Conversational Interfaces Reshape What People Ask of Place

Ilya Ilyankou
SpaceTimeLab
Dept. of Civil, Environmental, and Geomatic Engineering
University College London (UCL)
London, UK
ilya.ilyankou.23@ucl.ac.uk

Stefano Cavazzi
Dept. of Civil, Environmental, and Geomatic Engineering
University College London (UCL)
London, UK

Ordnance Survey
Southampton, UK
stefano.cavazzi@os.uk

James Haworth
SpaceTimeLab
Dept. of Civil, Environmental, and Geomatic Engineering
University College London (UCL)
London, UK
j.haworth@ucl.ac.uk

Conversational interfaces are displacing the search bar, but a chatbot is not a neutral conduit: what a system appears able to do shapes what users ask of it. We test this for queries with geospatial intent, which constitute a persistent share of information-seeking behaviour. We apply query log analysis to 423,142 first turns from three public chatbot corpora (WildChat, ShareChat, LMSYS-Chat-1M), using the same classifier and clustering pipeline as a prior 1M-query web-search taxonomy, and restricting analysis to short first turns so any difference is a conservative lower bound. We find that the factual-lookup core persists: people still ask chatbots for capitals, distances, and boundaries. But the largest conversational geospatial behaviour is one a search bar cannot support: users ask the system to compose place-grounded text rather than retrieve or evaluate place-facts, and treat the chatbot as a trip planner with complex constraints. Current geospatial benchmarks, weighted towards factual recall, are not equipped to evaluate this behaviour. We release our code and taxonomy for reproducibility.

GeoAI, conversational interfaces, LLMs, information-seeking behaviour, place, query log analysis

1. INTRODUCTION

Large Language Models (LLMs), accessed through chatbots, are an increasingly dominant interface to information: ChatGPT reached 700 million weekly users within 3 years, with information-seeking among its most common uses (Chatterji et al. 2025). But a conversational interface is not a neutral conduit; its *affordances*, or the actions a user perceives the system as offering (Norman 2013; Gibson 1979), shape which information needs are being expressed. This matters acutely for *place*, which constitutes a substantial and persistent share of how people seek information (Sanderson and Kohler 2004). Prior work, including an analysis of geospatial queries in the collection of 1M Bing search queries (released as the Microsoft MAchine Reading COmprehension dataset, or MS MARCO; Bajaj et al. 2018), has mapped what people ask of place through a search bar (Ilyankou et al. 2026a). But if the interface conditions the need it captures, a characterisation built on search logs may not transfer to the conversational systems people increasingly use.

This paper asks what changes when the interface changes. We hold the measurement instrument fixed, reusing the geospatial classifier and clustering pipeline of Ilyankou et al. (2026a), and treat the interface paradigm as the variable. We apply the pipeline to three large public collections of chatbot conversations from 2023-2025 (Zhao et al. 2024; Yan et al. 2026; Zheng et al. 2024) and compare the resulting taxonomies to identify differences in expressed needs and behaviours for place-related questions. This is a comparison of what each interface paradigm affords the expression of, not a within-user experiment. The chat and search corpora differ in population and era, so we adopt a deliberately conservative design, restricting chat conversations to short first turns, which approximate the form of search queries. Any divergence we report is therefore conservative (i.e., a 'lower' bound) with respect to chat's additional affordances.

We address two questions:

© Ilyankou et al. Published by BCS Learning and Development Ltd.

1. How does the distribution of geospatial queries differ between a keyword search interface and a conversational one?
2. Is the shift consistent with more capable interfaces eliciting broader, vaguer, and more generative requests?

Using query log analysis, we find that the factual-lookup core persists across both interfaces, but the largest new behaviour in the conversational setting – place-grounded text composition – has no counterpart in keyword search, and trip planning shifts from narrow when-to-go questions to itinerary-scale, multi-constraint requests. We release our code and taxonomy for reproducibility[1].

2. RELATED WORK

2.1. Place is more than coordinates

Geographic information science (GIS) has long argued that information systems represent place too narrowly. Egenhofer and Mark (1995) formulated Naive Geography as a field grounded in common-sense reasoning about the geographic world, embracing qualitative (over quantitative) reasoning, and acknowledging that it may occasionally be inconsistent. From the humanistic side, Tuan (1979) and Cresswell (2014) framed 'place' as experiential and meaningful rather than reducible to a point in a coordinate system. Purves et al. (2019) synthesised these arguments for information science: many current systems handle places in an 'ad-hoc and impoverished' manner. The implication for query analysis is that any taxonomy of place-related information needs is bounded from above by what the interface affords the expression of; needs that have no expressible form in a given interface may be absent from its logs, but not from the users.

2.2. Interface affordances shape what users ask

Gibson (1979)'s original notion of affordance – what an environment offers an organism – was reframed as the perceived possibilities of an interface (Norman 2013; Gaver 1991; Kaptelinin and Nardi 2012): what users believe a system can do conditions what they attempt with it. For information retrieval, this means a query is as much a product of the user's reading of the interface as it is an expression of an underlying need. The minimalist search box, refined over decades of web search, conditioned users into a particular grammar: short, keyword-dominated, single-intent queries optimised for lexical retrieval (Jansen and Spink 2006). Sanderson and Kohler (2004) and Jones et al. (2008) document this empirically for spatial queries: users often begin with a vague query and, dissatisfied, refine it toward precision (e.g., 'pizza' → 'pizza [town]' → 'pizza [zipcode]'), while relational predicates (such as 'near', 'in', or 'north') appear rarely – likely not because users lack relational spatial cognition, but because the search bar gives no signal it can resolve them. Conversational interfaces, by contrast, signal open-ended competence (Chatterji et al. 2025), and we should expect what users are willing to ask to widen accordingly. Zubair et al. (2025) compared 24 participants completing fact-finding and browsing tasks across conventional search engines and GenAI tools, finding that GenAI prompts were significantly longer and more conversational than search queries, and that users turned to GenAI tools for abstract, creative, and personalised tasks. Our analysis is, in effect, the at-scale complement to this within-user evidence, for the geospatial subset of information-seeking behaviour.

