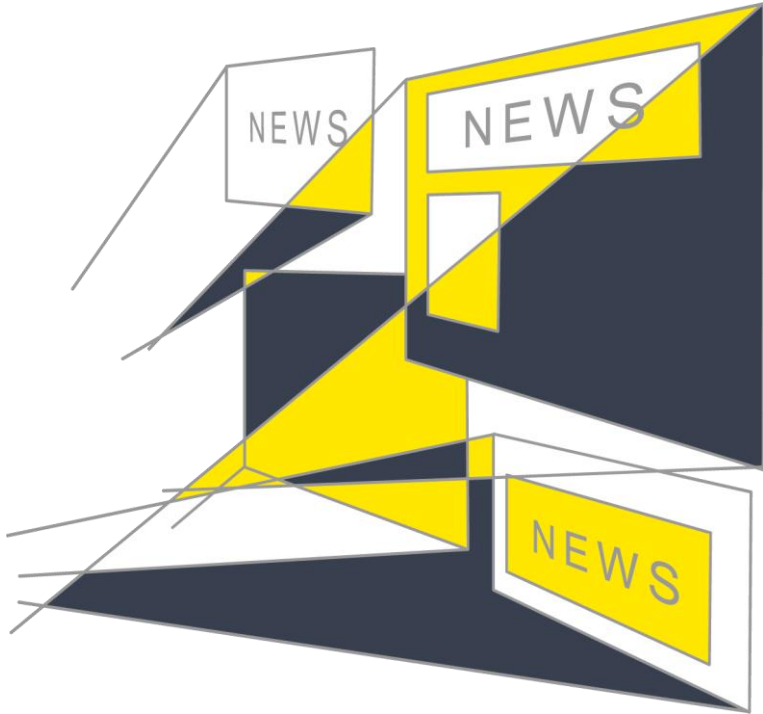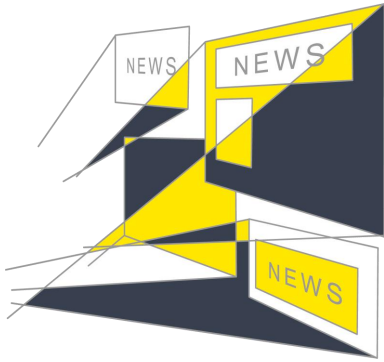# BBC online
local news analysis

Prepared for the Department for Culture, Media and Sport (DCMS)

December 2023

About the authors

Alma Economics combines unparalleled analytical expertise with the ability to communicate complex ideas clearly.

www.almaeconomics.com

About the commissioning organisation

This independent analysis was commissioned by the Department for Digital, Culture, Media & Sport (DCMS). The analysis and findings are those of the authors and do not represent the views of DCMS.

www.gov.uk

Table of contents

Executive summary

Alma Economics was commissioned by the Department for Digital, Culture, Media and Sport (DCMS) to carry out an independent study to examine the relationship between online local news content published by the BBC and local commercial news organisations. The main objective of this study was to analyse differences in the content and type of local news coverage provided by the BBC and the commercial sector.

The study was conducted from August to October 2022 and involved: (i) a desk-based review of local news data supplied by the Joint Industry Currency for Regional Media Research (JICREG), of BBC strategy documents and of previous research on the UK news media sector, (ii) quantitative analysis of 143,438 BBC and commercial local news articles published online between July 2018 and September 2022, and (iii) one-on-one interviews with 13 online news editors at commercial news providers and one senior BBC executive.

The main conclusions of the study are listed below:

- Our findings suggest that content duplication between the BBC and commercial providers in online local news exists in a way that is consistent with multiple providers covering some of the same news events. For example, news providers may request quotes from the same sources or describe the facts or key features of news events in a consistent manner.

- However, it is important to note that this conclusion has been reached on the basis of quantitative analysis of article content using text similarity measures. These measures cannot be benchmarked against other news-based studies (as no similar research has previously been carried out) or against a comparison between commercial providers (which was not possible due to dataset limitations).

- Stakeholders within the local news sector highlighted both (i) concerns about the market impact of further expansion of the BBC into online local news coverage and (ii) calls for greater collaboration with the BBC.

- Further research is required to understand the extent to which the BBC complements and competes with commercial providers in online local news provision.

BBC and commercial online local news coverage

- Overall, we identified 57,123 BBC online local news articles and matched these to 86,315 commercial articles reporting on the same events.

- Fewer than 42% of local news events are covered by both the BBC and commercial providers. This figure is likely to be an overestimate due to the limitations of our methodology.

- Fewer than 2% of local news events are covered only by the BBC (i.e. not also covered by commercial providers) and tend to be focused on broader issues faced by communities rather than specific events.

- For local news events covered by the BBC, on average a single local news event is covered by the BBC and three commercial providers. In low coverage areas, news events are usually only covered by the BBC and a single commercial provider.

- For local news events covered by the BBC, on average the BBC "breaks the story" (is the first provider to publish an article reporting on the news event) 19% of the time in our dataset, and these articles are most commonly about weather, the NHS/health or local government. On average, the BBC is the third or fourth outlet to report on an event.

- Across the BBC and commercial providers, law enforcement and crime is the most frequent article topic (31% of all events covered), followed by the NHS/health (24%), then local interest stories such as community events or issues (17%).

Differences in coverage of events by the BBC and commercial sector

- When reporting on the same event, commercial articles are usually longer than BBC articles, although article length varies substantially (some articles are around 100 words long, while others exceed 750 words).

- When reporting on the same event, BBC and commercial articles are somewhat similar (a 67% similarity score) and frequently draw on the same sources. However, there are differences in how the information is presented, with some topics explored in greater depth and others shortened or summarised.

- A relatively small number of articles include links to other websites, and a slightly greater number of BBC articles include external links compared to commercial articles about the same news event. External links tend to be further reading/resources providing more detail about the news event, though some also direct to local commercial providers who reported on the event earlier.

- Most articles reporting on the same event, whether published by the BBC or commercial providers, use the same set of images.

Interviews with media organisations

- Due to low response rates to interview requests (over 500 e-mail invitations were sent out), we were only able to carry out 13 interviews with editorial staff as part of this research (no interviews were conducted with commercial staff or staff from other divisions).

- The commercial news editors we spoke to generally assigned reporters to specific geographic regions, with news events identified from media invitations, social media, fieldwork or community-submitted tips/suggestions. The BBC relied on similar approaches to identify news events, though with greater collaboration across other BBC media (radio and TV).

- Commercial news editors sought to make their titles distinct by emphasising the local nature of their reporting and their relationship with the local community, building name recognition and a reputation for credibility and trustworthiness among local audiences.

- The BBC also emphasised the importance of covering events that mattered to local communities, with its online coverage focused more on educating and informing audiences. In addition, the BBC sought to distinguish its coverage from other providers through better planning, better research and greater editorial scrutiny.

- Few commercial news editors regularly interacted with the BBC except through the Local Democracy Reporting Service, and none of the news editors specifically considered BBC reporting when deciding which news events to cover.

- Almost all commercial news editors stated that content duplication with BBC local news websites was not considered a risk or had no financial impact on their organisation. News editors often mentioned that BBC local news teams had to cover much larger geographic areas and could not provide as much coverage of local events as their own titles.

- Around one-third of commercial news editors expressed concern that it was difficult to compete with a free source of local news, and that increased investment by the BBC in local news provision might bring the BBC into more direct competition with commercial providers.[1] In general, though, editors felt that online aggregators such as Facebook and Google had a greater financial impact or posed a greater risk to their organisation.

- However, commercial news editors were also optimistic about opportunities for greater collaboration with the BBC in the future, including in investigative/longform reporting or through an expansion of the Local Democracy Reporting Service.

---

[1] Interviews were conducted before the BBC's announcement in October 2022 to establish investigative reporting teams across England and boost daily online local news provision.

1. Introduction

One key aspect of the UK Government Mid-Term Review of the BBC Charter is examining how the BBC and Ofcom assess the market impact and public value of the BBC in an evolving marketplace and how that relates to the wider UK media ecology, including with regards to commercial radio and local news sectors and other content producers and distributors.

Previous research (KPMG 2015) found limited evidence that the BBC had either a significant negative market impact on, or crowded out, commercial news providers during the previous Charter period. However, concerns remain among news publishers that the dominance of the BBC makes it harder for them to compete in a digital market. Since the last Charter review in 2016, the acute financial challenges faced by news providers have worsened, particularly at the local level. As highlighted by the 2019 Cairncross Review, these challenges arise from the shift from print to online news consumption, a transition which has accelerated as a result of the Covid-19 pandemic. Driven by a decline in print circulation and a fall in advertising revenues, overall local press sector revenue declined by 37% between 2010 and 2018, with an increasing number of providers facing negative margins. Readership of regional print daily newspapers fell by 18% through the second half of 2020, and while the largest providers found that online readership and subscriptions had increased, it is still unclear if subscribers will remain in sufficient numbers to offset further decreases in print circulation (Economic Insight, 2021). These financial challenges have been a factor in driving the growing consolidation of the local news industry (Ramsay and Moore, 2016) and the decline in the informational role of local news over time (Firmstone 2016, Reardon et al. 2020).

In light of these changes in the UK news media sector, the Department for Digital, Culture, Media and Sport (DCMS) commissioned Alma Economics to carry out an independent study to examine the relationship between online local news content from the BBC and local commercial news organisations. The research questions for the study focus on understanding the extent to which the BBC acts as a complement to, rather than substitute for, commercial news. In particular:

- What are the differences in the content and type of local news coverage provided by the BBC and the commercial sector? What volume of local news events are covered by both commercial news providers and the BBC?

