Computational Humanities Research

www.cambridge.org/chr

Towards efficient and accessible geoparsing of U.K. local media: A benchmark dataset and LLM-based approach

Simona Bisiani[1], Agnes Gulyas[2] and Bahareh Heravi[1]

[1]Institute for People-Centred AI, University of Surrey, Stag Hill, Guildford, UK and [2]School of Creative Arts and Industries, Canterbury Christ Church University, Canterbury, UK

Research Article

Cite this article: Bisiani Simona, Agnes Gulyas and Bahareh Heravi. 2025. "Towards efficient and accessible geoparsing of U.K. local media: A benchmark dataset and LLM-based approach" *Computational Humanities Research*, **1**:e10, https://doi.org/10.1017/chr.2025.10012

Received: 31 January 2025
Revised: 23 May 2025
Accepted: 16 September 2025

Keywords:
geoparsing; large language models (LLMs); location extraction; prompt engineering; toponym disambiguation; local news

Corresponding author:
Simona Bisiani;
Email: s.bisiani@surrey.ac.uk

Abstract

Location mentions in local news are crucial for examining issues like spatial inequalities, news deserts and the impact of media ownership on news diversity. However, while geoparsing – extracting and resolving location mentions – has advanced through statistical and deep learning methods, its use in local media studies remains limited and fragmented due to technical challenges and a lack of practical frameworks. To address these challenges, we identify key considerations for successful geoparsing and review spatially oriented local media studies, finding over-reliance on limited geospatial vocabularies, limited toponym disambiguation and inadequate validation of methods. These findings underscore the need for adaptable and robust solutions, and recent advancements in fine-tuned large language models (LLMs) for geoparsing offer a promising direction by simplifying technical implementation and excelling at understanding contextual nuances. However, their application to U.K. local media – marked by fine-grained geographies and colloquial place names – remains underexplored due to the absence of benchmark datasets. This gap hinders researchers' ability to evaluate and refine geoparsing methods for this domain. To address this, we introduce the Local Media UK Geoparsing (LMUK-Geo) dataset, a hand-annotated corpus of U.K. local news articles designed to support the development and evaluation of geoparsing pipelines. We also propose an LLM-driven approach for toponym disambiguation that replaces fine-tuning with accessible prompt engineering. Using LMUK-Geo, we benchmark our approach against a fine-tuned method. Both perform well on the novel dataset: the fine-tuned model excels in minimising coordinate-error distances, while the prompt-based method offers a scalable alternative for district-level classification, particularly when relying on predictions agreed upon by multiple models. Our contributions establish a foundation for geoparsing local media, advancing methodological frameworks and practical tools to enable systematic and comparative research.

Plain Language Summary

Local news articles often mention specific places, which is important for understanding issues, such as unequal news coverage, areas lacking local journalism and the influence of media ownership on news diversity. However, studying these location mentions is challenging due to inconsistent methods and technical difficulties. Geoparsing – the process of extracting and identifying place names – has improved with advances in artificial intelligence but remains underused in local media research because of its complexity and lack of accessible tools. To address this, we developed a new dataset of U.K. local news articles with accurately annotated locations. We also propose a novel approach using large language models (LLMs) that relies on prompt engineering, eliminating the need for resource-intensive fine-tuning. Our method performs competitively compared to fine-tuned models: while fine-tuned models achieve higher accuracy in pinpointing exact coordinates, our approach offers a scalable and accessible solution for identifying broader geographic areas, especially when combining predictions from multiple models. This work provides valuable resources and methods to support more reliable and systematic research on geographic references in local media.

Introduction

Locations mentioned inside of local news articles constitute a critical unit of analysis for a wide range of timely research topics, including evaluating spatial inequalities in local news provision (Napoli and Weber 2020), identifying news deserts (Khanom et al. 2023), measuring the impact of digital practices and media ownership consolidation on local relevance (Firmstone and Whittington 2021; Vogler, Weston, and Udris 2023) and defining local media in the digital age (Hagar et al. 2020). Yet, studies extracting and analysing location mentions are fragmented

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

across geographic contexts and methodologies, resulting in varied and isolated practices. This underscores the need to revisit research applications, core concepts and the defining features of local media relevant to converting texts into geographic data. Addressing these gaps would enable a more systematic approach to analysing local media's spatial dimensions, advance theoretical frameworks underpinning the study of geography in journalism, ground methods in robust conceptual foundations and foster comparative studies to deepen our understanding of the geographic, social and institutional factors shaping local journalism.

The task of extracting and spatially resolve location mentions in texts is known as geoparsing. Methodologically, geoparsing has seen recent significant improvements, shifting from rule-based systems to statistical and deep learning approaches (Zhang and Bethard 2024). However, the technical expertise required to implement these methodologies creates barriers to adoption beyond computational departments. This overhead may partly explain the limited use of geoparsing in local media research (Madrid-Morales 2020), despite its potential for scalable, advanced and efficient location retrieval. Selecting the appropriate geoparsing approach for local media is challenging. Existing evaluations offer limited insight into performance across diverse geographic contexts. Benchmarking datasets typically cover broad domains like news, social media and historical documents (Hu et al. 2024), and even news-focused datasets often fail to represent the nuances of the contemporary local news sector. The only resource dedicated to local media, the Local–Global Lexicon (LGL) corpus, comprises 588 articles from 78 U.S. newspapers (Lieberman, Samet, and Sankaranarayanan 2010). Consequently, the performance of geoparsing techniques on local media remains largely unknown. This is concerning, given that geoparsing methods struggle to achieve consistent performance across domains and regions due to gazetteer geographic biases and the predisposition to favour popular locations in ambiguous instances, clashing with the localised, fine-grained nature of local news geography (Hu et al. 2023). Particularly, a gap remains in the development of generalisable models capable of robust language understanding on both fine-grained and high-level geographic entities (Hu et al. 2024). In this context, large language models (LLMs) offer a promising avenue. Research suggests LLMs can improve the efficiency, accessibility and scalability of language understanding and generation tasks (Bommasani et al. 2022), including geoparsing (Hu et al. 2024, 2023). LLMs can perform highly on general tasks without extensive training or fine-tuning (Sanh et al. 2022), increasing accessibility for researchers without specialised expertise. Leveraging open-source LLMs in geoparsing pipelines could offer a scalable solution for local media research. However, while LLMs are attuned to context, they have limitations and biases in their geographic knowledge (Mai et al. 2024) and can generate inaccurate information, known as "hallucinations" (Huang et al. 2024). These limitations, particularly in disambiguating local media, require further investigation. To advance geoparsing research in local media and address current gaps, we contribute:

- A structured framework for developing robust geoparsing methodologies in media studies that promotes best practices and enables systematic cross-study comparisons.
- A comparative review of local media geoparsing research, identifying key limitations, gaps and effective approaches.
- The Local Media UK Geoparsing (LMUK-Geo) dataset: a gold-standard, hand-annotated corpus of 182 U.K. local news articles designed to support development and evaluation of geoparsing models for this underexplored geographic context.
- A novel, scalable and accessible geoparsing approach based on prompt-engineering of LLMs, providing an alternative to fine-tuning techniques. This approach effectively addresses disambiguation and contextual challenges in U.K. local news.

Through these contributions, we aim to establish a foundation for scalable, accessible and robust geoparsing of local media.

The relevance of place, space and location in journalism

Over recent decades, journalism studies have undergone a "spatial turn," bringing novel theories and empirical evidence relating to the relationship between media and geography (Reese 2016). At the core of this transformation is a redefinition of the notions of place and space, stimulated by digital technologies' disruption of traditional boundaries between media and the audience (Usher 2019). Place is understood not as a fixed geographical construct but as a fluid and contested social construction, shaped by information flows, shared experiences and the cultural meanings attached to location (Hess and Waller 2017). Journalism plays an important role in shaping the cultural meanings attached to location by mediating perceptions of place through representations of events, communities and local voices, actively embedding geography into socio-cultural narratives and constructing a "sense of place" that informs identity, community and a sense of belonging (Hess and Waller 2017; Weiss 2018). Geographic concepts, such as location and proximity, are central to understanding journalism's connection to place. Location refers to the spatial context of a news event – its geographic coordinates or cultural setting – which shapes journalistic decisions and audience perceptions (Weiss 2018). Similarly, proximity, whether physical or psychological, influences newsworthiness and the relevance of content to specific audiences (Koetsenruijter and De Jong 2023).

Both notions of proximity and location have taken centre stage in discussions about the performative role of local journalism today. Across a number of media systems, technological disruption has manifested as falling newspaper circulation and advertising revenue, difficulties in monetising digital content and in making sufficient revenue from digital advertising (Cairncross 2019). These challenges have gravely impacted the viability of local news operations. In the United Kingdom, advertising income in the print sector has dropped drastically, from £3.1 billion in 2004 to £0.5 billion in 2020 (Majid 2023). Furthermore, the digital reader subscriptions and advertising revenues have done little to compensate for these losses. Since 2005, hundreds of local media operations have disappeared (Hunter 2024), accelerating a decline that started in the mid-1980s (Franklin and Cushion 2006). The disappearance of local news outlets, taking place at similar scale in North America and Europe, has generated the term "news deserts," used to describe geographical areas lacking a locally dedicated news operation (Abernathy 2020). National-level mappings of news deserts have been effective in showing the declining state of local journalism today across different countries (Metzger 2024; Public Interest News Foundation 2023). Yet, emphasis on outlet presence overshadows that other ongoing phenomena, namely, media ownership consolidation and an emphasis on digital audience reach, are impacting proximity and location in the news (McAdam and Hess 2024). This signifies that the mere availability of a local news outlet does not guarantee the provision of local news reporting.

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

While these conditions are shared across a variety of media systems, the United Kingdom provides a clear example of these dynamics. The U.K.'s local print and digital media sector is dominated by a handful of large companies, with three major firms (Reach, National World and Newsquest) owning over 57% of the market as of 2023 (The Media Reform Coalition 2023). Economic pressures have incentivised consolidation as a strategy to achieve economies of scale, centralising printing and production processes to reduce costs. Reach exemplifies this trend, leveraging regional monopolies to amalgamate newspapers under centralised digital and physical "hubs" (Moore and Ramsay 2024). While newsroom closures and consolidations are not new (Franklin 2006), these practices have accelerated in the wake of the Coronavirus pandemic (Sharman 2021; Waterson 2021). This dispersion of infrastructure and staff implies that local journalism is now less "visible" and less "sensible," that is, less attuned to understanding the community for which the news is produced (McAdam and Hess 2024). This is particularly concerning as ownership consolidation has promoted the repurposing of content across titles under the same regional hub, or even under the same publisher (Sjvaag 2014). While repurposing content reduces the need for field reporting, content syndication is often viewed negatively, as it dilutes local news coverage (Garz and Ots 2025). Furthermore, the struggle to capture and maintain audience attention has led local news outlets to focus more heavily on entertaining content, such as celebrity news and sensational stories, deprioritising more substantive and critical topics (Napoli and Weber 2020).

Meanwhile, three of the four major local news groups have made significant job cuts in response to declining revenues and rising automation. Industry press PressGazette estimates that two-thirds of editorial jobs across Reach, National World and Newsquest have disappeared, falling from 9,000 to 3,000 between 2007 and 2022 (Ponsford 2024). While this doesn't necessarily mean fewer communities have a dedicated reporter, it does suggest that reporters are now responsible for increasingly larger geographic areas, forcing them to produce a significantly higher volume of stories (Lewis, Williams, and Franklin 2008). As a result, we can expect an ever greater "distance" between reporters, coverage and the audience. As Franklin (2006, xxi) put it, "In the new millennium, local newspapers are local in name only; the town or city emblazoned on the newspaper's masthead may be one of the few remaining local features of the paper."