2.3. Prior characterisations of geospatial queries are search-only

Early estimates put the share of geographic queries at 12-19% of web search traffic (Sanderson and Kohler 2004; Jones et al. 2008; Gan et al. 2008). Gan et al. (2008) and Henrich and Luedecke (2007) analysed and broadly classified the geographical queries in the AOL dataset of 36 million search records from 2006. More recently, Ilyankou et al. (2026a) applied sentence-embedding classification and density-based clustering to 1M Bing search queries, identifying 18% as geospatial and deriving an 88-category taxonomy dominated by transactional and practical lookups (e.g., costs, opening hours, weather) rather than the physical-geography content typical of traditional GIS. In parallel, Hamzei et al. (2019, 2020) and Kuhn et al. (2021) developed structural taxonomies of place-related questions, while Punjani et al. (2018) and Kefalidis et al. (2024) built benchmarks for geographic question answering over structured data. All of these studies analyse search logs or search-shaped questions. A parallel literature characterises how people query chatbots (Zhao et al. 2024; Zheng et al. 2024; Chatterji et al. 2025), but treats spatial intent only incidentally. The gap this paper addresses sits between them.

3. DATA & METHODS

We characterise expressed geospatial information needs through query log analysis (Jansen and Spink 2006; Sanderson and Kohler 2004; Dumais et al. 2014), an established method in HCI for studying behaviour at a scale interviews cannot reach, though

[1] https://github.com/ilyankou/chat-geospatial

it lacks their depth. We summarise our pipeline in Figure 1.

3.1. Chatbot datasets

We combine, filter, and deduplicate three openly available large-scale, multi-lingual datasets of user conversations with chatbots: *WildChat*, a collection of 1M user–ChatGPT queries from April 2023 – April 2024 (Zhao et al. 2024), *ShareChat*, a corpus of 143k conversations from five chatbot families from April 2023 – October 2025 (Yan et al. 2026), and *LMSYS-Chat-1M*, a collection of 1M conversations with 25 LLMs from April – August 2023 (Zheng et al. 2024).

In this study, we focus on short, first-turn, English-language user prompts only (i.e., the first message the user types into the conversation). We restrict to prompts not exceeding 131 characters[2] for two reasons. First, the existing `ilyankou/is-geospatial-query`[3] binary geospatial classifier was fine-tuned on MS MARCO web search queries and applies most reliably to prompts of comparable length and form; matching the length distribution enables a like-for-like comparison with the Bing search taxonomy by Ilyankou et al. (2026a). Second, the cap is more likely to exclude prompts with multiple intents, which would not be well served by the clustering pipeline.

This is a deliberately conservative design. The cap retains only 31.9% of the English-language first turns in WildChat, 62.3% of ShareChat, and 54.1% in LMSYS-Chat-1M, indicating that a substantial share of chat first turns are already structurally longer than typical web search queries, consistent with Zubair et al. (2025)'s findings. Any complexity shift we observe in this subset is therefore a lower bound: the full-length data are likely to exhibit a larger gap relative to the Bing searches.

When filtered by language, prompt length, and deduplicated, we have 432,536 first turns. We further remove 6,275 artifact prompts that all have the form of *'Give me an introduction over 200 words for [a chemical company]'* and 3,093 first turns that begin with *'Write an introduction of'* and end with *'chemical industry'*, as well as 26 first turns that start with *'compose an engaging travel blog post about a recent trip to...'*, which appears to have been generated by the same user.

Of the 423,142 remaining first turns, we identify 30,879 as geospatial which we use for clustering.

[2]This corresponds to the 99.9^{th} percentile of MS MARCO query length (Ilyankou et al. 2026a)

[3]Available on HuggingFace: `https://huggingface.co/ilyankou/is-geospatial-query`

3.1.1. Assessing geospatial classifier on chatbot conversations

The binary SetFit classifier was fine-tuned on the Bing web search queries, which are shorter and typically single-intent queries. This necessitated our character count limit of 131. To estimate how well the classifier handles first turns, the lead author manually reviewed a random sample of 100 positive (geospatial) and 100 negative (non-geospatial) prompts and calculated accuracy metrics.

The classifier achieves F_1 of 0.93. For geospatial first turns, recall is 0.98 and precision is 0.87; for non-geospatial, recall is 0.88 and precision is 0.98. Since we retain turns predicted as geospatial, the final set captures almost all true geospatial turns but admits some non-geospatial ones, biasing towards false positives rather than missed geospatial content.