- Where the BBC and the commercial sector cover the same events, are there differences in how these events are reported? What are the positives, negatives and risks associated with this duplication/amplification for audiences (e.g. in terms of reaching different or diverse audiences) and media organisations (e.g. in terms of lost revenue or reach)?

The rest of this report is structured as follows:
- Section 2 provides an overview of the study methodology and summarises the dataset of online local news articles collected for analysis.

- Section 3 summarises our findings from the quantitative analysis of BBC and commercial online local news articles.

- Section 4 summarises our findings from interviews with industry stakeholders.

- Section 5 synthesises our quantitative and qualitative findings as well as how these address the study's research questions.

- Appendix A provides more details on the technical aspects of the study methodology.

- Appendix B lists the full set of questions used during our interviews with industry stakeholders.

2. Methodology and data

Data collection

To analyse differences in local news coverage between the BBC and commercial providers, we first needed to collate a dataset of UK local news articles, then match articles to specific news events (this allowed us to compare the frequency and type of coverage across providers for the same news event) and geography (so we could analyse the relationship between BBC and commercial online local news coverage separately for "high coverage" and "low coverage" areas). We built this dataset using a five-step process:

1. Develop a list of UK commercial local news titles and their website URLs.
2. Identify all articles published by each title across a variety of databases using an automated approach.
3. Develop a definition of local news and keep only the articles in our dataset reporting on local news.
4. Identify commercial articles reporting on the same news event as BBC articles using search engine results (sorted by relevance).
5. Match articles to local authorities based on local news titles or place names mentioned in the article's text.

Although our approach to data collection has several important limitations (discussed in the "Data limitations" section below) and limits the external validity of our analysis, we believe this approach is still valuable as it significantly expands the number of articles that can be included in our analysis compared to previous research that used manual methods to extract top articles on website homepages for content analysis. Our much-larger sample size thus provides the most comprehensive view of the online local media landscape in the UK to date, including both the magnitude of articles published as well as their content and structure.

A list of UK commercial local news titles was compiled from the Audit Bureau of Circulations (ABC), Hold the Front Page, the Independent Community News Network (ICNN) and the Joint Industry Currency for Regional Media Research (JICREG). From this list of titles, we identified local news articles using three different methods:

- A title-specific search in the News API database for articles published online by the title since July 2018.

- A Bing Web and Bing News search for each BBC local news article identified in the News API databases using the full BBC article title.[2] Search results were filtered by publication date (articles in search results were only included if they were published within a three-day window of the BBC article), and the list of search results was cross-referenced with the list of UK commercial local news title web domains (as many search results were from third-party news aggregator websites or commercial/trade publications).[3]

- A Twitter search (using the Twitter API) for each news title.

---

[2] The Bing News and Search API (application programming interface) allowed us to automatically collect search results returned for a specific search string (these are the same search results that would appear if an individual visited www.bing.com and entered the search string directly).

[3] We reviewed the top 100 search results for each Bing search, as the Bing Search API returns a maximum of 100 results.

We also visited each news title's URL to check if the title was still in use as a main domain name. After this check, we compiled news articles from 377 unique local news websites in total, including 289 in England, 50 in Scotland, 26 in Wales and 12 in Northern Ireland.

The terms and conditions of many multi-title providers specifically prohibit automated scraping or collection of news articles using the provider's website to build a database of news articles. In addition, many of these websites will slow webpage loading speeds or introduce CAPTCHAs/robot-detection checks if an automated script is detected, which makes this method of collecting articles extremely time consuming. To address this issue, we used three separate datasets (News API, Bing Web/News Search, Twitter) to take advantage of news articles indexed on search engines, which can be collected using an automated script.

While our full dataset included around 750,000 articles, we removed three types of articles to ensure we only focused on local news:

- Articles mentioning keywords related to the UK government or international affairs: "House of Commons", "prime minister", the names of specific prime ministers or names of other countries.

- Articles mentioning keywords related to entertainment, sports or pop culture: articles under the news provider's "Sports" or "Lifestyle" categories or which included the name of a TV show, film or sports team as prominent keywords.[4]

- BBC articles for which the URL did not include one of the 63 subdomains for BBC regions (for example, articles with a URL starting with https://www.bbc.co.uk/news/england/tees would be included, while articles with a URL starting with https://www.bbc.co.uk/news/politics would not be included).

For the first two types of articles, we only removed the article if it did not mention a place more granular than the regional level ("East of England", "West Midlands", etc.). For example, an article on MPs from the North East discussing the UK Government's levelling up pledges is included because it mentions specific geographies such as Teesside and Darlington that would be relevant to a local news audience. Our final dataset of articles included 143,438 articles published since July 2018:[5] 57,123 from BBC regional websites and 86,315 from commercial providers that could be linked to a BBC local news article. In addition, we identified 254,031 articles from commercial providers from the News API dataset that met the first two criteria above but did not appear in Bing Web/News search results or could not be automatically extracted due to website terms and conditions.

Before we could carry out event-level analysis, we first needed to classify articles to specific events so we could directly compare BBC and commercial provider articles reporting on the same news event. Different automated methods for article classification were tested (these are described in more detail in Appendix A); however, there were sufficient differences in the classification results between these automated methods, in part due to the limited information provided from article titles alone, that we did not feel these methods were robust. Instead, for the event-level analysis presented in the report, we automatically classified any news article identified in Bing Web/News searches in the same event as the BBC article used as the search keyword (in practice, this means

---

[4] More specifically, we removed any of the following subdomains used by news providers: /sport, /sports, /videos, /jobs, /life, /style, /podcasts, /showbiz, /entertainment, /puzzles, /travel, /arts, /technology, /world-news, /life, /woman, /lifestyle, /arts-and-culture, /gallery, /leisure, /drink, /arts, /events.

[5] July 2018 was the earliest publication date for articles included in the News API dataset.

classification was based on search result relevance assessed by Bing). While it is possible that not all articles returned as search results by Bing reported on the same news event as the BBC article, we adopted a narrow window of search results to consider, and a manual review of a sample of different Bing search results across geographies and primary topics confirmed that all search results from local providers were focused on the same news event.

Analysis by article primary topic

In addition to aggregate-level statistics, we also carried out sub-group analysis of articles (by primary topic and local authority) to understand variations in the relationship between BBC and commercial coverage of local news. Primary topics included law enforcement/crime, the NHS/health (including the Covid-19 pandemic), business and the economy, local government/politics, education and weather. The remaining articles could not be classified into one of the topics above and were placed into a category of "local interest stories", such as community events or issues. Classification was undertaken using a range of automated machine-learning techniques (described in more detail in Appendix A). Single topics were selected based on keyword weighing, with keywords highly suggestive or relevant to a primary topic given the greatest weight: for example, "sentencing" and "solicitor" are highly relevant to law enforcement/crime, the name of specific councils for local government/politics, etc. This was cross-checked against the outputs from the zero-shot classification models, which assigned a probability score for each primary topic (a news article would have separate scores for law enforcement/crime, the NHS/health, education, etc.).

Analysis by geography

To define local authorities with limited news coverage, we used JICREG data to map news titles to local authorities (looking up individual local authorities returns a list of news titles), in line with Ramsay and Moore (2016) and Lavender et al. (2020). There are two possible metrics that can be estimated from the JICREG data: (i) total number of local commercial news titles with some online reach in the local authority (threshold of 20% of adults who live in the local authority), or (ii) total proportion of readers in the local authority that access a local commercial news title online, summed across all local commercial news titles.

However, neither metric fully aligns with an intuitive understanding of "low coverage" or "high coverage" areas: news titles are often linked to individual towns, and local authorities with a number of small- and medium-sized towns would then be covered by a number of titles. For example, Eden, the least densely populated local authority in England, has four papers with at least 20% online reach (Chronicle Live, Cumberland & Westmorland Herald, News & Star, The Westmorland Gazette), while Manchester only has one such paper (Manchester Evening News). The total online reach of commercial titles in Eden is nearly 200%, while in Manchester total online reach is 82%. While Lavender et al. (2020) retain local titles for a local authority if a specific percentage of the title's circulation occurs in the local authority (10% for daily titles, 20% for weekly titles), JICREG does not report the proportion of online visits to a commercial news title by location.

For our geographic analysis, we define a "low-coverage" area as a local authority with two or fewer commercial news titles with at least 20% online reach in the local authority, removing all local authorities classified as "Urban with major conurbation", "Urban with minor conurbation" or "Urban

with city and town" by the ONS.[6,7] A "high-coverage" area is defined as a local authority classified as "Urban with major conurbation" or "Urban with minor conurbation" by the ONS or a local authority with five or more commercial news titles with at least 20% online reach. Intuitively, we would expect "low-coverage" areas to be more rural and "high-coverage" areas to be more urban, as news events in urban areas are more likely to be covered by national-level commercial providers such as the Guardian or the Independent.[8] Based on these definitions, we have identified 86 local authorities across the UK that could be considered low-coverage areas and 128 local authorities that could be considered high-coverage areas. High-coverage areas account for 45.2% of the UK's total population, while low-coverage areas account for 16.6%.

Finally, we matched news events to local authorities using two methods:

1. If a news title primarily covered a single local authority (based on JICREG online reach data), then we assigned this local authority to the news event.