The analysis of location mentions in local media

Amidst these transformations, the significance of location mentions in news content has grown. The places reported in the news and how they are portrayed offer valuable insights into the production of journalism and its societal impact. The distribution of geographic attention in news content reveals patterns of inequality across communities, such as the urban–rural divide in the availability, quality and accessibility of local journalism (Usher 2019). Beyond these social dimensions, the quantification of local news coverage itself has emerged as a key area of inquiry. This reflects a shared understanding of the vital role local journalism plays in fostering informed citizenship and community cohesion (Napoli and Weber 2020). As Lindgren (2009, 80) asserts: "From a social capital perspective [...] the amount of news coverage a community receives matters because information is an important determinant of community engagement and local democracy." Communities with weaker media infrastructure experience poorer public resource management, higher corruption and lower political engagement (Gao, Lee, and Murphy 2020; Hayes and Lawless 2018; PLUM Consulting 2020). As put by Ramsay and Moore (2016, 14), the consequences of dire news provision are "less engaged citizens, less scrutiny of authorities, poorer representation of shared concerns, less community cohesion, a sense of powerlessness and a lack of connectedness." However, the impact of current market conditions on news content, and subsequently, the effect of that content on democratic engagement, remains underexplored. Recently, the Public Interest News Foundation (PINF) has emphasised that, "the most important" research direction "would be to understand the quality and quantity of public interest news that each outlet produces – and whether that news covers the entire area that the outlet claims to cover" (Public Interest News Foundation 2023, 21). A number of studies have emerged in recent years, responding to these conditions, that have dissected local news content to measure: the decline of locally relevant news (Napoli and Weber 2020; Vogler, Weston, and Udris 2023); the scarcity of local constituency coverage in the context of general elections (Moore and Ramsay 2024); and the spatial distribution of location mentions (Khanom et al. 2023; Madrid-Morales, Rodríguez-Amat, and Lindner 2023). Together, these works illustrate a growing effort to quantify the geographic dimension of local journalism, particularly in response to challenges, such as news deserts and declining community news coverage. However, this body of research remains highly fragmented, with studies differing widely in geographic scope, methodological approaches and conceptual focus. For instance, while Madrid-Morales, Rodríguez-Amat, and Lindner (2023) relied on string matching against a gazetteer to extract locations, Vogler, Weston, and Udris (2023) employed a multi-step procedure that combined location identification with disambiguation to ensure the correct real-world entity was linked. These methodological variations reflect differences in technological choices and highlight the fragmented nature of geographic reference extraction in local media studies. Although geographic reference extraction is central to an established research domain known as geoparsing (Middleton et al. 2018), many studies neglect geoparsing theory and approaches in their methodologies. This oversight may stem from a lack of established frameworks for effectively applying geoparsing in media research, leading to inconsistent or inappropriate handling of key aspects of geographic reference extraction. As a result, both the systematic application of these methods and the ability to compare studies effectively are hindered. These challenges define our first two objectives:

O1 - To formulate a framework for applying geoparsing in media studies.

O2 - To conduct a comparative review of geoparsing in media studies to identify gaps and opportunities.

Geoparsing: State-of-the-art (SOTA) and key considerations

Extracting geographic information from texts, also known as geoparsing, involves two key tasks (Hu et al. 2023): 1) toponym recognition, *or geotagging*, which identifies location references in text and 2) toponym disambiguation, or *geocoding*, which resolves these references to real-world geographic entities, typically as coordinates or polygons (Middleton et al. 2018). Gazetteers and Knowledge Bases (KBs) are central to both tasks, serving as repositories of geographic names and their geopolitical or spatial attributes (Gritta, Pilehvar, and Collier 2020). Initial geotagging approaches were rule-based or gazetteer-based systems, and relied on predefined geographic knowledge sources, such as GeoNames or OpenStreetMap (Zhang and Bethard 2024). These methods

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

struggled with ambiguity, the finite knowledge and geographic bias of gazetteers, and the complexity involved in developing a set of precise yet generalisable custom rules (Liu et al. 2022; Quattrone, Capra, and De Meo 2015). The emergence of machine learning (ML)-based methods, with named entity recognition (NER) systems leveraging annotated corpora to statistically detect locations in texts, has since offered improvements in accuracy (Hu et al. 2023).

Despite advancements in geotagging, geocoding still poses several challenges (Hu et al. 2023; Karimzadeh et al. 2019), particularly in the ability of SOTA methods to generalise well across corpora within different domains (Hu et al. 2024). Geocoding seeks to resolve the ambiguity of place names – such as the numerous locations named "Paris" or "Station Road" in the world or within a geographical region – by linking them to the correct entry in structured databases like OpenStreetMap (OSM) or GeoNames, or by predicting geographic coordinates (Ardanuy et al. 2023). To address ambiguity, these systems often rely on heuristics – rules or patterns designed to guide the selection of the most likely match. Common heuristics include population-based prioritisation, which favours larger or more prominent places, and spatial minimality, which assumes that references within a text are geographically clustered (Leidner 2007; Zhang and Bethard 2024). Another frequently applied heuristic, "one sense per referent," suggests that identical place names in a document typically refer to the same location (Gale, Church, and Yarowsky 1992). While heuristics can simplify decision-making, their effectiveness varies depending on the context. For example, population-based heuristics are less suitable for local media, where smaller, niche locations are more common and may be overlooked. In contrast, spatial minimality and one sense per referent align more closely with the geographically coherent nature of local media content (Hu et al. 2024).

Various approaches to toponym disambiguation exist (Zhang and Bethard 2024). Rule-based systems rank potential candidates based on predefined criteria, such as geographic proximity or linguistic context. However, these systems often struggle to generalise effectively (Hu et al. 2023). ML models enhance ranking by learning patterns from annotated data, incorporating features like spatial distance and linguistic context. Despite their promise, ML models are limited by the availability of high-quality training datasets, which are scarce for local media (Hu et al. 2023). An ensemble method combining multiple approaches has recently achieved SOTA results (Hu et al. 2023). Nevertheless, even this approach achieves an accuracy of only 0.84, indicating that further refinement is needed for toponym disambiguation. Additionally, performance was found to vary across corpora, with limited benchmarking on local media datasets (Hu et al. 2024).

Geoparsing systems are evaluated on a limited number of datasets with varying domain, geographic scope and ambiguity (Zhang and Bethard 2024). Several datasets exist for the news domain, including GeoVirus (Kafando et al. 2023), GeoWebNews (Gritta, Pilehvar, and Collier 2020) and TR-News (Kamalloo and Rafiei 2018), all of which focus on global geographies and international media outlets. The LGL (Lieberman, Samet, and Sankaranarayanan 2010) is the only dataset encompassing local media, comprising 588 articles from 78 U.S. local newspapers. It was created to address the lack of datasets specific to local media (Lieberman, Samet, and Sankaranarayanan 2010) and has since become a prominent dataset in geoparsing research (Gritta, Pilehvar, and Collier 2020; Liu et al. 2022). Yet, the dataset has several limitations. Firstly, it is based exclusively on U.S. local newspapers, which limits its applicability to other contexts. Secondly, the dataset only represents print media, failing to account for the growing diversity of digital and online news formats, which can differ in how geographic references are presented, especially when targeting broader audiences (Lieberman, Samet, and Sankaranarayanan 2010). Lastly, LGL's annotation approach, which marks demonyms and place-referenced organisation names as toponyms, deviates from standard disambiguation practices (DeLozier, Baldridge, and London 2015; Matsuda et al. 2015). This misalignment is problematic as it limits the generalisation of model performance across different corpora and contexts. Overall, the lack of diverse benchmarking datasets of local media hinders the development and evaluation of accurate and contextually appropriate geoparsing models. This makes it difficult to develop or identify approaches that can effectively handle the unique linguistic and geographic nuances of local news across geographic contexts, including variations in toponym usage, the prevalence of lesser-known place names and the evolving conventions in reporting locally relevant news across media systems, thus impeding equal progress in geoparsing local media outside of the United States. A widespread availability of geoparsing benchmark datasets matters because validation is a critical component of geoparsing, due to the inherent challenges of toponym disambiguation (Gritta, Pilehvar, and Collier 2020). In this context, methods are developed, refined, and tested on gold standard annotations which provide evidence of an approach, a tool or a gazetteer's suitability to a particular corpus.

There are several critical considerations for geoparsing local media which stem from the literature. First, discrepancies between how place names are spelled in corpora and in gazetteers or KBs require consideration of linguistic variations (Leppämäki, Toivonen, and Hiippala 2024) and regional spelling differences (Ardanuy et al. 2020). Second, the absence of smaller, lesser-known toponyms in global gazetteers, compounded by geographic biases favouring larger locations (Matsuda et al. 2015; Quattrone, Capra, and De Meo 2015), stresses the importance of careful geographic database selection. Third, methods relying on population-based heuristics are limited in their ability to successfully classify fine-grained, lesser-known toponyms, which are common in local media (Hu et al. 2024). Because population-based approaches are commonly used across the most advanced geoparsing systems (Zhang and Bethard 2024), a gap persists for building a system attuned to local geographies.

Collectively, these challenges illustrate the limitations of selecting and validating geoparsing solutions on local media across diverse geographic and contextual settings. Overcoming them will require greater availability of benchmarking datasets suited to the unique linguistical and geographical characteristics of local media. We take a first step in this directions by defining our third objective:

O3 - To develop a novel gold dataset of U.K. local media articles for benchmarking geoparsing tasks.

Can LLMs improve toponym disambiguation?

The emergence of LLMs, models displaying high contextual understanding, presents promising opportunities for advancing toponym disambiguation. LLMs, trained on extensive text corpora and equipped with billions of parameters, excel at capturing deep semantic relationships within context (Bommasani et al. 2022; Mai et al. 2024). By default task agnostic, LLMs can be adapted for specialised applications through fine-tuning, few-shot learning or zero-shot learning (Bommasani et al. 2022). This versatility

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

has enabled their application across a range of tasks, including data annotation, NER (Dubourg, Thouzeau, and Baumard 2024; Goel et al. 2023), text classification (Yin, Hay, and Roth 2019), fact-checking (Chatrath, Lotif, and Raza 2024) and even spatial tasks (Hu et al. 2023; Hu et al. 2024; Mai et al. 2024). While LLMs have demonstrated limitations in directly resolving geographic coordinates due to their imperfect geographic knowledge (Mai et al. 2024), their capacity for contextual inference suggests potential for a more nuanced approach. We hypothesise that LLMs can effectively classify the administrative district associated with a given toponym due to their ability to interpret context. Local news articles often omit explicit details for locations presumed familiar to the target audience, relying on shared local knowledge. Conversely, when mentioning less familiar locations, the authors tend to provide additional context, such as the format "Ripley, Yorkshire," to aid readers in disambiguating between multiple locations or simply localising the place name for a geographically dispersed audience. This phenomenon reflects the concept of the "local lexicon" in local news media, as formulated by Lieberman, Samet, and Sankaranarayanan (2010). While traditional ML methods often struggle with such references due to ambiguity and a lack of contextual understanding, LLMs have shown promise in similar contextual tasks (Hu et al. 2024). Hu et al. (2024) fine-tuned open-source lightweight models (Mistral, Llama2, Baichuan2 and Falcon) to generate unambiguous toponym references (e.g., city, state and country) based on a given toponym and its surrounding text. The unambiguous reference generated by the LLM is then resolved to precise geographic coordinates using GeoNames, Nominatim and ArcGIS. Tested on seven benchmarking datasets, Hu et al. (2024)'s approach achieves significant improvements over the SOTA ensemble methods by Hu et al. (2023), offering a scalable, efficient and accessible solution for geospatial disambiguation. However, these advancements have yet to be tested on datasets specific to local media. Moreover, while fine-tuning is often leveraged to improve task-specific performance, LLMs have demonstrated remarkable capabilities in their out-of-the-box state (Brown et al. 2020). A growing body of research focuses on prompt engineering, a human–computer interaction technique that uses natural language *prompts* to steer model outputs (Dang et al. 2022). This approach eliminates the need for extensive technical overhead, broadening the accessibility of artificial intelligence across diverse research domains (Lee et al. 2025). Studies suggest that iterative testing and evaluation of prompt strategies can improve LLMs' outputs. Effective techniques include one-shot and few-shot prompting, which provide limited input–output examples to guide the model (Liu et al. 2023), and Chain-of-Thought prompting, which incorporates reasoning steps to enhance decision-making accuracy (Wei et al. 2022). Additionally, leveraging the collective intelligence of multiple LLMs through ensemble methods, such as majority or plurality voting, has been shown to improve robustness and reliability by selecting the most voted answer by a group of models (Trad and Chehab 2024; Zhao, Wang, and Peng 2024). Despite these advancements, the application of prompt-based strategies for toponym disambiguation remains underexplored.