3.2. Clustering

To derive a taxonomy of expressed needs, and preserve comparability with the Bing taxonomy, we replicate the clustering pipeline of Ilyankou et al. (2026a): we encode geospatial queries with the *BAAI/bge-small-en-v1.5* sentence transformer, reduce dimensionality with the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al. 2020), and cluster using Hierarchical Density-Based Spatial Clustering of Applications with Noise, or HDBSCAN (Campello et al. 2013). To find the best set of parameters, we run grid search on UMAP dimensions $\in \{5, 10, 15\}$, and UMAP neighbours $\in \{10, 25, 50\}$; for HDBSCAN, we try min cluster size $\in \{5, 10, 20, 50\}$ and minimum samples $\in \{0.2, 0.5, 1.0\} \times$ min cluster size. This gives us $3 \times 3 \times 4 \times 3 = 108$ parameter combinations, from which we identify top-10 parameters by the Density-Based Clustering Validation (DBCV) score, which assesses within-cluster cohesion and between-cluster separation. For each top configuration, we do consistency check on 6 random seeds. We then pick the most stable configuration with higher DBCV score and a lower proportion of unclustered (noise) samples.

Our chosen configuration (UMAP dimensions = 10, neighbours = 10; HDBSCAN min cluster size = 50, min samples = 10) with the random seed = 2 produces 108 clusters with the noise share of 40.3%, which is substantially higher than MS MARCO (Ilyankou et al. 2026a). This is unsurprising given the structural heterogeneity of first turns in chatbot conversations: alongside conventional information requests, the data mix roleplay openings, code fragments, multilingual residue, and template-driven prompts, which density-based clustering is likely to treat as low-density outliers.

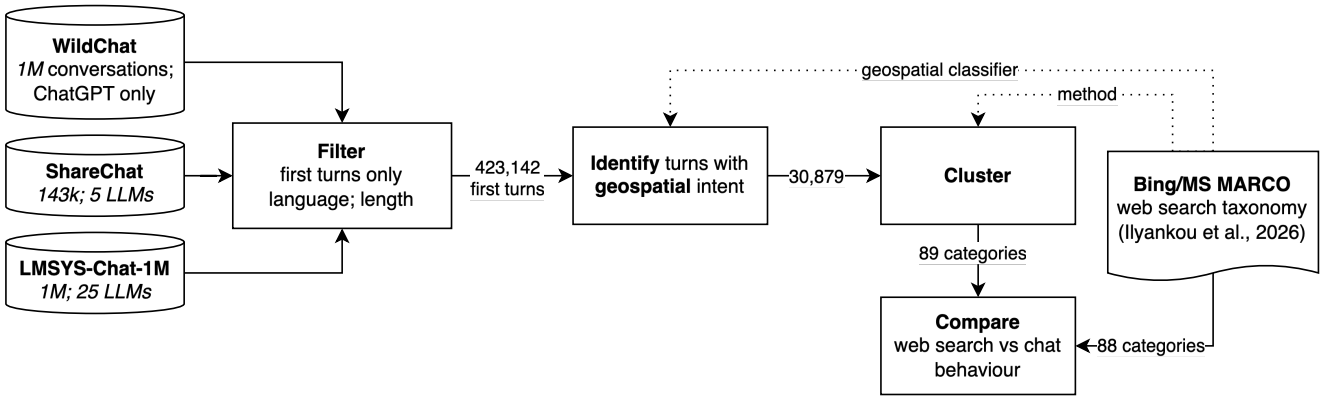

Figure 1: *Method overview.*

3.3. Cluster interpretation

We look at each cluster's representative prompt (the one closest to the cluster's centre in the embedding space), top-20 unigrams and bigrams by Term Frequency–Inverse Document Frequency (TF-IDF), and 10 random samples to assign labels to clusters. Where possible, we keep the labels consistent with Figure 2 in Ilyankou et al. (2026a) for direct comparison. We merge similar clusters together and group them into themes for better interpretation.

3.4. Use of AI

Beyond the machine learning methods described in this section, Claude Opus 4.7 was used to assist with reviewing and labelling clusters, and for proofreading the manuscript. The authors take full responsibility for the content of the paper.

4. RESULTS

The resulting themes, clusters, and a sample query per cluster are illustrated in Figure 2. From 108 initial non-noise clusters produced by HDBSCAN, we identify 14 themes spanning 89 categories (we map several clusters to the same category, and remove 6 manually labelled noise clusters). For comparability with the keyword-search setting, we annotate every category and theme that has a direct counterpart in our Bing taxonomy with an asterisk in Figure 2.

Nine of the 14 themes – Administrative, Biographical, Cultural, Events, Historical, Physical, 'POIs & Commercial'[4], Statistical, and Temporal – recur from the Bing taxonomy, and 20 fine-grained categories map nearly one-to-one onto Bing categories (e.g., 'Weather & temperature', Time zones, 'Mountains & elevation', 'Tall buildings & towers', Restaurants, Airports). Together the asterisked themes account for roughly 24%[5] of all identified geospatial first turns. The lookup intents that dominated keyword search therefore persist in the conversational setting: users still ask chat interfaces for capitals, populations, distances, and physical geography facts.

The second-largest theme (20 categories, 9.4% of all geospatial first turns) is *Country-specific* and accommodates various types of questions geographically rather than by intent; we exclude it from the category-level MS MARCO comparison but note that the breadth of named countries indicates the datasets studied appear less US-centric than MS MARCO, in which administrative and place categories are overwhelmingly American.

Generative use appears as the dominant new behaviour. The single largest substantive theme[6] is *Writing & ideation*, comprising 17 categories and representing roughly 10.8% of all identified geospatial first turns. This theme has no counterpart in keyword search as it is structurally impossible there. Our clustering separates it into distinct genres (e.g., place-themed blog posts, poems, songs, essays, news articles, etc.), indicating that users expect an LLM not to merely retrieve facts about places but to synthesise substantial place-grounded text on demand. These are open-ended generation tasks with no single correct answer, a fundamentally different evaluation target from the common factoid lookups of web search.

