world's most renowned universities, including Tsinghua University and Peking University	6. Art and culture: Beijing is home to a number of museums, galleries, and theaters, and is host to many cultural events throughout the year.	
	7. A hub for international diplomacy and trade	7. Parks and outdoor spaces: Beijing is home to many parks and outdoor spaces, including the Summer Palace, Temple of Heaven, and Beihai Park.	
	8. A major transportation center, with three international airports	8. Transportation: Beijing has an extensive public transportation network, making it easy to get around the city.	
	9. A vibrant food culture, with a wide variety of regional cuisines	9. Weather: Beijing has a temperate climate, with hot summers and cold winters.	
	10. An ancient city with a rich cultural heritage, including the birthplace of Peking Opera	10. People: Beijing is a friendly and welcoming city, with a diverse population of people from all over the world.	
New York	1. Iconic skyline	1. Diverse and cosmopolitan population	The streetscapes of New York City are a reflection of its cultural diversity, its vibrant energy, and its history. They represent the hustle and bustle of the city, the iconic architecture, and the unique atmosphere that can only be found in New York. The streetscapes of New York are an integral part of its identity and provide a sense of place for its citizens and visitors.
	2. Home to the Statue of Liberty	2. Iconic skyline and landmarks	
	3. Home to the Empire State Building	3. Arts and culture hub	
	4. Cultural melting pot	4. Financial and business center	
	5. Financial capital of the world	5. Home to the fashion industry	
	6. Home to world-renowned museums	6. Vibrant nightlife	
	7. Home of Broadway and the theater district	7. Rich history and heritage	
	8. Home to world-famous restaurants	8. World-renowned cuisine	
	9. Shopping mecca	9. Home to the world's most famous sports teams	
	10. Home of the Yankees and other professional sports teams	10. International gateway and melting pot	

This concept suggests that places have distinct personalities or identities that are constructed through the collective experiences and memories of the people who inhabit or interact with them. These identities can influence how individuals relate to and interact with their surroundings, shaping their sense of belonging, attachment, and pride in a place.

Place identity is often closely linked to notions of community, heritage, and cultural identity, as well as to broader issues such as urban planning, environmental conservation, and tourism development. Understanding and preserving place identity is important for maintaining the unique character and authenticity of different locations and for fostering a sense of connection and belonging among residents and visitors alike."

From this response, we were able to verify that ChatGPT introduces place identity as unique characteristics and meanings of a particular location constructed by collective experiences of people. Also, it successfully described the benefits of place identity in fostering a sense of belonging and attachment among people, which coincides with our specific focus in the definition of place identity in this study.

Then, we retrieved place identity responses generated by ChatGPT through asking three different prompt keywords, {*place*}, {*urban*} and {*street*} identity. Table 2 presents examples of responses in two large metropolitan cities, namely, Beijing and New York. Intuitively, GenAI models capture various aspects that align with people's general perceptions of these cities, ranging from the economy, education, culture, and history of cities. Moreover, several proper nouns that signify unique names are highlighted, such as Olympics, Tsinghua/Peking University and Tiananmen Square for Beijing or Yankees, Broadway and Empire State Building for New York, which further demonstrates ChatGPT's ability to generate contextually relevant place identity descriptions. To gain a better understanding into the characteristics of ChatGPT responses, we offer several basic statistics of the generated outputs. On average, each bullet point contains 11.98 words, with a standard deviation of 6.43. Descriptions of urban

Fig. 2 Example images of place identity generated by DALL·E2. a Beijing. **b** New York.

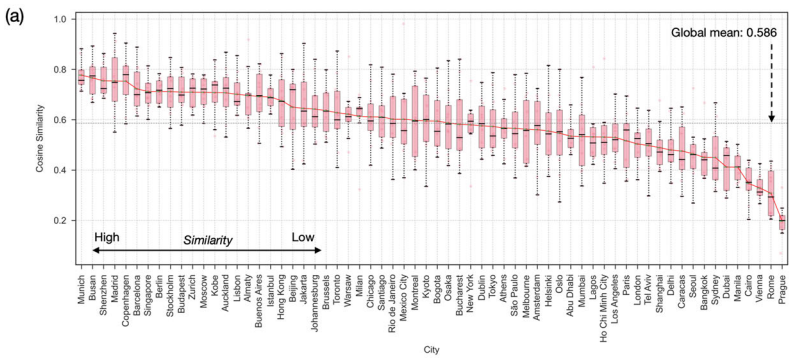
identity tend to be lengthier, with an average of 15.86 words per bullet point and a standard deviation of 5.83. Street identity, on the other hand, is typically presented in a paragraph format with an average of 19.65 words and a standard deviation of 5.05.

Results of place identity generated by DALL·E2. Figure 2 also demonstrates examples of place identity image outputs generated by DALL·E2 in Beijing and New York. These provide visual representations that align with people's general perceptions and common knowledge about these cities. For instance, in images depicting Beijing in Fig. 2a, we observe a combination of metropolitan cityscapes and classic Chinese architectural styles, such as *hutong* and *siheyuan*. Regarding images of New York in Fig. 2b, they reflect high density buildings, yellow traffic lights or fire escapes that align with our common perceptions of "The Big Apple". These differences between the two groups of images clearly illustrate the ability of GenAI models in capturing unique visual features of place identity in these cities.

Comparing place identity generated by ChatGPT with Wikipedia Corpus. To assess the accuracy and reliability of place identity generated by ChatGPT, we conducted a cross-validation with Wikipedia. Here, we intend to test whether AI-generated texts can provide a reliable representation of a city's place identity. This involves computing the cosine similarity between sentence embeddings of ChatGPT responses and Wikipedia corpuses, and presenting visual comparisons between pairs of word clouds. Overall, the average text similarity scores for {*place*}, {*urban*} and {*street*} identity responses were 0.59, 0.58, and 0.56, respectively. This suggests that the similarity between ChatGPT and Wikipedia descriptions of a place are non-varying with respect to the prompt used for the generative model. In this section, we particularly focus on results for the {*place*} prompt case while discussing the results of this study.