While LLMs exhibit considerable promise for toponym disambiguation, their inherent limitations may elucidate the reluctance of some researchers to adopt such methodologies. A primary concern lies in the limited interpretability of LLMs, as their black-box architecture obfuscates the decision-making processes, raising critical questions about their dependability in high-stakes applications like geoparsing (Bommasani et al. 2022). Moreover, LLMs are susceptible to geographic and linguistic biases, which often mirror the imbalances present in their training data. Models trained predominantly on English-language corpora, for example, may exhibit reduced efficacy when applied to texts featuring diverse linguistic structures or originating from underrepresented geographic contexts. Another issue pertains to hallucinations – instances in which LLMs generate erroneous or fabricated outputs.

Experimenting with prompt engineering offers a promising pathway to minimising the risk of hallucinations by steering LLM outputs through carefully crafted input prompts. More systematic approaches, such as fine-tuning, can further mitigate hallucinations by tailoring model weights to specific tasks, while retrieval-augmented generation (RAG) techniques can address both bias and contextual limitations by integrating external knowledge sources (Lewis et al. 2020). However, given the relatively limited exploration of prompt engineering in isolation, this study prioritises initiating research into LLM-driven geoparsing within a focused geographic context: the United Kingdom. This approach not only provides a manageable scope for examining the feasibility of prompt-based strategies but also lays the groundwork for broader investigations into LLM applications in geoparsing. This rationale leads directly to the formulation of our third objective:

O4 - To assess the performance of LLMs in disambiguating toponyms within the context of U.K. local journalism, with a focus on prompt-based strategies.

Evaluating location extraction approaches in local news: A framework and comparative review

The increasing availability of geoparsing approaches and resources has not translated into robust solutions for analysing local news texts. While a "one-size-fits-all" solution remains elusive, we can leverage insights from existing geoparsing research to support robust, context-aware designing of geoparsing pipelines in media studies. The framework in Figure 1 outlines key steps for designing geoparsing pipelines. Researchers should begin by clearly defining the research problem, specifying both geographic scope (e.g., city, region or country) and target entity types (e.g., administrative units and landmarks). Preliminary research informs method selection, enabling the identification of suitable techniques for geotagging (e.g., rule-based approaches and ML) and geocoding (e.g., rank-based systems and statistical disambiguation). A robust geotagging phase prioritises extracting relevant entities, while geocoding requires geographic databases with balanced coverage and specificity to ensure reliability. Validation, using test data and metrics, such as Mean Error Distance or F1-score, is crucial for iterative refinement, providing insights into pipeline performance (Gritta et al. 2018). These steps aim to support the robust extraction of geographic information from local media, aligned with research objectives. Taken together, the steps in this framework should guide local media researchers through key geoparsing stages (O1), emphasising critical considerations for successful location extraction from texts in line with research objectives.

To evaluate current practices in location extraction, we reviewed five key studies on local news geoparsing (summarised in Table 1). While these studies highlight the value of geographic analysis, gaps remain in problem definition, method selection and validation.

Notably, most studies (e.g., Napoli and Weber (2020) and Moore and Ramsay (2024)) rely on vague definitions of "local," limiting replicability. A key element of the framework is the emphasis on preliminary research to guide methodological choices. The impact

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

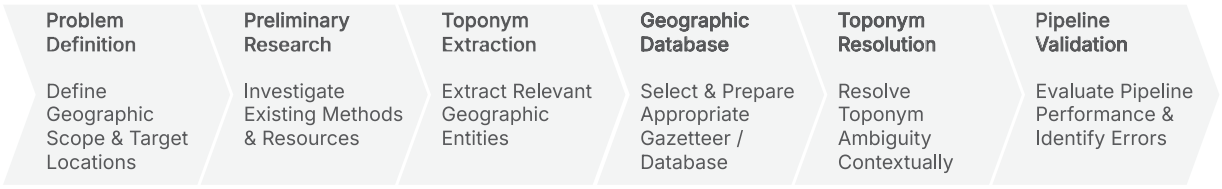

Figure 1. Methodological framework for evaluating and implementing geoparsing in local news studies.

Table 1. Summary of studies investigating geographic content in local media: Objectives and methodological approaches

Study	Number of articles	Research objective	Approach
Napoli and Weber (2020)	16,000	To quantify local coverage	Manual coding
Moore and Ramsay (2024)	5,233	To quantify local coverage	String matching
Madrid-Morales, Rodríguez-Amat, and Lindner (2023)	519,004	To map online news deserts	String matching with GeoNames gazetteer
Khanom et al. (2023)	3,564	To map location mentions in news articles	NER, custom rules, Google Maps API
Vogler, Weston, and Udris (2023)	15,254	To measure geographic news diversity	Gazetteer matching, NER, custom ranking

of this step is evident in the contrasting approaches of the reviewed studies. Napoli and Weber (2020)'s reliance on manual coding suggests a lack of engagement with existing geoparsing techniques. Similarly, Moore and Ramsay (2024)'s justification for using Steno, based on prior use, suggests a limited exploration of alternative methods. Conversely, Vogler, Weston, and Udris (2023) demonstrate the benefits of thorough preliminary research. Their review of geoparsing challenges in the Swiss context directly informed a tailored pipeline featuring a custom gazetteer, combined toponym detection methods, and a candidate ranking mechanism. While details of the ranking mechanism are lacking, this approach reflects an awareness of the complexities of geoparsing local news and the need for context-specific solutions. Others, such as Madrid-Morales, Rodríguez-Amat, and Lindner (2023), employ string-matching techniques with limited handling of toponym ambiguity, while Khanom et al. (2023) introduce innovative NER-based methods but conflate organisational entities with geographic references. Across all studies, validation efforts are underemphasised, with few explicitly quantifying errors or leveraging benchmarking datasets. A recurring limitation across reviewed studies is the lack of explicit toponym disambiguation. For instance, Napoli and Weber (2020), Moore and Ramsay (2024) and Madrid-Morales, Rodríguez-Amat, and Lindner (2023) omit this critical step, raising concerns about the accuracy of geographic data. While Khanom et al. (2023) incorporate the Google Maps API for geocoding, they do not detail its ambiguity resolution process, leaving questions about its reliability. In contrast, Vogler, Weston, and Udris (2023) address disambiguation through a custom gazetteer and ranking mechanism, though further details are needed to assess its efficacy. Similarly, the validation phase remains underdeveloped. Apart from Moore and Ramsay (2024), no studies quantify errors or benchmark their pipelines, limiting the reliability of findings. Addressing these gaps through robust validation and context-specific methodologies is critical for advancing geoparsing in local media research. Building on this review, our proposed objectives (O3 and O4) aim to address these gaps by introducing a benchmark dataset for validation and a straightforward, context-aware toponym disambiguation approach. These contributions are designed to enhance the reliability and scalability of geoparsing pipelines, providing researchers with robust tools to explore the geographic dimensions of local news. By bridging methodological gaps, this work advances the study of local news ecosystems, enabling more nuanced insights into media diversity and societal impact.

Methods

In this section, we describe the methodology for developing the novel benchmarking dataset and the prompt-based toponym disambiguation approach.

Dataset

For the purpose of this study, we selected the United Kingdom as our geographic focus. The United Kingdom is one of the better understood markets in local media research. However, it presents an interesting case study. Despite concerns about the diminishing local relevance of news content due to intensified media ownership consolidation, there is little empirical evidence of changes in content, particularly a shift in geographic focus (Bisiani et al. 2025). Furthermore, the U.K. context benefits from the availability of comprehensive datasets and well-curated media directories (Bisiani and Heravi 2023; Bisiani and Mitchell 2024). These resources enable robust sampling and rigorous validation. Upon selecting our geographic scope, we defined our geographic "granularity." The goal of the dataset is to facilitate benchmarking of geoparsing techniques on niche, fine-grained locations which are prevalent in local news. Therefore, our annotation efforts focus exclusively on U.K.-specific locations. We thus exclude larger-scale geographic references (e.g., nations and regions) because their geocoding typically relies on centroid coordinates. Centroid-based geocoding can introduce inaccuracies by misrepresenting these references as singular points or conflating them with entire administrative districts. We defined the upper boundary of the toponym tagging and resolution to be local authority districts (LADs), which are a commonly used administrative resolution level (N=361) for analyses of news deserts in U.K. studies (Bisiani and Heravi 2023; PLUM Consulting 2020; Public Interest News Foundation 2023).

We sourced the articles for our dataset from UKTwitNewsCor (Bisiani 2024), a corpus of over 2.5 million articles from 360 U.K. digital local media outlets (2020–2022). We chose this dataset due to its recency and comprehensive coverage across geography and providers (Bisiani et al. 2025). We first randomly sampled 100

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

articles from the dataset using simple random sampling. Given the size and inherent imbalances within UKTwitNewsCor, we then added a further 100 articles, sampled through stratification across time, geography and outlet. For time stratification, we divided the dataset into yearly quarters and aimed for proportional representation from each quarter. For geographic stratification, we used each outlet's LAD of coverage reported in UKTwitNewsCor's metadata. This stratified approach ensured that our dataset included a wider range of outlets and geographies, mitigating potential biases in the LLM geocoding process due to the distribution of articles in the original corpus or potential biases in the LLM training data. The sample size of 200 articles, although small, is not unlike other sizes found in geoparsing benchmarking datasets (see Zhang and Bethard (2024) for a review). Focusing on a manageable sample allowed us to ensure high-quality and accuracy at each stage. The validation work carried out throughout this pipeline was conducted by two annotators, a member of the research team and an externally recruited coder.