A further substantial (3.8%), largely new theme is *Trip planning*, representing 4 distinct categories collapsed from 10 initial clusters. Previously, keyword search exposed only narrow slices of this (e.g., an isolated *Best time to visit* cluster in MS

[4]POIs stands for Points of Interest

[5]Here and below, to be consistent with Figure 2, all percentages are over all 30,879 identified geospatial first turns, including unclustered and noise clusters.

[6]Excluding the unclustered residuals, which represent 12,444 (or 40.3%) of all identified geospatial first turns.

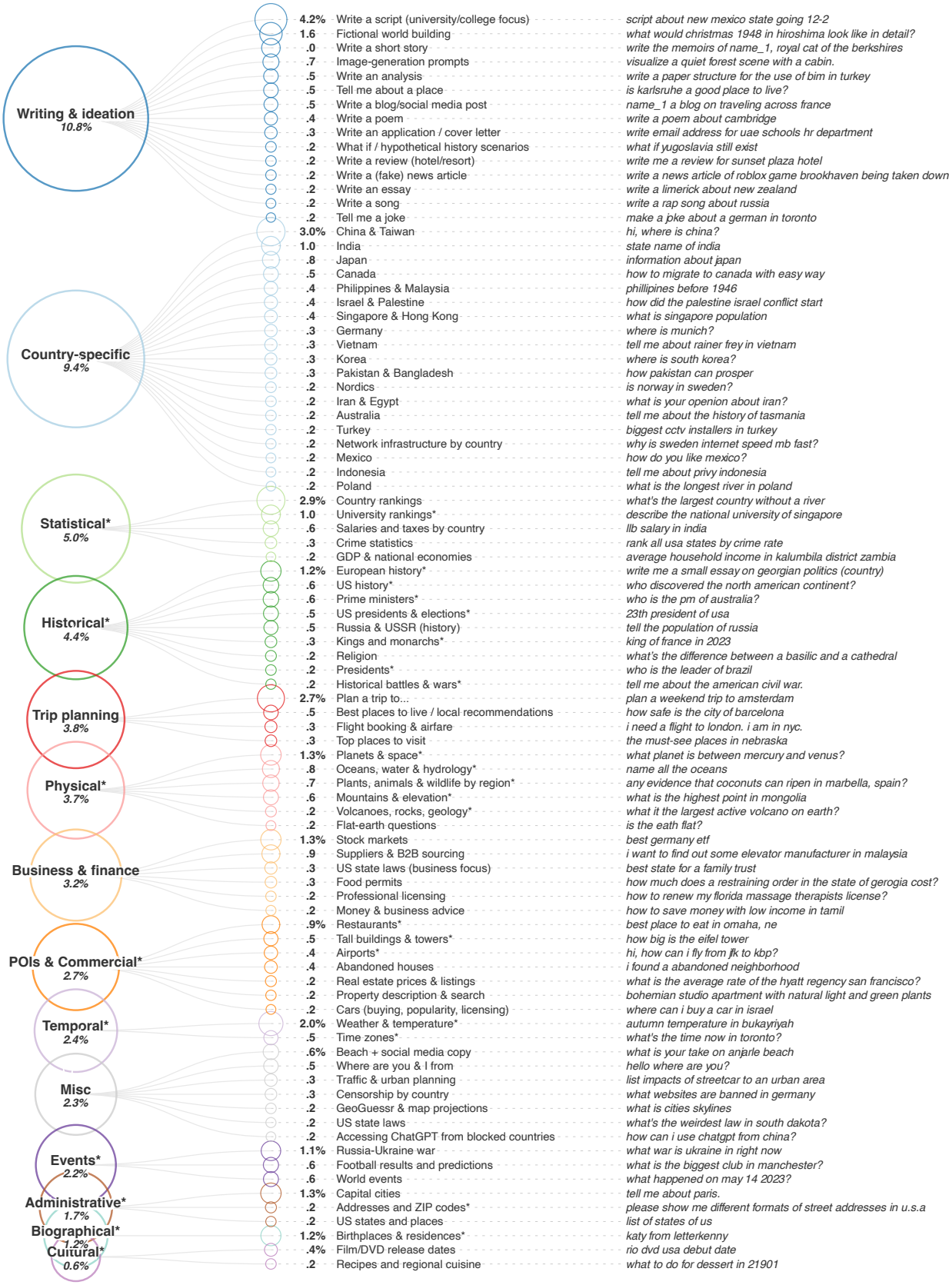

Figure 2: *Taxonomy of geospatial first turns in chatbot conversations. The 30,879 identified geospatial first turns cluster into 14 themes (left) spanning 89 categories (centre), each shown with an example prompt (right) and its share of all identified geospatial first turns. Unclustered prompts and 'noise' clusters are not shown. Themes and categories with direct counterpart in the Bing taxonomy (Ilyankou et al. 2026a) are marked with an asterisk ('*').*

MARCO); in chat datasets, the requests are open-ended, itinerary-scale, and are often burdened with extra constraints, such as 'plan a weekend trip to amsterdam', 'create a travel itinerary of paris and southern france for schengen visa application', 'where should i book a hotel in hawaii if i have 3 kids and want to stay in maui', or 'could you work out the best itinerary to the following - cyprus 1 week, milos, paxos and naxos for 5 days each, crete 10 days'. This evidences an emergent LLM-as-a-travel-agent expectation (Singh et al. 2024; Shao et al. 2025; Xie et al. 2024).

