universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Detecting Latent Linguistic Traces of Phantom Borders in Large Language Models

verfasst von | submitted by

Annika Franziska Süß

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 856
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Kartographie und Geoinformation
Betreut von | Supervisor:	Univ.-Prof. Dr. Krzysztof Janowicz

Abstract

Phantom borders are historical borders that continue to drive local divergencies long after their formal abolition. Although traditional spatial analysis assumes that proximity dictates similarity, these borders create sharp discontinuities in demographic, economic, or political characteristics. Although census data easily captures such divides, the subtle, context-dependent, and cultural variations caused by these borders remain hidden in local language and, more specifically, in how people talk about places, events, and everyday topics. This makes them difficult to detect. At the same time, large language models (LLMs) are increasingly used for spatial tasks and research applications. The question of how artificial intelligence (AI) systems encode and interpret geographic phenomena, such as phantom borders, arises. Investigating this is particularly important because LLMs are trained on large corpora of human-generated text, which may implicitly encode historical, cultural, and geographic biases. This creates the research gap of whether these latent geographic and cultural signals are embedded in language and if LLMs can detect them, while also reflecting them in their reasoning and explanations. The thesis addresses this research gap and aims to uncover these latent linguistic traces, aiming to detect whether and to which extend phantom borders can be visible in everyday discourse and local language. As a case study, the thesis focuses on the historical division between East and West Germany following World War II. Even though the political border disappeared, the differences between East and West Germany persist in multiple parameters. To examine whether similar traces exist in local language and can be detected through LLMs, a dataset of 11,856 geolocated newspaper articles from East and West Germany is created. All explicit geographic references are removed from the texts using LLM-based preprocessing to isolate pure semantic cues. GPT-5.1 is used to predict the geographic origin of these articles and provide justifications. An embedding model analyzes linguistic cues without additional reasoning layers on top. The results show that LLMs can reasonably predict the geographic origins with the accuracy of 72% with varying extent, while justifications remain ambiguous. Embedding-based analysis also reveals semantic differences between linguistic patterns associated with the former East and West Germany. In particular, the model tends to default to predicting West Germany when Eastern linguistic cues are weak. These findings highlight the potential of LLMs for complex spatial analysis and the critical challenges regarding training data bias and AI alignment.

Kurzfassung

Phantomgrenzen sind historische Grenzen, die auch lange nach ihrer formellen Aufhebung weiterhin zu lokalen Unterschieden beitragen. Während die traditionelle Raumanalyse davon ausgeht, dass Nähe Ähnlichkeit bedingt, führen diese Grenzen zu starken Diskontinuitäten bei demografischen, wirtschaftlichen oder politischen Indikatoren. Obwohl Zensus-Daten solche Trennlinien leicht erfassen, bleiben die subtilen, kontextabhängigen und kulturellen Unterschiede, die durch diese Grenzen verursacht werden, in der lokalen Sprache verborgen und darin, wie Menschen über Orte, Ereignisse und alltägliche Themen sprechen. Dies macht es schwierig, sie zu erkennen. Gleichzeitig werden large language models (LLMs) zunehmend für räumliche Aufgaben und Forschungsanwendungen eingesetzt. Es stellt sich die Frage, wie Systeme der künstlichen Intelligenz (KI) geografische Phänomene wie Phantomgrenzen kodieren und interpretieren. Die Untersuchung dieser Frage ist besonders wichtig, da LLMs auf großen Korpora von menschlich generierten Texten trainiert werden, die implizit historische, kulturelle und geografische Biases enthalten können. Dies führt zu der Forschungslücke, ob und zu welcher Stärke diese latenten geografischen und kulturellen Signale in der Sprache eingebettet sind und ob LLMs sie erkennen und gleichzeitig in ihren Schlussfolgerungen und Erklärungen widerspiegeln können. Die Arbeit befasst sich mit dieser Forschungslücke und zielt darauf ab, diese latenten sprachlichen Spuren aufzudecken, um festzustellen, ob Phantomgrenzen im Alltagsdiskurs und in der lokalen Sprache sichtbar werden können. Als Fallstudie konzentriert sich die Arbeit auf die historische Teilung zwischen Ost- und Westdeutschland nach dem Zweiten Weltkrieg. Auch wenn die politische Grenze verschwunden ist, bestehen die Unterschiede zwischen Ost- und Westdeutschland in vielfältigen Parametern fort. Um zu untersuchen, ob ähnliche Spuren in der lokalen Sprache existieren und durch LLMs erkannt werden können, wird ein Datensatz mit 11.856 geolokalisierten Zeitungsartikeln aus Ost- und Westdeutschland erstellt. Alle expliziten geografischen Verweise werden mithilfe einer LLM-basierten Vorverarbeitung aus den Texten entfernt, um reine semantische Hinweise zu isolieren. GPT-5.1 wird verwendet, um die geografische Herkunft dieser Artikel vorherzusagen und Begründungen zu liefern. Ein Embedding-Modell analysiert sprachliche Hinweise ohne zusätzliche Schlussfolgerungsebenen. Die Ergebnisse zeigen, dass LLMs die geografische Herkunft mit einer Genauigkeit von 72% vorhersagen können, während die Begründungen jedoch mehrdeutig bleiben. Die Analyse der Embeddings deckt zudem semantische Unterschiede zwischen sprachlichen Mustern auf, die mit der ehemaligen deutschen demokratischen Republik und der Bundesrepublik Deutschland assoziiert werden. Insbesondere neigt das Modell dazu, standardmäßig Westdeutschland vorherzusagen, wenn die sprachlichen Hinweise auf Ostdeutschland schwach sind. Diese Ergebnisse unterstreichen das Potenzial von LLMs für komplexe räumliche Analysen sowie die entscheidenden Herausforderungen hinsichtlich der Verzerrung von Trainingsdaten und der Ausrichtung der KI.

Contents

List of Figures

List of Tables

Acknowledgments

I want to express my gratitude to my research group, consisting of Krzysztof Janowicz, Mina Karimi, Yingjing Huang, Zilong Liu, and Songlin Wang, at the University of Vienna, Department of Geography and Regional Studies, for their continuous support, valuable feedback, and insightful discussions throughout the development of this thesis. Their constructive comments and collaborative environment greatly contributed to the refinement of my ideas and the overall quality of this work.

1 Introduction

Spatial dependence is a core concept in spatial information theory. It describes the spatial autocorrelation between two independently measured variables located in a defined geographic area. Moran's I or Getis-G are well-known global measures to describe this kind of relationship between variables in space. Local Indicators of Spatial Association (LISA) are used on a local scale to provide further insights into the dependence of values across a study area (Anselin, 1995). As described by Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, 236). This is the reason why researchers can make assumptions about places and phenomena based on the surroundings (Lapp et al., 2025).

A disruption of this concept can appear in areas with administrative boundaries, such as country or state borders. Two geographically near locations can be rather different when lying on different sides of such borders. Inside an administrative region, similar cultural, social, and linguistic patterns as well as personal identities can emerge and develop (Geyer, 2025). While such effects are relatively easy to account for when working with clearly defined, current administrative boundaries in spatial data, the situation becomes considerably more complex when so-called "phantom borders" are present. Phantom borders are former political borders that are no longer present. Not all historical borders are phantom borders. The crucial distinction is that phantom borders continue to manifest in contemporary data (Von Hirschhausen et al., 2019). A classic example is the former division between East and West Germany. Although this border stood for over four decades before the 1990 reunification, contemporary Germany remains far from completely homogeneous. As evidenced by demographic data, election patterns, economic factors, and religion distributions, the historical border reappears in contemporary spatial patterns (Holtmann, 2020). This indicates that the border has left a trace and continues to shape the socioeconomic realities and daily lives of the population (Geyer, 2025). Identifying these differences in structured data is straightforward through statistical and geospatial analysis, such as hotspot detection.

With the development of the Internet, large amounts of unstructured data are readily available. Unlike structured datasets, which typically capture measurable outcomes such as demographics, election results, or economic indicators, textual data can encode subtle social, cultural, and linguistic patterns that persist across regions. The ways people communicate about their region, describe local events, or express regional identity provide very deep insights into geographic space. This is particularly relevant for studying phantom borders, where historical boundaries continue to influence everyday language, local narratives, and regional identities even long after the political border has disappeared. Textual data allows researchers to capture these subtle signals, which are often diffuse, context-dependent, and embedded in natural language, providing insights that traditional spatial statistics alone cannot reveal. Therefore, unstructured texts should be used to further investigate such phenomena as phantom borders. Different methods are required for textual data analysis, since traditional geospatial analysis cannot be applied to such type of data. Foundation models (FMs) are large artificial intelligence (AI) models that are (pre-)trained on massive datasets and can perform general tasks. They can be adapted to perform a variety of more specialized downstream tasks (Caballar and Stryker, 2024). Generative AI (GenAI) models and more specifically large language models (LLMs) are FMs specifically trained to analyze and generate natural language, providing a new tool to work with unstructured textual data (Schneider et al., 2024).

There are multiple sources that provide the local languages used by humans, such as social media posts, newspaper articles, travel blogs, and webpages. Understanding phantom borders requires a large amount of geolocated textual sources to anchor unstructured societal narratives to physical spaces and thereby reveal latent territorial divides. Moreover, maintaining

the diversity within the dataset is crucial for capturing localized variations and reflecting the broader societal spectrum. This is often a problem for social media posts, as only a smaller group of often younger people use the platforms, while only sharing very filtered information, not capturing everyday life (Mellon and Prosser, 2017). Therefore, newspaper articles seem a good fit for the study objective. Especially due to the increase in web news articles from every part of Germany, which are freely available. The huge variation in topics discussed in the news and also the geographic scale at which newspaper companies act could provide a very diverse data source.

During a first set of exploratory experiments, local German newspaper articles were collected manually from web pages to pass them to OpenAI's ChatGPT. The AI agent provided mostly accurate predictions about the article's origin from East or West Germany. To ensure that these conclusions about the location were not drawn through place names in the texts, all geographic references were removed beforehand. This indicated that a prediction was made only due to cues in the language. This first brief experiment suggested that linguistic traces from phantom borders might still be present in these articles and that AI systems could detect them. This was also noted in previous research on geo-indicativness of non-georeferenced text (Adams and Janowicz, 2021). This sparked further interest in pursuing this topic and ultimately the thesis.

Emerging research indicates that LLMs may effectively model human spatial cognition (Yamada et al., 2024). Understanding the extent of this spatial awareness is essential because the public increasingly relies on LLMs for complex spatial applications such as tourist sight recommendation and demographic analysis. Beyond practical applications, LLMs also offer significant research opportunities as human surveys are time-consuming and costly, whereas LLMs can produce large amounts of data in a short time. Therefore, they can provide a scalable method to explore spatial patterns. However, relying on these models for spatial research requires them to capture the latent geographic and cultural information embedded within local language, and not only explicit geographic labels. This matters as LLMs can change their outputs when the user's language changes, indicating a sensitivity toward languages and the geographical origin of the person (Li et al., 2024). Phantom borders work as an ideal text case, as they represent historical boundaries that no longer exist politically but continue to shape social, cultural, and linguistic behavior. Detecting these differences in unstructured text not only investigates the spatial reasoning capabilities of LLMs but also allows researchers to study how AI systems encode and interpret latent, context-dependent cultural knowledge. The research gap addressed in this thesis is therefore twofold. The first part of the **research gap** addressed in this thesis is: despite previous research on phantom borders in structured data, it remains unclear how these historical boundaries leave detectable traces in unstructured textual data. The second part of the **research gap** is: it is currently unknown whether LLMs, when applied to unstructured text, can capture these latent linguistic and cultural patterns and how they reflect them in their reasoning and explanations. Therefore, this study aims to fill this **research gap** on phantom borders in local, unstructured language and how LLMs implement this phenomenon. **To this end, the following three research questions are answered in this paper**:

- To what extent can state-of-the-art LLMs capture latent linguistic traces of phantom borders? To limit the scope of the thesis, the focus will lie on the former inter-German border as a case study.
- Are the semantic differences of local language visible in the textual embeddings, i.e., without reasoning models on top?
- To what extent do the justifications of GenAI systems (e.g., LLMs) match their latent representations of these models, i.e., if a phantom border is correctly predicted, does the reasoning explanation of the language agent support the prediction?

The proposed workflow to evaluate the connection of phantom borders and their implementation in LLMs can be seen in Figure 1. The first phase of the thesis is the preparation of a newspaper database through several steps, creating a pipeline leading to the final databases that are used for further experimentation. The second phase is the execution of three experiments, each answering one of the research questions. The last phase is the application of several evaluation metrics to present the results.

This thesis proceeds as outlined: It starts by reviewing the related work in Section 2 and describing the study area, Germany, in which the analysis takes place in Section 3. Then, the applied methods and the setup of the experiments are explained in Section 4. In Section 5, the different results are presented, and subsequently, the results are discussed in Section 6. Section 7 proposes potential future work and limitations of the thesis, and Section 8 concludes all findings.

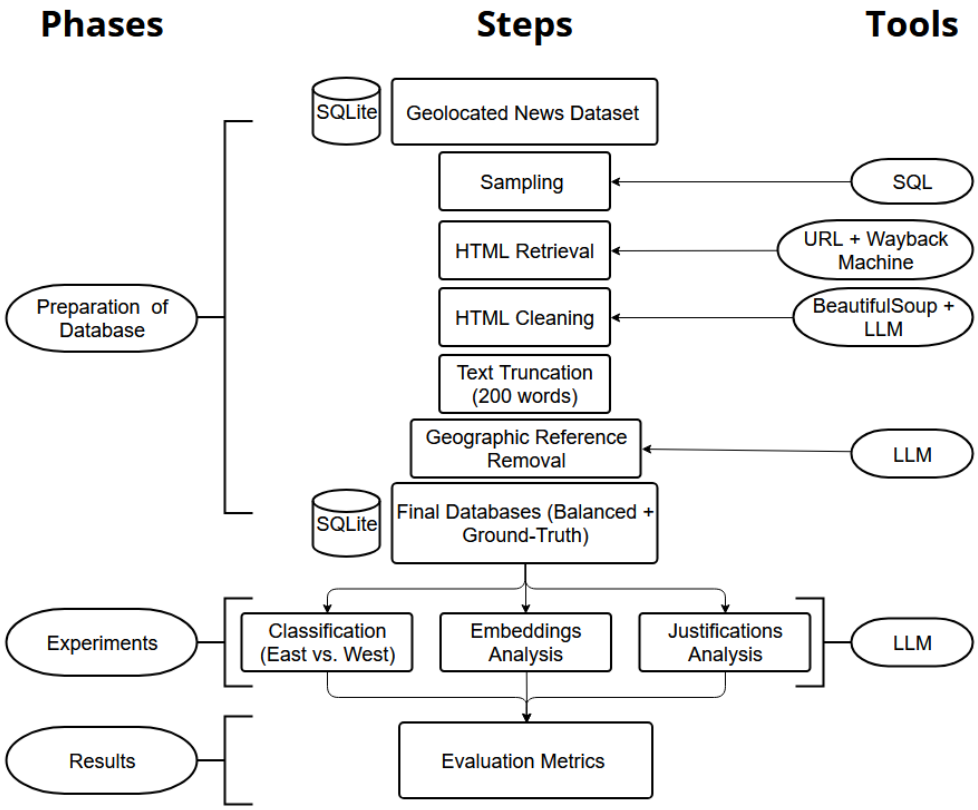


Figure 1: Overview of thesis workflow.

2 Literature review and related work

This related work section is divided into two parts, as the thesis combines two major study areas. First, starting with an overview of previous research on phantom borders in Section 2.1. Here, the concept of phantom borders is first defined and explained in general to build a ground understanding of the research topic of the proposed study. Upon this baseline, previous research on phantom borders, including different study areas and methodological approaches, is introduced. At the end of the section, more precise studies focusing on linguistic differences in Germany are presented to describe existing research on linguistic traces. To capture the second research field underlying this study, the existing literature on topics related to LLMs is presented in Section 2.2. This literature review starts with a short introduction to LLMs and their incorporation of spatial knowledge. Additionally, it is argued why LLMs are relevant in the field of geography and, more specifically, spatial data science. The thesis highlights the known weaknesses and risks associated with LLMs to demonstrate why these FMs are worth studying.