Toponym recognition

We manually geotagged the first subsample prior to deciding to increment the overall dataset size. To annotate this first subsample, we used Prodigy (Montani and Honnibal 2018), a popular data annotation tool. We obtained free access upon applying for a university research license. Prodigy provides an interface in which the user manually highlights entities in the text. Each annotation is converted into metadata providing information about the start and end position of a specific entity in a document, aligning with the typical formatting of NER-annotated data. We focused on three entities: geopolitical entities (GPEs), encompassing countries, cities and other administrative regions; natural landmarks (LOC), which include mountains, rivers and other geographical features; and infrastructure (FAC), covering buildings, roads and other man-made structures. These are the three spatially oriented entities in NER systems, and a wide range of study has focused on these to include both fine-grained locations and admin units (Berragan et al. 2023; Hu et al. 2023). For the second subsample, we opted to test the Spacy NER tool (Hess and Waller 2017), to annotate LOC, FAC and GPE entities. We chose SpaCy because it was built by the same creators of Prodigy and provides convenient comparison functionalities and similar file formats with Prodigy. SpaCy, despite not being among the top five NER tools identified in a comparative review by (Hu et al. 2023), has performed best in other contexts, including local media (Khanom et al. 2023). Recently, a new transformer-based model (en_core_web_trf) has been released, achieving SOTA results against its predecessors.[1] We benchmarked SpaCy's performance on the hand-annotated entities. We obtained an F1-Score of 0.94, Precision of 0.97 and Recall of 0.91. Satisfied with these results, we used SpaCy to annotate the second subsample.

Candidates generation

Once identified locations, the following step involved identifying potential real-world coordinates applying to each unique place name. We created a comprehensive set of location candidates by combining two key sources: Ordnance Survey (OS) Open Names gazetteer and OpenStreetMap (OSM) Nominatim API. OS was chosen because it is the national mapping agency of Great Britain,[2] and its Open Names dataset includes over 870,000 named and numbered roads, 44,000 settlements and 1.6 million postcodes.[3] Previous U.K.-based studies have leveraged OS for geocoding (Berragan et al. 2024). We performed exact string matching between the lowercased toponyms and entries from the OS gazetteer. A limitation of OS is its lack of coverage for Northern Ireland. To address this gap and improve coverage, we also queried Nominatim, an open-source geocoding service widely used in locally-oriented geospatial research due to its extensive coverage of fine-grained locations (Hu et al. 2024). We limited Nominatim queries to return only coordinates within the boundaries of the United Kingdom to avoid false positives (e.g., returning all instances of "Station Road" in the world, despite the lack of likelihood that a local news outlet in the United Kingdom refers to such a fine-grained location in a country other than the United Kingdom), as proposed in Hu et al. (2024). We retrieved all available results for each query. While OS matching was based on exact string comparison, Nominatim provides flexibility in handling variations, such as the presence of commas or structured formats (e.g., "street, town, region, country"). Combining results from OS and OSM, we generated a candidate list of 7,374 locations. However, 52 unique place names, appearing a total of 67 times in the corpus, did not return a match. For these, annotators manually searched for the correct coordinates in Google Maps, using article context to identify the appropriate real-world location. Next, we enriched the candidate list with additional geographical metadata. Although the OS gazetteer did not provide coordinates for its toponyms, it did offer outward postcodes. We used the PostcodesioR package (Walczak 2023) to perform postcode lookups and retrieve coordinates for each postcode. We also assigned each toponym to an LAD by performing a spatial join with a shapefile of U.K. LAD boundaries from the Office for National Statistics.[4] Eight entries failed to return a matching LAD, as their coordinates matched bodies of water off the coastline. We concluded this failure was due to imprecise coordinate assignment by PostcodesioR or OSM's API. For these cases, we applied a proximity-based matching technique to find the nearest LAD based on geodesic distance, leveraging the `sf` R package (Pebesma 2018). Through these steps, we created a robust collection of candidates for the place names identified during the geotagging stage. Figure 2 presents the level of ambiguity among toponyms, the contribution of various gazetteers to the candidate list, and the distribution of articles based on the number of districts associated with their candidates. Plot (a) indicates that the majority of toponyms and their candidates (94%) appear in only one document, particularly Facility (FAC) and Location (LOC) entities. FAC and LOC entities are also significantly more ambiguous than GPEs, with average candidate counts of 14.9 and 8.5 compared to 3.8 for GPE. Plot (b) shows that OpenStreetMap (OS) contributes the majority of candidates (over 60%), with a small proportion (just over 10%) provided by OpenStreetMap (OSM) but not present in OS. Finally, most articles include candidates spanning a limited number of districts, although it is rare for an article's candidates to be confined to a single district.

Toponym disambiguation

We proceeded to review all toponyms and their associated candidates, including cases with a single candidate, to account for instances where OS and OSM provided candidates but none of

[1] https://spacy.io/usage/facts-figures
[2] https://www.ordnancesurvey.co.uk/about
[3] https://www.ordnancesurvey.co.uk/products/os-open-names
[4] https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-may-2024-boundaries-uk-bfe-2/about

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

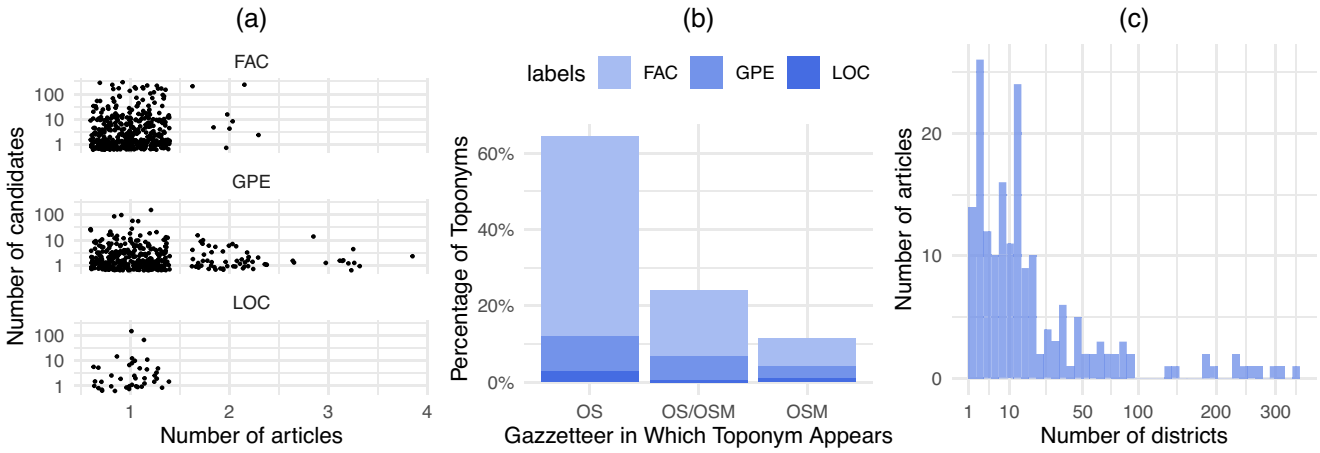

Figure 2. Overview of candidate list: (a) distribution across documents based on the number of candidates per toponym, (b) candidate source distribution, and (c) number of articles categorised by the number of districts.

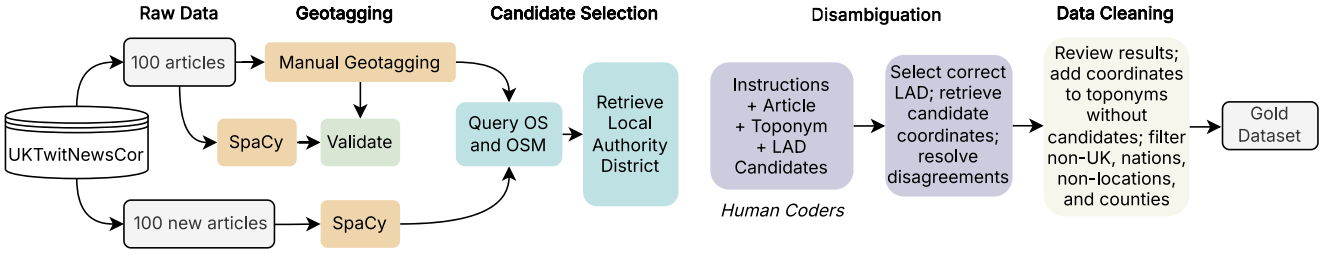

Figure 3. Overview of procedure for creating the dataset.

the candidates matched the correct real-world location. We began by merging the candidate dataset with the dataset of articles and their toponyms. To streamline the annotation process, we applied the one-sense-per-referent heuristic. When a place name appeared multiple times within a document, it was presented to annotators only once. This reduced the number of location mentions requiring resolution from 1,313 to 925. Each annotator, using the Label-Studio platform,[5] was shown the article, the toponym requiring resolution and the list of candidates in a multiple-choice format. Each option represented the LAD associated with the coordinates of a candidate. We structured the task as a toponym-to-district resolution rather than a toponym-to-coordinates mapping, as no place name had more than one set of coordinates per LAD. This approach simplified the task, as we found that our annotators, like models (Hu et al. 2024), find it easier to understand spatial relationships at a hierarchical level rather than at a coordinate level. The verification procedure involved researching each toponym, verifying which LAD it is situated in using Wikipedia pages of location names or Google Maps. Four supplementary options were provided to address edge cases: (1) the correct district was not among the listed options, (2) the toponym was not a location, (3) the toponym referred to a location outside the United Kingdom, or (4) the toponym spanned multiple districts (e.g., a region or nation). The annotation task achieved an inter-annotator agreement rate of 95%. Disagreements were resolved through discussion between the annotators. We then linked the toponym and the selected option back with the candidate list, to map spatial coordinates to each toponym. We performed a series of final data cleaning steps. We removed 21 instances in which SpaCy's NER erroneously classified non-locations (e.g., the surname *Port*) as toponyms. We removed 13 non-U.K. locations to maintain the dataset's focus on U.K. entities. We removed 20 place names representing large geographic areas (e.g., *Scotland* or *North Wales*) due to the inability to provide accurate coordinates or assign a singular district mapping. Finally, 149 toponyms did not correspond to any of the options provided by OS and OSM. We manually registered the correct LAD and Google Maps coordinates. Through these steps, we refined the dataset to ensure its accuracy and utility for benchmarking geoparsing techniques in the context of fine-grained, localised spatial references within U.K. local media. Figure 3 encapsulates the various steps within the gold standard dataset creation process.