2. Otherwise, we identified place names in the article text and cross-referenced these with the OS Open Names database.

Data limitations

Due to the limited scope of coverage across each of the three datasets used to collect local news articles, our dataset does not include all local news articles published since July 2018: we are limited to accessing the articles that News API has collected, or that news providers have optimised for search engine optimisation (SEO) or posted on social media. This means that while we can estimate the proportion of BBC news articles reporting on local events covered or not covered by commercial providers, we can only estimate an upper bound for the proportion of local events covered by commercial providers but not the BBC. This upper bound also has not been adjusted for articles published under the Local Democracy Reporting Service, which are typically shared across the BBC and commercial providers and account for a proportion of overlap in news coverage. In addition, because different providers might vary in their SEO resourcing, if a news title is only listed with a small number of articles in our dataset, this could be due to (i) the provider publishing a relatively small number of local news articles, (ii) the provider engaging to a more limited extent with social media/search engines, or (iii) the provider choosing titles for articles that are significantly different from the titles used by the BBC. As a result, our findings may be biased towards those providers who have invested time to ensure that their news articles appear higher up in search engine results. This means our findings are not fully generalisable; however, the very large number of articles in our dataset suggests we can be reasonably confident that our quantitative findings are indicative of the relationship between the BBC and commercial providers in online local news provision.[9]

---

[6]

https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification

[7] We use the equivalent classifications for Scotland and Wales, mapping from postcodes/output areas to local authorities and adjusting the classification for population size in 2021. Low- and high-coverage areas do not include any local authorities in Northern Ireland as JICREG does not include website audience data for Northern Ireland.

[8] One opposing argument that could be made is that smaller commercial titles do a better job of covering community issues/events while local coverage from national-level commercial providers focuses on "high-engagement" events (such as crime). We believe our definition is more robust for low-coverage areas, which tend to be of greater policy interest when assessing the health of the local news sector in the UK.

[9] One way of addressing this limitation is the use of weights proportional to total population, total site visits, total

Quantitative analysis

For each article, we calculated the following metrics:

- Article length (total number of words)
- Names, quotes and references (including hyperlinks)[10,11]
- Images provided
- Similarity in language (based on the specific words used in the article)
- Similarity in topics or themes discussed in the article (at the paragraph level)[12]

For hyperlinks, we exclude any references to "Read more about…" or links to general pages on the same provider's website (links to specific articles on the same provider's website are included if the link is part of the article text).

Engagement with industry stakeholders

To provide a broader understanding of the local news journalism ecosystem and understand the risks and impacts of content duplication, we carried out 13 interviews with editors and assistant editors at commercial news providers (all of whom were part of multi-title publishers) as well as one interview with a senior BBC executive. Interviews were semi-structured and lasted between 30-45 minutes, with questions focusing on the editorial process, relationship with the BBC, impacts of content duplication and the future role of the BBC in local news provision. A follow-up e-mail was sent to all interviewees requesting any evidence or data on the financial impact of duplication of local news content by the BBC. The full interview topic guide is included in Appendix B.

---

circulation or other aggregate statistics, though we did not end up using this approach as our desk-based review did not identify any evidence that suggested one or more of these aggregate statistics were correlated with total articles published.

[10] We count the number of quotes using the number of opening quotation marks in the article as a proxy, though this is an indicative rather than precise measure as many news articles summarise what sources had discussed without providing direct quotes.

[11] Unlike Cushion (2019), we do not include hyperlinks that direct readers to social media platforms.

[12] One potential limitation of this analysis is that news articles frequently include paragraphs with just one or two sentences, as individual quotes or sources tend to be split up into separate paragraphs.

3. Analysis of local news articles

Our final dataset included 143,438 local news articles first published between 1 July 2018 and 31 August 2022: 57,123 from BBC regional websites and 86,315 from commercial providers. We identified an additional 254,031 articles from commercial providers that either (i) the full article text could not be extracted from the news website, or (ii) the article did not appear in Bing search results and could not be matched to a corresponding BBC article on the same news event. For the remainder of the analysis, we focused specifically on the 143,438 articles that appeared in Bing search results and for which the full article text could be extracted (unless otherwise indicated).

Among the 57,123 BBC articles, around 17% reported on news events in low-coverage areas and 45% reported on news events in high-coverage areas. In contrast, 13% of the 86,315 articles from commercial providers focused on low-coverage areas and 45% on high-coverage areas.

Provider	All areas	High-coverage areas	Low-coverage areas
BBC	57,123	25,705	9,710
Commercial (total)	86,315	37,115	11,307

The average number of articles in our dataset by commercial provider is 230, while the median number of articles is 53 (suggesting larger commercial providers invest more resources indexing their articles on search engines/third-party databases or sharing these articles on social media).

At the regional level, Scotland (43,923 total articles across the BBC and commercial providers) and the West Midlands (13,835) had the greatest number of local news articles in our dataset, while the North East (2,693) and East Midlands (3,532) had the fewest number of local news articles.[13]

Number of articles by geography

---
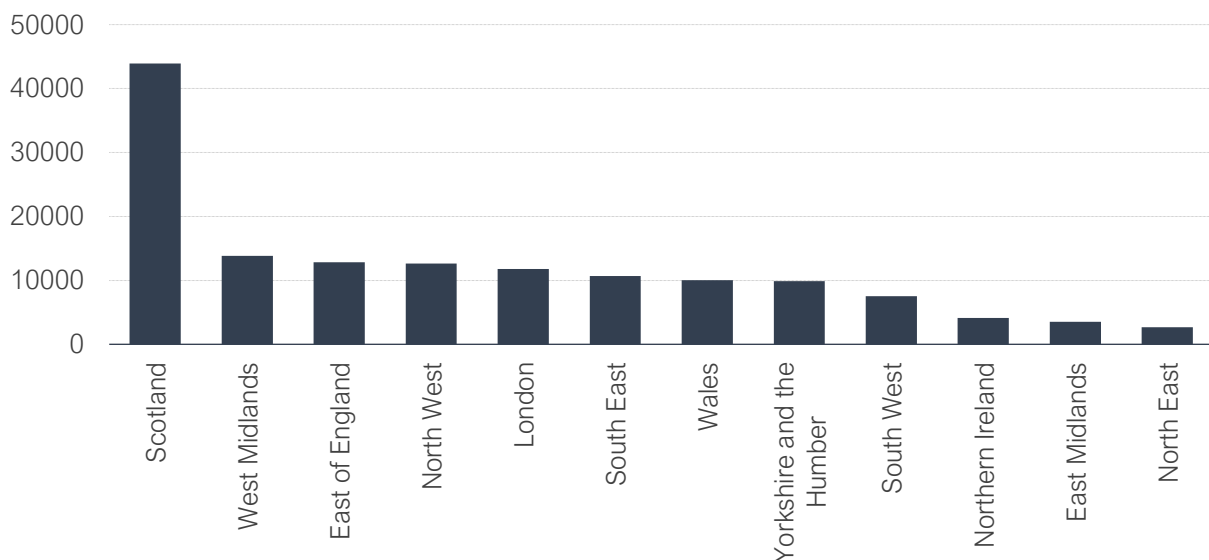
[13] The relatively large proportion of articles from Scotland in our dataset is likely a consequence of our approach to data collection and is not necessarily representative of the true distribution of total online local news articles across UK regions/countries. This result should not be interpreted as stating that Scotland is the focus of a disproportionately large amount of online local news coverage.

Out of the days of the week, the greatest proportion of local news articles are published on Thursday, with Saturday and Sunday having the lowest proportion of articles published. May and June have the greatest proportion of local news articles published across months of the year, with April and September having the lowest proportion. Both the distribution across days of the week and months of the year are consistent across the BBC and commercial providers as well as high- and low-coverage areas.

Share of articles by primary topic
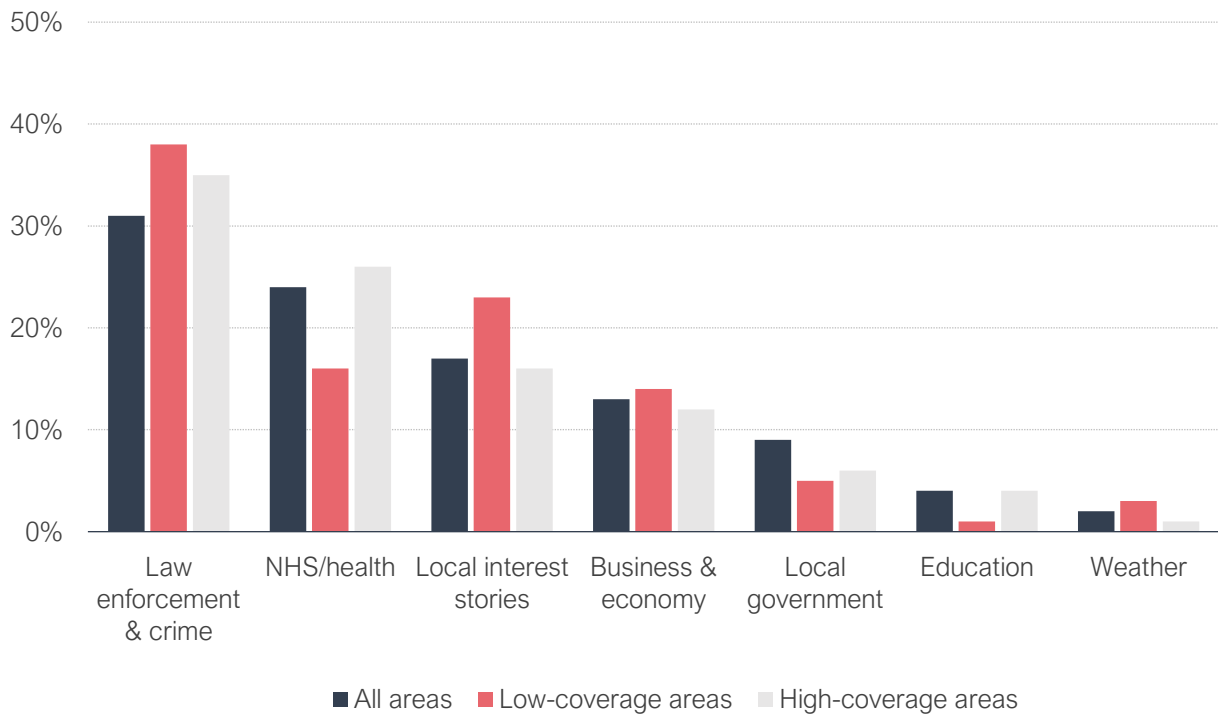
Across the BBC and commercial providers, law enforcement/crime was the most frequent primary topic of local news articles (31% of all news events covered), followed by the NHS/health, including the Covid-19 pandemic (24% of articles), business and the economy (13%), local government and politics (9%), education (4%) and weather (2%). 17% of all articles were placed into a category of local interest stories such as community events or issues. The distribution of articles by primary topic remained mostly consistent across high-coverage and low-coverage areas, though low-coverage areas had a slightly higher proportion of law enforcement/crime articles (38% of all news events covered) and local interest stories (23% of all news events covered), with relatively fewer articles on local government and politics (5%).[14] Note that one consequence of our specific approach to classifying news articles to events (through search results) is that we could not robustly examine differences in the distribution of BBC and commercial articles across article primary topic.[15]