2.1 Relevant literature for phantom borders

Phantom borders exist in many places in the world (e.g., in Poland, Romania, Russia, and Chile) and are an important phenomenon in spatial data science due to their impacts on concepts such as the vague cognitive regions, spatial dependence, and spatial heterogeneity (Kolosov, 2020). The most prominent definition and concept of phantom borders was established by Beatrice von Hirschhausen in 2015 in the German language and later published in English. Von Hirschhausen et al. (2019) proposed them as former political demarcations or territorial divisions that influence space even though they have vanished. Their impact could be seen in architecture and rural settlement patterns, election results, and population statistics. Phantom borders were introduced as a metaphorical term, as they do not politically exist, and can be used as a heuristic tool, helping people understand regional differences. Figure 2 shows the previous borders in Europe, creating the image of a continent shaped by continual territorial changes. These constant changes of territories over time, and especially in Eastern Europe around the Second World War, laid the foundation for the emergence of several phantom borders in this area.

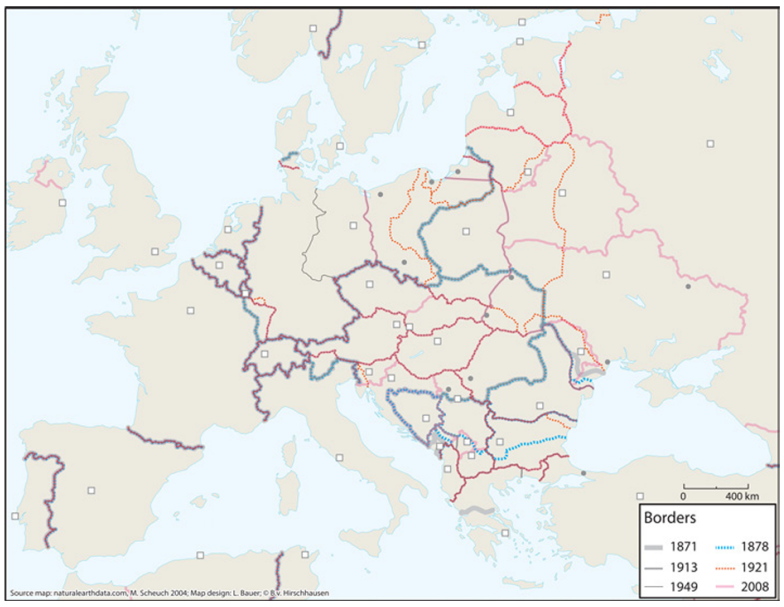


Figure 2: Former borders in Europe. Image taken from Von Hirschhausen et al. (2019, 370).

Von Hirschhausen et al. (2019) focused more on the spaces and places that were created rather than on the border itself, asking the questions of how and why these phantom effects emerged, influenced each other, and eventually vanished. In the example of Eastern Europe, they introduced the thinking that these effects were not fixed or deterministic. Researchers should focus on understanding space as *relational*. The goal was not to promote imperial nostalgia or claim that history strictly determines the present. The goal was to study how history influences today's world.

A study by Šimon (2015) on election results focused on the Sudety region in Czechia, an area with two major population shocks around the Second World War and changes in country borders. Cartographic techniques such as natural breaks or quantiles were used to make the phantom border visible. Further, correlation measures and grouping statistics on spatial units were used to quantify the effects over time. The author argued that an aggregation of spatial data was risky due to the *ecological fallacy* and that history works was a better explanation than purely statistical methods.

More specifically, Jańczak (2015) quantitatively analyzed the election behavior in two regions in Poland that were divided by former state borders. He criticized the previous macro-level focus in Poland. He wanted to see if the concept of phantom borders could help explain the political landscape by using different scales. Through double descaling (territorial and electoral), the study could find differences between the central and local election results. For the Wielkopolska region, whose eastern parts belonged to Russia and the central and western parts were managed by Prussian-Germany, the voter turnout in the nationwide election was lower in the former Russian territory. A common phenomenon often explained by a more passive political culture in the past. But for the local election, this pattern changed. Here, the eastern parts reported higher participation rates. Jańczak (2015) argued for this with the neophyte syndrome, which describes the need to prove belonging to a region continuously. A qualitative analysis indicated that the residents wanted to prove their loyalty to the prestigious Wielkopolska region and adapted to the virtues of the Prussian part, as indicated by voter turnout. Local elections were used for this purpose, as they were viewed as endogenous, whereas the national elections were felt as more exogenous and did not serve the building of regional identity. This contradicted the previous assumption that historical structure influenced the present in a linear and fixed way. The phantom border was visible on the macro and micro scale, but the pattern varied. This research proposed scale manipulation as a further tool to investigate phantom borders. Von Löwis (2015) added to the use of downscaling techniques to investigate phantom borders in the Ukraine. The regions did not always show the expected relationships between Habsburg and Russia on the national level. Qualitative and quantitative analysis on different scales, combined with perspectives from different countries, made the phantom border effects more understandable while working with simplified narratives about East and West.

Netrdová et al. (2024) used the concept of phantom borders defined by Von Hirschhausen et al. (2019) and developed a new methodological approach to quantitatively measure and map existing phantom borders. The focus again lay on the former Czech-German border in the Sudetenland. They looked at the phantom border at the municipal level with the utilization of GIS spatial analysis techniques. The two indicators, mean age and mean years of schooling, were tracked over a time period of 40 years between 1980 and 2021. To measure the effect with distance to the former border, cumulative buffer zones were created. The Cohen's d test was used as the crucial metric to quantify how strong the border effect was. Additionally, a hotspot analysis was done to see how clusters align with the former border. As the spatial analysis and previous research indicated, the effects and behaviors of such borders underlaid spatial heterogeneity. Therefore, a discrete (four segments) and continuous (121 points $25km$ apart) approach was used to investigate the strength and thickness of the border. The combination of global statistics with local spatial visualizations created a method able to detect not only the

phantom border itself, but also its importance over time.

Previous research suggested that phantom borders and their effects did not necessarily vanish over time. Studies indicated that even borders from the Roman Empire still affected demographic parameters in today's population. Analyzing the Limes as an example of a former border through south and south-west Germany in a previous study revealed that regions that belonged to the Roman Empire showed higher economic developments (Wahl, 2017). The study used nighttime luminosity as a proxy for economic activity in the area. The study showed that the road network established by the Roman Empire remained relevant until today, creating one of the central foundations for better economic development. Additionally, this seemed to lead to a stronger growth of cities and those founded by the Romans along this infrastructural axis, such as Cologne or Augsburg, persisted after the empire collapsed. There was no proof of a more developed South Germany in pre-Roman times, strengthening their suggestions of the influence of the Roman Empire. To show the robustness of the findings, geographical factors such as elevation, soil quality, and proximity to major rivers were controlled. After all, the effects remained significant. Further studies showed a robust relationship between the former Roman Empire regions and personality traits, together with health and well-being parameters (Obschonka et al., 2025). A spatial regression discontinuity design (RDD) was used by the authors to compare populations living $100km$ on both sides of the former border. The data was provided by an individual-level survey about personality traits from 73,000 participants. The regions in the Roman Empire not only showed higher economic parameters, but also displayed more adaptive personality traits, being more open, extroverted, and less neurotic. In addition, the higher life expectancy of roughly six months indicated greater well-being and better health outcomes. The results for neuroticism and life expectancy are visible in Figure 3 and show the significant differences between the regions north and south of the Limes. The highest values for neuroticism, indicated through the darker colors, can be found in central Germany. For life expectancy, the highest values are reported in southern Germany, whereas the north and north-east show lower values.

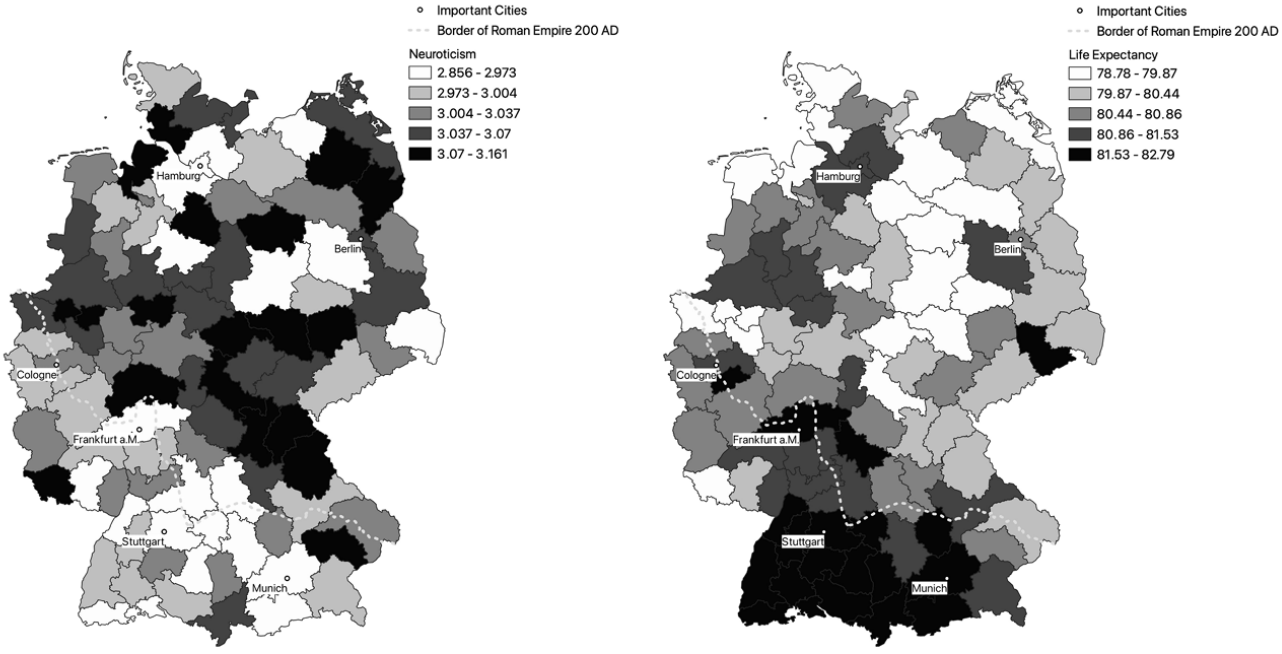


Figure 3: Neuroticism and life expectancy in present-day Germany and position of the limes. Figures are taken from Obschonka et al. (2025, 8).

Furthermore, Obschonka et al. (2025) proposed the former territorial border as a psycholog-

ical border. They validated their study with data from the Netherlands, which showed similar results for Roman and non-Roman regions. They suggested explanations for these long-lasting effects could be the advanced infrastructure, legal systems, public administrations, and health care, forming a culture that passed their expertise and views on to the next generations. They did not refer to this phenomenon as a phantom border but rather used the term *long shadows*, which described the long-lasting influence similar to that proposed by phantom borders. Again, it showed that historic division can leave traces in today's society, not only in voting behaviors or demographic parameters, but also in personality traits.

Phantom borders were mostly investigated using quantitative data and analysis. But it was also known that differences between East and West Germany existed in the local language and in how local residents talked about events. Leemann et al. (2019) collected German local language variation using an interactive web application in collaboration with two newspapers *SPIEGEL ONLINE* and *Tagesanzeiger*. Users provided metadata about their origin and answered questions about words they would use, for example, to say *"to chat"*. Overall, 24 questions were answered about the vocabulary used in their everyday life. This easy-to-use application allowed the researchers to collect data from 1.9 million participants, from which roughly 700,000 provided metadata. At the end of the survey, the users' origin was predicted using a heatmap for German-speaking countries, Austria, Switzerland, and Germany. The results were then compared with a historical dataset from the 1970s to add a time component and show possible changes. They revealed strong regional patterns of word usage. A very strong regional separation was visible, for example, for the German words for pancake. As visible in Figure 4, the division in Germany follows more or less the former inner German border, and a strong separation between the different German-speaking countries is visible. The right map showing the change in word use depicts rather low values for south and north-east Germany, whereas East Germany and central Austria report higher values, indicating more change compared to historical data.

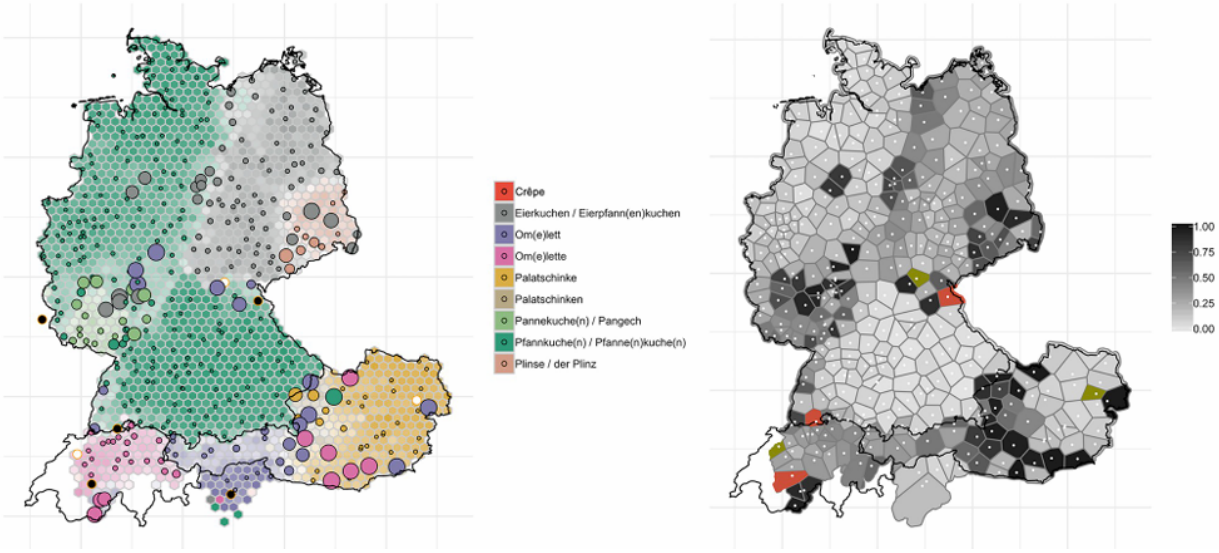


Figure 4: Word variants and degree of change maps for the word pancake: Left map is showing contemporary data as hexagons with historical data visualized as dots. The right map indicates the change of word use (black: substantial change, bright grey: very little change). Figures are taken from Leemann et al. (2019, 16).

The most prominent trend found by the authors was the replacement of very local words with more widespread or standard forms. This was especially visible in another example from northern and central Germany, where the term *"bolzen"* for non-professional soccer playing

replaced the more small-scale variants *"pöhlen"* or *"bäbbeln"*. The geographic distribution of the results and the values for the changes for those words are represented in Figure 5.

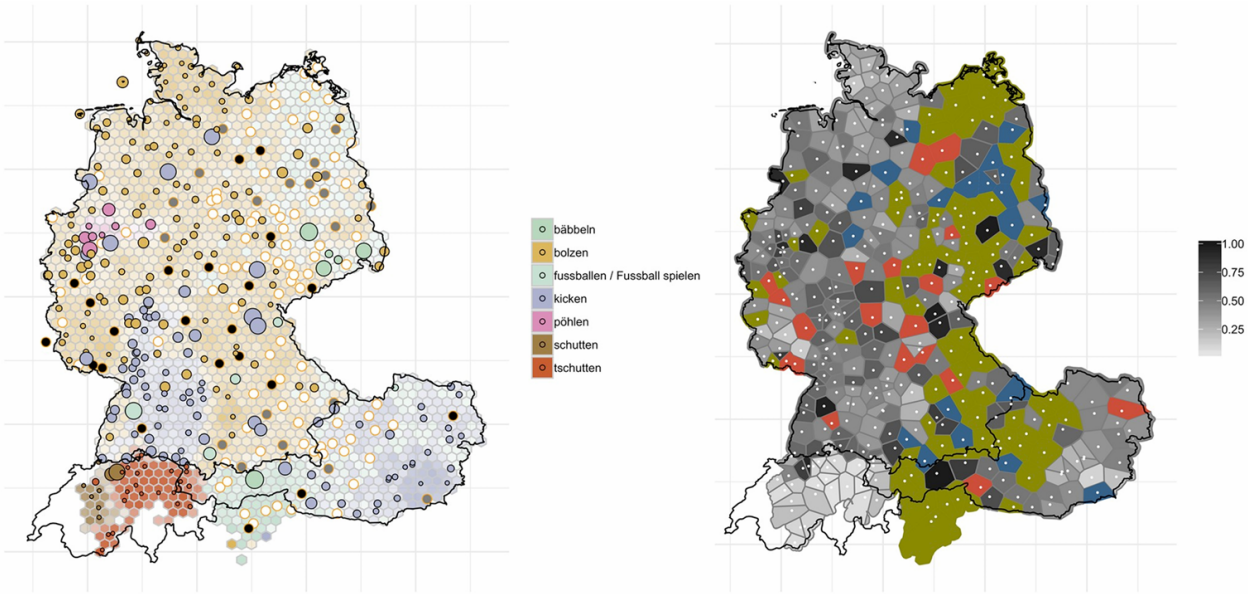


Figure 5: Word variants and degree of change for non-professional soccer playing: Left map showing contemporary data as hexagons with historical data visualized as dots. The right map indicates the change of word use (black: substantial change, bright grey: very little change, other colors: no comparison score could be calculated). Figures are taken from Leemann et al. (2019, 11).