Prompt-based geoparsing approach

To develop an accessible and efficient approach, we utilised open-source LLMs in their lightweight configurations, running them locally via the Ollama framework within an R environment, ensuring reproducibility with a fixed random seed. Ollama is a platform that simplifies the deployment and local running of LLMs. It allows users to leverage optimised versions of popular open-source LLMs on personal devices without requiring cloud access, promoting privacy, accessibility and reproducibility (Liu, Kang, and Han 2025). We selected four LLMs for our experiments, which ranked highest in popularity in the Ollama[6]: Gemma2 (9B), Llama3.1 (8B), Qwen2 (7B) and Mistral (7B). Their size range (7B–9B parameters) enables exploration of the trade-off between computational efficiency and performance, addressing whether smaller models can achieve satisfactory results. Their diverse architectures, training

[5] https://labelstud.io/

[6] https://ollama.com/search

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

data and origins (e.g., Meta, Alibaba and Mistral AI) further provide a broad basis for assessing generalisability across LLM implementations. We tested four temperature settings (0, 0.25, 0.5 and 1). Temperature, a key parameter in LLM inference, controls the randomness or "creativity" of the model's output (Karjus 2025). A temperature of 0 makes the output deterministic, always producing the most probable response, which is desirable for classification tasks where consistency is crucial. Higher temperatures introduce more randomness, leading to more diverse and potentially creative outputs. We reasoned that while a lower temperature might be optimal for maximising accuracy in a classification-like task, exploring higher temperatures could reveal the model's confidence levels or its ability to handle ambiguous or challenging cases. To facilitate automated processing of the LLM's responses, we guided the models towards generating machine-readable output in JSON format. We repeatedly clarified the desired JSON structure in our prompts, instructing the LLMs to adhere to a specific schema consisting of the chosen option, or correct district, and the reasoning behind the decision. We investigated two problem setups:

- *Contextual toponym disambiguation from KB:* This approach aims to guide the LLMs by providing a structured KB of potential LADs, thereby minimising the risk of hallucinations or the generation of inaccurate LADs due to the LLMs' imperfect geographic knowledge. We treated the LLMs as annotators, providing them with context (the article text, metadata and the same list of potential LADs given to the human annotators) and prompting them to select the correct LAD for a given toponym from the candidate list generated in the "Candidates Generation" section. We tested one concise prompt, prioritising speed and ease of parsing (Karjus 2025), alongside a more detailed prompt with one-shot examples, hypothesising that richer context, analogous to chain-of-thought prompting (Wei et al. 2022), might enhance disambiguation performance, despite the increased computational cost. We also explored the impact of varying metadata context on LLM geoparsing performance, hypothesising that richer metadata would generally enhance accuracy, up to a point. For each prompt, we systematically tested all seven possible combinations of the following metadata fields: the name of the outlet publishing the story, the main district of coverage of the outlet and the names and district candidates of other toponyms within the same article. The latter was intended to mimic the spatial minimality heuristic in geocoding, which uses the distance between candidates within the same document to resolve toponyms (Leidner 2007). This approach allowed us to investigate the trade-off between aiding the LLM with contextual information and potentially introducing extraneous or even conflicting data.
- *Few-shot LAD classification:* This approach attempts to address the potential lack of comprehensive geographic coverage in existing gazetteers and databases, while also reducing the computational overhead associated with providing extensive candidate lists. We prompted the LLMs to identify the LAD for a given toponym using only the article text and the name of the publishing website, without providing a list of options. Moreover, we used exclusively this approach for the 52 place names for which no candidates were found in OS and OSM during the "Candidates Generation" section.

We ran these combinations of parameters (model, temperature, prompt and metadata) on each toponym. We then compared results across models and approaches and used these to narrow down our approach to one prompt and metadata configuration. The prompts used in the experiments are provided in the Supplementary Material 1 for replication and transparency.

Geocoding evaluation employs both binary and continuous error metrics (Gritta, Pilehvar, and Collier 2020). For entity linking tasks, where the output is a binary (correct/incorrect) classification, the F1-score is commonly used (Ardanuy and Sporleder 2017). Given the classification nature of the task, we followed a binary approach to assess the model predictions: A true positive was defined as a case where the LLM's predicted LAD matched the annotator's chosen LAD. A true negative was recorded when both the LLM and the annotator agreed that the location represented an edge case. A false positive occurred when the LLM's predicted LAD was different from the annotator's choice. A false negative was recorded when the LLM predicted an LAD, but the annotator identified the location as an edge case. Based on these classifications, we calculated Precision (the proportion of correct LLM LAD predictions), Recall (the proportion of correctly identified annotator-chosen LADs), the F1-score (the harmonic mean of precision and recall) and overall Accuracy (the proportion of correct classifications). During our analysis, we observed a tendency for the LLMs to identify the correct LAD even when this option was not provided. While human annotators correctly marked such instances as "LAD not in option," the LLMs occasionally returned the correct LAD. We developed a flexible evaluation strategy, marking these instances as True Positive. This leniency will not make a difference in terms of the final outcome of the geoparsing procedure, which involves joining the model option with the candidate list generated in the "Candidates Generation" section. Due to the model candidate missing from OS and OSM to begin with, no coordinates will be assigned to these cases. We account for these errors in our spatial error analysis (more about this in the "Evaluation" section). Yet, within this context, this approach provides a more accurate reflection of the LLMs overall performance in terms of spatial linking, regardless of the limitations of gazetteers. Using the F1-Score to identify the best prompt-metadata configuration, we kept the best performing temperature for each model.

We then implemented a majority voting system to enhance robustness (Trad and Chehab 2024; Zhao, Wang, and Peng 2024). This approach mitigates the risk of individual model errors by selecting the most frequent prediction among the LLMs, effectively promoting a more reliable consensus-based outcome, akin to the "wisdom of the crowd" heuristic. We evaluated the voting system performance under varying consensus thresholds (e.g., requiring agreement from a majority or all models). This analysis allowed us to assess the impact of model uncertainty and determine whether a less stringent consensus criterion could still yield accurate predictions (Figure 4).

Evaluation

We present four sets of performance metrics on our approach: (1) prompt-based, which is our baseline approach, where we select the most voted option and default to a random one if no majority is present; (2) prompt-based (excluding edge-cases), where we report on the same data but exclude instances where the model predicted an edge-case. This case aims to capture the realistic error of the subset of disambiguated toponyms, excluding cases where no prediction would be returned; (3) prompt-based (majority filtered), where we filter out instances where all models disagree; and (4) prompt-based (unanimous filtered), where we retain instances where all models agree. These last two results aim to capture the

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

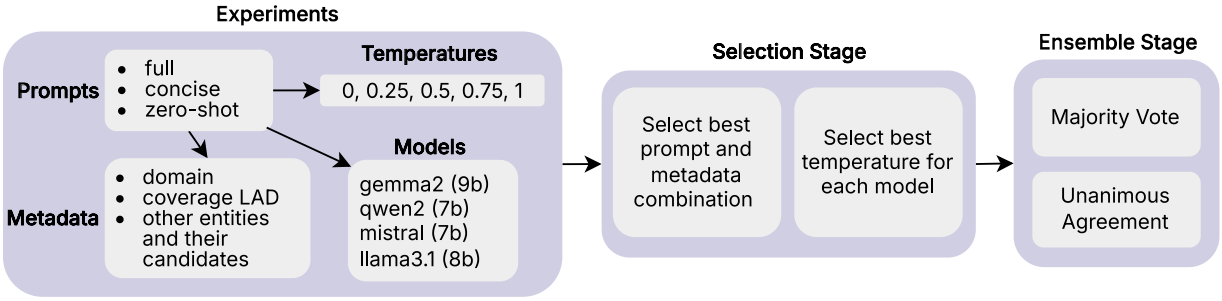

Figure 4. Schematic representation of the proposed LLM prompt-disambiguation approach.

extent to which performance improves under stringent consensus conditions at the loss of some data, which we quantify. For the purposes of benchmarking our approach against preexisting SOTA approaches, our toponym dataset was submitted to Hu et al. (2024) for processing via their fine-tuned LLM-based geoparsing pipeline, as described in their paper. The resulting geocoordinates from their pipeline were then provided to us.

We analysed results at two levels: at a classification level, using the LAD predicted as the unit of analysis, and at a spatial level, calculating the distance between the geocoordinates of the gold and predicted annotations. As the data provided by Hu et al. (2024) only contained coordinates, we performed a spatial join using the same approach defined in the "Candidates Generation" section in order to obtain the respective LAD. By using a variety of metrics, we aim to provide a comprehensive evaluation of the performance of our approach:

- *Classification metrics*: We counted as TRUE any instances where the fine-tuned or prompt-based LAD matched the gold data LAD, and calculated Accuracy as the proportion of TRUE predictions across all predictions. In the event of an edge-case prediction in our approach, we provide two results: (1) Task Accuracy, where we count a result as TRUE if the LLM's prediction matched the answer by the annotators and (2) System Accuracy, where we penalise our approach for failing to lead to a correct LAD identification where no correct candidate was provided. These two cases aim to gauge the extent to which the models predict correctly, given the information at hand, and the extent to which the system performs well overall.
- *Distance metrics*: We calculated three metrics: Mean Error Distance, Accuracy@20km and Accuracy@161km. Mean Error captures the average Haversine distance between predicted and true coordinates (DeLozier, Baldridge, and London 2015; Gritta, Pilehvar, and Collier 2020), while Accuracy@20km and Accuracy@161km measure the percentage of predictions within 20 km and 161 km of the true coordinates, respectively (Ardanuy et al. 2023). These thresholds provide insights at different levels of granularity: Accuracy@161km reflects broader regional accuracy, whereas Accuracy@20km assesses performance at a local scale. The 20 km threshold is particularly relevant to our focus on local news, as many outlets serve areas within this radius, enabling us to evaluate the LLMs' ability to capture the shift away from truly local news coverage – a phenomenon often linked to media consolidation and content syndication. Before calculating the Haversine distance, we accounted for nuances in the LLM-predicted coordinates. When no coordinates were provided due to an edge-case selection, the distance defaulted to the maximum possible error on Earth (20039 km), following Gritta, Pilehvar, and Collier (2020). For predictions where a district is given (correct or incorrect) but coordinates are missing due to the absence of a toponym-LAD match in the OS+OSM candidate list, we employed two methods: (1) Max Error Defaulting, assigning the maximum error distance and (2) Centroid Assignment, where the centroid of the predicted LAD was used as the coordinate. Finally, when predictions did not correspond to an LAD (e.g., "London"), the maximum error distance was applied.

Ethical considerations

The use of LLMs raises ethical considerations regarding accessibility, data privacy, transparency, potential misuse and environmental impact (Bommasani et al. 2022). As the articles in our dataset are already publicly available online, no additional anonymisation was necessary. Transparency is prioritised through detailed documentation of our methods and parameters, ensuring reproducibility. Our approach is designed with clearly defined, ethical use cases to prevent misuse. We aim to make our proposed method accessible by leveraging open-source LLMs. Finally, we optimised computational workflows by using lightweight, locally run models to minimise energy consumption. These measures aim to ensure the responsible and accessible application of LLMs.

Results and discussion

In this section, we first provide an overview of the gold standard dataset. Finally, we use the dataset to evaluate the proposed LLM approach.

Gold standard dataset

The gold standard dataset developed for this study, henceforth referred to as the LMUK-Geo dataset, is summarised in Table 2. It encompasses 182 unique articles sourced from 142 distinct outlets representing 32 publishers. Despite an initial selection of 200 articles, 18 articles were found not to have any toponyms. The 182 articles span 133 districts, with each district contributing an average of 1.4 articles (median = 1, SD = 0.7). The dataset comprises 38,619 words (mean per article = 212). It includes 1,313 toponyms, 838 of which (64%) are unique. Toponyms are further categorised into named entity types: locations (LOC), facilities (FAC) and GPE entities. LOC entities account for 45 instances, 89% of which are unique, with an average occurrence of 1.5 per article (median = 1, SD = 0.9). FAC entities appear 512 times, with 81% being unique, averaging 3.8 per article (median = 2, SD = 4.6). GPE entities are the most frequent, with 756 occurrences, though only 54% are unique, averaging 4.7 per article (median = 3, SD = 5.5). The data highlights the unique contributions of LOC, FAC and GPE references to

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

Table 2. Descriptive statistics of the novel dataset LMUK-Geo

	N	*N* Unique (%)	Mean, median, sd
Articles		182	
Outlets		142	1.3, 1, 0.6
Publishers		32	4, 1, 10
District		133	1.4, 1, 0.7
Words	38,619	9,992	212, 188, 144
Toponyms	1,313	858 (64%)	7, 5, 8
LOC	45	40 (89%)	1.5, 1, 0.9
FAC	512	413 (81%)	3.8, 2, 4.6
GPE	756	405 (54%)	4.7, 3, 5.5

understanding the geographic scope of local media content. LOC features rarely in local media articles. In contrast, FAC entities are both numerous and highly distinct, thus playing an important role in providing the geographic context of local media. Their high uniqueness underscores the importance of specific buildings and infrastructure in grounding stories within everyday community spaces. Focusing solely on GPEs, as we noted, is common practice in local media geoparsing research, risks overlooking the granularity provided by LOCs and FACs, which could play a critical role in describing hyperlocal dynamics and situating narratives within tangible community settings. Such a narrow focus may lead to overgeneralisation, reinforce top-down perspectives, and misrepresent the true geographic diversity of local reporting. This underscores the need for gazetteers and methodologies that encompass all three entity types, ensuring a comprehensive and balanced analysis of local media content.