---

[14] Articles published under the Local Democracy Reporting Service fall under the primary topic of local government and politics, though due to inconsistencies in how these stories are attributed across providers we could not estimate the proportion of LDR articles in our dataset.

[15] More specifically, since 52% of BBC articles could not be matched to a commercial news provider article covering the same event, a distribution analysis of primary topic for the 57,123 BBC articles and 86,315 commercial articles in our dataset would largely be biased towards news events for which BBC and commercial articles had similar titles. For the 2,500 randomly selected BBC articles, since we matched 97% of these articles to at least one commercial article, the distribution of primary topics would be nearly identical between the BBC and commercial providers (as for our analysis we assume that primary topics are determined by the news event, not the specific article).

Coverage of BBC and commercial online local news

Out of the 27,139 BBC articles matched to at least one commercial news article covering the same event (based on Bing Web/News search results), the average number of commercial providers covering the event was 3.18. This means that the average number of providers covering a local news event was around four (the BBC and three commercial providers).

Proportion of news events covered by the BBC but not commercial providers

Based on Bing Web/News search result matches using the BBC article title alone, 52% of BBC articles could not be matched to a commercial news provider article covering the same event. Because it is plausible that the choice of language and keywords in the article title by BBC authors/editors will differ from those used by commercial providers (based on differences in editorial processes and guidelines), we carried out a manual Bing Web/News search of 2,500 randomly selected BBC articles that could not be matched to a commercial news provider and reviewed the first 100 results for each search. Out of these 2,500 articles, 2,433 could be matched to an article from a commercial news provider covering the same event. The remaining 67 articles were news articles that focused more on community issues or concerns rather than specific events.[16] It is possible that commercial providers covered these issues or concerns outside the six-day window we searched for (three days before and after the BBC article was first published) or used a significantly different set of keywords in the article title than the BBC, though we could not systematically test this assumption except through our manual review of Bing search results. These results suggest it is likely that the total proportion of BBC online local news events not covered by commercial providers is less than 2% of total events.

Proportion of news events covered by commercial providers but not the BBC

Estimating the total proportion of online local news events covered by commercial providers but not the BBC was more challenging as (i) website terms and conditions means it is not possible to collect article links directly from commercial provider websites, and (ii) the three datasets used in our analysis (News API, Bing Web/News, Twitter) do not cover all articles published by commercial providers. Since we know there are at least 254,031 additional local news articles published by commercial providers, dividing this by the average number of articles per news event (3.18) yields 79,884 news events. As each of the 57,123 BBC articles in our dataset covers a separate news event, we divided   79,884 by (79,884 + 57,123), or the total number of local news events covered by the BBC divided by our estimate of the total number of local news events identified in our dataset. This calculation yields an upper bound of 42%, and we can plausibly conclude that no more than 42% of local news events covered by commercial providers are also covered by the BBC. In practice, this is likely to be an overestimate: many local news articles from commercial providers could not be identified/analysed and anecdotal evidence from interviews with editorial staff suggests the volume of local news articles published by commercial providers on a daily basis is around three to five times larger than the volume published by the BBC.

---

[16] For example, https://www.bbc.co.uk/news/uk-england-south-yorkshire-63032887, which is a quote-centric article about the impact of a council scheme on a specific set of individuals.

Note that this estimate is an upper bound, not a point estimate, and relies on several assumptions: (i) each of the 57,123 BBC articles constitutes a separate news event, and (ii) the average number of providers covering a news event is consistent across events.[17] For example, if local news events not covered by the BBC are smaller in scale, they may also have fewer commercial providers covering the news event. As an indicative example, if we assume the average number of articles per news event is 33% lower (2.13 instead of 3.18 articles), then the proportion of local news events covered by commercial providers as well as the BBC would decrease from 42% to 32%.

Across primary topic of article, weather had the largest number of providers covering an individual news event (3.85 commercial providers on average), as individual weather events (such as flooding or heat waves) were likely to impact many different areas of the UK at the same time. Law enforcement/crime had the second largest number of providers covering an individual news event (3.71 providers on average), which aligns with the widely held belief that these articles tend to generate more clicks and engagement than other news articles. In contrast, there was relatively little overlap in local interest stories (1.46 providers on average) or education (1.71 providers on average).

### Number of news providers covering an event
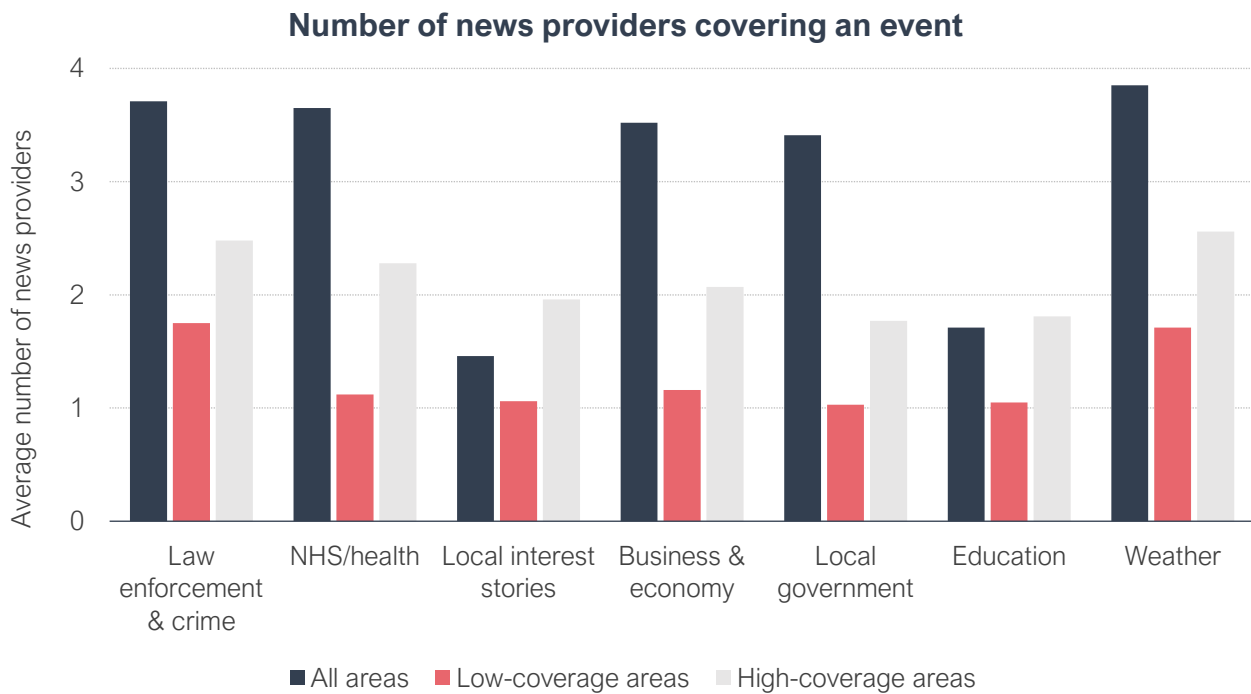
As expected, the average number of providers covering individual news events in low-coverage areas was much lower than across the UK as a whole (1.36 commercial providers on average). There was relatively less variation across article primary topic, with most news events covered by between 1-2 providers on average (law enforcement/crime news events were covered by the most providers on average, at 1.75). Surprisingly, when examining high-coverage areas, the average number of commercial providers covering a news event was also quite low (2.23 commercial providers on average). This in part reflects a limitation with JICREG data, which tends to include just one or two providers in large urban areas. In practice, news events taking place in urban areas may also be covered by commercial titles which are national in scope, but national titles are not included in JICREG data and articles from national titles were not collected as part of this study.