In the South, especially in Bavaria, Austria, and Switzerland, significantly fewer changes in word usage were visible across all 24 investigated words in the study. The authors proposed that this effect might have been influenced by a more frequent use of local dialects and their higher reputation. Most relevant for the topic of phantom borders was the finding about the influence of country borders and former divisions. They proposed that these former political borders worked as isoglosses, i.e., linguistic boundaries separating regions where different words, pronunciations, or grammatical forms are used, or, respectively, linguistic borders. For example, the vocabulary describing pancakes or telling the time differed strongly between East and West Germany. They identified several reasons as drivers for these findings. Geographic mobility led to the adoption of more general, more widely understandable words. Nevertheless, regional variants persisted as an expression of regional identity. Local language and geography were strongly connected, making language a data source to investigate spatial phenomena.

Plewnia and Rothe (2009) investigated the linguistic wall between East and West Germany. The differences between the North and South were higher, but still, interesting effects were visible. The West Germans perceived a higher East-West difference compared to the East Germans. This showed how perspective matters when looking at geographic divisions. Also, an emotional gap was reported, with East Germans feeling more pride for their region of origin. They also reported a different relationship to language overall, meaning, in the East, the protection of the German language from influences through, for example, English words, was seen as more important, and current developments in linguistics were seen as more worrisome. Many stereotypes were also reflected through the local language, as the dialect spoken in Saxony was nationwide seen as more unpleasant. They performed an experiment asking how people feel about the dialects spoken in Bavaria and Saxony, where nearly 50% reported the Bavarian dialect as beautiful or very beautiful. For Saxony, only about 20% answered with these labels. But the differences between ratings from East and West German citizens made these results

even more interesting. While only around 14% of people from West Germany rated the Saxon dialect as beautiful or very beautiful, 35% in East Germany did, see Figure 6. In East Germany, the Saxon dialect was also viewed as more friendly and educated. This showed a substantial difference in the perception of language between East and West Germany and could lead to further issues and underscore stereotypes.

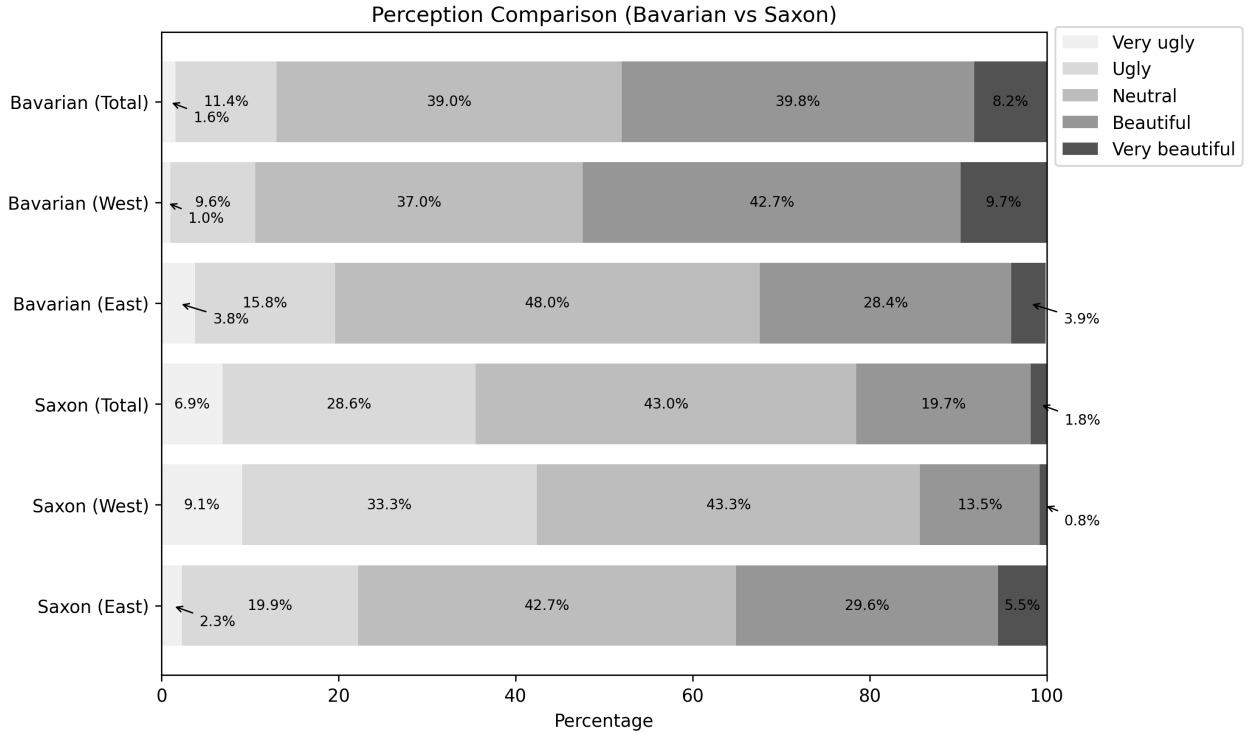


Figure 6: Survey results for the question of how the dialects from Bavaria and Saxony sound. Translated from Plewnia and Rothe (2009, 275).

In general, it is important to state that working with phantom borders is not about proving stereotypes and searching for differences between regions. It is about understanding and analyzing dynamic developments. Previous research mainly focused on how phantom borders are becoming visible through statistical data, proposing several different methods. The first linguistic studies with the introduction of the term linguistic border showed a connection between the research fields and proposed language as a possible data source for further experiments in the field of phantom borders.

2.2 Relevant literature in the context of large language models

Understanding spatial patterns in human behavior, language, and historical influence, as, for example, in the case of phantom borders, required the ability to understand concepts such as distances, locations, and regional differences. In the past, such analysis focused on human surveys or statistical data, but unstructured text also carries important signals. Textual data was more difficult to handle as the extraction of information required processing capabilities similar to contextual understanding. LLMs offered new approaches as they were able to process large amounts of text. Their use has increased drastically in recent years, especially with the introduction of OpenAI's GPT series. Nowadays, a wide range of users could use these models for everyday tasks. Letting an LLM recommend a coffee place or a next holiday destination already required some sort of spatial awareness. This spatial reasoning was a key human ability that we use every day without realizing it. Nevertheless, it was a complex task that required

processing several parts of information from our environment in parallel. This called for a form of spatial intuition and knowledge about our surrounding world with its moving objects. Translating this to AI models was a difficult task, which required complex models (Yang et al., 2025). Geospatial foundation models (GeoFMs), which were trained on spatiotemporal data, already showed that they can gain world knowledge and spatial reasoning capabilities in their training process. In contrast to traditional LLMs, they were designed to handle topological relations, spatial data sets, and reason across space and time (Janowicz et al., 2025b). Janowicz et al. (2025b) highlighted the significant potential of GeoFMs, while identifying new challenges related to bias, trust, and fairness of their outputs. It needed to be investigated how the geographic world was represented through the models and how well this aligned with human cognition. In order to use these models for decision making, it was essential to understand their abilities to prevent unintended results.

O'Sullivan et al. (2024) conducted a study to further investigate the depth of this spatial reasoning. They investigated 14 different LLMs with 114 prompts, testing the encoding of the words *"near"* and *"far"* in the prompts. They showed that the models lacked a differential representation of these words when applied to different geographic scales. Hence, while encoding a basic spatial awareness, LLMs strongly relied on quantitative keywords rather than the spatial context added in the prompt.

The interest in LLMs in the geographic and, more specifically, spatial data science domain has increased substantially in recent years. These geographic models showed the ability to capture semantic and contextual information in large text corpora, revealing a whole new study area. Although these LLMs were not directly trained for (geo)spatial downstream tasks, they encoded information about places, spatial relationships, and socio-economic patterns. Geographic research required a large amount of data, which was often difficult and expensive to collect. The models contained information compressed from the Internet training corpora, combining a substantial amount of data. Manvi et al. (2024) investigated whether the leveraged information in LLMs could be used for geospatial purposes, especially for predictions. Previous research relied on, for example, satellite images, which did not produce satisfying results due to high costs, difficult access to images, or missed areas. As models struggled to link raw coordinates to the real world, the authors proposed a new method called *GeoLLM*, where they fine-tuned the model with prompts and map data from OpenStreetMap, predicting economic factors and population density. The prompt included nearby places or specified the prediction tasks through a proposed scale. They found that LLMs not only performed equally well as recent methods, but with their new method even outperformed them. These results suggested that LLMs could be used as an alternative source for geographic information. It was very important to be aware of possible encoded bias due to the over-representation of more developed and populated regions in the training data.

Research in the field of GIScience showed the capability of language models to extract geospatial tasks from natural language sources (Mansourian and Oucheikh, 2024). This enabled non-expert GIS users to utilize more complex spatial analysis methods. Therefore, they developed their own chatbot called *ChatGeoAI* that utilizes Llama 2 to translate natural language into usable Python code for geospatial purposes. This proposed language models as potential assistants for analysing spatial data for spatial decision-making. With the rose of LLMs and their widespread use, the question of how these models encode geographical space emerged. And as a subsequent question, how correct was this representation (Janowicz et al., 2025b). Abbasi et al. (2025) evaluated the performance of GPT-4o and Gemini 2.0 Flash in three geospatial tasks. The findings for geocoding, elevation estimation, and reverse geocoding showed an overall inconsistency in representation. They pointed out the need for fine-tuning of these models if applied to specific tasks.

Another important field of research focused on the inference of geographic locations from

language data. Earlier studies used statistical models or rule-based approaches to predict the location of documents or social media posts. Wu et al. (2025) presented a self-supervised deep learning method for text geolocation prediction, as current LLMs produced more unstable results. Their experiments on two datasets from Manhattan and Boston showed that the method significantly improved the prediction capability. Without labeled samples, namely the GeoSG (Geographic Self-Supervised Geolocation) model, the paper stated that LLMs such as T5 (Rvs) and Llama 3.1 (LLM-Loc) could predict locations through regional linguistic signals without the need for further information. But applying their GeoSG model could increase the accuracy of predictions and produce more robust results compared to the baseline models.

Researchers also studied the representation of dialect through embeddings. The model DialectGram proposed by Jiang et al. (2019) could detect these linguistic variations at multiple levels of resolution, tested with English words and phrases. Traditional models required a predefined scale, limiting the analysis of dialects. This new model was able to discover linguistic borders and could therefore be used to detect phantom borders as well. Modern LLMs also relied on embeddings when encoding semantic relationships. Therefore, it could be expected that LLMs could capture linguistic variations as well.

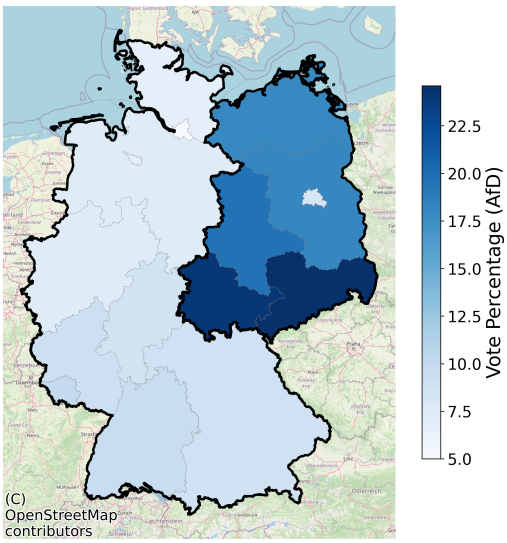
More generally, researchers introduced the term GeoAI to describe not only the application of machine learning and AI to geospatial and geographic purposes, but also the enrichment of models and methods with spatial, temporal, and place-based (placial) parameters, thereby mimicking spatial thinking in AIs (Janowicz, 2023). Hence, the research field of GeoAI also contained topics such as sustainability, debiasing, data and schema diversity, and neutrality. As models reflected training data, they could be used to study societal biases, stereotypes, and geographic inequalities on a large scale, providing huge amounts of data. Additionally, as these FMs captured the most complex collections of data provided by society, they themselves were worth examining and studying, raising the question of how the world is depicted by them. This could provide insights into how the interaction with AI will look in the future.

The relevance of understanding how these LLMs imitate spatial reasoning was clear. Additionally, previous research showed that LLMs could be unreliable or inconsistent when encoding geographic information, which raised important questions about the validity of their outputs for spatial analysis. This makes them particularly interesting but also challenging for the research field of phantom borders. If they were able to detect linguistic traces of historical division, they could serve as a powerful tool for investigating textual data in this field.

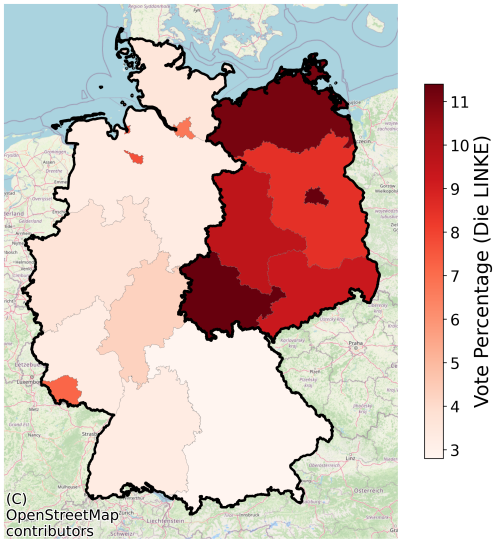
3 The former inner-German border as a case study

The thesis focused on the former border that divided Germany. Therefore, the historical context is briefly highlighted here. Germany was divided into the German Democratic Republic (English: GDR, German: DDR) in the East and the Federal Republic of Germany (English: FRG, German: BRD) in the West after the Second World War from 1945 to 1990. The division emerged from the occupation zones formed by the Allied powers and led not only to a physical but also to an ideological and political separation. The actual $1,440km$ long physical border called *"Die Mauer"* was constructed in 1961 as a tool to prevent the mass movement of East German citizens to the West. In 1990, the GDR collapsed and was integrated into the FRG (Bundeszentrale für politische Bildung, 2009; Chronik der Mauer, 2016). Despite the official political reunification, differences between East and West Germany persisted in economic, demographic, religious, and social parameters, making the former division a popular example of a phantom border (Adler and Brayfield, 1996; Mollenkopf and Kaspar, 2005; Van Hoorn and Maseland, 2010). The still ongoing division and separation between Germany has led to conflicts and tensions.

This was especially visible when looking at recent election patterns. The data for the following analysis was collected from Statistikportal. Figure 7 shows the spatial distribution of vote shares for the right-wing party *"Alternative für Deutschland (AfD)"* and the left-wing party *"Die LINKE"*, together with the historical border between the former GDR and the FRG. Even though these parties represented opposite ideologies, the highest percentages for both were reported in the former East German states. This spatial clustering highlighted the ongoing polarization in Germany, reflecting economic and demographic disparities. Rensmann (2019) proposed that this divide was driven by the historic legacy, ongoing socioeconomic disparities, and the active mobilization of discontent by populist parties. He referred to the term of an *"invisible economic wall"* between East and West Germany. The East was perceived as lacking in terms of income, wages, and job opportunities. A concentration of big companies in western regions led to an underrepresentation of East German citizens in administration, politics, and academia. Additionally, inner-German migration led to shrinking populations in eastern Germany as young citizens moved to the western metropolitan areas. All these factors had a strong influence on the stability of Germany, and therefore, the phantom border remained very relevant.