Prompt-based approach performance

We now present the results from our LLMs experiments. After running our initial experiments, we consistently found the full prompt, with only domain metadata provided, to perform best across models. Predictions were impacted, in order of significance, by the model, the prompt and metadata configuration, and finally temperature. Temperature, in particular, did not impact predictions greatly, for a given model and configuration (Figure 5). We found that Mistral's performance was relatively poorer across the board compared to other models, and particularly so for the few-shot model. The few-shot approach was weakest across the board, indicating that aiding the models by presenting a list of options improved performance. However, the few-shot approach of the best performing models, Gemma2 and LLama3.1, still outperformed the best results (regardless of prompt) in Qwen2 and Mistral. Adding metadata did not improve model performance, contrary to what we hypothesised. Removing contextual spatial information, including the district candidates of other toponyms within the same article and the coverage district of the publishing website consistently improved performance, regardless of prompt or model. Our results signal the importance of trialling various models to identify models which due to their training and architecture, are adept to this task and its particular setup.

Due to Mistral's poor performance, we excluded this model from the subsequent voting ensemble. We retained the predictions from the best performing temperature for each remaining model for the full prompt with the name of the publishing website as the sole metadata. We proceeded to select as a final prediction the most recurring answer across Gemma2, LLama3.1 and Qwen2. In instances where all models disagreed, we defaulted the final prediction to a random selection between the three models. While we initially considered defaulting to Gemma2, we noticed Gemma2 had much lower Recall than other models, driven by its propensity to erroneously predict edge cases in instances where other models instead returned the correct answer. With this dataset at hand, we then calculated evaluation results for the subset of observations where all or some models agree, as outlined in the "Evaluation" section.

Table 3 summarises the performance of our prompt-based approach across different handling strategies, compared to the fine-tuned model by Hu et al. (2024).

Our prompt-based approach, using majority voting among three LLMs (Gemma2, Llama3.1 and Qwen2), demonstrates competitive performance. While it does not surpass the fine-tuned baseline, it achieves notable results, particularly when applying Centroid Assignment to handle missing coordinates. Significant improvements were observed in both classification and distance-based metrics when applying stringent criteria for retaining predictions. The filtering approaches ("Majority" and "Unanimous") reflect the trade-off between data coverage and accuracy. Unanimous Filtering achieves the highest A@20 (0.91) and A@161 (1.00) within our prompt-based approach, but it significantly reduces the number of geoparsed instances (down to 68%) due to the strict consensus requirement. In contrast, Majority Filtering offers a compromise, maintaining reasonable accuracy while retaining a larger portion of data points (90%).

When comparing the different missing coordinate handling strategies, we found a clear improvement in results when assigning the predicted LAD centroid coordinates to predictions that fail to link to our candidate list. In the best case, the subset of data where all three models agree reached a Mean Error Distance of 75.46 km. Meanwhile, the difference between Task Accuracy and

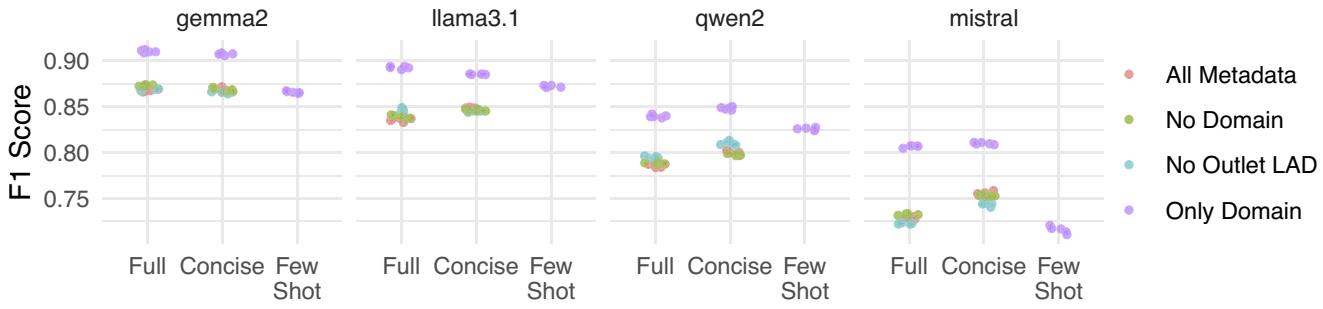

Figure 5. Performance of the LLMs across prompts and metadata configurations. Each dot represents a temperature.

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

Table 3. Evaluation results on LMUK-Geo with different handling of missing coordinates

				Max error			Centroid		
Approach	Coverage	Task Acc.	System Acc.	A@20	A@161	ME (km)	A@20	A@161	ME (km)
Fine-tuned (Hu et al. 2024)			0.80	**0.90**	**0.99**	**22.11**			
Prompt-based		0.87	0.78	0.70	0.79	4070.43	0.76	0.86	2667.17
Excl. edge-cases	0.89	0.87	0.87	0.78	0.88	2179.77	0.86	0.96	610.37
Majority	0.90	0.93	0.85	0.76	0.84	3182.83	0.81	0.89	2035.20
Unanimous	0.68	0.97	**0.97**	0.86	0.95	1079.36	**0.91**	**1.00**	**75.46**

Note: Best results are marked in bold.

System Accuracy highlights the challenge of handling edge cases where no valid LAD can be assigned due to missing data from the gazetteers used in this study.

Collectively, the LLMs performed well on the task, providing the same answer as human annotators for 87% of toponyms. However, as a system, the accuracy against the ground truth was 78%, just behind the fine-tuned approach. We view this as a positive sign that our prompt-based approach could be refined by incorporating additional gazetteers. The extent to which gazetteer choice influences results is difficult to discern, as our approach used OS and OSM, while the fine-tuned approach used OSM, ArcGIS and GeoNames. As noted by Ardanuy et al. (2020), the candidate selection stage plays a crucial role in determining the success of geoparsers, yet it remains an underexplored area in geoparsing research. In future work, we aim to explicitly account for gazetteer discrepancies to better isolate the performance of different geoparsing approaches.

At the pipeline level, our method does not currently achieve SOTA results compared to the fine-tuned approach by Hu et al. (2024). However, it demonstrates the promising potential of readily available LLMs for geoparsing without the need for resource-intensive fine-tuning. In summary, our results demonstrate that a prompt-based LLM approach can provide competitive geoparsing performance, particularly when combined with appropriate missing coordinate handling and filtering strategies. While fine-tuning remains highly effective, our method offers a viable and efficient alternative, especially when fine-tuning resources are limited. Furthermore, our analysis of filtering strategies provides valuable insights into the relationship between consensus thresholds and geoparsing performance, highlighting the importance of balancing accuracy and data coverage.

Future directions

This study provides a foundation for accessible and robust geoparsing of local news, showcasing the potential of prompt-based LLM approaches. However, several avenues for future research remain. First, hallucinations – instances where the model generates place names or locations that do not actually exist – pose a distinct challenge compared to typical errors, such as misidentification or ambiguity, as they introduce entirely fabricated geographic references. Although our current study did not formally quantify hallucinations, we did not observe clear examples during manual examinations of the LLMs-driven predictions. Nonetheless, given their potential impact, future work will prioritise the systematic identification and mitigation of hallucinations to ensure more reliable geoparsing outcomes.

Second, given our focus on resolving all location mentions, future work could expand this analysis to the document level, accounting for the relative importance of location mentions and their contribution to the article's overall geographic narrative. Although document-level geoparsing is less common (Teitler et al. 2008), it offers valuable insights into the thematic and spatial coherence of news content. We posit that LLMs, given their contextual understanding abilities, might perform well in such discrimination tasks.

Third, we consider the difference in approach between our method and that of Hu et al. (2024). We provide options to the LLMs, which, relative to not giving any candidates, improves performance (as seen in Figure 5). The fine-tuned approach instead prompts LLMs to elaborate on mentioned locations before querying gazetteers, aiming to reduce candidate selection errors. We hypothesised that our more controlled environment would support more robust results. By focusing on district-level understanding rather than requiring precise place name knowledge, we aimed to leverage the LLM's contextual abilities effectively while controlling for their lack of spatial knowledge (Mai et al. 2024). Altogether, these system differences point to several potential architectures for further improvement. For example, incorporating RAG could enhance inference and output reliability by providing models with additional context programmatically. Another promising direction is the incorporation of human-in-the-loop workflows to correct and refine LLM outputs. Furthermore, LLMs could be used as judges in subsequent steps, reviewing and refining other models' decisions (Zheng et al. 2023).

Fourth, while this study focuses on U.K. local media and open-source LLMs, future studies should examine the applicability of the approach to different geographic contexts, media systems and LLM architectures. Adapting the methodology to other countries will necessitate the careful selection of relevant KBs, administrative boundaries, and potentially fine-tuning or adapting LLM prompts to account for linguistic and cultural variations. It is important to acknowledge that LLMs, including those used in this study, might perform significantly better on material in English and concerning countries or regions more extensively represented in their training data. This imbalance likely influences the quality and reliability of geoparsing outputs, particularly in less represented or multilingual contexts. Our focus on U.K. local media, primarily English-language content, means the models benefit from relatively rich training signals. Future research should explicitly investigate the impact of training data biases on geoparsing performance and explore methods to mitigate these effects, such as incorporating diverse regional corpora, using multilingual LLMs or developing localised gazetteers.

Exploring the performance of more powerful, proprietary LLMs is also a natural extension of this work. Comparative analyses across diverse languages and territories would provide

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

valuable insights into the cross-cultural performance of LLMs for geoparsing and the influence of varying media landscapes. Finally, we hope that our detailed approach to creating our gold dataset will inspire future efforts to expand the number of benchmarking datasets of local media in different media systems, offering key resources to advance both geoparsing theory and applied geoparsing in local media studies.

Conclusion

In pursuit of establishing a foundation for robust and accessible local news geoparsing, this study has achieved four objectives: the development of a framework to guide future media geoparsing research, a comparative review of existing local media geoparsing studies, the creation of an annotated dataset for geoparsing U.K. local news and the implementation of a novel, scalable prompt-engineering LLM geoparsing approach to achieve robust performance without requiring resource-intensive fine-tuning. These contributions, taken together, provide valuable resources and insights to researchers working at the intersection of geoparsing and local journalism. Building on these foundations, our work enables the exploration of downstream applications of geoparsed local news data. At a spatial level, researchers could analyse how the use of different toponym types (LOC, GPE and FAC) in local news varies spatially and temporally, exploring theoretical implications for understanding geographic and socio-political narratives. This could reveal how political events, regional development and media ownership shifts influence geographic references, offering insights into how media reflects and shapes geographic hierarchies and power structures. Another potential use case is analysing the relationship between media ownership consolidation and the geographic distribution of news content. By tracking location mentions across outlets with varying ownership, researchers can explore whether consolidation leads to homogenised or skewed geographic reporting, potentially exacerbating pre-existing inequalities. Finally, our approach can be used to examine the impact of news deserts from a content perspective, quantifying the spatial distribution of location mentions and its influence on audience engagement, community identity and democratic participation. By enabling these types of analyses, our work can contribute to a deeper understanding of the relationship between local news, geographic narratives and socio-political dynamics. This, in turn, can inform policy decisions related to media ownership, local journalism and community development. In conclusion, this study provides valuable methodological tools, data resources and empirical insights that advance the field of local news geoparsing and open up exciting new avenues for investigating the interplay between media, geography and society.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/chr.2025.10012.