---

[17] A manual review of a sample of BBC articles suggested the first assumption is likely to be true while the second assumption is not as the BBC tends to cover the largest or most prominent local news events.

Breaking news

We define "breaking" news as the first article published covering a specific news event, with the publication date determined by the webpage source code.[18] Out of the 27,139 BBC articles matched to at least one commercial news article covering the same event, 5,047 of the events covered in articles were first published by the BBC (19% of the sample of articles). The commercial providers with the greatest number of "breaking" news articles included the Daily Record (1,285 articles), the Herald (1,167 articles) and the Scotsman (1,152 articles), all three of which are national news titles covering all of Scotland, followed by the Express & Star (860 articles). The average placement of the BBC article for each event is 3.30. Since the average number of providers (including the BBC) covering a news event is 4.18, this suggests that most news events are first covered by local commercial news providers before being picked up by the BBC.[19] This pattern is consistent across primary topic of article as well as low-coverage areas: the BBC article is published first in between 10% (business and the economy) and 24% of news events (weather), and in low-coverage areas the BBC article is usually published second after the commercial provider article has been published.

% of events for which BBC publishes first article

■ All areas   ■ Low-coverage areas   ■ High-coverage areas

---
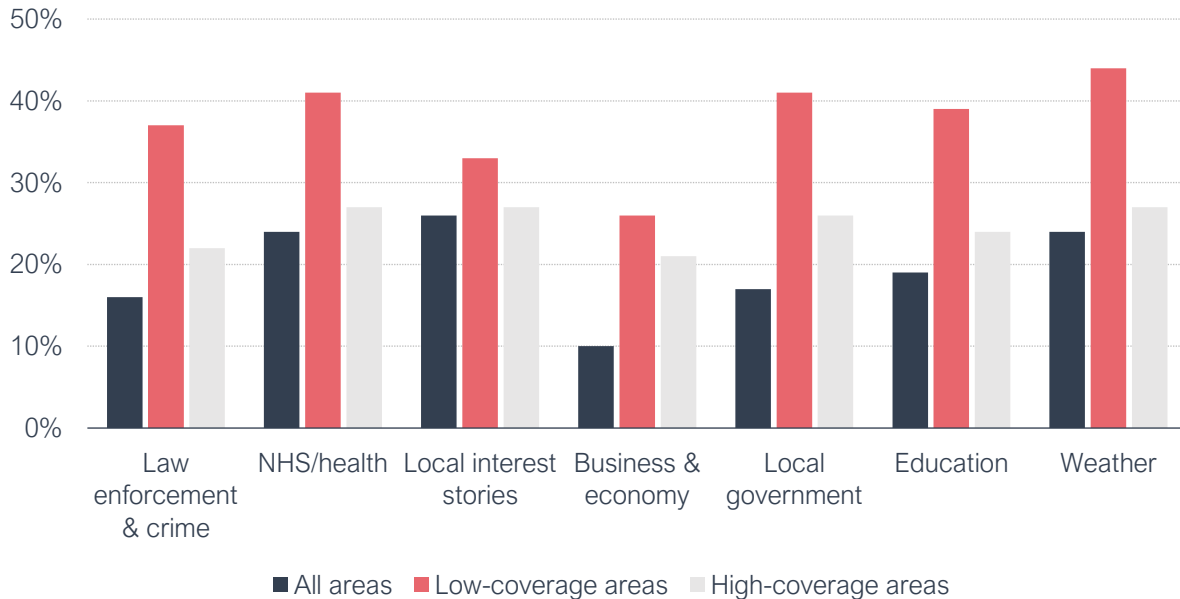
[18] Article webpages include one <time> HTML tag showing when the webpage was created and another <time> HTML tag showing when the article was last updated. In general, these timestamps are automatically generated (instead of being manually entered by authors or webmasters), so we can be reasonably confident in the accuracy of these dates.

[19] When asked about this, the senior BBC executive interviewed for this project emphasised how a greater amount of planning, research and editorial scrutiny made BBC reporting distinctive from reporting by commercial news providers.

Differences in events covered by the BBC and the commercial sector

For each of the four categories of quantitative metrics below, a case study is presented to highlight how differences in these metrics between BBC and commercial articles are displayed in the structure, language and key elements of the text of these articles themselves.

Article length

The average length of BBC online local news articles in our dataset was 355 words, while the average length of commercial online local news articles was 389 words. However, the distribution for this metric was very wide, with the standard deviation of BBC articles at 250 words and commercial articles at 207 words (many articles are fewer than 200 words long). For both the BBC and commercial news providers, articles about community events/issues tended to be the longest (an average of 409 words for BBC news articles and 439 words in commercial news articles), while articles about law enforcement/crime were the shortest on average (317 words for the BBC and 366 words for commercial providers). Articles published in low-coverage areas were broadly similar in length compared those published across the UK as a whole (349 words for BBC news articles and 386 words for commercial news articles).

Average article length

BBC ■ Commercial providers

alma economics
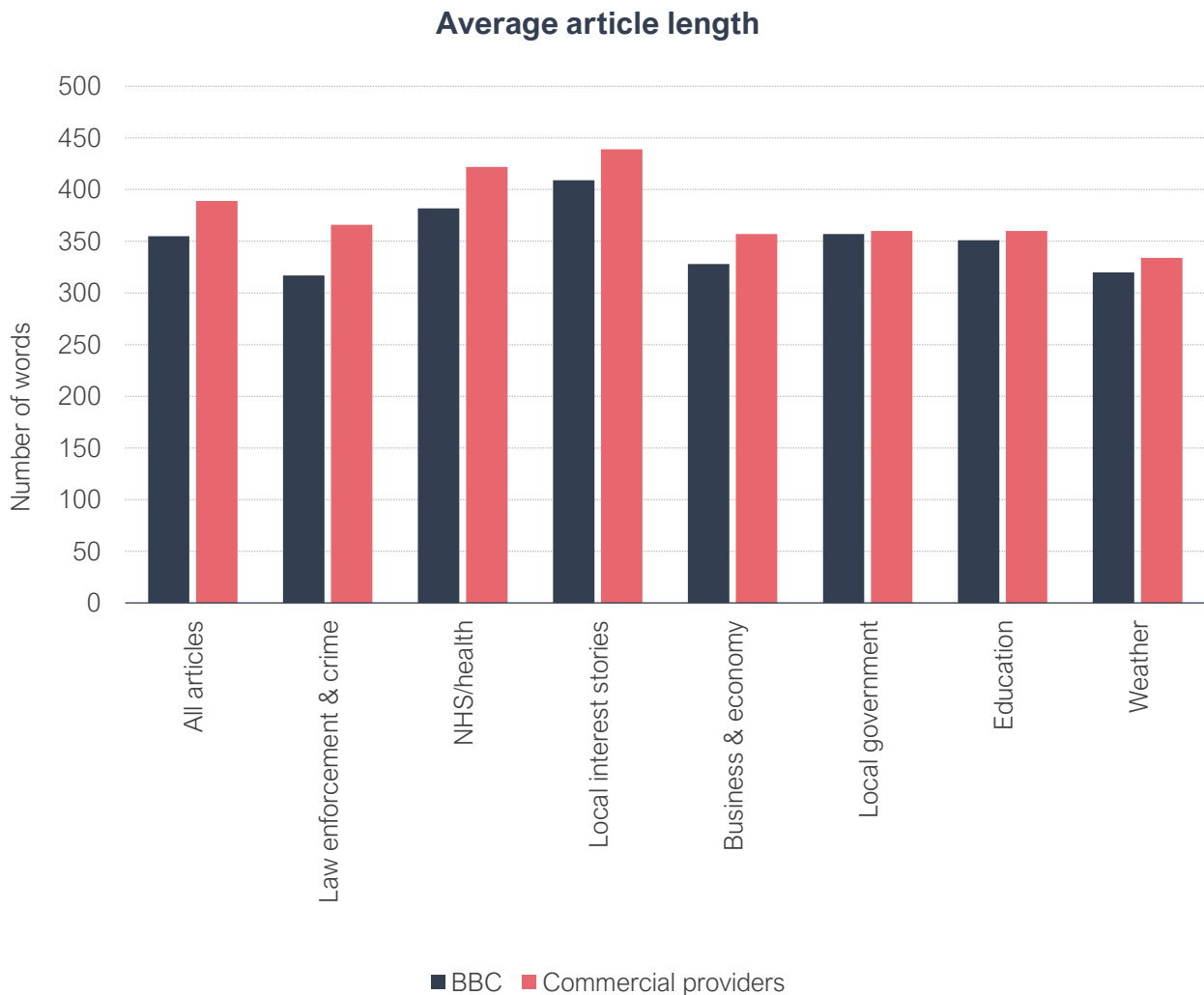
Case study 1:

Arrests after police officer hit by glass bottle

BBC Sussex and The Argus (a local newspaper covering Brighton, Hove and Sussex which is part of Newsquest's local newspaper network) published news articles on how local police responded to reports of antisocial behaviour by a large group of individuals on Hove Lawns and Brunswick Square on 5 June 2021.

The BBC article is 238 words long, while the Argus article is 681 words long. Both articles were published on the same day (6 June 2021, one day after the incident), but the Argus article was published four hours earlier than the BBC article (12:03 compared to 16:19). Both articles reported on the same key elements of the event (a bottle was thrown at a group of police officers, four people were arrested in connection with the incident, the injured officer was treated in hospital).
The primary reason for the observed difference in article length is the length of the quotation from Detective Sergeant Chris Lane of Brighton and Hove Criminal Investigations. Where the BBC article summarised one important line from the quotation, the Argus printed much of the quotation in full, including Det Sgt Lane's request for the public to aid the investigation.