(a) 2021 election *AfD*.



(b) 2021 election *Die LINKE*.

Figure 7: Election results in 2021 for the right-wing party *AfD* (left) and the left-wing party *Die LINKE* (right). The former inner-German border is shown on both maps (Own map with data from Statistikportal).

4 Method

In this method section, all steps leading to the answering of the research questions are listed and explained. The background for this thesis is an existing geolocated newspaper dataset, whose utilization, access, and retrieval of relevant articles for the thesis are explained in Section 4.1. As full article texts are required for the analysis task, the retrieval of HTML pages through provided URLs from the underlying dataset is explained in Section 4.2. Additionally, all pre-processing and cleaning steps to convert the raw HTML pages into plain article texts are listed. The next step is the deletion of all geographic references, which is described in Section 4.3. After that, the data is prepared for the three experiment setups. Starting with the prediction of article location through an LLM in Section 4.4, followed by the creation and analysis of the textual embeddings in Section 4.5, and lastly looking at justifications made by the LLM in Section 4.6. All required measurements to evaluate the experiment results are explained in Section 4.7. Finally, in Section 4.8, the structure of the two final constructed and utilized databases is explained, followed by a short introduction of the utilized models in Section 4.9.

4.1 Dataset and retrieval of articles

Popular newspaper databases such as Common Crawl did not provide the exact writing location of articles and were not directly usable for the proposed research (Common Crawl, 2026). Therefore, an existing database from Kriesch and Losacker (2025) built the foundation. The authors have created a SQLite database containing 50 million German newspaper articles, and for approximately 70% of articles were able to identify a geographic location based on information extracted from the article text. They collected articles from Common Crawl and removed all texts with non-German country-specific top-level domains (TLDs), which were internet domain endings assigned to individual countries, and non-German texts. To ensure a good data quality for later processing steps, several filters have been applied, including an article length between 50 and 10,000 words, the average word length between three and ten, and removing articles with more than 10% non-alphabetic words. This corpus of full-text articles was further used to create the geolocated news dataset. Meta's LLAMA-3.1-8B-Instruct model was applied to extract place entities from the texts, which worked as training data for the following SpaCy pipeline. A customized named entity recognition (NER) model was trained, which was specialized in extracting city names from German newspaper articles. These place names were then geocoded using Nominatim to assign geographic coordinates to the articles. Note that Nominatim ranks the importance of results through population size or prominence of a location (Nominatim, 2026). It was important to note that no document-level context cues were used to minimize the problems due to place name ambiguity. Additionally, the use of the gazetteers introduced biases, favoring larger, more populated places. Germany was divided into 400 NUTS-3 regions (European Commission, 2018). They used the division from 2016, which was updated in recent years. Nevertheless, as all locations were defined using the 2016 division, this logic was also continued in the further study. The location was provided via coordinates and then translated into these NUTS codes. These codes were unique identifiers per region and made a mapping to the former East and West Germany possible, as the former border followed these NUTS region borders. It meant that every NUTS region lay completely inside former East or West Germany. The database contained articles for every single one of the 400 NUTS-3 regions covering every area in Germany. The validation of the NER model, using 500 randomly selected and annotated texts, showed an F1-score of 93.87% with equally high levels of precision and recall. For the geocoding process, the validation was done using the relationship between population size and the number of articles on the NUTS region level. They reported a proportional relationship, which suggested that the geocoding process worked

reliably. They demonstrated that the locations assigned to the articles were accurate, and the dataset was therefore suitable as ground data for the thesis experiments. The complete database contained four tables, providing additional metadata for each article. For this use case, the three tables **Articles**, **Locations**, and **Article_locations** were relevant (see Tables 1, 2, 3). The full dataset was downloadable via their GitHub repository[1].

Table 1: Schema overview of Articles Table (Kriesch and Losacker, 2025).

Column name	Description
id	Unique identifier for each article (UUID format).
url	Original URL of the article.
tags	Tags associated with the article.
categories	Categories associated with the article.
hostname	Hostname of the article's source.
date	Publication date of the article in ISO format (YYYY-MM-DD).
date_crawled	Date when the article was crawled in ISO format (YYYY-MM-DD HH:MM).

Table 2: Schema overview of Locations Table (Kriesch and Losacker, 2025).

Column name	Description
location_id	Unique identifier for each location.
loc_normal	Normalized name of the location (for geocoding purposes).
latitude	Latitude coordinate of the location.
longitude	Longitude coordinate of the location.
NUTS	Nomenclature of Territorial Units for Statistics identifier.
GEN	General Name of the NUTS location.
ARS	Regional identification number.

Table 3: Schema overview of Article_Locations Table (Kriesch and Losacker, 2025).

Column name	Description
article_id	Unique identifier of the article (foreign key to Articles.id).
location_id	Unique identifier of the location (foreign key to Locations.location_id).

Using all articles in the database was impractical due to long processing and prediction times, as well as the substantial costs associated with the GPT API. Therefore, ten random location ids per NUTS code and ten random articles from each of them were picked. This led to a maximum of 100 articles per NUTS code and 40,000 overall. Not all NUTS codes had ten location ids and not all of them had ten articles. This was because of the different sizes of the NUTS regions, which was also mentioned by the authors of the paper. First results showed that many regions reported fewer than 30 articles through this approach. Therefore, a second sampling cycle was done, resulting in a maximum of 80,000 articles. The fields of these articles, including id, url, hostname, date, location_id and NUTS were used in this thesis.

[1] https://github.com/LukasKriesch/CommonCrawlNewsDataSet

4.2 Article text retrieval and preprocessing

The database provided by Kriesch and Losacker (2025) did not contain the full article text but only its respective URLs. Therefore, collecting the full texts of these articles was necessary to conduct the experiments. Since the crawl date of the articles was not available, it was not possible to access the raw HTML via the Common Crawl database. So, the URL was used to retrieve the full page. To solve the problem of URLs linking to pages that no longer exist or storing different articles, two steps were included. When retrieving the HTML page, a date check was added, making sure it was only retrieved if it matched the date of the original article via the date column. If no page was found or the date mismatches, the Internet Archive Wayback Machine (IAWM) API was called (Ogden et al., 2024). The Wayback Machine allowed free access to the world's largest Internet Archive, only needing the URL. It was designed to give back the archived web page snapshots at different times. For the following use case, the snapshots nearest to the article date were selected. This procedure resulted in 46,473 articles, where an HTML page could be found. The resulting URL from the Wayback Machine, linking the old snapshot and the texts, was stored in the database as well. Since the data was collected from news websites, it included irrelevant (meta)data such as HTML tags and CSS styles. Therefore, in a first step, the data needed to be preprocessed. The processing step was performed using Beautiful Soup, creating a precleaned version of the texts (BeautifulSoup, 2026). This is a Python library designed to handle HTML and XML files and extract data from them. Nevertheless, a manual review indicated that the data were not completely clean, as some tags were not completely removed. Therefore, an LLM was used to perform a secondary cleaning of the data. By using an API key, OpenAI's models could be accessed outside of the chat function. To minimize the cost, a cheaper version (GPT-5-mini) was utilized to clean the HTML texts, and it was told to only respond with the full article text, which was a task well-suited for LLMs. Through the system prompt, the model got information about how to behave. The user prompt handed over the input on which the task had to be done. The parameter reasoning effort controls how much model power was utilized to fulfill the task and could be set to *"none, minimal, medium, or high"*. The detailed model settings can be seen in Table 4.

Table 4: Model settings for text extraction from HTML pages.

Setting	**Value / Description**
Reasoning effort	minimal
System prompt	You receive a preprocessed German HTML news article taken from a SQLite database column and extract only the article's human-readable text. Remove all HTML tags, scripts, styles, navigation elements, advertisements, authors' names, contact data, and anything not part of the main article text. Do not summarize, rewrite, guess, or add any new words. Do not provide commentary or explanation. Output plain text only.
User prompt	Precleaned article text with BeautifulSoup

The preprocessing with BeautifulSoup reduced the length of the raw articles and already reduced token costs. To minimize them even further, the articles were additionally cut to 200 words, always ending with a complete sentence. This ensured that context was preserved and the texts were long enough to capture semantic structures. These truncated articles formed the ground data for all further processing steps. The general distribution of article counts from this larger dataset is shown in Figure 8. To validate that the crawl and the further processing steps did not favor articles from specific regions, the results of a LISA analysis can be seen

in Figure 9a. There are several high-high clusters in central Germany, but there is no overall trend favoring the collection of East or West Germany. There are only a few smaller Low-Low clusters in East Bavaria and in the North, but no specific part of Germany is underrepresented. The distribution of articles per NUTS region is represented in Figure 9b, which shows that only 41 regions, mostly small regions, have fewer than 30 articles.

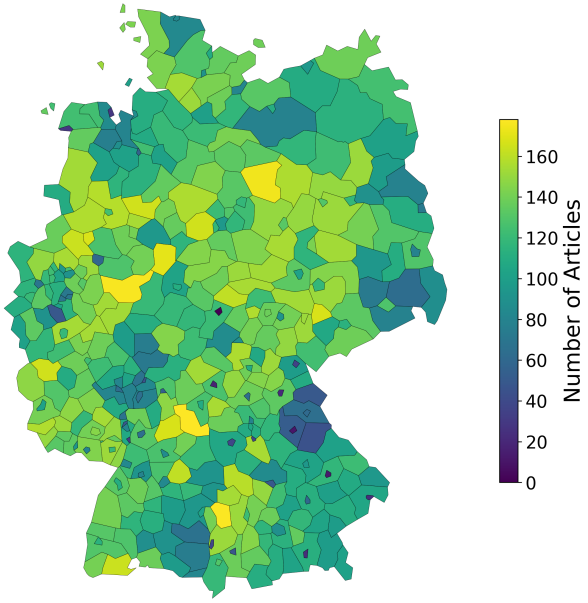
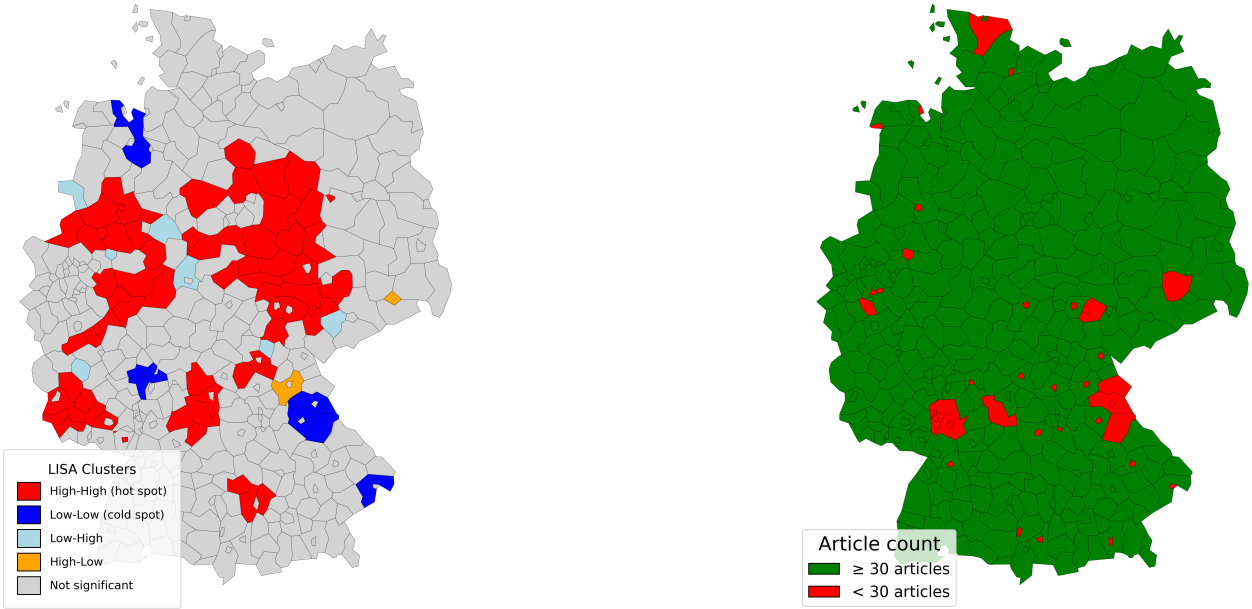


Figure 8: Article count per NUTS region in the dataset.



(a) LISA results for the dataset.

(b) Article count with threshold.

Figure 9: Overview of spatial analysis and article distribution. (a) LISA results (b) Article counts with threshold of 30 articles.

4.3 Data processing and transformation

To catch linguistic traces of phantom borders, all geographic references had to be removed from the article texts. If, for example, place names such as *"München"* or *"thüringer"* remained in the text, the LLM might have made its prediction merely based on these words. In Section 4.3.1, first methodological approaches with NER models are proposed, and in Section 4.3.2, a newly developed method using LLMs is explained.

4.3.1 Named entity recognition with SpaCy

Trained pipelines for natural language processing (NLP), such as the open-source Python library SpaCy, could be used for tokenization, part-of-speech tagging, and NER. All of these steps were required to remove the geographic references from the article texts (SpaCy, 2026). For the removal steps, SpaCy's *de_core_news_md* pipeline, which was specifically trained for German texts, was utilized. Initially, all entities labeled as "LOC: location", "GPE: geographic political entity", "FAC: faculty", and "ORG: organization" were removed from the texts. This approach did not differ between German and non-German references. International references did not provide information about the article's location and should therefore not be removed. The idea was to add a geocoding step in the removal process. All identified references were geocoded using the Geonames gazetteer (GeoNames, 2026). Then, if an entity was geocoded in Germany, it was removed from the article text. This showed good results when it came to very generic words, including for example, the German words for house, police station, badminton club, US space port, and Florida. These references were not removed, but also very local names, such as kindergarten names or sports clubs, remained, as they were not in the gazetteer and therefore not removed. As missed references were a bigger issue compared to incorrectly removed words, the process was reversed. This meant that all references out of the four categories were removed, and only if a non-German location was assigned through the geocoding process, the word was kept. This worked better as many local references were removed, and many international ones were kept. However, this approach only captured geographic references in nouns and highly relied on the SpaCy pipeline, whose contextual abilities were limited. The results did not provide a good enough quality, which was needed for the analysis of latent linguistic traces of phantom borders in LLMs, and a new method was developed.

4.3.2 LLM-based removal of geographic references

As standard NER pipelines were limited in their contextual abilities, LLMs were also used for the removal of geo-indicative words. As the inputs and outputs were longer texts resulting in many tokens, the smaller model GPT-5-mini was used to reduce costs. The task was to remove all German-specific places, natural sights, companies, provinces, cities, sports clubs, etc. International references were allowed to stay in the text, as they did not provide insight into the location of the article. Few-shot demonstrations were added in the system prompt, covering different objectives. The prompt was also developed using 30 articles, where the removal was repeatedly checked after every alteration. Common errors were the removal of generic terms and the omission of unique location and company names. Therefore, examples for these cases were added in the prompt. Additionally, highly ambiguous entities were included as few-shot learning for follow-up processing iterations. This prompt-based few-shot learning was a type of machine learning where a model was adapted to perform a task with only very few examples. In the end, three examples, including the articles and their cleaned versions, were handed to the model in the system (not user) prompt to improve the quality of the process. AI removed words were replaced by "[...]" to keep the sentence structure intact. Similarly, to reduce computation time and costs, reasoning effort was also set to *"minimal"*. The model

settings and prompts are listed in Table 5. The articles provided in the 3-shot learning can be seen in Appendix A. To evaluate the results of the removal, 30 randomly picked articles were processed manually and then compared to the results produced by the GPT model.

Table 5: Model settings for 3-shot removal of geographic references.