Acknowledgements. We would like to thank Xuke Hu for benchmarking our method against their fine-tuned approach and the thoughtful feedback.

Data availability statement. The dataset created in this study, LMUK-Geo, is publicly available on the Harvard Dataverse and can be accessed via its DOI: https://doi.org/10.7910/DVN/SGVXIU. This dataset includes U.K. local media news articles annotated with geographic references and is provided in standard formats (CSV and JSON) suitable for geoparsing and location-based analysis. It adheres to the FAIR principles of Findability, Accessibility, Interoperability and Reusability. The dataset is findable through its persistent identifier and detailed metadata, accessible via the Harvard Dataverse platform, interoperable with various tools due to its standard formats, and reusable under clear terms of use with proper attribution.

Source code for this study can be accessed on GitHub and are stored at the following persistent link: https://doi.org/10.5281/zenodo.14783454.

Author contributions. S.B.: Conceptualisation, methodology, software, data curation, formal analysis, investigation, visualisation, writing–original draft. A.G.: supervision, writing–review and editing. B.H.: supervision, writing–review and editing.

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

Funding statement. The first author was supported during their doctoral studies by the Surrey Institute for People-Centred AI at the University of Surrey.

Ethical standards. The authors affirm that this research did not involve human participants.

Disclosure of use of AI tools. The authors wrote this manuscript. Generative AI tools (ChatGPT and Claude) were used only to suggest improvements to narrative flow, structure, and to help resolve minor coding issues. No sections of text were generated verbatim by these tools. All ideas, analyses, and code are the work of the authors, who take full responsibility for the content.

References

Abernathy, Penelope Muse. 2020. *News Deserts and Ghost Newspapers: Will Local News Survive?* Chapel Hill, NC: UNC Hussman School of Journalism and Media.

Ardanuy, Mariona Coll, Kasra Hosseini, Katherine McDonough, Amrey Krause, Daniel Van Strien, and Federico Nanni. 2020. "A Deep Learning Approach to Geographical Candidate Selection Through Toponym Matching." In C.-T. Lu, F. Wang, G. Trajcevski, Y. Huang, S. Newsam & L. Xiong (Eds.), *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL GIS 2020 (pp. 385–388). Association for Computing Machinery, New York, NY.

Ardanuy, Mariona Coll, Kasra Hosseini, Katherine McDonough, Amrey Krause, Daniel Van Strien, and Federico Nanni. 2023. "The Past is a Foreign Place: Improving Toponym Linking for Historical Newspapers." In A. Šeļa, F. Jannidis, & I. Romanowska (Eds.), *Proceedings of the Computational Humanities Research Conference* 2023 (Vol. 3558, pp. 123–134). CEUR Workshop Proceedings.

Ardanuy, Mariona Coll, and Caroline Sporleder (2017). "Toponym Disambiguation in Historical Documents Using Semantic and Geographic Features." In *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage DATeCH2017*, 175–80. Association for Computing Machinery, New York, NY.

Berragan, Cillian, Alex Singleton, Alessia Calafiore, and Jeremy Morley. 2023. "Transformer Based Named Entity Recognition for Place Name Extraction From Unstructured Text." *International Journal of Geographical Information Science* 37, no. 4: 747–66.

Berragan, Cillian, Alex Singleton, Alessia Calafiore, and Jeremy Morley. 2024. "Mapping Cognitive Place Associations Within the United Kingdom Through Online Discussion on Reddit." *Transactions of the Institute of British Geographers* 49, no. 3: e12669.

Bisiani, Simona. 2024. "UKTwitNewsCor." Harvard Dataverse. Creative Commons Attribution Non Commercial 4.0 International. https://dataverse.harvard.edu/citation?persistentId=doi:10.7910/DVN/R5XTEO

Bisiani, Simona, Agnes Gulyas, John Wihbey, and Bahareh Heravi. 2025. "UKTwitNewsCor." A Dataset of Online Local News Articles for the Study of Local News Provision. *Proceedings of the International AAAI Conference on Web and Social Media*, 19(1), 2371–2384. https://doi.org/10.1609/icwsm.v19i1.35940

Bisiani, Simona, and Bahareh Heravi. 2023. "Uncovering the State of Local News Databases in the UK: Limitations and Impacts on Research." *Journalism and Media* 4, no. 4: 1211–31.

Bisiani, Simona, and Mitchell, Joe. 2024. "UK Local News Mapping Report—April 2024." Public Interest News Foundation. https://www.publicinterestnews.org.uk/_files/ugd/cde0e9_31f2ee78fff64c3e8616e7eafdf28f99.pdf

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill Erik, Brynjolfsson Shyamal Buch, Dallas Card, Rodrigo Castellon Niladri, Chatterji Annie, Chen Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, LaurenGillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2022. "On the Opportunities and Risks of Foundation Models." Preprint. http://arxiv.org/abs/2108.07258

Brown, Tom Benjamin, Mann Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. "Language Models are Few-Shot Learners." In *Advances in Neural Information Processing Systems*, Vol. 33, 1877–901. Curran Associates, Inc.

Cairncross, Frances. 2019. "The Cairncross Review. A Sustainable Future for Journalism." https://assets.publishing.service.gov.uk/media/5c6bfcd4e5274a72b933311d/021919_DCMS_Cairncross_Review_.pdf

Chatrath, Veronica, Marcelo Lotif, and Shaina Raza. 2024. "Fact or Fiction? Can LLMs be Reliable Annotators for Political Truths?" Preprint. http://arxiv.org/abs/2411.05775

Dang, Hai, Lukas Mecke, Florian Lehmann, Sven Goller, and Daniel Buschek. 2022. "How to Prompt? Opportunities and Challenges of Zero- and FEW-SHOT LEARNING for Human-AI Interaction in Creative Applications of Generative Models." http://arxiv.org/abs/2209.01390

DeLozier, Grant, Jason Baldridge, and Loretta London. 2015. "Gazetteer-Independent Toponym Resolution Using Geographic Word Profiles." In B. Bonet & S. Koenig (Eds.), *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 1074–1080). AAAI Press.

Dubourg, Edgar, Valentin Thouzeau, and Nicolas Baumard. 2024. "A Step-by-Step Method for Cultural Annotation by LLMs." *Frontiers in Artificial Intelligence* 7.

Firmstone, Julie, and Rebecca Whittington. 2021. "Local Political Journalism: Systematic Pressures on the Normative Functions of Local News." In J. Morrison, J. Birks, & M. Berry (Eds.), *The Routledge Companion to Political Journalism* (1st ed., pp. 84–93). London: Routledge.

Franklin, Bob. 2006. "Attacking the Devil? Local Journalists and Local Newspapers in the UK." In B. Franklin & D. Murphy (Eds.), *Local journalism and local media: Making the local news* (pp. 3–15). London: Routledge.

Franklin, Bob, and Stephen Cushion. 2006. "Downgrading the 'Local'in Local Newspapers' Miscing of the 2005 UK General Election." In B. Franklin & D. Murphy (Eds.), *Local journalism and local media: Making the local news* (pp. 3–15). London: Routledge.

Gale, William A., Kenneth W. Church, and David Yarowsky. 1992. "One Sense Per Discourse." In D. H. Pallett (Ed.), *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York*, February 23–26, 1992 (pp. 233–237). Association for Computational Linguistics.

Gao, Pengjie, Chang Lee, and Dermot Murphy. 2020. "Financing Dies in Darkness? The Impact of Newspaper Closures on Public Finance." *Journal of Financial Economics* 135, no. 2: 445–67.

Garz, Marcel, and Mart Ots. 2025. "Media Consolidation and News Content Quality." *Journal of Communication* 75, no. 3: 195–206, https://doi.org/10.1093/joc/jqae053

Goel, Akshay, Almog Gueta, Omry Gilon, Chang Liu, Sofia Erell, Lan Huong Nguyen, Xiaohong Hao, Bolous Jaber, Shashir Reddy, Rupesh Kartha, Jean Steiner, Itay Laish, and Amir Feder. 2023. "LLMs Accelerate Annotation for Medical Information Extraction." In S. Hegselmann, A. Parziale, D. Shanmugam, S. Tang, M. N. Asiedu, S. Chang, T. Hartvigsen & H. Singh (Eds.), *Proceedings of the 3rd Machine Learning for Health Symposium* (pp. 82–100). Proceedings of Machine Learning Research, Vol. 225. PMLR.

Gritta, Milan, Mohammad Taher Pilehvar, and Nigel Collier. 2020. "A Pragmatic Guide to Geoparsing Evaluation." *Language Resources and Evaluation* 54, no. 3: 683–712.

Gritta, Milan, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. 2018. "What's Missing in Geographical Parsing?" *Language Resources and Evaluation* 52, no. 2: 603–23.

Hagar, Nick, Jack Bandy, Daniel Trielli, Yixue Wang, and Nicholas Diakopoulos. 2020. "Defining Local News: A Computational Approach." In *Computational + Journalism Symposium 2020*. https://cj2021.northeastern.edu/files/2020/02/CJ_2020_paper_40.pdf

Hayes, Danny, and Jennifer L. Lawless. 2018. "The Decline of Local News and its Effects: New Evidence from Longitudinal Data." *The Journal of Politics* 80, no. 1: 332–6.

Hess, Kristy, and Lisa Waller. 2017. *Local Journalism in a Digital World*, (First published ed.). London: Macmillan Education/Palgrave Macmillan.

Honnibal, Matthew, and Ines Montani. 2017. "spacy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing."

Hu, Xuke, Tobias Elßner, Shiyu Zheng, Helen Ngonidzashe Serere, Jens Kersten, Friederike Klan, and Qinjun Qiu. 2024. "DLRGeoTweet: A Comprehensive Social Media Geocoding Corpus Featuring Fine-Grained Places." *Information Processing & Management* 61, no. 4: 103742.

Hu, Xuke, Jens Kersten, Friederike Klan, and Sheikh Mastura Farzana. 2024. "Toponym Resolution Leveraging Lightweight and Open-Source Large Language Models and Geo-Knowledge." *International Journal of Geographical Information Science*, 1–28.

Hu, Xuke, Yeran Sun, Jens Kersten, Zhiyong Zhou, Friederike Klan, and Hongchao Fan. 2023. "How Can Voting Mechanisms Improve the Robustness and Generalizability of Toponym Disambiguation?" *International Journal of Applied Earth Observation and Geoinformation* 117: 103191.

Hu, X., Z. Zhou, H. Li, Y. Hu, F. Gu, J. Kersten, H. Fan, and F. Klan. 2023. "Location Reference Recognition From Texts: A Survey and Comparison." *ACM Computing Surveys* 56, no. 5: 112:1–37.