- The Argus: "A member of the crowd threw a glass bottle at officers, which hit one directly in the face and knocked her unconscious. As she fell to the floor, her head also hit the police car, and she was unconscious and injured in what was clearly a very dangerous situation. Thankfully her injuries are not believed to be serious and she is now resting at home after being discharged from hospital. The officer's body camera footage shows that the bottle was thrown from a considerable distance away, and it could have seriously injured any one of a number of police officers or the person they were arresting. To indiscriminately throw a bottle into a group of police officers is deplorable and could have caused serious injury…"[20]

- BBC Sussex: "To indiscriminately throw a bottle into a group of police officers is deplorable and could have caused serious injury."

Other key differences between the two articles included:

- The BBC article title only mentioned a police officer being injured, while the Argus article title also included a reference to arrests made as a result of the injury.

- The BBC article led with a concise summary of the event ("A police officer has been knocked unconscious after a glass bottle was thrown at her and other officers during a disturbance in Hove."), while the Argus article's first sentence instead focused on an image of the glass bottle captured by the officer's body camera ("This shocking image shows the bodycam of an officer moments before she was struck in the face by a glass bottle, knocking her unconscious.")

- The BBC article embedded a Twitter link to the Sussex Police's tweet containing the body camera image (this meant that the image was not shown directly on the article webpage due to Twitter's sensitive media policy), while the Argus article included the body camera image as well as six other images of the scene on the article webpage.

[20] The full quotation from Det Sgt Lane in The Argus is 348 words long.

Text similarity

We used two metrics to quantify the similarity of text between BBC and commercial online local news articles. The first was cosine similarity, which measures the frequency of words and the total number of distinct words used in an article on a 0 to 1 scale (a score of 0 would mean that the two articles used very little of the same language, and a score of 1 would mean that the two articles are identical). The second was soft cosine measure, which also links words that are related in meaning (for example, "road" and "street" would have a low cosine similarity but high soft cosine measure).

The average cosine similarity between an individual commercial news article and BBC article reporting on the same local news event is 0.67 (with a standard deviation of 0.06), and the average soft cosine measure is 0.83 (with a standard deviation of 0.05).[21] The distribution of the text similarity scores tended to be similar for all article primary topics except community events/issues, which had an average cosine similarity of 0.51 and average soft cosine measure of 0.75. For low-coverage areas, the average cosine similarity was slightly higher (0.71), with a similar average soft cosine measure (0.84). Finally, for news events published first by the BBC, the average text similarity was slightly lower (0.62), with an average soft cosine measure of 0.74.

When we carried out a topic/thematic analysis focusing on the main ideas presented across each paragraph of the article, there was a very high degree of overlap in topics/themes: in 97% of news events, the same topics/themes identified in the BBC article were also uncovered in the commercial news articles (in the remaining 3% of articles, the BBC quoted an additional individual or included a short analytical section that did not appear in commercial news articles).

Article text similarity by primary topic
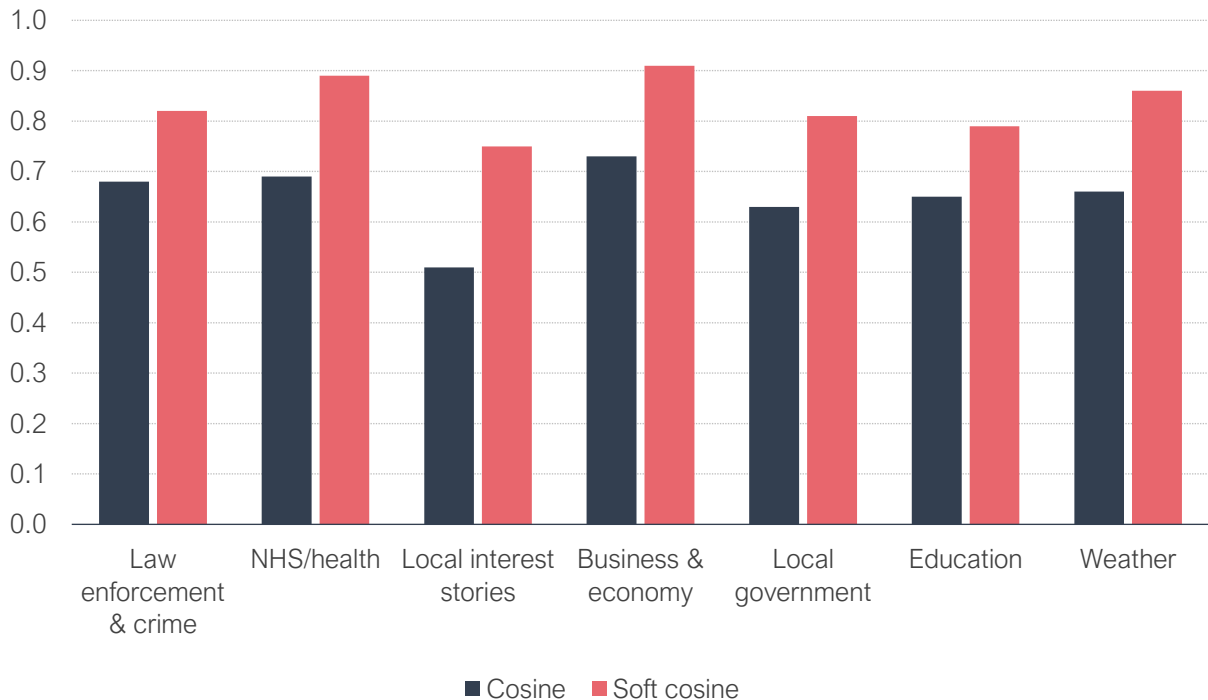
[21] To estimate this figure, we calculated the cosine similarity between the BBC news article and each of the commercial news articles reporting on the same event, then averaged the cosine similarity across all commercial news articles.

We would expect the text and thematic similarity scores to be relatively high if the two articles were reporting on the same news event (which means the same names, places, quotes and general event descriptions would be used). However, an average cosine similarity of 0.67 means that the BBC and commercial news articles are not using duplicate language in their reporting (a score close to 1 would indicate a high degree of overlap in text between articles, while a score closer to 0.5 would suggest some similarities in text but also sections of articles that are meaningfully different in syntax). If applied to the sentence level, a cosine similarity of 0.67 could be interpreted as similar to a score between 3 ("The two sentences are roughly equivalent, but some important information differs/missing)" and 4 ("The two sentences are mostly equivalent, but some unimportant details differ") in the semantic text similarity scale used by Agirre et al. (2013).[22] Instead, this suggests that each article takes a similar approach in how the event is described, with variation in editorial guidelines or style guides possibly driving the differences in specific language observed.

Case study 2:

Jobs lost as Dalkeith home improvement firm collapses

BBC Scotland and three Scotland-wide commercial news providers (The Herald, The Scotsman and the Edinburgh Evening News) reported on the liquidation of the Terry Healy Group, a building and home improvements business based in Dalkeith. Despite differences in the length of these articles (the BBC article was the shortest at 125 words, while The Scotsman article was the longest at 268 words), all three articles from commercial news providers scored highly on cosine similarity and soft cosine measure when compared to the BBC article. This is because all articles included the same key elements:

- A quote from FRP Advisory describing "unsustainable cash flow problems".

- A quote from FRP Advisory discussing "soaring labour and raw material costs".

- Mention of 51 staff made redundant.

- Mention that the Terry Healy Group was founded in 2014 to serve "trade and domestic clients".

- A quote from FRP Advisory partner Callum Carmichael praising the Terry Healy Group's legacy: "Terry Healy Group had grown rapidly into one of the high profile and respected multi-trades home improvement businesses in the east of Scotland."

- A quote from FRP Advisory partner Callum Carmichael on next steps: "Unfortunately, the business has been unable to overcome very serious financial problems and closure was the only option."

Article sources, links and references

For each article in the dataset, the total number of names, quotes, hyperlinks and references to other news providers were calculated. Out of the 86,315 articles by commercial providers in the dataset, around 6% included links to other websites and 3% referenced another news provider, and out of the 57,123 articles by the BBC, 8% included links to other websites and 7% included links to

---

[22] https://aclanthology.org/S13-1004

other news providers.[23][24] When links to other websites were included, these links were almost always the same between the BBC and commercial provider articles. There did not appear to be a specific pattern to the position of links within the article, with links embedded at the beginning, middle and end of articles depending on the structure of the article.

On average, articles included the names of around four individuals and six quotes, though quotes could not be mapped to specific individuals as many quotes were only introduced with personal pronouns (i.e. "he", "she"), with the individual mentioned earlier on in the article (it is also likely that many individuals mentioned in the article do not provide the quotes but instead were taking a specific action described by the news article). In 96% of news events, both the BBC and commercial news articles include the same names (aligning with our finding of relatively high similarity scores between BBC and commercial news articles), and for the remaining 4% of articles, there is no clear pattern by primary topic or geography on whether BBC or commercial news articles include more names (for example, on articles reporting on law enforcement/crime, a commercial news provider may speak to more eyewitnesses, while on articles reporting on community events/issues, the BBC article may include a quote from an umbrella group spokesperson).