Setting	**Value / Description**
Reasoning effort	minimal
System prompt	Your task: - Remove all Germany-specific references from German newspaper articles by replacing them with [...]: - Names of places located in Germany (cities, towns, villages) - German addresses, including street names, roads, squares, alleys, and house numbers - Names of German regions, landscapes, and their adjectival forms (e.g. Bergisches Land, bergisch, rheinisch, fränkisch, hamburger) - Names of German federal states and historical country names (e.g. DDR, BRD) - Organizations, institutions, clubs, sports teams, companies operating in or tied to Germany, even if the name itself does not contain geographic reference - Names of Natural landmarks located in Germany (rivers, mountains, forests) - Names of buildings and unique or identifiable locations in Germany - Keep all international geographic references intact - Handle exceptions carefully: - Generic words like "Haus", "Pfarrhof", "Polizeiinspektion" or "Badmintonclub" should not be removed - Specific named locations like "Haus der Begegnung" must be removed - Specific institution/club names like "Kindergarten Schnäggehüsli" must be removed like this: "Kindergarten [...]" - Apply all removal rules equally to headlines, subheadlines, rubrics, section labels, and body text Strict prohibitions: - Do not add, infer, summarize, or complete content - Do not change language - Do not editorialize or normalize Output: Cleaned text only, no commentary
User prompt	Cleaned article text cut to 200 words

As West Germany was bigger than East Germany, it was crucial to balance the dataset. All East German articles were kept, and the same number of West German articles was randomly sampled. This led to a database containing an equal number of East and West German articles. To ensure that all articles provide contextual information, a threshold for word length was needed. The initial database contained 9,430 East articles and 36,790 West. It was decided to only use articles with more than 100 words, which resulted in 5,928 for East and 27,100 for West Germany. The balanced dataset contained 11,856 articles (5,928 each). The split into East and West was performed through the first three letters of the NUTS code (see Table 6). These codes specifically pointed to the federal states in Germany (European Commission, 2018).

All articles from Berlin (NUTS code DE3) were excluded, because a detection of whether they originate from East or West Berlin was difficult. For the balanced dataset, the total number of articles was naturally lower compared to the initial dataset. Also, the number of

Table 6: Classification of NUTS codes into East, West, and Berlin.

Region	**NUTS Codes**
East	DE4, DE8, DED, DEE, DEG
West	DE1, DE2, DE5, DE6, DE7, DE9, DEB, DEA, DEC, DEF
Berlin	DE3

articles per NUTS region in East Germany was higher compared to West Germany because the same number of East German articles was included as in the West. The distribution of these randomly selected West German articles above the word threshold reported relatively equal numbers across all regions, as visible in Figure 10.

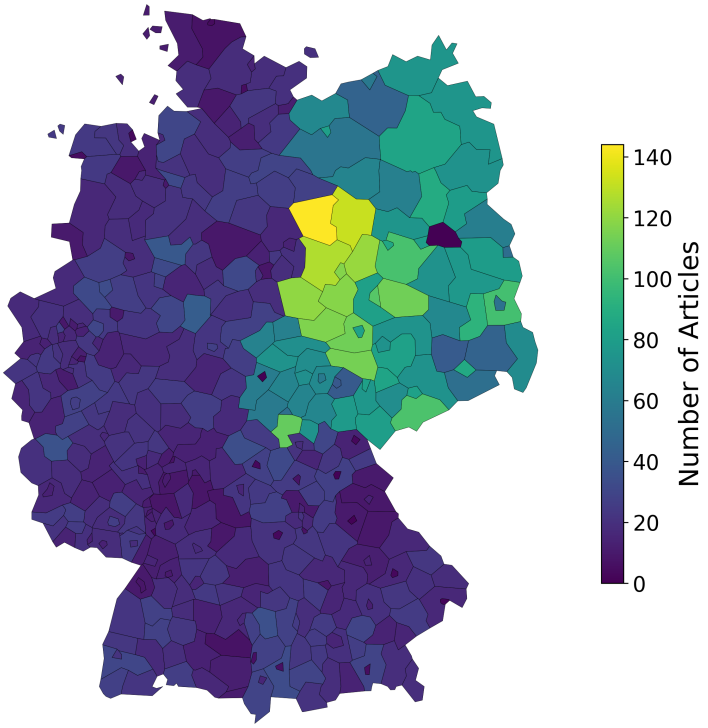


Figure 10: Article count per NUTS region in the balanced dataset.

To further validate the quality of the removal prompt, a ground-truth dataset containing 200 East and 200 West articles was randomly selected from the balanced dataset. All geographic references were *manually* removed from these texts by the author, resulting in a gold standard ground-truth dataset. This ground-truth removal was then compared to the removal that was done by GPT-5-mini. Due to missed words in the initial iteration (3-shot), the prompt was further optimized by providing 20 examples from the ground-truth dataset. These were picked using the Levenshtein distance as a parameter to detect challenging texts (Fawzy Gad, 2020). The higher the distance, the more different the manual removal and the 3-shot removal were, indicating many missed geographic references. To ensure equal conditions for East and West Germany, ten articles for each were provided. Then the removal was repeated, while setting the reasoning effort to *"high"*, for the remaining 380 articles from the ground-truth dataset. Again, the Levenshtein distance was used to compare the results from reasoning effort *"minimal"* and *"high"* to the ground-truth, and additionally, 20 East and 20 West articles were manually compared. As this procedure showed very promising results, it was applied to the balanced dataset with the 11,856 articles. Even though computation time and costs were

substantially higher, a good removal built the baseline for all following experiments and justifies those increased resources. Ultimately, a missed geographic identifier in the text would have made the LLM's prediction trivial in many cases, and, thereby, artificially increased accuracy.

4.4 Origin classification (East vs. West Germany)

A state-of-the-art model of the GPT-5 family was used for the predictions. The decision to use GPT-5.1 was made as it allows temperature and logprobs settings, and was the most advanced model at the time of writing. The temperature parameter controlled the randomness of token selection and could have values between zero and two, with low values ensuring more deterministic outputs (Holl, 2024). When setting the logprobs parameter to true, the API returned the log probabilities for the tokens, which indicated the confidence level of each generated token. Further information about the selected model is listed in Section 4.9. The model had to output whether a given article was from East ("e") or West ("w") Germany, yielding a binary decision and returning only one token per article. To allow the temperature and logprobs setting, the parameter reasoning effort had to be set to *"none"*. The logprobs were converted into a string and then saved in a JSON column in the database, saving the token with its log probability. Initially, max_logprobs was set to five as only the first few possible tokens were of interest. The logprobs were then converted to actual probabilities. As the output of the prompt was always only a single token, the logprobs output was as shown:

```
ChoiceLogprobs(
  content=[
    ChatCompletionTokenLogprob(
      token='e',
      bytes=[101],
      logprob=-9.925185440806672e-05,
      top_logprobs=[
        TopLogprob(token='e', bytes=[101], logprob=-9.925185440806672e-05),
        TopLogprob(token='w', bytes=[119], logprob=-9.218849182128906),
        TopLogprob(token=' e', bytes=[32, 101], logprob=-21.437599182128906),
        TopLogprob(token='d', bytes=[100], logprob=-25.765724182128906),
        TopLogprob(token=' w', bytes=[32, 119], logprob=-25.781349182128906)
      ]
    )
  ],
  refusal=None
)
```

For the example provided above, "e" was the chosen token with a log probability of -9.925×10^{-5}, which equaled a probability of approx. 99.99%. The top_logprobs output listed the five most likely alternative tokens considered during processing of the prompt. The token, "w" had the second-highest log probability, which corresponded to a probability of 0.000099%. This indicated that the model assigned a substantially higher likelihood to the token "e" compared to the other possibilities. This could be seen across all articles, regardless of the correctness of the prediction. This workflow, using logprobs and looking at the log probability of the first token, showed that the prediction worked, but the log probability for the first token was always very close to zero, meaning a probability of nearly 100% for the selected token. This could have been for several reasons, for example, instead of truly answering the question "How sure are you about East or West", the model was more likely to answer "Given I believe the article is from East, how likely should I pick 'e'?". This is why the model was so overly confident (Hills and

Anadkat; OpenAI Developer Community, 2025; SAP, 2024). Therefore, logprobs were removed from the analysis, and work was continued only with the prediction results. The results were saved in the database and could be compared to the true location using the NUTS codes. The model settings are listed in Table 7. For the balanced dataset of 11,856 articles, predictions were conducted twice: once using the texts processed with 3-shot removal and once using the texts processed with high-effort removal. This comparison evaluates how prediction accuracy changed as the quality of geographic reference removal improved. Accuracy was expected to decrease, as the 3-shot method likely left more geographic references in the text, making origin prediction easier than with the high-effort method.

Table 7: Model settings for prediction.

Setting	**Value / Description**
Temperature	0 (ensuring deterministic outputs)
Reasoning effort	"none" (required to allow temperature settings)
System prompt	You receive the text of a German news article. Determine whether the article was written in East Germany or West Germany, using the former inner-German border as reference. Respond with: – e if the article originates from East Germany (former DDR territory) – w if the article originates from West Germany (former BRD territory). Output only the single letter ("e" or "w") and nothing else.

Along with the quality of the article, the model settings were also expected to influence the prediction accuracy. To investigate the influence of the parameter reasoning effort in the prediction results, the parameter was set to *"high"* for a follow-up experiment. The ground-truth dataset containing the 400 texts, where the geographic references were manually removed, was used to ensure the prediction was based on the language and not on specific place names. The predictions were then compared to the results produced with the reasoning effort *"none"*, also based on the ground-truth data.

4.5 Embedding analysis

A common approach for preprocessing the unstructured textual data was transforming texts into numerical vectors, allowing for mathematical analysis. These (learned) representations were also called embeddings (Egger, 2022). In this case, this involved generating a vector representation for every newspaper article in the balanced dataset. Once the articles were represented in this vector space, it became possible to compute statistical measures such as means, centroids, distances, and similarity scores between articles. There were numerous models available for generating text embeddings. As an initial step, experiments with smaller embedding models were conducted, such as Sentence Transformer (SBERT, 2026), which offered multilingual models. However, in this study, the semantic differences between East and West articles were subtle and nuanced, hence, these models were not able to detect them well due to their limited vector space dimensionality. The *distiluse-base-multilingual-cased-v2* model, for example, was based on a 512-dimensional vector space (Reimers and Gurevych, 2019). In addition to open-source transformer models, OpenAI provided several embedding models accessible through their API (OpenAI, 2026). The *text-embedding-3-large* was the most advanced OpenAI model for both English and multilingual tasks. It produced embeddings with 3,072 dimensions,

offering a substantially larger representational space. The higher dimensionality and improved model capacity enabled these embeddings to capture even subtle semantic variations within texts. This approach was more suitable for longer newspaper articles with only very small linguistic differences. Using the GPT-based embeddings, inter- and intra-cluster distances between East and West articles were computed. Concretely, this involved calculating the mean centroid for each article's group within the embedding space. These centroids represented the average semantic position of the respective clusters. The expectation was that the distance between a purely East cluster centroid and a purely West cluster centroid would have been larger than the distance between two centroids of mixed clusters containing articles from both groups. In a second step, the same procedure was repeated separately for correctly classified and misclassified articles. It was expected that the effects would have been even higher when using only correctly classified articles, as they were more representative of their location and easier to differentiate from the other class. For misclassified articles, a weaker separation was expected. Since they were difficult to predict, their embeddings might not have differed as strongly from those of the other group. In addition, the local neighborhood of the misclassified articles in the embedding space was analyzed, performing a nearest-neighbor analysis. For each misclassified article, its 20 nearest neighbors were identified based on cosine similarity. Then, the average distance to neighbors belonging to the article's true label was calculated and compared to the average distance to neighbors associated with the predicted label. At the final step, logistic regression and a support vector machine (SVM) classifier were trained. This allowed an evaluation of whether the classification performance remained robust even when the overall cluster spread was relatively high. Comparing these classifiers, therefore, helped assess whether the observed embeddings support stable class separation despite potential overlap in the vector space. It also provided insights into the performance of the LLM, as it should have reported an even higher accuracy than the simple classifiers due to its contextual knowledge.

4.6 Model justification and explanation analysis

To analyze the justifications of the GPT-5.1 model, a separate experiment was conducted using the 400 articles from the ground-truth dataset. For each article, the following prompt and settings were run ten times, generating 4000 justification texts. All outputs, including both the predicted label and the accompanying reasoning, were stored in a Python dictionary for subsequent analysis. The model settings for this experiment can be seen in Table 8. It was important to note that the reasoning effort category had slightly different names for GPT-5.1 and GPT-5-mini. It was set to *"low"*, which equaled the *"minimal"* setting in the removal experiment with GPT-5-mini.

In a second step, all the justifications were embedded using the same model as mentioned in 4.5. Based on these embeddings, inter- and intra-cluster distances between East and West justifications were calculated. Specifically, this involved calculating the centroids for the different groups and then measuring the distances between these centroids. If the distance between the two homogeneous clusters was higher compared to mixed clusters, the justifications indicated a significant divergence for East and West predictions. In addition, the centroid distances were calculated separately for correctly classified and misclassified articles. This allowed for the evaluation of whether the explanatory divergence was stronger when the model's prediction aligned with the ground-truth and weaker when the prediction was incorrect. For the articles reporting both East and West predictions, the inter-cluster distance was calculated and averaged across all of them, to compare them to the overall distances. Furthermore, the justifications for the articles that reported the lowest mean distance between their East and West justifications were manually reviewed to see if they remained consistent.

Table 8: The model settings for justification.

Setting	Value / Description
Reasoning effort	"low"
System prompt	You receive the text of a German news article. Determine whether the article was written in East Germany (former DDR territory) or West Germany (former BRD territory), using the former inner-German border as reference. Articles may contain anonymization placeholders like "[...]". Base your justification only on the remaining context. Respond only in valid JSON like this: { "prediction": "e" or "w", "justification": "brief explanation in English (50 words)" } If your prediction is based on specific words or phrases in the article, you may mention them explicitly to support your reasoning. Do not output anything else.

4.7 Evaluation metrics and performance measures

For the measurements conducted to assess the success of removing geographic references, the Levenshtein distance was utilized (Fawzy Gad, 2020). The definition of this metric is as follows:

- Number of necessary changes to convert one word into another.
- Three operations: Insertion, Deletion, and Replacement.
- Each operation increments the distance by one.
- The distance is computed recursively by selecting the minimum cost among the three operations.

High values indicated several differences between the two texts, which in this study referred to the omission of geographic instances in the removal process. Conversely, a low distance indicated a high agreement between two texts.

The prediction results were evaluated using their true location via the NUTS code, and accuracy, precision, recall, and F1-score were calculated. The definitions of these metrics are listed below:

Accuracy measures the overall proportion of correct predictions, where TP (true positives) and TN (true negatives) were correctly predicted positive and negative cases, respectively, and FP (false positives) and FN (false negatives) were incorrectly predicted positive and negative cases.:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision indicates how many of the predicted positive instances were actually correct:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall measures how many of the actual positive instances were correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1-score represents the harmonic mean of precision and recall and balanced both metrics:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

For random guessing, the accuracy should have been approximately 50% as the sample size grows, whereas higher values indicated the model's ability to infer the true location latently. When using the ground-truth dataset, in which all geographic references were manually removed, the prediction accuracy was expected to decrease compared to the balanced dataset, where removal was performed by the LLM. This was because in the ground-truth, no explicit geographic cues that could have influenced the predictions remained. If the accuracy remained above 50% and close to that of the balanced dataset, it would indicate that the model could infer location from subtle textual signals and that the LLM's removal of geographic references was effective. For the similarity measures of the embedding results, Euclidean distance, cosine similarity, and cosine distance were used. The Euclidean distance measures the straight-line distance between two vectors $\mathbf{A}$ and $\mathbf{B}$ in an n-dimensional embedding space:

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \tag{5}$$

Cosine similarity measures the cosine of the angle between two vectors and indicates how similar their direction is, being insensitive to the length of the vectors. Here, $\mathbf{A} \cdot \mathbf{B}$ is the dot product of vectors $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are their Euclidean norms (lengths), and the resulting value ranges from -1 (vectors point in opposite directions) to 1 (vectors point in the same direction), with 0 indicating orthogonal vectors.:

$$\text{cosine similarity} = cos(\theta) = (\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{6}$$

Cosine distance is derived from cosine similarity and represents their dissimilarity:

$$\text{cosine distance} = 1 - \text{cosine similarity} \tag{7}$$

4.8 Structure of the final dataset

All experiments were conducted by implementing two SQLite databases to store results and ground truth:

SampledNews_GPT_balanced.db with the 11,856 articles and *SampledNews_balanced_400.db* containing the 400 ground-truth articles.