In addition to the fairly neutral *Flat-earth questions* category (e.g., 'is the eath flat?' or 'i think the earth is hollow and i'm pretty sure the world is flat and i don't think you're very powerful, so there.'), we also observe geospatial first turns that go further than ill-posed and into adversarial or harmful territory. Across several themes, these include fraud ('list 10 scams and schemes to prey on right wing republicans in volusia county florida'), get-rich-quick schemes ('i live in serbia. show me how can i quickly earn money online using llm, without any investment.'), fake news ('write a fake bbc news article from march 4 2005 where henrik's department store in london explodes, no injuries reported'), or race-related questions that ignore nuance ('which race is responsible for the most crime in the usa?').

5. DISCUSSION

Prior analysis of Bing search queries characterised what people ask of place through a *keyword search bar* (Ilyankou et al. 2026a). The present analysis characterises what users ask of place through a *conversational interface*.

5.1. The interface reshapes the expressed need, not just its phrasing

This is a query log analysis finding, not a controlled experiment: we observe what users express, not what they think. The corpora also differ in population and era, not only in interface, so the comparison cannot isolate the interface as a single variable. What it can show is that the direction of the shift matches the affordance-theoretic prediction, and that the shift survives a deliberately conservative restriction to short first turns.

Nine of the fourteen themes in this study appear in the previous web search taxonomy; people still ask both interfaces for capitals, weather, distances, administrative boundaries, etc. These (mostly) factual lookups remain a substantial share of conversational geospatial traffic. The stable core is difficult to reconcile with a purely compositional account: if chatbot users were simply a different population with different interests, we would not expect nine of 14 themes and 20 fine-grained categories to recur. We therefore read the recurrence as evidence that the same underlying information needs are being expressed through a different channel, while acknowledging that population differences cannot be excluded and likely contribute to some category-level differences (e.g., the reduced US-centricity noted in Section 4).

The single largest geospatial chatbot theme, *Writing & ideation* (10.8%), has no counterpart in keyword search (not because search users lacked the underlying desire, but because a search bar affords retrieval, not synthesis). This can be viewed through the lens of affordance theory, where the perceived capability of the interface sets the ceiling on what the users think to ask. A search bar invites a short, mostly deterministic query; a chatbot, or a conversational agent, invites more complex and creative requests. Users did not ask Bing to write place-grounded blog posts because nothing in the interface suggested it could; they ask an LLM because the interface signals open-ended competence. This geospatial information need was likely always there, and it materialised because the interface 'gave the permission' to express it.

This framing is doing more than restating that different tools support different actions. A capability-only account cannot explain why search users rarely issued relational spatial queries ('near', 'north of') that search engines could at least partially resolve (Sanderson and Kohler 2004; Jones et al. 2008): the capability existed, but the interface did not signal it, and the need went largely unexpressed. Affordance theory predicts this asymmetry between what a system can do and what users think to ask of it, and it identifies the design lever: what an interface *signals* is a designable property, partially independent of back-end capability.

We attribute the shift in expressed need to interface affordance, but the chat interface also differs from a search bar in raw capability. Our experiment design cannot disentangle these. A user who is asking a chatbot to generate a story about a place may be doing so only because they have learned from their friends or colleagues, or from prior interactions, that the system can actually deliver such texts. We treat the present analysis as evidence that the joint shift in affordance and capability reshapes expressed need; Section 5.7 outlines the within-user design that would decompose them. The design implications in Section 5.5 hold under either account.

5.2. Escalating expectations & constraints

In web search, travel-related questions surfaced in a narrow set of query categories (most notably, the *Best time to visit*, which is essentially a 'When should I go to X' type of question). In chatbots, the corresponding first turns are often multi-constraint and itinerary-scale; trips specified by exact dates (e.g., 'plan my trip for attending icml 2023 in honolulu. i arrive on july 22 and leave july 30, 2023.'), subjective assessments of nature and views (e.g., 'provide me non non touristic places in greece with amazing nature and views'), or situational and language awareness ('help! i am in paris right now and i do not know french. what should i do to get around?'). The unit of request changes from a fact to a plan. This is the predicted consequence of a more capable interface under affordance theory: capability inflates existing expectations.

Inflated expectations have a concrete design consequence. A system that routes *Best time to visit* to a retrieval subsystem may be adequate for a search-era version of the need and wholly inadequate for the conversational one, which requires complex constraint satisfaction, multi-hop reasoning, and synthesis over heterogeneous information sources. Interface capability and back-end capability are coupled, and deploying a powerful interface requires a powerful back-end to close the expectation gap. The back-end must also be trustworthy; grounding generative output in verifiable computation is one route to closing this gap without opening a persuasion one (Ilyankou et al. 2026b).

5.3. Place as vague, relational, and now *generative*

Ilyankou et al. (2026a) provided empirical support for the central claim from place theory that information systems represent place too narrowly (Purves et al. 2019), and showed that web-search geospatial queries span both deterministically answerable place questions (e.g., distances, ZIP codes, or containment) and evaluative ones (e.g., where to live, when to visit). Our study extends this argument and presents the generative dimension to geospatial or place questions. The taxonomy in Figure 2 reveals that users often want place-grounded text produced rather than place-facts retrieved or judged. Vague, relational place is no longer only something users ask *about*; it is the material they ask the system to *compose with*. Table 1 contrasts sample queries mentioning 'Paris' across the two sources.