Case study 3:

Torbay Children's Services good after years of inadequacy

BBC Devon, DevonLive and Torbay Weekly reported on Ofsted rating Torbay Children's Services "good" after a series of poor and inadequate ratings since 2011. While all three articles shared key elements such as quotes from the Ofsted report and Torbay Council Leader Steve Darling, there were several significant differences. Most notably, the BBC article included direct links to the most recent Ofsted inspection as well as previous reports in 2011, 2016 and 2018, while the DevonLive article included a link to the Torbay Council's webpage on Ofsted reports and the Torbay Weekly article did not include any links at all.

Other differences across the three articles included:

- The BBC and Torbay Weekly articles discussed the previous history of poor performance in Torbay's Children's Services, but the DevonLive article did not. The BBC referenced the resignation of then-director of children's services Carol Tozer in 2011, while both the BBC and Torbay Weekly articles mentioned the transfer of control to Plymouth City Council in 2018 as part of a recovery plan.

- The DevonLive article provided an extensive bulleted list summarising key strengths and areas of improvement mentioned in the Ofsted report.

- The BBC article included one paragraph quoting current Director of Children's Services Nancy Meehan, while the DevonLive article included two paragraphs and the Torbay Weekly article included three paragraphs.

---

[23] Anecdotally, the BBC seems to post slightly more links to charities/other organisations/support resources relevant to the news event, though we did not formally analyse this as it would involve manually reviewing a large number of URLs.

[24] This contrasts with the findings from Cushion (2019), which found that 72% of BBC online local news articles included hyperlinks (albeit with a sample size of 150 articles). We also do not include general (not article-specific) internal links in our analysis; Cushion (2019) similarly found that 84% of all hyperlinks in their sample of local news articles were internal.

The BBC and DevonLive articles included an image of Torbay Town Hall credited to Local Democracy Reporter Joe Ives. The BBC article also included a header image from PA Media depicting a child holding her head in her hands. The Torbay Weekly article did not include any images in common with the other two articles, instead posting a header image of children credited to Torbay Council (though not appearing on any Torbay Council webpage) as well as images of Nancy Meehan and Cabinet member for children's services Councillor Cordelia Law, who was also quoted in the article.

Article images

All articles in our dataset had at least one image included in the article webpage. For 72% of news events in the dataset, the BBC and all commercial providers covering the event included the same image (or identical set of images). For 23% of news events, there were at least three unique images used across the BBC and commercial provider articles, suggesting that each provider used a distinct image in the article webpage.[25] Our analysis only examined whether images were identical, as estimating the subject or similarity of images is a much more computationally intensive process.

Case study 4:

Disused golf course 'rewilding' only due to faulty mower

BBC East Midlands and LeicestershireLive both reported on a walker who had e-mailed Leicester City Council to thank them for rewilding the Western Golf Course and was disappointed to learn that the grass had not been cut due to a broken mower. Because reporting on the news event was largely reliant on a single source (a highly-engaged local resident named Steve Walters), the images included in the article reflect this. Both articles included a set of four images: two images depicting the golf course after grass had been allowed to grow, one image after the grass had been cut and one image of a buzzard chick (representing local wildlife). All four images were credited to Steve Walters.

In addition, both articles included the following elements:

- A series of quotes by Steve Walters, who had campaigned for the golf course to be kept as a wildlife haven.
- The Leicester City Council's response in full: "Sorry to disappoint you, but there have been no changes to our mowing policy, our mower has broken down and we have had difficulty in sourcing parts."
- A statement by a council spokesperson on its maintenance responsibilities with respect to the golf course.

However, there were also a number of key differences between the articles:

- While both article titles mention the golf course "rewilding" or being "returned to nature",
- the BBC article title noted the role of a faulty mower while the LeicestershireLive article title only stated that the council had responded to an e-mail.

---

[25] For the 23% of news events with at least three unique images used in articles, it would be interesting to examine the source of these images. It is plausible that a single organisation may be taking a set of official photographs distributed to local press (for example, a council may publish a set of official photographs when a new leisure centre is opened) and news providers are choosing different photographs to use. Alternatively, each provider covering the event may send their own staff photographer, so the different images would come from different sources.

- The BBC article included five paragraphs with quotes from Steve Walters, while the LeicestershireLive article included 10 paragraphs (some of which were paraphrased by the BBC article).

- While both articles cited "grass on the bus shelters" as an example provided by Steve Walters of a "headline grabber" with minimal impact, the BBC article included the link to the Leicester City Council webpage describing the project in question, while the LeicestershireLive article did not.

4. Interviews with media organisations

To provide a broader understanding of the local news journalism ecosystem as well as the risks and impacts of content duplication, we carried out 13 interviews with editors and assistant editors at commercial news providers (all of whom were part of multi-title publishers) and one interview with a senior BBC executive.

Editorial process

Interviewees generally described similar editorial processes for their respective organisations: journalists tended to be assigned to specific geographic regions (patches), with news events identified from media invitations, social media, fieldwork or community-submitted tips/suggestions. Larger news providers had separate journalists specialising in sports, local government, business or courts, though dedicated investigative teams were rare unless coordinated at the national level from the broader multi-title publisher. The BBC relied on similar approaches to identify news events, though with greater collaboration across other BBC media (radio and TV). While the BBC set out both to inform and entertain viewers in its news coverage, in general its online news provision focused more on educating and informing audiences compared to radio and TV, which were more balanced across the two aims.

When asked to describe how they sought to differentiate their title from other local news providers, interviewees emphasised the local nature of their reporting and their relationship with their local community. Some mentioned they worked for titles with extensive histories (dating back to the 19th century or earlier) and this provided them with name recognition and a reputation for credibility/trustworthiness among local audiences. Others noted that audiences appreciated their effort to place national news in a local context and highlight local issues or perspectives not addressed by larger publications due to their broader reach. Most interviewees did not perceive any challenges or risks in determining which events to report on. They largely acknowledged that (i) the ease of accessing different news websites meant that most of their audience read both their own titles and the BBC's online local news reporting, and (ii) major local events would most likely be covered by all providers in the community. In practice, interviewees felt they could better differentiate themselves by paying attention to hyperlocal issues or adopting a distinct local voice.

Relationship between the BBC and commercial providers

When asked about their relationship with the BBC, almost all editors characterised this relationship as minimal or non-existent. For most interviewees, their primary interaction with the BBC was through the Local Democracy Reporting Service. All interviewees who were currently working with (or had previously worked with) this service praised their experience with the news provider's Local Democracy Reporter(s), as this provided them the opportunity to expand coverage of public interest news (such as local council meetings) where they would otherwise not have been able to do so due to lack of resources. However, interviewees raised issues around the tendering process for Local Democracy Reporter contracts, with providers placed in a difficult situation if the contract was won by a competing local news provider.

Potential risk and impact of content duplication

While interviewees read the BBC local news website out of personal interest and occasionally were invited to participate in BBC radio or television broadcasts, no interviewee stated that they specifically considered the events the BBC local news websites had reported or were likely to report on when deciding what their own journalism team should cover. In addition, almost all editorial staff interviewed stated they felt that content duplication with BBC local news websites was not considered a risk nor did it financially impact their organisation in any way, although it is important to note that interviews did not include commercial or other non-editorial staff and the sample size was very small (due to low e-mail response rates).[26] A common theme emphasised across interviews was that the BBC local news teams had to cover a larger geographic area due to limited resources and therefore could not provide as much coverage of local events, with one editor estimating that his team would publish 15 to 20 articles on any given day while the BBC local news website would only publish four to six articles. They felt that the BBC had instead chosen to focus on television and radio media, and these did not directly compete with online local news provision (although this research was conducted before the BBC's announcement in October 2022 to reprioritise £19 million from broadcast services towards online and multimedia production to keep pace with changing audience expectations). Instead, more than half of interviewees identified online aggregators (including Facebook and Google) as having much more of an impact on their financial situation. These interviewees mentioned that at least half of their web traffic came from clicks through Facebook and Google (one interviewee said this figure was as high as 80%), and frequent changes in the algorithm around what news was displayed to users meant that web traffic fluctuated on a monthly or even weekly basis independent of news cycles or number of articles published.

Despite the low level of perceived risk or impact posted by content duplication, four interviewees expressed concern about the differences in business models between the BBC and commercial providers. In particular, they felt it was difficult competing with a free source of local news in an environment with shrinking resources, even if they sought to differentiate themselves from the BBC in their coverage. Interviewees were worried that the BBC might choose to expand or invest greater resources in online local news coverage, which would bring them in more direct competition with commercial providers. At the same time, the remaining seven interviewees expressed a desire for greater collaboration with the BBC in local news coverage, in particular developing longform pieces or carrying out investigations. Interviewees also generally favoured expanding the Local Democracy Reporting Service to cover court reporting.

---

[26] This view may not align with those from stakeholders in the wider sector: for example, the News Media Association (NMA) expressed concern that the BBC "poses a direct threat to local journalism both through its ambitions to launch local news services…and its use of their content without proper attribution" in response to a 2021 consultation on how Ofcom regulates the BBC.