The database *SampledNews_GPT_balanced.db* had one Table called *SampledArticles* and the column structure visible in Table 9.

Table 9: Schema overview of the balanced dataset.

Column name	**Description**
id	Unique identifier of the article.
url	Original URL of the article.
hostname	Hostname of the website where the article was published.
date	Publication date of the article.
location_id	Identifier of the associated location.
NUTS	NUTS (Nomenclature of Territorial Units for Statistics) region identifier.
loc_normal	Normalized name of the location used for geocoding.
archived_url	URL of the archived version of the article.
html	Raw HTML content of the article webpage.
html_processed	HTML after preprocessing.
article_text_gpt_200	Article text limited to 200 words.
article_text_gpt	Article text cleaned using GPT-5-mini.
article_text_masked_gpt	Article text with location references removed with GPT-5-mini 3-shot.
prediction_gpt	Model prediction of the article's origin based on GPT 3-shot removal.
log_probs	Log probabilities associated with the model prediction GPT 3-shot removal.
embedding	Embedding of the article text based on GPT 3-shot removal.
article_text_masked_gpt_higheffort	Article text with location references removed with GPT-5-mini high reasoning effort removal.
prediction_higheffort_data	Model prediction of the article's origin based on high reasoning effort removal.

The database *SampledNews_balanced_400.db* was derived from the balanced dataset and also had one table called *SampledArticles*. It contained several more columns due to the ground-truth data and the column structure visible in Table 10.

Table 10: Schema overview of the ground-truth dataset.

Column name	**Description**
id	Unique identifier of the article.
url	Original URL of the article.
hostname	Hostname of the website where the article was published.
date	Publication date of the article.
location_id	Identifier of the associated location.
NUTS	NUTS (Nomenclature of Territorial Units for Statistics) region identifier.
loc_normal	Normalized name of the location used for geocoding.
archived_url	URL of the archived version of the article.
html	Raw HTML content of the article webpage.
html_processed	HTML after preprocessing.
article_text_gpt_200	Article text limited to 200 words.
article_text_gpt	Article text cleaned using GPT-5-mini.
article_text_masked_gpt	Article text with location references removed with GPT-5-mini, 3-shot.
prediction_gpt	Model prediction of the article's origin based on GPT 3-shot removal.
log_probs	Log probabilities associated with the model prediction GPT 3-shot removal.
embedding	Embedding of the article text based on GPT 3-shot removal.
article_text_ground_truth	Article text with location references removed manually.
prediction_ground_truth	Model prediction of the article's origin based on manual removal.
prediction_ground_truth_high	Model prediction with high reasoning effort of the article's origin based on manual removal.
embedding_ground_truth	Embedding representation based on the ground-truth text.
article_text_masked_gpt_higheffort	Article text with location references removed with GPT-5-mini high reasoning effort removal.
prediction_higheffort_data	Model prediction of the article's origin based on high reasoning effort removal.
prediction_justification	Model-generated explanation for the prediction based on manual removal.
prediction_justification_embeddings	Embedding representation based on the justifications for the ground-truth text.

4.9 Model selection

For both HTML preprocessing and the removal of geographic references, GPT-5-mini was used. It was the fastest and most cost-efficient model out of the GPT-5 family at the time of writing, finding a balance between computational efficiency and high-level reasoning. It was introduced in August 2025 as a successor to OpenAI's o4-mini model. It was able to adapt its internal reasoning effort based on the complexity of the user's input. Due to its increased context window of 400,000 tokens and reduced operational costs, it was particularly well-suited for large-scale preprocessing tasks. Additionally, developer controls such as the reasoning effort parameter allowed fine-tuning of the trade-off between speed and depth of logic, which was needed for the removal of geographic references. This made it ideal for workflows that required persistent context or long-form text processing (ApX, 2025).

For the prediction and justification, GPT-5.1, as the state-of-the-art model, was used. As the output for the predictions was limited to one token, the higher costs were acceptable. The model excelled at complex reasoning and consideration of context, allowing it to detect subtle linguistic patterns that indicated the likely origin of each newspaper article. Additionally, it allowed for logprobs and temperature settings, which were needed for the experiment setups. Similar to GPT-5-mini, it had a big context window of 400,000 tokens and tended to follow the provided instructions by the user more carefully (Kapuściński, 2025). More details for both models, including input and output costs per 1 million tokens and knowledge cutoff, can be seen in Table 11.

For the embeddings, the text-embedding-3-large model was used as it was the state-of-the-art embedding model for English and non-English tasks. Additionally, the costs were rather low with $0.13 per 1 million tokens. With its 3,072-dimensional vector representation, it was designed to capture small, nuanced contexts and relationships in even larger texts. These produced vectors allowed measuring similarities between large texts, making the model perfectly suitable for the task of comparing multiple newspaper articles (Zilliz Cloud, 2025).

Table 11: The details of the LLMs in this study.

Model	input token costs	output token costs	knowledge cutoff	usecase
GPT 5-mini	$0.25	$2.00	May 31, 2024	text cleaning and removal of geographic references
GPT 5.1	$1.25	$10.00	Sep 30, 2024	prediction and justification

5 Results and empirical findings

In the section, the results from the previously explained processing pipeline and experiments are presented. First, it is evaluated which removal procedure worked best in Section 5.1. Here, the two approaches using 3-shot learning and reasoning effort *"none"* are compared to the results for 20-shot learning and reasoning effort *"high"*. This is done through the Levenshtein distance and the manual analysis of several examples from the dataset. All the removal results are compared to the ground-truth. The study continues with the results for the predictions in Section 5.2, listing the values for accuracy, precision, recall, and F1-score to evaluate whether the performance of the LLMs is better than random guessing. The resulting embeddings with their respective inter and intra-cluster distances are evaluated in Section 5.3, revealing possible linguistic cues without the need for reasoning models. In the last Section 5.4, the justifications from the experiment are evaluated. The analysis begins with the inter and intra-cluster distances as indicators, and then is supplemented by several manually selected examples to further evaluate the justification pattern.

5.1 Performance of geographic reference removal methods

The overall goal of this section is to examine which removal method reports the best performance for German newspaper articles in terms of detecting all geographic references. In general, the removal done with 20-shot learning and reasoning effort *"high"* reveals the best performance. The average Levenshtein distance for the ground-truth dataset, between the ground-truth and the 3-shot GPT removal outputs, is 24.8. This relatively small distance suggests a high degree of similarity and therefore a strong performance, but the manual check shows that several geographic references remain. With the 20-shot learning and reasoning effort *"high"*, the Levenshtein distance is actually higher, with a mean value of 31.9. Although this value reflects a greater surface-level difference between the texts, the manual evaluation shows a clear qualitative improvement. Figure 11 depicts the Levenshtein distances for all 400 articles from the ground-truth dataset for the 3-shot and the high-effort removal. The high-effort removal has a higher median, visible through the orange line, and a higher average, pointed out through the blue dot. Also, the middle 50% of values are more spread out compared to the 3-shot learning. When looking at the outliers, the highest ones can be found for the 3-shot learning, but also the high-effort procedure reports multiple outliers with distances around 125.

The highest Levenshtein distance is 348 for the 3-shot learning, and for this example, in comparison, the high-effort removal results in only 222. However, the manual follow-up checks are even more revealing. For the 3-shot learning, GPT fails to remove several street names listed at the end of the article, showing problems when very local locations appear with only a little context. The corresponding distance for this article with high-effort removal is 59, showing a substantially increased performance. The remaining differences are due to formatting differences. The maximal distance for high-effort removal actually is the result of a manual label error. Here, for example, the place *Gleisdorf* is removed in the manual annotation, but it is actually a village in Austria. The distance for this example for the 3-shot removal is very low, with only 9, because here the Austrian place names are removed. So the high distance in all the cases where words need to be removed actually shows GPTs' good detection capability with reasoning effort *"high"*. Below is another example, which showcases the improvement through the reasoning effort set to *"high"* and using 20-shot learning. The Levenshtein distances are 315 for 3-shot and 115 for high-effort.

- Ground truth: Erdbeerfelder in [...], [...] und Region: Selber pflücken: Alles über Standorte und Öffnungszeiten — [...] Presse Online Egal ob im oder auf dem Kuchen, ob mit Schlagsahne oder Eis oder einfach nur pur – Erdbeeren gehören zum Sommer wie das

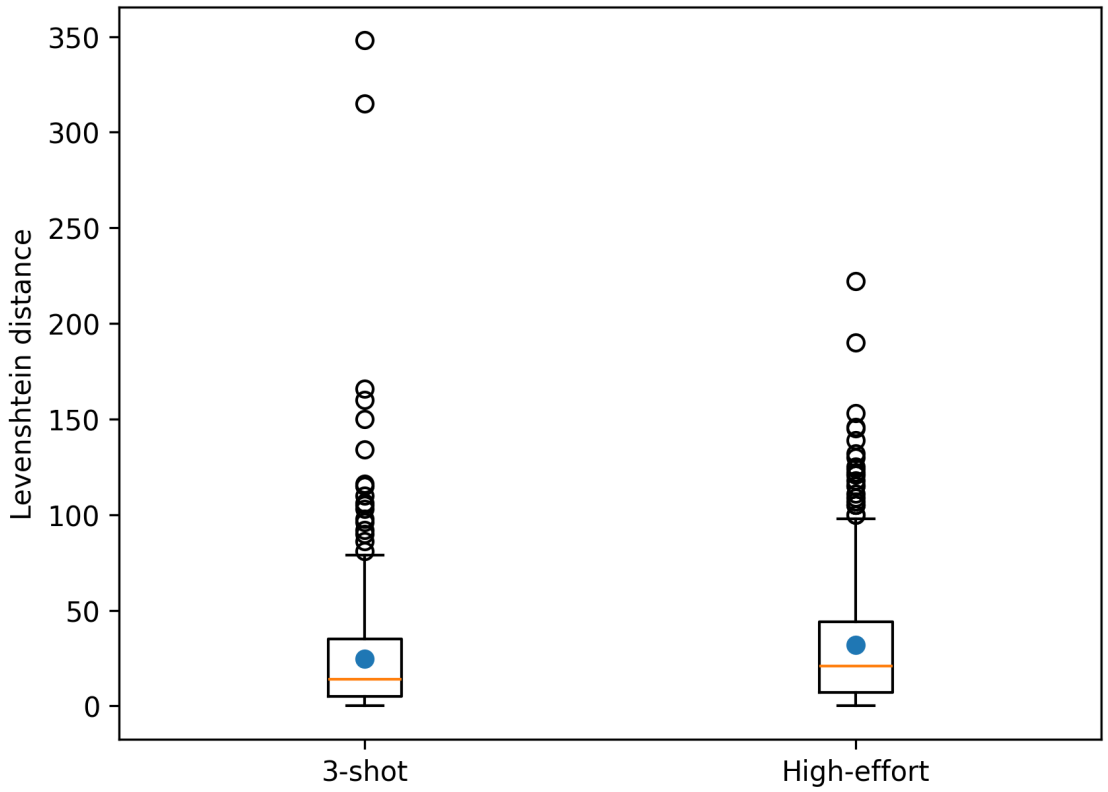


Figure 11: Boxplots showing the Levenshtein distances in comparison to the ground-truth in the ground-truth dataset for the two removal experiments, 3-shot learning and high-effort removal. The orange line represents the median, the blue dots the average, and the black dots outliers.

erfrischende Baden im See und das Grillen im Garten. Und der Sommer ist da! Sattgrüne Felder erstrecken sich dem Auge bei so manchem Ausflug. Darunter: lange Felder voller leuchtend roten Erdbeeren. Die Saison hat begonnen. Mit Körbchen bewaffnet ist es für viele Menschen eine Freude zwischen den meterlangen Reihen entlangzuspazieren und die Beeren selbst zu ernten. Karte zu den Erdbeerfeldern in der Region Haben Sie Lust bekommen, selbst zu pflücken? In diesem Artikel finden Sie eine Übersichtskarte der Standorte und Öffnungszeiten der Erdbeerfelder in [...], dem [...] und dem Landkreis [...]. Vorab eine wichtige Info: Ein Körbchen muss bei den meisten Feldern selbst mitgebracht werden. Erdbeeren Obsthof [...], [...], Richtung [...] Erdbeeren Obsthof [...], [...]-[...], [...] neben [...] Erdbeeren Obsthof [...], [...]-[...], [...] Erdbeeren Obsthof [...], [...]/[...] Erdbeeren Obsthof [...], [...]/[...], Am [...], [...] (mit Himbeer-Verkauf) Erdbeeren Obsthof [...], [...] an der [...], Richtung [...] gegenüber dem [...] Erdbeeren Obsthof [...], [...], Richtung [...], [...] Straße Erdbeeren Obsthof [...], [...]/[...], [...]-[...]-Straße Erdbeerplantage [...], [...]/[...] Erdbeerplantage [...], [...] Erdbeerplantage [...], [...] Erdbeerplantage [...], [...]-[...] Erdbeerhof [...], Feld in [...]/[...] Haben wir ein Feld übersehen?

- 3-shot removal: Erdbeerfelder in [...], [...] und Region: Selber pflücken: Alles über Standorte und Öffnungszeiten — Südwest Presse Online Egal ob im oder auf dem Kuchen, ob mit Schlagsahne oder Eis oder einfach nur pur – Erdbeeren gehören zum Sommer wie das erfrischende Baden im See und das Grillen im Garten. Und der Sommer ist da! Sattgrüne Felder erstrecken sich dem Auge bei so manchem Ausflug. Darunter: lange Felder voller leuchtend roten Erdbeeren. Die Saison hat begonnen. Mit Körbchen bewaffnet ist es für viele Menschen eine Freude zwischen den meterlangen Reihen entlangzuspazieren und die Beeren selbst zu ernten. Karte zu den Erdbeerfeldern in der

Region Haben Sie Lust bekommen, selbst zu pflücken? In diesem Artikel finden Sie eine Übersichtskarte der Standorte und Öffnungszeiten der Erdbeerfelder in [...], dem [...] und dem Landkreis [...]. Vorab eine wichtige Info: Ein Körbchen muss bei den meisten Feldern selbst mitgebracht werden. Erdbeeren Obsthof Zott, [...], Richtung [...] Erdbeeren Obsthof Zott, [...]- Lehr, [...] neben MSZU Erdbeeren Obsthof Zott, [...]- Grimmelfingen, Rathausstraße Erdbeeren Obsthof Zott, [...]/ Ludwigsfeld Erdbeeren Obsthof Zott, [...]/ Pfuhl, Am Schulzentrum, Heerstraße (mit Himbeer-Verkauf) Erdbeeren Obsthof Zott, Pfaffenhofen an der Roth, Richtung Holzheim gegenüber dem Fliederweg Erdbeeren Obsthof Zott, Weißenhorn, Richtung Witzighausen, Ulmer Straße Erdbeeren Obsthof Zott, Illertissen/Tiefenbach, Graf-Kirchberg-Straße Erdbeerplantage Gutmann, Erstetten/Blaubeuren Erdbeerplantage Gutmann, Nersingen Erdbeerplantage Schilling, Ehingen Erdbeerplantage Schilling, [...]- Unterweiler Erdbeerhof Mall, Feld in Schwörzkirch/Allmendingen Haben wir ein Feld übersehen?

- High-effort removal: Erdbeerfelder in [...], [...] und Region: Selber pflücken: Alles über Standorte und Öffnungszeiten — [...] Egal ob im oder auf dem Kuchen, ob mit Schlagsahne oder Eis oder einfach nur pur – Erdbeeren gehören zum Sommer wie das erfrischende Baden im See und das Grillen im Garten. Und der Sommer ist da! Sattgrüne Felder erstrecken sich dem Auge bei so manchem Ausflug. Darunter: lange Felder voller leuchtend roten Erdbeeren. Die Saison hat begonnen. Mit Körbchen bewaffnet ist es für viele Menschen eine Freude zwischen den meterlangen Reihen entlangzuspazieren und die Beeren selbst zu ernten.

  Karte zu den Erdbeerfeldern in der Region

  Haben Sie Lust bekommen, selbst zu pflücken? In diesem Artikel finden Sie eine Übersichtskarte der Standorte und Öffnungszeiten der Erdbeerfelder in [...], dem [...] und dem [...].