We first investigate the relevance between two datasets. Figure 3a is a box plot showing the distribution of cosine similarity scores, where each point denotes a comparison of each bullet point in ChatGPT responses with the most relevant match within Wikipedia. Also, note that cities are arranged in descending order of mean similarity, from left to right. Here, we observe a wide range of similarities, which reflect both similar and dissimilar descriptions of place identity by ChatGPT. Several examples of high and low similarity cases are further listed in Fig. 3b. For instance, Munich and Busan were cities with the two highest mean scores, whose contexts related to either its political importance or geographical conditions were successfully generated. In contrast, however, descriptions of Rome and Prague resulted with similarity levels that were far lower than the global average. While we requested ChatGPT to generate "in ten bullet points" and conducted a sentence-by-sentence comparison with the Wikipedia corpus to obtain uniformity in length, its descriptions for both cases were much shorter than sentences from Wikipedia. The examples suggest that low similarity results may be partially due to the length of texts being compared, and therefore, a more concrete way to minimize the discrepancy in length is crucial for the effectiveness of GenAI models in capturing the complex nuances of place identity.

We also present a visual comparison between pairs of word clouds created for ChatGPT answers and Wikipedia to understand the primary contents from both textual sources. Figure 3c shows example results for four different cases: Seoul, Singapore, Barcelona, and Almaty. First, ChatGPT described Seoul's place identity through topics including *culture, vibrant*, and *modern*, while Wikipedia introduction of Seoul covered keywords including *soul, life, human, spirit* and *belief*. We find that both results emphasize intangible aspects of the capital of South Korea, which correspond to the 'meaning' element of place identity models as defined in the fields of environmental psychology and geography (Canter 1977; Relph 1976). Recalling that 'meaning' refers to individual or group sentiments created through people's experiences, this indicates that ChatGPT captures the subjective atmosphere and cultural values as the most salient characteristics of Seoul. From word cloud comparison for Singapore, we observe keywords such as *diverse*, *multiculturalism* and *melting pot* from ChatGPT responses. These are supported by keywords such as *Singaporean*, *Malaysia*, *British* and *Chinese* in Wikipedia word cloud, implying that the text-to-text model identified Singapore's diverse and polyethnic culture. Barcelona and Almaty are the cases whose identities are described in relation to broader ethnographic or national contexts. The most notable keywords in word clouds generated based on their ChatGPT responses are



(b)

City	Source	Description	Similarity Score
Munich	ChatGPT	Bavarian capital: Munich is the capital of the German state of Bavaria and one of the most important cities in the country.	0.865
	Wikipedia	As the capital of Bavaria, Munich is an important political centre for both the state and country as a whole.	
Busan	ChatGPT	Coastal City: Busan is located on the southeastern tip of the Korean peninsula, making it a coastal city with beautiful beaches and stunning views of the sea.	0.893
	Wikipedia	Busan is located on the southeastern tip of the Korean Peninsula.	
Rome	ChatGPT	Ancient city with a long and rich history.	0.436
	Wikipedia	Buildings such as huts and longhouses have been used for living since the late Neolithic.	
Prague	ChatGPT	One of Europe's most beautiful cities.	0.330
	Wikipedia	Dover states there is a distinction of aesthetic and the moral senses of the term; "It must be emphasized that the Greeks did not call a person 'beautiful' by virtue of that person's morals, intelligence, ability or temperament, but solely by virtue of shape, colour, texture and movement".	

Fig. 3 Text similarity results. a Box plot of cosine similarity scores between {*place*} identity responses generated by ChatGPT and Wikipedia corpuses. Each city includes ten points, each indicating the highest cosine similarity per ChatGPT sentence. From left to right, cities are in descending order of their mean cosine similarity. Red line indicates the mean similarity level of individual cities. For box plots based on {*urban*} and {*street*} prompts, see Supplementary Fig. S1. **b** Examples of high (Munich and Busan) and low (Rome and Prague) text similarity scores. **c** Comparison of word clouds between ChatGPT's outputs (left) and Wikipedia corpuses (right): from top to bottom, Seoul, Singapore, Barcelona and Almaty.

Catalan and *Kazakhstan*, respectively. Likewise, word clouds of Wikipedia corpus also highlight both keywords, from which we infer that the place identity of Barcelona and Almaty are deeply intertwined with either the ethnographic or national contexts.

Comparing place identity generated by DALL·E2 with Google images. We measured the image similarity between images generated by DALL·E2 and those collected via Google search. Parallel to the text similarity analysis, here, we examined the generative capability of GenAI in producing realistic representation of place-specific scenes of cities. In particular, we computed the Learned Perceptual Image Patch Similarity (LPIPS) that evaluates the distance between image patches and has been widely used in previous studies for aligning well with human judgment (Cheon et al. 2021; Zhang et al. 2018). A value equivalent to 1 – LPIPS is defined as an image similarity score ($S_{i,j}$) to quantitatively assess the perceptual similarity of images, where a higher score indicates greater similarity, and vice versa.

Figure 4a provides a box plot showing the distribution of image similarity score ($S_{i,j}$) in ascending order, from left to right. Here, we observe variability in image similarity across different cities. Overall, the average is 0.575 and the standard deviation is 0.066. We further explore specific examples selected from two contrasting cases identified with the highest and lowest mean perceptual similarities between their generated and real-world scenes. In Fig. 4b, it is evident that DALL·E2 successfully depicted the decorative Baroque-style guildhalls on the Grand-Place in Brussels. In contrast, images generated for the place identity of Tokyo were dissimilar from real-world scenes shown in Google images. As shown in Fig. 4c, the repetitive generation of mundane streets without strong visual cues may be a sign of placelessness in the urban landscapes of Tokyo. Yet, we also point out that lighting conditions may have influenced the outcome. While DALL·E2 is strongly inclined to generate daytime images, certain cities include more images of night scenes in their Google search data. This tendency is more apparent in cities that are well known for their vibrant nighttime economy. Such differences in the time of day being illustrated in DALL·E2 outputs and Google images may contribute to low perceptual similarity.

Furthermore, we aimed to verify if this computational approach corresponds with human responses, by conducting a survey where a total of 30 respondents rated the similarity between a given pair of generated and Google search images using a 7-point Likert Scale (see Supplementary Table S1). The average similarity score of all image pairs was 3.406 with a standard deviation of 0.606. At an individual city level, the top three similarities rated by human responses were Chicago (4.967), Madrid (4.867) and Montreal (4.267), whereas the bottom three results were Seoul (2.367), Auckland (2.467) and Kobe (2.467). This coincides with the previous finding in Fig. 4, in that Chicago and Montreal are among the fourth quarter (above the third quartile) in their LPIPS-based similarity, while Seoul is among the first (below the first quartile). Yet, we also noted contrasting cases such as Madrid, Auckland and Kobe, which presented mid-level similarities in Fig. 4. The Pearson correlation between the two similarities was 0.229, with a p-value of 0.071. While this result is not statistically significant at the conventional 0.05 level, it is significant at the 0.1 level. Given the exploratory nature and the inherent subjectivity in human survey responses with a relatively small sample size, we consider a significance level of 0.1 to be appropriate (Jackson 2006; Stevens 2002). The correlation result warrants further investigation. Despite the positive relationship, the weak correlation indicates a disparity between the two similarity scores, suggesting that LPIPS-based evaluation may not fully capture the nuances of human perception of how well