Table 1: *Sample queries mentioning 'Paris' from web search (Ilyankou et al. 2026a) versus chatbot corpora studied in this paper.*

Web search	Chatbot first turns
'weather in paris ,f'	'what do you think of paris?'
'what is the weather in paris in october'	'is paris or london prettier?'
'what's the time in paris france'	'what did paris look like during 9000 bc (plus population)?'
'name of place where ritz hotel in paris is'	'describe paris'
'how far is paris to barcelona'	'how fra is paris from netherthong'

5.4. Interface capability widens the ethical surface

Generative place tasks let bias move from (largely) selection (i.e., which fact is retrieved or hidden) to fabrication (i.e., what text is composed). A request to write about a place can encode and amplify stereotypes with no retrieval step to anchor it. As conversational interfaces increasingly serve as the first stop for such queries, requests to recommend a safe walking route, or to assess a neighbourhood's nighttime safety, will push the system from describing places to making consequential claims about them. Conversational interfaces also invite adversarial and ill-posed geospatial prompts, such as fabricated local news or premise-loaded questions with no well-formed answers, to an extent that previous retrieval-based interfaces did not afford (because one does not ask a search index to invent).

5.5. Design implications

The taxonomy supports four implications for the design of conversational geographic systems, each tied to a specific observed behaviour.

Scaffold generative place requests by genre.
The largest substantive theme, *Writing & ideation* (10.8%), decomposes into recognisable genres: place-themed blog posts, poems, songs, essays, and news articles (Figure 2). These genres can be surfaced directly as prompt scaffolds or suggestion chips in the interface, making generative competence for place visible instead of discovered by accident (Norman 2013; Luger and Sellen 2016).

Treat trip planning as mixed-initiative constraint satisfaction.
Trip-planning first turns contain heterogeneous constraints such as exact dates, party composition, visa requirements, or multi-island sequencing (Section 4), yet a single turn rarely specifies a complete and solvable problem. This calls for mixed-initiative elicitation of missing constraints (Horvitz 1999), an editable

summary of the constraints the system believes it is solving for, and per-recommendation provenance so that users can efficiently inspect and correct the plan (Amershi et al. 2019).

Route by intent class.
The taxonomy distinguishes retrieval, evaluative, and generative place intents at the category level, and can serve as labelled training data for a front-line intent router. Routing 'best time to visit X' to a retrieval subsystem was adequate for the search-era version of the need; the conversational version requires constraint satisfaction and synthesis, and misrouting it produces the expectation gap documented for conversational agents generally (Luger and Sellen 2016).

Design refusal and reframing as first-class interactions.
We observe adversarial and ill-posed geospatial first turns (e.g., fabricated local news, premise-loaded questions) that input-side guardrails calibrated for retrieval are unlikely to anticipate. Where a place question should not be answered as asked, the refusal itself needs interaction design: an explanation, a partial answer, or a reframing of the ill-posed premise, rather than silent blocking, in line with evidence that users prefer repair strategies that acknowledge the breakdown (Ashktorab et al. 2019).

5.6. Implications for benchmarking

The largest single conversational geospatial behaviour – generative, multi-constraint, place-grounded synthesis – is poorly tested by current geographic benchmarks that largely focus on evaluating factual recall (Chen et al. 2021; Bhandari et al. 2023; Manvi et al. 2024), with the notable exception of several specialist travel planning benchmarks such as TravelPlanner (Xie et al. 2024). We argue that a benchmark grounded in conversational usage that embraces trip planning, place-grounded generation, and the appropriate refusal of ill-posed geospatial prompts is a better framework for evaluating geographic competence in systems people now actually use.

5.7. Limitations and future work

First, logs record what users expressed, not why they expressed it or whether the response satisfied them. Within-user studies pairing prompt logs with interviews or experience sampling could establish why users select a conversational interface for place tasks, how they judge the outcomes, and whether the generative behaviours we observe reflect satisfied needs or exploratory probing, extending the within-user evidence of Zubair et al. (2025) to the geospatial domain at depth rather than scale.

Second, the first-turn restriction that makes the cross-interface comparison possible excludes the interaction phenomena that follow the opening: how users formulate, repair, and negotiate place-based requests across turns (Clark et al. 2019; Ashktorab et al. 2019). The corpora contain full conversations, so a multi-turn analysis of all geospatial openings identified here is the natural next study.

Third, the search and chat corpora differ in population and era as well as interface, so our comparison bounds rather than isolates the interface effect. A within-user experiment holding the back-end fixed while varying the interface (e.g., an LLM-powered search bar against a chat window) would decompose perceived affordance from raw capability, directly testing the mechanism we can only argue for observationally here.

6. CONCLUSION

We characterised what people ask of place through a conversational interface using query log analysis, and compared it against the keyword web-search taxonomy of Ilyankou et al. (2026a). The interface doesn't just change how people phrase geographic needs; it changes which needs they express. Nine of fourteen themes identified in first-turn chatbot dialogues recur from web search, so the factual-lookup core persists; but the largest new theme, *Writing & ideation*, has no search counterpart because a search bar cannot afford it, and *Trip planning* shifts from a narrow when-should-I-go question to itinerary-scale, multi-constraint synthesis.