5. Discussion of findings

In its six-year blueprint *The BBC Across the UK* published in 2021, the BBC laid out ambitious objectives to transform the future portfolio of BBC local services to "touch more lives, in more communities, more often", aiming to cover the stories that matter most to local audiences and effectively represent different voices and perspectives. Through its Local News Partnerships, the BBC has sought to provide coverage of local democratic institutions and processes (through the News Hub, Shared Data Unit and Local Democracy Reporting Service) that would not have otherwise been covered. Building on the six-year blueprint, in an announcement made in October 2022, the BBC also committed to (i) creating 11 investigative reporting teams across the country, (ii) implementing multimedia news operations across radio, TV and online local news teams and (iii) boosting online local news provision by reinvesting in online and multimedia production, with the broader aim to "reshape its local services across England to deliver greater online impact and provide more original journalism". However, this strategy has come into conflict with news media industry stakeholders, with NMA chief executive Owen Meredith expressing concern that the blueprint would "create a democratic deficit by putting local publishers out of business".[27] To date, limited analysis testing these dynamics has been undertaken. This report has set out an innovative, large-scale quantitative methodology to answer one specific question that might shed further light on this issue: by collecting and analysing the text of 143,438 local news articles, what can be concluded about differences in local news coverage and the differences in type of coverage provided by the BBC and the commercial sector?

Taken together, our quantitative analysis of online local news articles published by the BBC and commercial providers, our review of online reach data for individual news titles from JICREG and our discussions with regional news editors provide evidence that the BBC does not serve as a one-to-one substitute for commercial providers in online local news provision. Content duplication between the BBC and commercial providers in online local news does exist, but in a way that is consistent with multiple providers covering some of the same news events, requesting quotes from the same sources and describing the facts or key features of the news events in a consistent manner. While the BBC tends not to be the first publication to cover a local news event, BBC articles tend to be shorter, relying more on paraphrasing and synthesis rather than lengthy quotations from sources; the language and text used is similar but does not duplicate commercial provider articles. More importantly, the BBC rarely covers local news events not also reported on by commercial providers, but commercial providers report on a much larger number of local events than the BBC (which holds true across a broad range of article topics), a claim supported by our quantitative analysis as well as interviewees' understanding of their local news media ecosystems. While we could not consider all local news articles published by commercial providers due to challenges around data collection, the large number of articles in our dataset and our use of three different data sources means we can be reasonably confident in our estimate of the lower bound of the distinctiveness of the commercial news sector's online provision of local news.

---

[27] More specifically, in response to a 2021 consultation on how Ofcom regulates the BBC, the NMA stated, "The BBC proposals regarding 'BBC News across the UK' and 'Supporting local communities' will likely undermine local news and stifle press plurality by exacerbating current problems in BBC conduct in these areas", citing the BBC's "taxpayer-funded resources that no local operator could hope to match" and its existing "overwhelming brand dominance in news supply".

While our research has noted both concerns raised by interviewees about further expansion of the BBC into online local news coverage and their views on scope for greater collaboration, our underlying methodology is less well-suited to exploring the extent to which the BBC complements commercial providers in online local news provision. Two key stakeholders beyond the scope of this research but critical to understanding the BBC's relationship to commercial providers in online local news provision are news audiences and commercial staff, and engagement with these two stakeholders provide an opportunity to explore this area further:

- Audiences visit news websites from a broad range of sources, such as Google Search, Facebook News, e-mail newsletters, retweets by celebrities, links texted by colleagues, etc. (Ofcom 2022). The online news experience and patterns of news consumption may not be consistent across news events or even times of day.[28] If individuals differ in how they choose to access and consume local news, this means they might also differ in the relative value they place on different providers: one person might only read articles published by one provider, while another person might learn about a news event from one provider and search for the same event across other providers. Building off existing surveys by Ofcom and the Reuters Institute for the Study of Journalism, one potential area of research might be to explore how audiences consume local news content and how consumption patterns relate to their perceptions and beliefs about local news providers. This could take place through in-depth interviews and focus groups with consumers, diaries of online behaviour or a browser plug-in that tracks navigation across individual news websites.

- Commercial staff at news providers are likely to have an in-depth understanding of how online news consumption might be influenced by different business models adopted by providers (free with ads, paywall with no free articles, paywall with a limited number of free articles, microtransactions, etc.) Drawing on data provided by commercial staff, areas for further research could involve estimating the price elasticity of demand for online news based on existing financial data, measuring consumers' willingness to pay for online local news through a field experiment/survey or analysing the types of consumers that choose to pay for online local news in an environment with freely-accessible public news media (similar to Fletcher and Nielsen 2016). In addition, commercial staff may have insights on which articles audiences are more likely to engage with, how article reach is impacted by multiple providers reporting on the same news event and the financial impacts of this content overlap. These insights are relevant because they could shed light on how the BBC potentially impacts non-paying online traffic that commercial providers seek to monetise as ad revenue. Understanding how providers vary in their ability to reach audiences in different communities/regions or with different demographic characteristics has implications for the broader objective of creating a more diverse and plural media environment in the UK.

---

[28] For example, see Vermeer et al. (2020), which tracked how Dutch online news consumers navigated a high-choice media environment and the role of context and content features in driving individuals' news experiences.

Appendix A: Methodology details

Calculation of metrics

For text and topic similarity, we first calculate a similarity score for each commercial article compared to the BBC article on the same local news event, then average the similarity score across all commercial articles reporting on the same news event (so each event had one average similarity score). The remaining metrics are estimated in one of two ways: (i) the average metric across all articles, and (ii) the difference between the BBC and average metric across all commercial articles on the same news event, averaged across all articles (where possible). As a hypothetical example:

- News event A was reported on by the BBC and Commercial Providers A, B and C.

- For similarity in language, the similarity score is as follows:

Article pair	Similarity score
BBC – Commercial Provider A	0.6
BBC – Commercial Provider B	0.7
BBC – Commercial Provider C	0.8
Average similarity score	**0.7**

- When reporting an average similarity score across all articles, we include the 0.7 as the similarity score for news event A.

- For average text length, the individual article metrics are as follows:

Article	Article length (words)
BBC	250
Commercial Provider A	200
Commercial Provider B	270
Commercial Provider C	400
Average article length (all articles)	**280**
Average article length (commercial articles)	**290**

- In addition to the average article length across all articles (280 words), we can report the difference between the BBC article length and average length of articles published by Commercial Providers A, B and C (290 – 250 = 40 words).

Text similarity

We calculate the similarity between two articles using the cosine similarity metric. Because the cosine similarity measures the cosine of the angle between two vectors, we first need to convert sentences to vectors using term frequency-inverse document frequency (TF-IDF). The TF-IDF value increases proportionally to the number of times a word appears in an article and decreases based on the number of articles in our dataset containing that word. Then, for two vectorised articles A and B with dimension n, we can calculate

$$similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

The soft cosine measure is calculated based on a matrix S indicating similarity between features:

$$soft\_cosine(a,b) = \frac{\sum_{i,j}^{N} s_{ij} a_i b_j}{\sqrt{\sum_{i,j}^{N} s_{ij} a_i a_j} \sqrt{\sum_{i,j}^{N} s_{ij} b_i b_j}}$$

where $s_{ij} = similarity(feature_i, feature_j)$. Features can either be individual words, but we have also tested n-grams and syntactic n-grams for robustness.

Classification of articles to news events and primary topics

Our primary method of classifying articles to news events and primary topics is zero-shot learning. This is a type of problem in machine learning where a model needs to predict the class (the news event or primary topic) that an article belongs to without having observed any examples of classes before. In lieu of labelled data, the auxiliary information we provide to the model is textual description: primary topics (such as "law enforcement/crime") are augmented with definitions and other descriptive text.

For this project, we use a pre-trained MultiNLI (Natural Language Inference) sequence-pair classifier based on the open-source Bart model developed by Facebook. In general, NLI models work by considering two sentences, a premise and a hypothesis. The classifier then tries to determine whether the hypothesis is true (entailment) or false (contradiction) given the premise. NLI models work with sequence-pair classification: the premise and hypothesis are both fed into the model together to create a single classification head. For example, to check if an article can be classified as "law enforcement/crime", we pass the entire news article as the premise and the label "law enforcement/crime" as the hypothesis. We then check if the NLI model predicts that the premise entails the hypothesis, which means the label is true.

More formally, if we have a sequence embedding model Φ_{sent} and a set of possible class names C, we classify a given sequence x as follows:

$$\hat{c} = \arg max_{c \in C} \cos(\Phi_{sent}(x), \Phi_{sent}(c))$$

where $\cos$ is the cosine similarity.

We can then increase the accuracy of our model by learning an additional least-squares projection matrix based on the embeddings of available labels to their corresponding data embeddings. To avoid overfitting, we add L$_2$ regularisation based on the identity matrix rather than the norm. In other words, we have a prior belief that the best classification for our data is $\Phi(X)I_d = \Phi(X)$, and we update that belief after adding more labelled training data.

Appendix B: Interview topic guide

1. Could you please start by telling me about yourself and your experience with local news?

2. Can you tell me more about your organisation's editorial process?

    a. How are staff and resources allocated across events/beats?

    b. What challenges or risks do you face in determining which events to cover?

    c. How do you seek to differentiate yourself from other local news providers?

3. What is your organisation's current relationship with the BBC?

    a. How does the BBC's local news content impact the decision-making process around which events your organisation covers?

    b. How has the relationship changed since the last Charter Review in 2016, if at all?

    c. Do you feel there are any differences between how your organisation and the BBC cover local news?

    d. To what extent do you think your audience perceives these differences?

4. To what extent does duplication of content with the BBC considered a risk?

    a. To what extent does it have a financial impact on your organisation? How would you describe the financial impact on your organisation?

    b. How would you characterise the overall impact of the BBC's online local news on your organisation?

5. Can you tell me more about what role the BBC should play in local news provision going forward?

    a. How can the BBC better support local news providers?

6. Those are all the questions I have for you. Is there anything we haven't talked about that you would like us to know?

alma economics