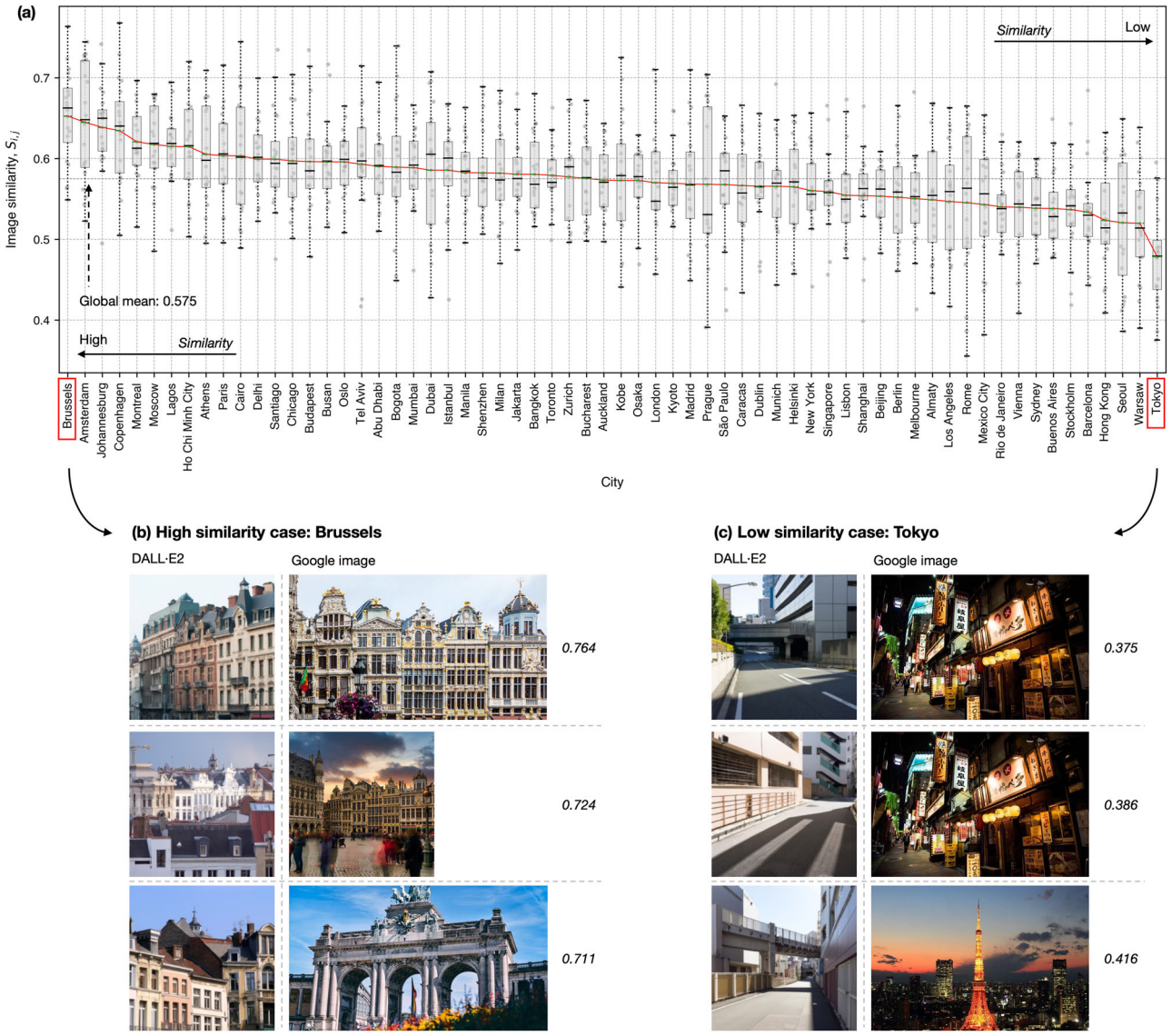


Fig. 4 Image similarity results. a Box plot of LPIPS scores between DALL·E2 generated and Google search images by cities. Each city includes twenty points, each indicating the highest image similarity score (equivalent to lowest LPIPS) per DALL·E2 generated image. From left to right, cities are in descending order of their mean perceptual similarity. Red line indicates the mean similarity level of individual cities. **b** High image similarity example: Brussels. **c** Low image similarity example: Tokyo.

GenAI represented the identity of cities. Therefore, it is necessary to involve more human opinions rather than relying on machine-based metrics. This discrepancy could be due to sample variability; the given pair might not represent the entire scenes of cities, while the 30 respondents might not represent the entire population. Meanwhile, this provides a valuable attempt to bridge the gap between quantitative and qualitative assessments of GenAI. We note that our goal was to provide a preliminary insight into the relationship between computational and human evaluations, not to conduct a comprehensive human study. Further research should incorporate a larger sample size and alternative computational techniques for a more robust estimation of the reliability of GenAI models based on human perception.

City-by-city pairwise similarity between DALL·E2 generated place identity. Finally, we compared the DALL·E2-generated outputs across different cities to examine whether GenAI can identify them distinctively. We aim to test two hypotheses throughout such comparisons: (1) Similarity between generated outputs of the same city is greater than that of different cities; and (2) similarity between generated outputs is greater in cities that are geographically and culturally close than in cities that are geographically and culturally distant. Figure 5a illustrates the similarity matrix constructed based on normalized Chamfer distance (CD) between sets of DALL·E2 generated images of a given city pair. Each cell is assigned with a value of 1–CD, so that higher value indicates higher similarity, and vice versa. We also note that cities were sorted by decreasing longitude to reveal geographical patterns of similarities represented by GenAI.