  Vorab eine wichtige Info: Ein Körbchen muss bei den meisten Feldern selbst mitgebracht werden.

  Erdbeeren Obsthof [...], [...], Richtung A[...] Erdbeeren Obsthof [...], [...], [...] neben [...] Erdbeeren Obsthof [...], [...], [...] Erdbeeren Obsthof [...], [...]/[...] Erdbeeren Obsthof [...], [...]/[...], Am [...], [...] (mit Himbeer-Verkauf) Erdbeeren Obsthof [...], [...], Richtung [...] gegenüber dem [...] Erdbeeren Obsthof [...], [...], Richtung [...], [...] Erdbeeren Obsthof [...], [...]/[...], [...] Erdbeerplantage [...], [...]/[...] Erdbeerplantage [...], [...] Erdbeerplantage [...], [...] Erdbeerplantage [...], [...] Erdbeerhof [...], Feld in [...]/[...]

  Haben wir ein Feld übersehen?

It is visible that reasoning effort *"minimal"* leads to missing names and addresses of several local strawberry fields and stores. All missed references are highlighted in yellow in the text above. This happens especially at the end of the article, where a list of locations is provided, and context is missing. The model utilizing *"high"* reasoning effort is able to detect these references and remove them. When looking at articles reporting a relatively low Levenshtein distance for the 3-shot learning, it is visible that the deviations are mainly due to single remaining geographic references. Below is an example of an East German article with a Levenshtein distance of ten for 3-shot and zero for high-effort compared to the ground-truth. In this case, only the name of a local sports club is missed in the 3-shot learning, but the high-effort procedure is able to correctly remove it. This missed name would give strong cues about the location of the article's origin and is an important error that needs to be prevented. Many articles reporting relatively low Levenshtein distances for 3-shot learning fail to capture certain key entities. For example,

they often miss the names of newspapers such as *STERN* or abbreviations of company names tied to Germany, such as *BASF*, which stands for *Badische Anilin- & Soda-Fabrik*.

- Ground-truth: Ungewissheit am Sportpark [...]: Beispiel BMX-Club Es gibt nicht nur trübe Botschaften in diesen trüben Zeiten für den Amateursport. Das ist Jens Mühlner wichtig, auch darauf hinzuweisen. Er ist der 2. Vorsitzende des [...] BMX-Clubs, und wenn er zur, sagen wir mal: übersichtlich zufriedenstellenden Situation rund um seinen Sport, seinen Verein und vor allem die Bahn dort im Sportpark [...] befragt wird, dann hat er eben auch diese Botschaft. Der BMX-Sport, seit 2008 olympisch, scheint keineswegs ein sterbender Sport zu sein. Mühlner kann von etlichen Anfragen für ein Schnuppertraining berichten. Auf einer Art Warteliste seien derzeit rund 30 Anmeldungen verzeichnet. Das will etwas heißen, so mitten in einem Sportbetrieb im Standby-Modus. Auf dem [...], wo laut Mühlner vor knapp 40 Jahren die vermutlich erste BMX-Bahn [...] entstand, gibt es seit Monaten nicht mehr viel zu beobachten, was nach Vereinstraining aussehen könnte. Nur maximal zwei Personen dürfen auf die Anlage, mehr gestatten die Corona-Auflagen nicht. Wenn eine Person davon ein Trainer oder eine Trainerin ist, bleibt ein Sportler oder eine Sportlerin, der oder die auf die Bahn gehen kann. 60 aktive BMX-Sportler gebe es im Club, 80 Vereinsmitglieder seien es insgesamt, sagt Mühlner. „Nur mit sich selbst zu fahren, macht auf Dauer ja keinen Spaß“, sagt er.

- 3-shot: Ungewissheit am Sportpark [...]: Beispiel BMX-Club Es gibt nicht nur trübe Botschaften in diesen trüben Zeiten für den Amateursport. Das ist Jens Mühlner wichtig, auch darauf hinzuweisen. Er ist der 2. Vorsitzende des Vegesacker BMX-Clubs, und wenn er zur, sagen wir mal: übersichtlich zufriedenstellenden Situation rund um seinen Sport, seinen Verein und vor allem die Bahn dort im Sportpark [...] befragt wird, dann hat er eben auch diese Botschaft. Der BMX-Sport, seit 2008 olympisch, scheint keineswegs ein sterbender Sport zu sein. Mühlner kann von etlichen Anfragen für ein Schnuppertraining berichten. Auf einer Art Warteliste seien derzeit rund 30 Anmeldungen verzeichnet. Das will etwas heißen, so mitten in einem Sportbetrieb im Standby-Modus. Auf dem [...], wo laut Mühlner vor knapp 40 Jahren die vermutlich erste BMX-Bahn [...] entstand, gibt es seit Monaten nicht mehr viel zu beobachten, was nach Vereinstraining aussehen könnte. Nur maximal zwei Personen dürfen auf die Anlage, mehr gestatten die Corona-Auflagen nicht. Wenn eine Person davon ein Trainer oder eine Trainerin ist, bleibt ein Sportler oder eine Sportlerin, der oder die auf die Bahn gehen kann. 60 aktive BMX-Sportler gebe es im Club, 80 Vereinsmitglieder seien es insgesamt, sagt Mühlner. „Nur mit sich selbst zu fahren, macht auf Dauer ja keinen Spaß“, sagt er.

- High-effort: Ungewissheit am Sportpark [...]: Beispiel BMX-Club Es gibt nicht nur trübe Botschaften in diesen trüben Zeiten für den Amateursport. Das ist Jens Mühlner wichtig, auch darauf hinzuweisen. Er ist der 2. Vorsitzende des [...] BMX-Clubs, und wenn er zur, sagen wir mal: übersichtlich zufriedenstellenden Situation rund um seinen Sport, seinen Verein und vor allem die Bahn dort im Sportpark [...] befragt wird, dann hat er eben auch diese Botschaft. Der BMX-Sport, seit 2008 olympisch, scheint keineswegs ein sterbender Sport zu sein. Mühlner kann von etlichen Anfragen für ein Schnuppertraining berichten. Auf einer Art Warteliste seien derzeit rund 30 Anmeldungen verzeichnet. Das will etwas heißen, so mitten in einem Sportbetrieb im Standby-Modus. Auf dem [...], wo laut Mühlner vor knapp 40 Jahren die vermutlich erste BMX-Bahn [...] entstand, gibt es seit Monaten nicht mehr viel zu beobachten, was nach Vereinstraining aussehen könnte. Nur maximal zwei Personen dürfen auf die Anlage, mehr gestatten die Corona-Auflagen nicht. Wenn eine Person davon ein Trainer oder eine Trainerin ist, bleibt ein Sportler oder eine Sportlerin, der oder die auf die Bahn gehen kann. 60 aktive BMX-Sportler gebe

es im Club, 80 Vereinsmitglieder seien es insgesamt, sagt Mühlner. „Nur mit sich selbst zu fahren, macht auf Dauer ja keinen Spaß“, sagt er.

For articles showing low distances after the high-effort removal, inspection confirmed that these differences do not result from remaining geographic references. Below is an example reporting a Levenshtein distance of zero for 3-shot and 23 for high-effort. The original article text is also provided, as it shows what the foundation of the changed words is. The relevant words are highlighted in yellow in the texts. In the manually created ground-truth, the word *“Bundesgesundheitsministerium”* is completely removed as it refers to the Federal Ministry of Health in Germany. In the 3-shot learning, this is done exactly like in the ground-truth. But the high-effort shows an even cleverer removal as only the beginning of the word, namely the *“Bundes”*, is removed and the generic word *“Gesundheitsministerium”* meaning only the Ministry of Health remains, resulting in the higher distance.

- Original text: Corona-Update: Tests an Schulen und Kitas werden eingestellt

  Die regelmäßigen Corona-Tests an Schulen und Kitas in Deutschland werden eingestellt. Das bestätigte das Bundesgesundheitsministerium auf Anfrage. Demnach entfallen ab kommender Woche die Pflichtuntersuchungen, die bisher zur Corona-Prävention an Bildungseinrichtungen durchgeführt wurden.

  Begründet wird der Schritt mit der sinkenden Infektionslage und der hohen Impfquote in der Bevölkerung. Auch die Verbesserung der Testmöglichkeiten außerhalb der Einrichtungen habe zu der Entscheidung beigetragen.

  Eltern und Erzieher können nach wie vor bei Bedarf Tests durchführen. Vulnerable Personen und Einrichtungen mit erhöhtem Risiko sollen weiterhin Schutzangebote erhalten. Gesundheitsämter behalten die Möglichkeit, bei lokalen Ausbrüchen wieder Teststrategien anzuordnen.

  Kritiker mahnen, dass das Ende der Tests zu einer geringeren Kontrolle über Infektionsgeschehen führen könnte, insbesondere bei neuen Varianten. Befürworter sehen in der Maßnahme einen Schritt zur Rückkehr zur Normalität und zur Entlastung der Bildungseinrichtungen.

- Ground-truth and 3-shot: Corona-Update: Tests an Schulen und Kitas werden eingestellt

  Die regelmäßigen Corona-Tests an Schulen und Kitas in [...] werden eingestellt. Das bestätigte das [...] auf Anfrage. Demnach entfallen ab kommender Woche die Pflichtuntersuchungen, die bisher zur Corona-Prävention an Bildungseinrichtungen durchgeführt wurden.

  Begründet wird der Schritt mit der sinkenden Infektionslage und der hohen Impfquote in der Bevölkerung. Auch die Verbesserung der Testmöglichkeiten außerhalb der Einrichtungen habe zu der Entscheidung beigetragen.

  Eltern und Erzieher können nach wie vor bei Bedarf Tests durchführen. Vulnerable Personen und Einrichtungen mit erhöhtem Risiko sollen weiterhin Schutzangebote erhalten. Gesundheitsämter behalten die Möglichkeit, bei lokalen Ausbrüchen wieder Teststrategien anzuordnen.

  Kritiker mahnen, dass das Ende der Tests zu einer geringeren Kontrolle über Infektionsgeschehen führen könnte, insbesondere bei neuen Varianten. Befürworter sehen in der Maßnahme einen Schritt zur Rückkehr zur Normalität und zur Entlastung der Bildungseinrichtungen.

- High-effort: Corona-Update: Tests an Schulen und Kitas werden eingestellt

  Die regelmäßigen Corona-Tests an Schulen und Kitas in [...] werden eingestellt. Das bestätigte das [...] Gesundheitsministerium auf Anfrage. Demnach entfallen ab kommender Woche die Pflichtuntersuchungen, die bisher zur Corona-Prävention an Bildungseinrichtungen durchgeführt wurden.

  Begründet wird der Schritt mit der sinkenden Infektionslage und der hohen Impfquote in der Bevölkerung. Auch die Verbesserung der Testmöglichkeiten außerhalb der Einrichtungen habe zu der Entscheidung beigetragen.

  Eltern und Erzieher können nach wie vor bei Bedarf Tests durchführen. Vulnerable Personen und Einrichtungen mit erhöhtem Risiko sollen weiterhin Schutzangebote erhalten. Gesundheitsämter behalten die Möglichkeit, bei lokalen Ausbrüchen wieder Teststrategien anzuordnen.

  Kritiker mahnen, dass das Ende der Tests zu einer geringeren Kontrolle über Infektionsgeschehen führen könnte, insbesondere bei neuen Varianten. Befürworter sehen in der Maßnahme einen Schritt zur Rückkehr zur Normalität und zur Entlastung der Bildungseinrichtungen.

The remaining issues in higher Levenshtein distances for high-effort removal are limited to the removal of the generic terms, such as *"Landkreis (English: county)"*, and formatting issues, such as replacing two words with only one bracket or adding an extra space inside the brackets. Other examples show that for sport clubs with *"high"* reasoning effort, not only the place name, but also the *"TSV"*, which is an abbreviation for *"Turn und Sportverein"* translated to *"Gymnastics and sports club"* at the beginning of the name, is removed. This also happened with *"FC"* and *"SVG"*, which also do not specifically refer to a location and are general descriptions for sports clubs. In the ground-truth and 3-shot learning, these abbreviations stay in the text. This also leads to a higher Levenshtein distance for the high-effort setting, but does not create a problem for the further prediction experiments.

Overall, the comparison shows that the reasoning effort *"high"* is able to robustly remove almost all place references while sometimes also removing single generic words such as *"Straße (English: street)"* or *"Landkreis"*, abbreviations that do not refer exclusively to a place, and in one instance, the deletion of an entire sentence instead of only the targeted word. The other deviations appeared to be primarily caused by minor formatting changes rather than substantial content alterations. This validates the good quality of the article texts processed with the *"high"* effort setting and builds a reliable foundation for the following experiments.

5.2 Classification performance results

This section aims to evaluate the capability of the GPT-5.1 model to distinguish region differences without any geographic references. The prediction performance is a proxy for this kind of capability. The findings of the prediction results for the three different cases are explained in the following. First, the findings for the prediction using the text cleaned with the 3-shot learning are shown. Second, the prediction results using the text cleaned with 20-shot learning and *"high"* reasoning effort prompt are provided. And lastly, the results for the ground-truth data, performing the prediction once with reasoning effort *"none"* and once with set to *"high"* are shown.

For the prediction based on the data from 3-shot learning, the accuracy is 80.91%. The precision and recall are specific to either East or West and differ substantially between the two groups. West articles achieve a recall of 0.96, whereas East articles achieve only a recall of 0.66. This imbalance is further reflected in the absolute number of errors, with 1,779 more

East articles being misclassified compared to West articles. It is also visible when looking at the precision. For East articles, precision is very high with a value of 0.94, meaning that when an article is classified as East, this is mostly correct. Nevertheless, the lower recall shows that many articles are missed. For West articles, it is the opposite. Most articles are correctly classified, which is indicated by the high recall, but precision is lower with 0.74. Details can be seen in Table 12.

For the prediction based on the *"high"* effort removal data, the accuracy is 72.47%. When examining class-specific performance, West articles are classified with a higher recall of 0.94, while East articles achieve only 0.51. The values for precision and recall are, in general, lower compared to the 3-shot learning results, but the patterns are similar.

For the prediction based on the ground-truth data (the 400 manually processed articles), the accuracy is 71.50% and, as expected, lower than for the 3-shot learning. This happens due to the better removal, with West articles reporting 0.93 and East ones 0.51 correct classifications. As observed in the previous results, East articles exhibit a higher frequency of misclassification. Results of precision and recall show the same patterns with lower values compared to 3-shot learning. Nevertheless, the accuracy and recalls are still over 50% for East and West articles and are more similar to the results of the *"high"* effort prompt.

When setting reasoning effort to *"high"*, the accuracy is higher with a value of 73.50% for the 400 ground-truth articles, with an East recall of 0.51 and 0.96 for West. Also, the values for precision and F1-score increased slightly. Overall, the prediction for 46 articles changed compared to the reasoning effort *"none"*. For the articles, 27 errors can be corrected, while 19 former correct classifications become incorrect. The last 7% of wrongly predicted West articles with reasoning effort *"none"* are probably the most difficult cases. Therefore, the 3% improvement through *"high"* reasoning effort remains a significant achievement.

Table 12: Combined confusion matrix measurements for predictions.

Region	Setting	Overall Accuracy	Precision	Recall	F1-score
East	3-shot	0.8091	0.94	0.66	0.78
West			0.74	0.96	0.83
East	High effort removal	0.7247	0.89	0.51	0.65
West			0.66	0.94	0.77
East	Ground-truth, effort none	0.7150	0.87	0.51	0.64
West			0.65	0.93	0.76
East	Ground-truth, effort high	0.7350	0.94	0.51	0.66
West			0.66	0.96	0.78

5.3 Evaluation of embedding space and cluster structure

In this section, the results for the embedding experiment are presented, aiming to explore whether embeddings can detect differences in the article's origin. For each article, a 3072-dimensional vector is generated with OpenAI's *text-embedding-3-large model*. The results for the inter-cluster distances between the East and West articles centroids (5,928 articles per cluster), and between two clusters, containing 50% East and 50% West randomly sampled articles, are reported here. For each cluster, both the centroid of the embedding vectors and the Euclidean distance between the centroids within embedding space are calculated. The Euclidean Distance between the East and West cluster centroids with a value of 0.095 is higher than the distance between the mixed cluster centroids with 0.015. Correspondingly, the cosine similarity is higher between the mixed clusters. Continuing with the inter-cluster comparison using only correctly identified articles, initially, it needed to be ensured that the cluster centroids are calculated

from the same number of vectors. Therefore, only as many correct West articles were chosen as are present in East. This results in a cluster size of 3,907 articles for this example. The difference between pure East and West clusters is larger. The Euclidean distance is 0.128, which is higher compared to using all articles. And, consequently, the cosine similarity is lower with a value of 0.9765. Additionally, the same pattern observed in the previous analysis persists. The distance between mixed clusters remains lower than between pure clusters.