Hu, Yingjie, Gengchen Mai, Chris Cundy, Kristy Choi, Ni Lao, Wei Liu, Gaurish Lakhanpal, Ryan Zhenqi Zhou, and Kenneth Joseph. 2023. "Geo-Knowledge-Guided GPT Models Improve the Extraction of Location Descriptions from Disaster-Related Social Media Messages." *International Journal of Geographical Information Science* 37, no. 11: 2289–318. Publisher: Taylor & Francis. https://doi.org/10.1080/13658816.2023.2266495

Huang, Lei, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2024. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." *ACM Transactions on Information Systems* 43 no. 2: 1–55.

Hunter, Thomas. 2024. "UK Local Newspaper Closures Update: 293 Now Gone Since 2005." Press Gazette. https://pressgazette.co.uk/publishers/regional-newspapers/local-newspaper-closures-uk-2022-to-2024/

Kafando, Rodrique, Rémy Decoupes, Mathieu Roche, and Maguelonne Teisseire. 2023. "SNEToolkit: Spatial Named Entities Disambiguation Toolkit." *SoftwareX* 23: 101480.

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

Kamalloo, Ehsan, and Davood Rafiei. 2018. "A Coherent Unsupervised Model for Toponym Resolution." In P.-A. Champin, F. Gandon, M. Lalmas, & P. G. Ipeirotis (Eds.), *Proceedings of the 2018 World Wide Web Conference (WWW '18)* (pp. 1287–1296). New York, NY: ACM.

Karimzadeh, Morteza, Scott Pezanowski, Alan M. MacEachren, and Jan O. Wallgrün. 2019. "GeoTxt: A Scalable Geoparsing System for Unstructured Text Geolocation." *Transactions in GIS* 23, no. 1: 118–36.

Karjus, Andres. 2025. "Machine-Assisted Quantitizing Designs: Augmenting Humanities and Social Sciences with Artificial Intelligence." Humanit Soc Sci Commun 12, 277. https://doi.org/10.1057/s41599-025-04503-w

Khanom, Asma, Damon Kiesow, Matt Zdun, and Chi-Ren Shyu. 2023. "The News Crawler: A Big Data Approach to Local Information Ecosystems." *Media and Communication* 11, no. 3: 318–29.

Koetsenruijter, Willem, and Jaap De Jong. 2023. "The Rhetoric of Trust in Local News Media: Proximity as a Quintessential News Quality." *Rhetoric and Communications* no. 55: 9–37.

Lee, Sam Yu-Te, Aryaman Bahukhandi, Dongyu Liu, and Kwan-Liu Ma. 2025. "Towards Dataset-Scale and Feature-Oriented Evaluation of Text Summarization in Large Language Model Prompts." *IEEE Transactions on Visualization and Computer Graphics* 31, no. 1: 481–91.

Leidner, Jochen L. 2007. "Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding." *SIGIR Forum* 41, no. 2: 124–6.

Leppämäki, T., T. Toivonen, and T. Hiippala. 2024. "Geographical and Linguistic Perspectives on Developing Geoparsers with Generic Resources." *International Journal of Geographical Information Science* 38, no. 10: 2039–60.

Lewis, Justin, Andrew Williams, and Bob Franklin. 2008. "Four Rumours and an Explanation: A Political Economic Account of Journalists' Changing Newsgathering and Miscing Practices." *Journalism Practice* 2, no. 1: 27–45.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems* 33: 9459–74.

Lieberman, Michael D., Hanan Samet, and Jagan Sankaranarayanan. 2010. "Geotagging with Local Lexicons to Build Indexes for Textually-Specified Spatial Data." In *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*, 201–12.

Lindgren, April. 2009. "News, Geography and Disadvantage: Mapping Newspaper Coverage of High-Needs Neighbourhoods in Toronto, Canada." *Canadian Journal of Urban Research* 18, no. 1: 74–97.

Liu, Fei, Zejun Kang, and Xing Han. 2025. "Optimizing RAG Techniques for Automotive Industry PDF Chatbots: A Case Study with Locally Deployed Ollama Modelsoptimizing RAG Techniques Based on Locally Deployed Ollama Modelsa Case Study with Locally Deployed Ollama Models." In *Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Intelligent Information Processing AIIIP'24*, 152–59. New York, NY: Association for Computing Machinery.

Liu, Pengfei, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. "Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing." *ACM Computing Surveys* 55, no. 9: 1–35.

Liu, Zilong, Krzysztof Janowicz, Ling Cai, Rui Zhu, Gengchen Mai, and Meilin Shi. 2022. "Geoparsing: Solved or Biased? An Evaluation of Geographic Biases in Geoparsing." *AGILE: GIScience Series* 3: 1–13.

Madrid-Morales, Dani. 2020. "Using Computational Text Analysis Tools to Study African Online News Content." *African Journalism Studies* 41, no. 4: 68–82.

Madrid-Morales, Dani, Joan Ramon Rodríguez-Amat, and Peggy Lindner. 2023. "A Computational Mapping of Online News Deserts on African News Websites." *Media and Communication* 11, no. 3: 330–42.

Mai, Gengchen, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, Chris Cundy, Ziyuan Li, Rui Zhu, and Ni Lao. 2024. "On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper)." *ACM Transactions on Spatial Algorithms and Systems* 10, no. 2: 1–46.

Majid, Aisha. 2023. "Who Owns UK Local News Media? Print and Digital Consolidation Charted." Press Gazette. https://pressgazette.co.uk/media-audience-and-business-data/who-owns-the-uk-regional-media-print-and-digital/

Matsuda, Koji, Akira Sasaki, Naoaki Okazaki, and Kentaro Inui. 2015. "Annotating Geographical Entities on Microblog Text." In *Proceedings of the 9th Linguistic Annotation Workshop*, edited by **Adam Meyers, Ines Rehbein, and Heike Zinsmeister**, 85–94. Denver, United States: Association for Computational Linguistics.

McAdam, Alison, and Kristy Hess. 2024. "Re-asserting the Value of Local "News Presence" for Small-Town News Outlets in a Digital Era." *Journalism Practice* 1–16.

Metzger, Zach. 2024. "The State of Local News 2024." Published by Local News Initiative, Medill School of Media, Journalism, and Integrated Marketing Communications, Northwestern University. https://localnewsinitiative.northwestern.edu/projects/state-of-local-news/2024/

Middleton, Stuart E., Giorgos Kordopatis-Zilos, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2018. "Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging." *ACM Transactions on Information Systems* 36, no. 4: 1–27.

Montani, Ines, and Matthew Honnibal. 2018. "Prodigy: A New Annotation Tool for Radically Efficient Machine Teaching." https://explosion.ai/blog/prodigy-annotation-tool-active-learning

Moore, Martin, and Gordon Neil Ramsay. 2024. "Local News in National Elections: An "Audit" Approach to Assessing Local News Performance During a National Election Campaign." *Digital Journalism* 1–20.

Napoli, Philip M., and Matthew S. Weber. 2020. "Local Journalism and at-Risk Communities in the United States." In A. Gulyas & D. Baines (Eds.), *The Routledge Companion to Local Media and Journalism* (pp. 368–378). London: Routledge.

Pebesma, Edzer J. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10, no. 1: 439.

PLUM Consulting. 2020. "Research into Recent Dynamics of the Press Sector in the UK and Globally." https://assets.publishing.service.gov.uk/media/5f7b4673e90e070dec5d9e29/Plum_DCMS_press_sector_dynamics_-_Final_misc_v4.pdf.

Ponsford, Dominic. 2024. "Colossal Decline of UK Regional Media Since 2007 Revealed." Press Gazette. https://pressgazette.co.uk/publishers/regional-newspapers/colossal-decline-of-uk-regional-media-since-2007-revealed/

Public Interest News Foundation. 2023. "Deserts, Oases and Drylands." https://www.publicinterestnews.org.uk/_files/ugd/cde0e9_97c4fe55ab0c49a0a29f40937e71d216.pdf

Quattrone, Giovanni, Licia Capra, and Pasquale De Meo. 2015. "There's no Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-Sourcing Datasets." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing CSCW'15*, 1021–32. Association for Computing Machinery.

Ramsay, Gordon, and Martin Moore. 2016. "Monopolising Local News: Is There an Emerging Local Democratic Deficit in the UK Due to the Decline of Local Newspapers?" Publisher: King's College London. https://www.kcl.ac.uk/policy-institute/assets/cmcp/local-news.pdf.

Reese, Stephen D. 2016. "The New Geography of Journalism Research." *Digital Journalism* 4, no. 7: 816–26.

Sanh, Victor, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M. Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Tali Bers, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. 2022. "Multitask Prompted Training Enables Zero-Shot Task Generalization." In the Tenth International Conference on Learning Representations (ICLR), 2022. Spotlight Paper.

Sharman, David. 2021. "Reach plc to Close all Bar 15 of its Newspaper Offices." Hold The Front Page. https://www.holdthefrontpage.co.uk/2021/news/publisher-to-close-all-bar-15-offices-leaving-dailies-without-base-on-patch/

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press

Sjvaag, Helle. 2014. "Homogenisation or Differentiation?: The Effects of Consolidation in the Regional Newspaper Market." *Journalism Studies* 15, no. 5: 511–21.

Teitler, Benjamin E., Michael D. Lieberman, Daniele Panozzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperling. 2008. "NewsStand: A New View on News." In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems GIS'08*, 1–10. Association for Computing Machinery.

The Media Reform Coalition. 2023. "Who Owns the UK Media?" https://www.mediareform.org.uk/wp-content/uploads/2023/10/Who-Owns-the-UK-Media-2023.pdf

Trad, Fouad, and Ali Chehab. 2024. "To Ensemble or Not: Assessing Majority Voting Strategies for Phishing Detection with Large Language Models." Preprint. http://arxiv.org/abs/2412.00166

Usher, Nikki. 2019. "Putting "Place" in the Center of Journalism Research: A Way Forward to Understand Challenges to Trust and Knowledge in News." *Journalism & Communication Monographs* 21, no. 2: 84–146.

Vogler, Daniel, Morley Weston, and Linards Udris. 2023. "Investigating News Deserts on the Content Level: Geographical Diversity in Swiss News Media." *Media and Communication* 11, no. 3: 343–54.

Walczak, Eryk J. 2023. "Postcodesior: An R Package for UK Geocoding." *Journal of Open Source Software* 8, no. 84: 5334.

Waterson, Jim. 2021. "Mirror Owner to Tell Most Journalists to Permanently Work From Home." The Guardian. https://www.theguardian.com/business/2021/mar/19/mirror-owner-tell-most-journalists-permanently-work-from-home-reach

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. 2022. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." *Advances in Neural Information Processing Systems* 35: 24824–37.

Weiss, Amy Schmitz. 2018. "Journalism Conundrum: Perceiving Location and Geographic Space Norms and Values." *Westminster Papers in Communication and Culture* 13, no. 2: 46–60.

Yin, Wenpeng, Jamaal Hay, and Dan Roth. 2019. "Benchmarking Zero-Shot Text Classification: Datasets, Evaluation and Entailment Approach." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, edited by **Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan**, 3914–23. Stroudsburg, PA: Association for Computational Linguistics.

Zhang, Zeyu, and Steven Bethard. 2024. "A Survey on Geocoding: Algorithms and Datasets for Toponym Resolution." *Language Resources and Evaluation* 59: 1775–96.

Zhao, Xiutian, Ke Wang, and Wei Peng. 2024. "An Electoral Approach to Diversify LLM-based Multi-Agent Collective Decision-Making." In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 2712–2727, Miami, Florida, USA. Association for Computational Linguistics.

Zheng, Lianmin, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena." *Advances in Neural Information Processing Systems* 36: 46595–623.

https://doi.org/10.1017/chr.2025.10012 Published online by Cambridge University Press