Overall, we observe two distinct results. First, relative high similarity scores (in red) appear along the diagonal. This shows that DALL·E2 outputs were more similar within itself than compared across cities, which corroborates the first hypothesis. In other words, the generative model produced images that may successfully represent the place identity of individual cities. For

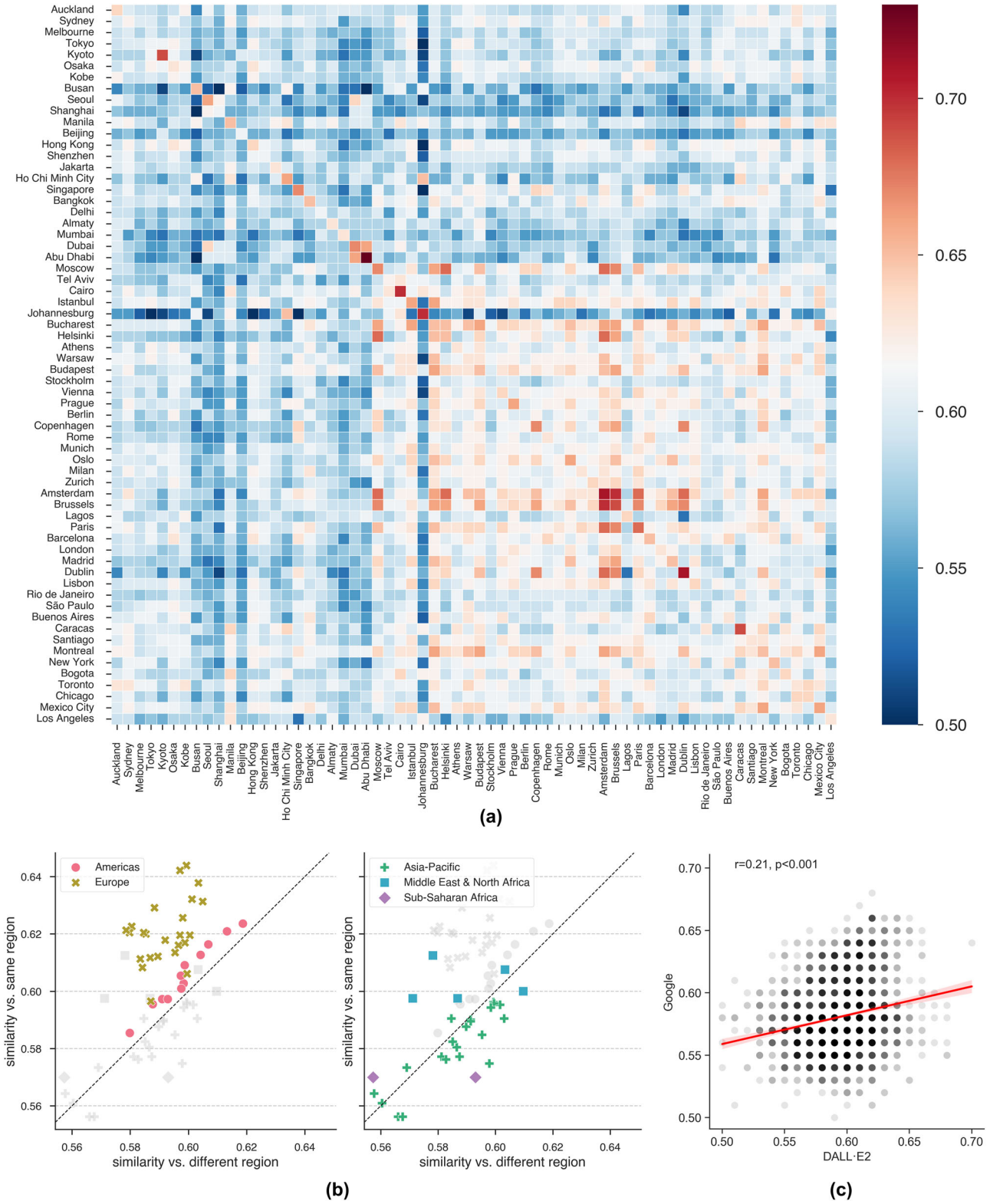


Fig. 5 City-by-city comparison of DALL·E2 generated place identity. a Pairwise similarity matrix. Normalized Chamfer distance (CD) is measured for sets of DALL·E2 images of a given city pair. Cities are sorted in orders of longitude. Each cell is colored based on a value of 1-CD, where red indicates strong and blue indicates weak similarity between the generated place identity of two cities. **b** The West (left) vs. the non-West (right). For each city, similarity with cities in different regions are plotted against that with cities in the same region. Symbol denotes the continent in which the city is located. **c** Pearson correlation between 1-CD for DALL·E2 and Google images of a given city pair.

Fig. 6 Examples of DALL·E2 generated streetscapes of residential areas of Boston with one different keyword. a "white community". **b** "black community".

example, Abu Dhabi, Amsterdam, Dublin, Cairo, Johannesburg, Brussels, Kyoto, Caracas, Paris and Dubai are top 10 cities with strongest identity captured by DALL·E2. In particular, the contrast between on- and off-diagonal values is most apparent for Kyoto, Abu Dhabi, Cairo and Johannesburg, indicating that DALL·E2 identified these cases as the most visually distinctive cities.

Another notable observation is the grouping of high similarity scores in the lower-right section which consists of Moscow, Istanbul, and cities from Bucharest and to the west. We view this as an indication of the dichotomy between place identity in the Western and non-Western worlds. On the one hand, cities in American and European countries are found to share visual similarities among themselves, where Amsterdam-Brussels is identified as the highest similarity pair (0.7) in all 4,160 pairwise comparisons. On the other, cities in Asia-Pacific, Middle East and African countries present relatively low similarities across most comparisons (except for the Abu Dhabi-Dubai pair with a similarity score of 0.66). This coincides with the lack of local identity in urban developments in non-Western megacities during the past decades (Choi & Reeve 2015; Shim & Santos 2014). Previous findings have pointed out the tendency of these cities to copy imported Western design, resulting in a chaotic mixture of urban and rural landscapes and failing to achieve the intended level of success (Al-Kodmany & Ali 2012; Yokohari et al. 2000). This contrast is further verified when similarity within the same region is compared against that with different regions. As illustrated in Fig. 5(b), all cities in Americas and Europe were presented with clear intraregional similarities (plotted above the reference line), whereas their non-Western counterparts showed irregular patterns across cities. Therefore, we conclude that our second hypothesis ---pairs of cities that are geographically and culturally closer are more similar --- is partially true for American and European cities, while DALL·E2 captures evidence of placelessness (Relph 1976) for the rest of the world.

These findings are supported by the positive correlation in Fig. 5c, which demonstrates that the similarity between generated images of a given city pair is consistent with that between the actual urban scenes shown through Google images. This provides empirical evidence of the effectiveness of GenAI in capturing the visual distinctiveness of cities through such pairwise comparisons and verifies its capabilities in representing place identity in response to place-related prompts.