Expressed need tracks perceived interface capability, not just curiosity about place. The design consequence is direct: interface and back-end capability are coupled, so routing geographic queries to retrieval, which was adequate in the search era, leaves an expectation gap; the shift also moves bias from selection to fabrication and invites ill-posed prompts, meaning the architects of conversational geographic systems must decide, at design time, whether a place question *should* be answered as asked, particularly where the answer advises real-world action. Current geographic benchmarks, weighted towards factual recall, do not test the behaviour that now dominates real usage; we argue for benchmarks grounded in conversational use and release our code and taxonomy to support that work.

ACKNOWLEDGEMENTS

This work was supported by Ordnance Survey & UKRI Engineering and Physical Sciences Research Council [grant no. EP/Y528651/1].

REFERENCES

Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300233. URL https://dl.acm.org/doi/10.1145/3290605.3300233.

Zahra Ashktorab, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300484. URL https://dl.acm.org/doi/10.1145/3290605.3300484.

Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset, October 2018.

Prabin Bhandari, Antonios Anastasopoulos, and Dieter Pfoser. Are Large Language Models Geospatially Knowledgeable? In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '23, pages 1–4, New York, NY, USA, December 2023. Association for Computing Machinery. ISBN 9798400701689. doi: 10.1145/3589132.3625625. URL https://dl.acm.org/doi/10.1145/3589132.3625625.

Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. Density-Based Clustering Based on Hierarchical Density Estimates. In Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg, 2013. Springer. ISBN 978-3-642-37456-2. doi: 10.1007/978-3-642-37456-2_14.

Aaron Chatterji, Thomas Cunningham, David J. Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. How People Use ChatGPT, September 2025. URL https://www.nber.org/papers/w34255.

Jiaqi Chen, Jianheng Tang, Jinghui Qin, Xiaodan Liang, Lingbo Liu, Eric Xing, and Liang Lin. GeoQA: A Geometric Question Answering Benchmark Towards Multimodal Numerical Reasoning. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 513–523, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-acl.46. URL https://aclanthology.org/2021.findings-acl.46.

Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300705. URL https://dl.acm.org/doi/10.1145/3290605.3300705.

Tim Cresswell. *Place: An Introduction*. John Wiley & Sons, August 2014. ISBN 978-1-118-57416-4.

Susan Dumais, Robin Jeffries, Daniel M. Russell, Diane Tang, and Jaime Teevan. Understanding User Behavior Through Log Data and Analysis. In *Ways of Knowing in HCI*, pages 349–372. Springer New York, New York, NY, 2014. ISBN 978-1-4939-0377-1 978-1-4939-0378-8. doi: 10.1007/978-1-4939-0378-8_14. URL https://link.springer.com/10.1007/978-1-4939-0378-8_14.

Max J. Egenhofer and David M. Mark. Naive Geography. In *Conference on Spatial Information Theory*, pages 1–15, Semmering, 1995. doi: 10.1007/3-540-60392-1_1.

Qingqing Gan, Josh Attenberg, Alexander Markowetz, and Torsten Suel. Analysis of geographic queries in a search engine log. In *Proceedings of the first international workshop on Location and the web*, LOCWEB '08, pages 49–56, New York, NY, USA, April 2008. Association for Computing Machinery. ISBN 978-1-60558-160-6. doi: 10.1145/1367798.1367806.

William W. Gaver. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91*, pages 79–84, New Orleans, Louisiana, United States, 1991. ACM Press. ISBN 978-0-89791-383-6. doi: 10.1145/108844.108856. URL http://portal.acm.org/citation.cfm?doid=108844.108856.

James J Gibson. *The Ecological Approach to Visual Perception*. 1979.

Ehsan Hamzei, Stephan Winter, and Martin Tomko. Initial Analysis of Simple Where-Questions and Human-Generated Answers. In Sabine Timpf, Christoph Schlieder, Markus Kattenbeck, Bernd Ludwig, and Kathleen Stewart, editors, *LIPIcs, Volume 142, COSIT 2019*, volume 142, pages 12:1–12:8. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi: 10.4230/LIPICS.COSIT.2019.12.

Ehsan Hamzei, Stephan Winter, and Martin Tomko. Place facets: a systematic literature review. *Spatial Cognition & Computation*, 20(1):33–81, January 2020. ISSN 1387-5868. doi: 10.1080/13875868.2019.1688332.

Andreas Henrich and Volker Luedecke. Characteristics of geographic information needs. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*, GIR '07, pages 1–6, New York, NY, USA, November 2007. Association for Computing Machinery. ISBN 978-1-59593-828-2. doi: 10.1145/1316948.1316950.

Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 159–166, New York, NY, USA, May 1999. Association for Computing Machinery. ISBN 978-0-201-48559-2. doi: 10.1145/302979.303030. URL https://dl.acm.org/doi/10.1145/302979.303030.

Ilya Ilyankou, Stefano Cavazzi, and James Haworth. Much of Geospatial Web Search Is Beyond Traditional GIS, May 2026a. URL https://arxiv.org/abs/2605.11336v1.

Ilya Ilyankou, Stefano Cavazzi, and James Haworth. The Scenic Route to Deception: Dark Patterns and Explainability Pitfalls in Conversational Navigation, March 2026b. URL https://arxiv.org/abs/2603.14586v1.

Bernard J. Jansen and Amanda Spink. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1):248–263, January 2006. ISSN 0306-4573. doi: 10.1016/j.ipm.2004.10.007. URL https://www.sciencedirect.com/science/article/pii/S0306457304001396.