Discussion

In the previous sections, we presented a computational framework that employed GenAI models to generate place identity results. We further computed text and image similarity scores between generative model responses and corresponding Wikipedia and Google image search data to test the reliability of their outputs for representing place identity in different cities. GenAI models capture salient characteristics of cities and could be utilized as a valuable data resource to advance our knowledge of place. However, their future directions as well as ethical issues and limitations should also be discussed. Here, we list several takeaways to offer implications for the future use of GenAI in urban studies pertaining to understanding place identity.

Generative AI for urban studies. In this study, we attempted to provide GenAI with prompts on place concepts that contain subjective meanings and verify its reliability in generating textual and visual outputs that capture place identity of cities. Future studies may extend this by using GenAI to construct a valuable dataset of place meanings at a larger spatiotemporal scale. For instance, we conducted a comparison among cities that best represent the countries in which they are located in. In addition, the results were obtained based on data before September 2021, the knowledge cutoff date officially announced by OpenAI (2023). Therefore, the approach in this paper can be revisited by adding more cities within the same country for an intranational study or rerunning in different years with updated data to reveal how place identity changes over time. These not only allow researchers to model the subjective nature of urban experiences (i.e., place identity, cognition, perception, etc.) but also provide a promising baseline for the use of GenAI tools in future urban studies.

GenAI can enhance our urban imagination and simulation by incorporating socioeconomic and subjective aspects of the urban environment in future studies. For instance, we can prompt GenAI to render urban scenes (or place identity) of different demographic attributes, such as age, education and race/ethnicity, leading to a question of how well the generated outputs align with different communities' perception of the urban landscape or whether they are skewed towards certain social classes or culture. Figure 6 presents examples of generated streetscapes of Boston using the same prompts except for one keyword. In Fig. 6a, residential areas of the "white community" include brownstone houses along roads whose pavement and streetlights are well-maintained; while Fig. 6b illustrates a degraded built environment for the "black community" with bumps and cracks on the road, overgrown bushes, and building architecture that is simple to the bare minimum. This indicates that what GenAI models predict is based on social stigmatization about certain urban populations as well, with risk of reinforcing this discriminatory lens, although there is no legal or infrastructural ground for such narratives. Future studies can examine cities from low- and middle-income

cities (LMICs) that often lack quality data to train GenAI models. This enables discussing the fairness of GenAI models, particularly for the social context of marginalized areas that have disproportionately low representation in the training datasets. Moreover, using query keywords that specify the perceptual qualities of the urban environment can help us understand the defining characteristics of safe, lively, wealthy, active, beautiful, and friendly cities (Dubey et al. 2016). While existing applications of generative models have mostly focused on automating the planning processes on a two-dimensional plane (Park et al. 2023; Wang et al. 2023), the proper use of GenAI models can help planners and designers obtain more realistic and imaginative urban scenes that are more relevant to human perception and experience.

Finally, we also raise concerns regarding the "black-box" deep learning approaches. Our results indicated that GenAI models possess varying capabilities in representing place-specific characteristics of cities depending on their output format. However, we have minimal information on the data used for training generative models at the current stage. Wikipedia is known to be one of the sources of training data for ChatGPT (Shen et al. 2023), which may overlap with that used in this study, raising concerns regarding circularity in evaluating GenAI results using its own training data. Despite such limitations, Wikipedia and Google Images have been considered valuable sources of collective place-specific meanings, considering the lack of large-scale ground-truth dataset about the identity of global cities (Choi et al. 2007; Coghlan et al. 2017; Jenkins et al. 2016). Therefore, their usage can still be informative when particularly focused on specific domains that require qualitative assessment of generated outputs. For instance, one of our main objects of interest in this study was to identify varying degrees of similarity in representations, through which we revealed intrinsic biases and errors for different global city cases. This provides a consistent baseline for assessing the reliability of GenAI results against commonly accepted and easily accessible information. In the meantime, it remains necessary to develop more explainable AI approaches that can better elucidate the reasoning behind the generated outputs. This can be addressed in future studies in two ways. First, data from different sources that is less likely to be included in the training of generative models might be considered for their real-world counterparts. Social media or automated online surveys are two alternative platforms to crowdsource direct opinions of people at scale (Dubey et al. 2016; Jang & Kim 2019). Second, it is necessary to customize the models for domain-specific applications. Although large language models have been effective in producing general human-like responses, researchers have recently demonstrated that 'smaller' language models could achieve high performance with greater efficiency when fine-tuned for a particular domain or context (Fu et al. 2023; Schick & Schütze 2020; Turc et al. 2019).

Place-specific Scenes vs. Generic City View. By asking DALL·E2 with prompts regarding place identity of streetscapes of cities, we obtained a collection of images that depicted various street scenes. These images were then assessed to measure their similarity with images of the real-world. We could observe subtle differences among different cities regarding the architectural style, street design, or vegetation type. For instance, as shown in Fig. 7, New York images created by DALL·E2 primarily showed prewar apartment buildings in Manhattan with wrought-iron fire escapes; images in Paris are represented by its Haussmannian architecture with stone facades, balconies, and double windows; and images in Singapore are emphasized by either its typical high-rise apartments or shophouses along with rain trees that grow in this region. All of these indicate that GenAI could capture the unique place identity, particularly related to architectural style, of each city.

However, it is worth noting that DALL·E2 has also generated a series of images that depict generic city views rather than specific to any particular place, thus failing to capture the unique characteristics of individual cities. Figure 8a shows a collection of images for New York, Tokyo, Seoul, London, Sydney, and Melbourne generated by DALL·E2. Generated images for different cities mostly depicted common urban features such as buildings, road signs, streetlights and pavements. These reflect the generic concept of a *city*, rather than *identity*, and fall short in representing the attributes that distinguish a particular city from the rest. As shown in the Sydney example in Fig. 8b, the generated place identity images do not capture landmarks of the city (Opera House and Harbour Bridge) or its scenic waterfront. Instead, a generic landscape of an urban environment is rendered, which makes it difficult to tell what the salient characteristics are from the generated images. Moreover, a pseudoword on a signpost, *Hork Str Sox*, hardly functions as a visual cue for the identity of streetscape in Sydney. These observations pose questions regarding the reliability of these generated images. Researchers need to carefully evaluate the quality of these AI-generated images before considering their practical use in research and real-world applications.

The observation of both generic and place-specific from generated images connects to the discourse of *space* and *place* that constructs the nature of geographical disciplines. As opposed to *space* which is an abstract and undifferentiated physical setting, *place* is given unique personalities over time to become locations with visual impact that brings sense of place among people (Tuan 1977). On the one hand, DALL·E2 produced scenes and images of placeless urban landscapes (see Fig. 8); "a scene may be of a place but the scene itself is not a place" (Tuan 1979, p. 411). On the other, results in Fig. 7 showed its promising capabilities in representing the *place* of different cities. This is particularly intriguing because unlike places such as monument buildings, religious spaces or public plazas that are easily identifiable as 'public symbols' of the city, places as 'fields of care' in an everyday setting (e.g., park, home, drugstore street corner, marketplace) have been discussed to lack visual identity and be barely discernible through physical or structural appearances without repeated experience of the place (Tuan 1979; Wild 1965). Yet, we were able to distinguish DALL·E2-generated streetscape scenes of New York, Paris, and Singapore from elements such as streetlights, vegetation or architectural style, implying the possibility of uncovering inconspicuous places with the use of GenAI without repetitive interaction with the physical environment. This can further contribute to urban planning and design practice, considering the importance of cultural heritage and identity of a place to foster as sense of belonging among city dwellers (Hernandez et al. 2010; Manzo & Perkins 2006). Particularly, GenAI tools can be effective in collecting multiple development scenarios or design options instantly from the public that better reflect the preferences and priorities of the community. Thus, we may expect GenAI to assist in not only generating visual representations rooted in the cultural contexts of a place but also in facilitating community engagement in the urban design process and developing placemaking strategies that enhance the sense of place and attachment. Returning to Tuan's (1979) conclusion, spatial analysis from the positivist perspective tends to simplify the underlying assumptions of people, space and place, whereas the humanist must take into account the intricacy of human nature—so must, and can, GenAI.

Fig. 7 Place-specific scenes produced by DALL·E2. a New York. **b** Paris. **c** Singapore.

Fig. 8 Generic city views produced by DALL·E2. a Generated images for New York, Tokyo, Seoul, London, Sydney, and Melbourne. **b** Sydney example of comparison with Google images.

Opportunities and Challenges. Looking forward, we close by outlining technical challenges and opportunities to be further explored for the application of GenAI in future urban research. First, to obtain more reliable results that represent place-specific attributes of different cities, researchers may develop more careful prompt engineering. The importance of appropriate prompt designs has been commonly emphasized in previous research to enhance the consistency of GenAI models for domain-specific applications (Hase et al. 2021; Kang et al. 2023). By discussing the results of this study, we found this is more imperative for the text-to-image model compared to its text-to-text counterpart. As suggestions to design effective prompts for DALL·E2 to yield relevant responses to the place identity of cities, we can specify the point of view (POV), perspective, and captured objects in output images as in the following format:

> "*What is the place identity of {city}? Show me a {perspective} focused on {object} with point of view pitch angle at {pitch}.*"

As DALL·E2 produced image results with different directions and angles, parameters to set specific POV headings and pitch, {*heading*} and {*pitch*}, can be added to provide consistent viewpoints. Also, clarifying whether to show a bird-eye view or street-level scene using a {*perspective*} parameter can reduce variation in terms of the image perspective. Moreover, to minimize unpredictability in scenes being rendered, an {*object*} parameter would let resulting images focus on specific urban elements of interest. As discussed earlier in the previous section, whether to generate either day or nighttime image may also be an effective parameter to control the lighting conditions being rendered. Examples of DALL·E2 results when different parameters were used in this prompt format are shown in Fig. 9.

Another future direction lies in the improvement of methods for evaluating the reliability of generative model outputs. Here, we suggest two potential approaches for this purpose, multi-source data fusion and advanced similarity analysis. The AI-generated outputs are not always consistent with Wikipedia corpus and Google image search results as found in this study. We could incorporate social media texts and images as valuable data sources in capturing users' various information related to places. Such data enable us to compare generative model outputs with people's direct opinions that can better represent the identity of places (Jang & Kim 2019). In the meantime, we observed uncertainties in the similarity analysis results led by the subjective nature of perception. That is, why differences in similarity scores are observed, what contributes to high or low similarity results, and which scene is more relevant to the place identity of specific cities. This can be further refined by defining a more concrete threshold for interpreting the cosine similarity and LPIPS metric used in this study. Furthermore, different methods can be adopted for comparison purposes. For instance, more advanced algorithms can be applied, such as object detection and image segmentation, to retrieve object occurrences from DALL·E2 outputs and verify their correspondence with real-world urban scenes.

It is also noteworthy that prompts and outputs in this study were created only in English, overlooking the performance of

Fig. 9 DALL·E2 results showing scenes that represent the place identity of Boston with different parameters added. a {*pitch*} (**b**) {*perspective*}. **c** {*object*}.

GenAI models in other linguistic settings. While a few previous studies have highlighted the potential of GenAI in overcoming language barriers from being built on billions of inputs and parameters (Gottlieb et al. 2023; Sajjad & Saleem 2023), it remains important to examine the generalizability of outputs through a critical lens when conducting a multicity comparison. In its technical report, OpenAI (2023) has demonstrated the outperformance of GPT models when using English or major European languages, likely because they were designed and built primarily with data from English sources without robust multilingual testing. In addition, the English Wikipedia has both the most number articles and page views, making non-English speakers less capable of contributing to the online encyclopedia. This disproportionate representation could be a plausible explanation for the high intraregional similarities between DALL·E2-generated images of cities in the Americas and Europe in contrast to those among non-Western cities (see Fig. 5). Hence, this raises the question of from whose perspective are outputs being generated. For instance, in Table 2, it is plausible to interpret that ChatGPT's description of Beijing has a nuance toward a foreign audience, whereas that of New York assumes a US-centric audience with prior knowledge about American culture. Considering the subjective nature of place identity, we offer future research directions to inquire whether GenAI outputs paint us a picture of the local people's knowledge, of foreign tourists and journalists' experience, or the local authorities' official statements by testing variations of multicultural and multilanguage prompts.

Last, acknowledging the difficulty in overcoming the limitation regarding the "black-box" nature of the generative models, a potential solution could involve comparing their outputs with actual human responses. This could be achieved by conducting a survey to how individuals assess the quality of the GenAI descriptions of different cities. GenAI outputs could be graded in terms of to what extent they are representative of people's place identity for a certain place. Also, a focus-group interview could be helpful to gather more detailed opinions on how participants from similar demographic or experiential backgrounds perceive the validity of generated results. Meanwhile, the rapid advancements in the development of new GenAI models call for regular updates to the results for improved relevancy and contribution of the work. Potentially repeating the experiments with the latest GPT-4 or GPT-4o models and DALL·E3 may help us reveal the up-to-date performance of GenAI models in understanding and depicting place identity without relying on deliberate efforts of OpenAI targeted on these particular abilities.