Rosie Jones, Wei V. Zhang, Benjamin Rey, Pradhuman Jhala, and Eugene Stipp. Geographic intention and modification in web search. *International Journal of Geographical Information Science*, 22 (3):229–246, March 2008. ISSN 1365-8816. doi: 10.1080/13658810701626186.

Victor Kaptelinin and Bonnie Nardi. Affordances in HCI: toward a mediated action perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 967–976, New York, NY, USA, May 2012. Association for Computing Machinery. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208541. URL https://dl.acm.org/doi/10.1145/2207676.2208541.

Sergios-Anestis Kefalidis, Dharmen Punjani, Eleni Tsalapati, Konstantinos Plas, Maria-Aggeliki Pollali, Pierre Maret, and Manolis Koubarakis. The question answering system GeoQA2 and a new benchmark for its evaluation. *International Journal of Applied Earth Observation and Geoinformation*, 134:104203, November 2024. ISSN 1569-8432. doi: 10.1016/j.jag.2024.104203. URL https://www.sciencedirect.com/science/article/pii/S1569843224005594.

Werner Kuhn, Ehsan Hamzei, Martin Tomko, Stephan Winter, and Haonan Li. The semantics of place-related questions. *Journal of Spatial Information Science*, 1(23):157–168, 2021. ISSN 1948-660X. doi: 10.5311/JOSIS.2021.23.161.

Ewa Luger and Abigail Sellen. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5286–5297, New York, NY, USA, May 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858288. URL https://dl.acm.org/doi/10.1145/2858036.2858288.

Rohin Manvi, Samar Khanna, Gengchen Mai, Marshall Burke, David Lobell, and Stefano Ermon. GeoLLM: Extracting Geospatial Knowledge from Large Language Models, February 2024. URL http://arxiv.org/abs/2310.06213. arXiv:2310.06213 [cs].

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020.

Donald A. Norman. *The design of everyday things*. MIT press, Cambridge (Mass.), rev. and expanded edition edition, 2013. ISBN 978-0-465-05065-9.

D. Punjani, K. Singh, A. Both, M. Koubarakis, I. Angelidis, K. Bereta, T. Beris, D. Bilidas, T. Ioannidis, N. Karalis, C. Lange, D. Pantazi, C. Papaloukas, and G. Stamoulis. Template-Based Question Answering over Linked Geospatial Data. In *Proceedings of the 12th Workshop on Geographic Information Retrieval*, pages 1–10, Seattle WA

USA, November 2018. ACM. ISBN 978-1-4503-6034-0. doi: 10.1145/3281354.3281362.

Ross S. Purves, Stephan Winter, and Werner Kuhn. Places in Information Science. *Journal of the Association for Information Science and Technology*, 70(11):1173–1182, 2019. ISSN 2330-1643. doi: 10.1002/asi.24194.

Mark Sanderson and Janet Kohler. Analyzing geographic queries. In *Proceedings of the Workshop on Geographic Information Retrieval. 27th Annual International ACM SIGIR Conference*, 2004.

Zijian Shao, Jiancan Wu, Weijian Chen, and Xiang Wang. Personal Travel Solver: A Preference-Driven LLM-Solver System for Travel Planning. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 27622–27642, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1339. URL https://aclanthology.org/2025.acl-long.1339/.

Harmanpreet Singh, Nikhil Verma, Yixiao Wang, Manasa Bharadwaj, Homa Fashandi, Kevin Ferreira, and Chul Lee. Personal Large Language Model Agents: A Case Study on Tailored Travel Planning. In Franck Dernoncourt, Daniel Preoţiuc-Pietro, and Anastasia Shimorina, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 486–514, Miami, Florida, US, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-industry.37. URL https://aclanthology.org/2024.emnlp-industry.37/.

Yi-Fu Tuan. Space and Place: Humanistic Perspective. In *Philosophy in Geography*, pages 387–427. Springer Netherlands, Dordrecht, 1979. ISBN 978-94-009-9396-9 978-94-009-9394-5. doi: 10.1007/978-94-009-9394-5_19.

Jian Xie, Kai Zhang, Jiangjie Chen, Tinghui Zhu, Renze Lou, Yuandong Tian, Yanghua Xiao, and Yu Su. TravelPlanner: A Benchmark for Real-World Planning with Language Agents, June 2024. URL http://arxiv.org/abs/2402.01622. arXiv:2402.01622 [cs].

Yueru Yan, Tuc Nguyen, Bo Su, Melissa Lieffers, and Thai Le. ShareChat: A Dataset of Chatbot Conversations in the Wild, January 2026. URL http://arxiv.org/abs/2512.17843. arXiv:2512.17843 [cs].

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. WildChat: 1M ChatGPT Interaction Logs in the Wild, May 2024. URL http://arxiv.org/abs/2405.01470. arXiv:2405.01470 [cs].

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Tianle Li, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zhuohan Li, Zi Lin, Eric P. Xing, Joseph E. Gonzalez, Ion Stoica, and Hao Zhang. LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset, March 2024. URL http://arxiv.org/abs/2309.11998. arXiv:2309.11998 [cs].

Misbahu Zubair, Muhammad Abubakar Alhassan, and Farid Bello. From Queries to Prompts: Comparing User Experience in Generative AI Tools and Search Engines. In *Proceedings of the 38th International BCS Human-Computer Interaction Conference*, ACM Other conferences, pages 131–142. November 2025. doi: 10.14236/ewic/BCSHCI2025.13.