Conclusions

We have recently witnessed the capabilities of GenAI models in various domains. Their capabilities in generating realistic texts and image outputs with only simple prompts have enabled collecting human-like responses in an efficient and cost-effective manner. In this study, we attempted to investigate the potential of using generative models in understanding *place identity*, an important concept in the field of urban design and geography. While place identity is subjective and closely tied with an individual's perception of cities, many studies have attempted to discover the collective identity that better explains both the physical and non-physical attributes of the urban environment. We departed from two aspects, languages and visual representation, and asked two GenAI models, ChatGPT and DALL·E2, with prompts related to the place identity of different cities. We further tested the reliability of their responses by measuring their similarity with fact-based datasets, Wikipedia and Google images, that depict the real urban settings. Moreover, we conducted a pairwise comparison to verify if GenAI can also capture the visual distinctiveness or similarity between cities. Our results indicate that GenAI models have the potential to generate outputs that represent salient characteristics of cities that make them distinguishable and can serve as a valuable data source and tool for urban studies. This study is among the pioneering attempts to investigate GenAI in urban design research before applying them into planning and design practices. While exploring the

capabilities of GenAI in representing the place identity of cities, we contribute to existing literature by discussing potential limitations and future research opportunities for further studies. The overall framework is expected to aid planners and designers in utilizing such tools to evaluate characteristics of cities for place-making and city branding purposes, and in turn, shaping more imageable cities.

Data availability

All relevant data used and generated in the research are publicly available in the Figshare Repository at: https://doi.org/10.6084/m9.figshare.25041452.v1. The detailed data management information can be found in the supplementary information.

Received: 2 January 2024; Accepted: 20 August 2024;
Published online: 07 September 2024

References

Al-Kodmany K, Ali MM (2012) Skyscrapers and placemaking: supporting local culture and identity. Archnet-IJAR: Int J Archit Res 6(2):43

Anantrasirichai N, Bull D (2022) Artificial intelligence in the creative industries: a review. Artif Intell Rev 55(1):589–656

Atkins C, Girgente G, Shirzaei M, Kim J (2024) Generative AI tools can enhance climate literacy but must be checked for biases and inaccuracies. Commun Earth Environ 5(1):226

Bolojan D, Vermisso E, Yousif S (2022) Is language all we need? a query into architectural semantics using a multimodal generative workflow. *POST-CARBON*. Proc 27th Int Conf Assoc Computer-Aided Architectural Des Res Asia (CAADRIA) 1:353–362

Canter D (1977) *The psychology of place*. Architectural Press

Cheon M, Yoon S-J, Kang B, Lee J (2021) Perceptual image quality assessment with transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 433–442

Choi HS, Reeve A (2015) Local identity in the form-production process, using as a case study the multifunctional administrative city project (Sejong) in South Korea. Urban Des Int 20:66–78

Choi S, Lehto XY, Morrison AM (2007) Destination image representation on the web: Content analysis of Macau travel related websites. Tour Manag 28(1):118–129

Coghlan A, McLennan C-L, Moyle B (2017) Contested images, place meaning and potential tourists' responses to an iconic nature-based attraction 'at risk': the case of the Great Barrier Reef. Tour Recreat Res 42(3):299–315

Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*

Dubey A, Naik N, Parikh D, Raskar R, Hidalgo CA (2016) Deep learning the city: Quantifying urban perception at a global scale. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 196–212

Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Baabdullah AM, Koohang A, Raghavan V, Ahuja M (2023) "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag 71:102642

Ford M (2015) *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books

Fu Y, Peng H, Ou L, Sabharwal A, Khot T (2023) Specializing Smaller Language Models towards Multi-Step Reasoning. In *Proceedings of the 40th International Conference on Machine Learning*, PMLR, 202, p 10421–10430

Gao S, Janowicz K, Montello DR, Hu Y, Yang J-A, McKenzie G, Ju Y, Gong L, Adams B, Yan B (2017) A data-synthesis-driven method for detecting and extracting vague cognitive regions. Int J Geogr Inf Sci 31(6):1245–1271

Gao Y, Chen Y, Mu L, Gong S, Zhang P, Liu Y (2022) Measuring urban sentiments from social media data: a dual-polarity metric approach. J Geogr Syst 24(2):199–221

Goodchild MF (2010) Formalizing place in geographic information systems. In *Communities, neighborhoods, and health: Expanding the boundaries of place* (pp. 21–33). Springer

Gottlieb M, Kline JA, Schneider AJ, Coates WC (2023) ChatGPT and conversational artificial intelligence: Friend, foe, or future of research? Am J Emerg Med 70:81–83

Hase P, Diab M, Celikyilmaz A, Li X, Kozareva Z, Stoyanov V, Bansal M, Iyer S (2021) Do language models have beliefs? methods for detecting, updating, and visualizing model beliefs. *ArXiv Preprint ArXiv:2111.13654*

Hernandez B, Martin AM, Ruiz C, del Carmen Hidalgo M (2010) The role of place identity and place attachment in breaking environmental protection laws. J Environ Psychol 30(3):281–288

Hu Y, Deng C, Zhou Z (2019) A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. Ann Am Assoc Geogr 109(4):1052–1073

Hull RBIV, Lam M, Vigo G (1994) Place identity: symbols of self in the urban fabric. Landsc Urban Plan 28(2–3):109–120

Jackson D (2006) The power of the standard test for the presence of heterogeneity in meta-analysis. Stat Med 25(15):2688–2699

Jang KM, Kim Y (2017) Collective Place Identity. *2017 International Conference of Asian-Pacific Planning Societies (ICAPPS 2017)*, 096

Jang KM, Kim Y (2019) Crowd-sourced cognitive mapping: A new way of displaying people's cognitive perception of urban space. Plos One 14(6):e0218590

Jenkins A, Croitoru A, Crooks AT, Stefanidis A (2016) Crowdsourcing a collective sense of place. PloS One 11(4):e0152932

Kang Y, Jia Q, Gao S, Zeng X, Wang Y, Angsuesser S, Liu Y, Ye X, Fei T (2019) Extracting human emotions at different places based on facial expressions and spatial clustering analysis. Trans GIS 23(3):450–480

Kang Y, Zhang Q, Roth R (2023) The ethics of AI-Generated maps: A study of DALLE 2 and implications for cartography. *ArXiv Preprint ArXiv:2304.10743*

Kim J, Lee J (2023) How does ChatGPT Introduce transport problems and solutions in North America? *Findings*. https://doi.org/10.32866/001c.72634

Larsen SC (2004) Place identity in a resource-dependent area of northern British Columbia. Ann Assoc Am Geogr 94(4):944–960

Latif E, Mai G, Nyaaba M, Wu X, Liu N, Lu G, Li S, Liu T, Zhai X (2023) Artificial general intelligence (AGI) for education. *ArXiv Preprint ArXiv:2304.12479*

Lee H-K (2022) Rethinking creativity: creative industries, AI and everyday creativity. Media, Cult Soc 44(3):601–612

Lewicka M (2008) Place attachment, place identity, and place memory: Restoring the forgotten city past. J Environ Psychol 28(3):209–231. https://doi.org/10.1016/j.jenvp.2008.02.001

Liu L, Silva EA, Wu C, Wang H (2017) A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Computers, Environ Urban Syst 65:113–125

Mai G, Huang W, Sun J, Song S, Mishra D, Liu N, Gao S, Liu T, Cong G, Hu Y (2023) On the opportunities and challenges of foundation models for geospatial artificial intelligence. *ArXiv Preprint ArXiv:2304.06798*

Manzo LC, Perkins DD (2006) Finding common ground: The importance of place attachment to community participation and planning. J Plan Lit 20(4):335–350

Mishkin P, Ahmad L, Brundage M, Krueger G, Sastry G (2022) DALL· E 2 Preview-Risks and Limitations. Noudettu 28:2022

Nasar JL (1990) The evaluative image of the city. J Am Plan Assoc 56(1):41–53

OpenAI. (2023) GPT-4 Technical Report. *ArXiv E-Prints*, arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774

Paananen V, Oppenlaender J, Visuri A (2023) Using text-to-image generation for architectural design ideation. Int J Archit Comput, 14780771231222783. https://doi.org/10.1177/14780771231222783

Paasi A (2003) Region and place: regional identity in question. Prog Hum Geogr 27(4):475–485

Park C, No W, Choi J, Kim Y (2023) Development of an AI advisor for conceptual land use planning. Cities 138:104371

Peng J, Strijker D, Wu Q (2020) Place identity: how far have we come in exploring its meanings? Front Psychol 11:294

Proshansky HM, Fabian AK, Kaminoff R (1983) Place-identity: Physical world socialization of the self. J Environ Psychol 3(1):57–83

Relph E (1976) *Place and placelessness* (Vol. 67). Pion London

Sajjad M, Saleem R (2023) Evolution of healthcare with ChatGPT: A Word of Caution. Annals Biomed Eng, 1–2

Schick T, Schütze H (2020) It's not just size that matters: Small language models are also few-shot learners. *ArXiv Preprint ArXiv:2009.07118*

Seamon D, Sowers J (2008) Place and placelessness, Edward Relph. Key Texts Hum Geogr 43:51

Seneviratne S, Senanayake D, Rasnayaka S, Vidanaarachchi R, Thompson J (2022) DALLE-URBAN: Capturing the urban design expertise of large text to image transformers. In *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, pp 1–9. https://doi.org/10.1109/DICTA56598.2022.10034603

Shen X, Chen Z, Backes M, Zhang Y (2023) In chatgpt we trust? measuring and characterizing the reliability of chatgpt. *ArXiv Preprint ArXiv:2304.08979*

Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, Moy L (2023) ChatGPT and other large language models are double-edged swords. In *Radiology* (Vol. 307, Issue 2, p. e230163). Radiological Society of North America

Shim C, Santos CA (2014) Tourism, place and placelessness in the phenomenological experience of shopping malls in Seoul. Tour Manag 45:106–114
Stevens J (2002) *Applied multivariate statistics for the social sciences* (Vol. 4). Mahwah, NJ: Lawrence Erlbaum Associates
Stewart WP, Liebert D, Larkin KW (2004) Community identities as visions for landscape change. Landsc Urban Plan 69(2–3):315–334
Sun Y, Dogan T (2023) Generative methods for Urban design and rapid solution space exploration. Environ Plan B: Urban Analytics City Sci 50(6):1577–1590
Tuan Y-F (1977) *Space and place: The perspective of experience*. U of Minnesota Press
Tuan YF (1979) Space and place: humanistic perspective. In *Philosophy in geography* (pp. 387-427). Dordrecht: Springer Netherlands
Turc I, Chang M-W, Lee K, Toutanova K (2019) Well-read students learn better: On the importance of pre-training compact models. *ArXiv Preprint ArXiv:1908.08962*
Turchi T, Carta S, Ambrosini L, Malizia A (2023) Human-AI Co-creation: Evaluating the Impact of Large-Scale Text-to-Image Generative Models on the Creative Process. *International Symposium on End User Development*, 35–51
Van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL (2023) ChatGPT: five priorities for research. Nature 614(7947):224–226
Wang D, Lu C-T, Fu Y (2023) Towards automated urban planning: When generative and chatgpt-like ai meets urban planning. *ArXiv Preprint ArXiv:2304.03892*
Wang S, Chen JS (2015) The influence of place identity on perceived tourism impacts. Ann Tour Res 52:16–28
Wang W, Wei F, Dong L, Bao H, Yang N, Zhou M (2020) Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Adv Neural Inf Process Syst 33:5776–5788
Wild J (1965) Existence and the World of Freedom
Yokohari M, Takeuchi K, Watanabe T, Yokota S (2000) Beyond greenbelts and zoning: A new planning concept for the environment of Asian mega-cities. Landsc Urban Plan 47(3–4):159–171
Zhang F, Zhang D, Liu Y, Lin H (2018) Representing place locales using scene elements. Comput Environ Urban Syst 71:153–164
Zhang F, Zhou B, Ratti C, Liu Y (2019) Discovering place-informative scenes and objects using social media photos. R Soc Open Sci 6(3):181375
Zhang R, Isola P, Efros AA, Shechtman E, Wang O (2018) The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595

Acknowledgements

The authors would like to thank the members of MIT Senseable City Lab who provided feedback on this project. The authors would like to acknowledge the financial support of open access from MIT Libraries. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

Author Contributions

KJ: Conceptualization, Methodology, Data Curation, Formal analysis, Writing—Original Draft. JC: Methodology, Software, Data Curation, Formal analysis. YK: Conceptualization, Methodology, Data Curation, Writing—Original Draft, Corresponding author. JK: Conceptualization, Resources. JL: Conceptualization, Validation. FD: Conceptualization, Supervision, Funding acquisition, Corresponding author. CR: Resources, Supervision, Funding acquisition. Writing—Review & Editing: all authors.

Competing interests

FD was a Collection Guest Editor for this journal at the time of acceptance for publication. The manuscript was assessed in line with the journal's standard editorial processes, including its policy on competing interests. Other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-03645-7.

Correspondence and requests for materials should be addressed to Yuhao Kang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.