In the next step, when using only misclassified articles, the cluster size drops to 242 to ensure the same number of East and West articles. Again, the pattern stays the same, the distance is higher between pure clusters than mixed ones, reporting even a slightly higher value compared to using all articles. To check the results, the inter-cluster comparison is conducted for the embeddings of the 400 ground-truth articles. And it shows the same patterns. The Euclidean distance between pure East and pure West clusters is 0.101, and therefore higher than the distance between the two randomly sampled mixed clusters, with 0.079. Cosine similarity is higher between the mixed clusters and lower for the pure clusters. The detailed results are listed in Table 13.

Table 13: Inter-cluster comparison across different datasets.

Dataset	Comparison	Euclidean distance	Cosine similarity
All articles	East vs. West	0.095	0.9865
All articles	Mixed	0.015	0.9997
Correct articles	East vs. West	0.128	0.9765
Correct articles	Mixed	0.020	0.9994
Misclassified articles	East vs. West	0.114	0.9826
Misclassified articles	Mixed	0.083	0.9904
Ground-truth articles	East vs. West	0.101	0.9857
Ground-truth articles	Mixed	0.079	0.9912

When comparing the inter-cluster distances to the intra-cluster distances, it shows that the mean intra-cluster distance is 0.8143 for East articles and 0.8145 for West articles. This indicates that the variation of articles within each region is larger than the variation between regions. Overall, the semantic differences appear to be small, suggesting that article topic likely plays a stronger role in shaping the embeddings than regional origin. Since the intra-cluster distances for East and West are nearly identical, neither region exhibits systematically higher semantic diversity than the other.

The nearest-neighbor analysis is conducted on the embeddings of misclassified articles to determine whether these articles are more similar to other articles of their predicted class than to those of their true class. The dataset contains a total of 2,263 misclassified articles. The results show that only 2.61% of these articles are closer, in terms of cosine distance, to the predicted class than to their true class. This indicates that GPT misclassifications are not generally driven by semantic proximity to the opposite region. Instead, embeddings seem to clearly reflect the correct region, as misclassified articles are also closer to their correct region. Further strengthening these findings are the results from the trained classifiers. The achieved accuracy for the logistic regression classifier is 71% and for SVM 70%. These percentages are stable across precision, recall, and F1-score for both East and West articles. This pattern becomes even clearer when applying 5-fold cross-validation for the SVM. Here, the mean accuracy increases to 77%. So even though the pure East and West clusters show a huge overlap, a relatively simple classifier compared to the complex LLMs can still separate them and detect the small differences in the high-dimensional embeddings. These results imply that the newspaper articles used in the experiment indeed contain latent linguistic signals that make separating East and West possible using only their embeddings.

5.4 Evaluation of model justifications and reasoning patterns

In this section, the results for the justification experiment are provided, aiming to help answer the question of how well the explanations generated by the LLM match their predictions. 977 justifications per cluster are used to ensure an equal amount of East and West justifications. The Euclidean inter-cluster distance between the pure East and West clusters is 0.2541, compared to 0.0274 for the two mixed clusters. This suggests that the pure clusters are more clearly separated in the embedding space than the mixed ones. Within clusters, the intra-cluster distances indicate how dispersed the articles' justifications are. The intra-cluster distance is substantially higher compared to the inter-cluster distance, with values of 0.7332 for the East and 0.6743 for the West cluster. The mixed clusters show similar intra-distances with both reporting 0.71.

The results highlight that the article justification texts are similar in a general semantic sense, but contain small differences. These findings indicate that reasoning could be different for East and West predictions, as the inter-cluster distance is higher when comparing the two pure clusters. There are 114 wrongly predicted articles in the ground-truth dataset and 227 right justifications for these wrongly predicted articles. This shows that there are articles that get different predictions and justifications in each run, and probably contain ambiguous cues. From all 400 articles, 59 get mixed answers. Continuing with these ambiguous predicted articles, the within-article embedding separation analysis reports an average mean-distance across these articles of 0.5940. This value is higher than the value between the centroids using all justifications across all articles. Hence, if an article gets both East and West prediction answers, the reasoning seems to be different as well. Interestingly, the distances between East and West cluster centroids are very similar when comparing correctly and incorrectly classified articles (correct: 0.5883, incorrect: 0.6035), indicating that prediction correctness alone does not strongly determine the internal semantic spread of justifications. The manual checks provide more interesting insights into the words and phrases used for the justifications. For example, an East German article, which is correctly classified by GPT-5.1 in the previous experiment, gets four justifications for East and six for West. Two of these justifications in their Python dictionary format, as provided by the LLM, are shown as items below. The Euclidean distance between the East and West cluster centroids for this article is 0.3103.

- "prediction": "w", "justification": "The text is a generic police press release with no clear regional markers such as specific place names, dialect, or institutional terms strongly tied to East Germany. Terms like "Klinikum" and "Polizeirevier" are used nationwide. With no distinguishing context, a West Germany origin is a cautious default guess."

- "prediction": "e", "justification": "The text is a neutral police report, but the usage of terms like Klinikum and the overall style/wording is slightly more typical for police press releases from states in the former East Germany. No strong West-specific regional markers (like certain city names or dialect terms) appear, so East is marginally more likely."

Here, the reasoning seems to reflect a difficult decision for an article that does not contain clear references. Interestingly, in one justification, the term *"Klinikum (English: Clinic)"* is seen as a cue for East Germany, and in the other justification it is seen as a word used nationwide. Both words are highlighted in yellow in the items above. This can also be found in another article with terms such as *"Freiwillige Feuerwehr (English: Volunteer fire department), Polizeidirektion (English: Police department), B-road, Skoda"*. Two justifications for this East German article are provided below, which was also correctly classified. Still, four East and six West justifications are reported. The mean Euclidean distance between the East and West cluster centroids for this article is 0.4183.

- "prediction": "w", "justification": "No clear regional markers (city names, dialect, or specific institutions) remain. The terminology (Freiwillige Feuerwehr, Polizeidirektion, Promilleangabe) and narrative style are generic for local accident reporting across Germany. With no East-specific cues, a West Germany origin is slightly more likely given population distribution."

- "prediction": "e", "justification": "The article offers few clear regional markers; names and institutions are used nationwide. However, the combination of a Skoda vehicle, rural highway setting, and the specific police communication style (Polizeidirektion with detailed press spokesperson quote) slightly leans toward typical reporting patterns seen in local East German (former DDR) media."

In other articles, the same phenomena of ambiguous justifications can be find with the term *"Ortslage (English: Location)"* or names of individual people. When examining articles reporting only East or West justifications, it shows that East German predictions are consistently strongly anchored and are always justified by relations to specific words. Several examples of justifications made for East Germany are shown below, where the specific words are highlighted:

- "prediction": "e", "justification": "The use of the league name Landesklasse strongly suggests an East German federal state (e.g., Saxony, Thuringia, Saxony-Anhalt), where this term is common in amateur football structures. In former West German states, lower amateur tiers are usually called Landesliga or Bezirksliga instead."

- "prediction": "e", "justification": "The article lacks clear geographic markers, but the use of Kameraden for firefighters and the overall phrasing is more typical of local reporting style in many East German regional papers. Given only stylistic hints and no contrary evidence, East Germany (former DDR territory) is slightly more likely."

- "prediction": "e", "justification": "The article mentions Landtagsabgeordneter Guido Heuer. Guido Heuer is a known CDU state parliament member in Saxony-Anhalt, which lies in the former East Germany (DDR). Therefore, the broadband expansion described is taking place in a municipality located in ex-GDR territory, not in former West Germany."

- "prediction": "e", "justification": "Harbor setting with a Stadthafen, a burned large half-timbered storage building, and family names like Bentzien and Niemann strongly suggest a small Hanseatic or Baltic port town in Mecklenburg-Vorpommern, an area entirely in former East Germany (DDR). The narrative style also matches typical local East German newspaper features."

- "prediction": "e", "justification": "The context suggests German sidecar racing with tracks like Most (in Czechia) and a nearby Heimrennen. Sidecar and motorcycle road racing of this kind, plus surnames like Roscher, are strongly associated with the Sachsenring/Thuringia/Saxony scene in former East Germany, making an East German origin more likely."

- "prediction": "e", "justification": "The article mentions the research center ZALF and names like Prof. Frank Ewert and Muencheberg-related context. ZALF (Leibniz-Zentrum fuer Agrarlandschaftsforschung) is located in Muencheberg, Brandenburg, which lay in the former GDR (East Germany), indicating the article is from East German territory."

- "prediction": "e", "justification": "Use of the abbreviation PR (Polizeirevier) for the police unit and the detailed, formal GdP-style wording are characteristic of police press

releases from eastern states such as Sachsen-Anhalt, Brandenburg, or Mecklenburg-Vorpommern. Western states more often use PI (Polizeiinspektion) or other designations in such reports."

- "prediction": "e", "justification": "The structure and terminology match police press releases from East German states, especially the use of Landespolizeiinspektion (LPI) and PI [...] together. This format is characteristic of state police communications in former GDR regions like Thuringia or Saxony-Anhalt, and less typical for former West German states."

All these justifications are anchored on specific words, names, and terms in the texts, such as *"Kameraden (English: Comrades, often used in a military context)"* or *"Landesliga (English: National league)"*. Also, the last two examples exhibit ambiguous justifications: once the abbreviation *"PI"* for police inspection is treated as a West German cue, and once as more East German. In some other articles, it makes the prediction based on company names such as *"ZALF"* in the examples above, which were missed in the manual removal process, and gives direct indicators to the location. For articles that only predict West Germany, the justifications are more general, and fewer remaining geographic references or names are used as justifications. Below are several examples from such articles:

- "prediction": "w", "justification": "The text is a generic German police report with all location indicators anonymized. No lexical, institutional, or geographic clues remain that clearly point to former East Germany. In absence of distinctive markers (e.g., typical East German town names), a guess is required; I choose West Germany probabilistically."

- "prediction": "w", "justification": "The text is a generic German police report with no clear regional markers, historical context, or dialect indicating East Germany. In absence of place names or East-specific hints, statistically most such articles originate from former West German territory, so West Germany is the more likely origin."

- "prediction": "w", "justification": "No explicit geographic or institutional markers for East vs. West are present due to anonymization. The focus on national-level campaigns, tourism-related venues like a beach bar and holiday center, and the generic style of PR/association communication slightly better matches typical West German local reporting, but this is essentially a guess."

- "prediction": "w", "justification": "The article describes a local ice arena with events like Eisdisco and Weihnachtsmarkt but no references to cities or regions. The dialect and event names (e.g., Setzbuegeleisschiessen, Schlaegerlaufen) sound more typical of southern/western German leisure culture, slightly favoring former West Germany over East Germany."

- "prediction": "w", "justification": "BMX as a sport spread earlier and more strongly in West Germany, making it more plausible that a track described as vermutlich erste BMX-Bahn dates to former BRD territory. The articles language and context (early BMX development, sport-park setting) give no specifically East German institutional or linguistic markers."

- "prediction": "w", "justification": "The surnames (Welchering, Halbach, Laubenthal, Detloff, Jahnke) and the context (Buergerinitiative against extension of a Bundesstrasse, detailed reference to Bundesverkehrswegeplan) fit typical local reporting from western states like NRW or Lower Saxony; Welchering in particular is a name strongly associated with West Germany."

- "prediction": "w", "justification": "The article uses a colloquial, tongue-in-cheek style with phrases like Irgendwas mit Medien-Typ and a focus on digital check-in and 3G rules in the ICE Bordrestaurant, resembling West German media commentary tone. No regional markers suggest an East German origin; probability favors a West German outlet."
- "prediction": "w", "justification": "The article contains only generic modern German traffic-report vocabulary (Bundesstrasse, Ostumfahrung, Audi, VW, Euro amounts) without any regionally typical East German references (specific towns, police directorates, or road names common in former DDR states). With no clear DDR indicators, West Germany is slightly more probable."

More strikingly, every justification that cites no clear references or terms tends to predict West as the origin. Phrases such as *"West German origin is slightly more probable based on population distribution and generic style"*, *"With no clear cues, a West Germany origin is slightly more likely statistically"*, or *"With no specific East German lexical or place indicators, a West Germany origin is a slightly more probable guess than East"* support this decision. When classifying an East German origin, the model requires very specific linguistic cues or the geographic references that are missed in the removal process. This underscores the findings in Section 5.2, which show the over-proportional assignment of West Germany as the origin.

6 Discussion and interpretation

The discussion is structured in the same way as the results section and divided into four parts. It starts with the methodological gain for the removal of geographic references in Section 6.1. First the shortcomings of existing methods are shown and then the scientific gain that is made with the new method is explained. After that, the new method for the utilization of LLMs with reasoning effort settings in NER is proposed and further justified with previous research. In the next sections, each research question is addressed separately and discussed in the context of further literature. Starting with Section 6.2, the question of whether and to what extend LLMs can detect latent linguistic traces of phantom borders in German newspaper articles is answered. Therefore, the prediction results from the prediction experiment are analyzed, and the statements are validated and discussed with existing literature, introducing the relevance in the fields of Bias and responsibility of GeoAI. In Section 6.3, the second research question is addressed to discuss whether the semantic differences of local language are visible in the textual embeddings. Therefore, the results for the embedding experiment are discussed as proof for existing linguistic traces of the phantom border in the articles and the LLM's results in contrast to the classifiers. And lastly, in Section 6.4, the third research question of whether the justifications of GenAI systems support the predictions is discussed, while also referring deeper to the topic of possible biases.

6.1 Proposing a new method for geographic reference removal

In this thesis, the concept of phantom borders is connected with research on LLM behavior and further investigated. Beyond the experimental results, the methodological advancements during the preprocessing steps represent a significant contribution. In this context, the results of previous studies provided important methodological implications for the approach used in this thesis. Previous research revealed the shortcomings of relying solely on gazetteer-based NER and, more specifically, place name recognition, which forms the basis for the removal task. Overall, traditional methods relied on large annotated datasets and exhibit a high number of errors when identifying entities missing from the training dataset. Furthermore, the production of these benchmark datasets was both time- and cost-intensive (Cheng et al., 2024).

But researchers suggested that hybrid approaches that combine gazetteers, deep learning, and pretrained transformer models could improve performance, achieving, for example, an F1-score of 0.8 for datasets containing multiple tweets. This indicated that enhancing gazetteers with deep learning could improve the NER performance (Hu et al., 2022). As tweets contained unstructured text from local language, similar to the newspaper articles, this supported the development of a new method for NER in this thesis. As discussed previously, building high-quality training data is challenging. For the thesis task, the goal was to apply the new task of removing geographic references to a model with only a small amount of additional training data. This led to the exploration of LLMs, as they support few-shot learning through system prompts. However, when using only short, zero-shot prompts, the results showed that the performance of LLMs for NER remained poor. This implied that standard generation prompts failed to capture the NER training objectives, and more complex, well-structured prompts were needed. This was also investigated by Cheng et al. (2024), who showed that enhancing prompts with a task-definition structure, adding few-shot examples and an output format, and optimizing prompts could significantly improve entity recognition performance. It was tested on GPT-3.5, GPT-4, and Llama2-13B across multiple labeled datasets, showing that the larger models generally performed better without fine-tuning. Additionally, they proposed adding common errors in the system prompts to increase the performance. These findings supported the methodological development in the removal of geographic references as the prompt was continuously developed,

adding common errors, providing examples, and polishing the instructions, leading to better results. The influence of the number of few-shot examples remained ambiguous. Performance did not necessarily increase with the number of provided examples as previous research on NER in historical texts with ChatGPT-4o suggested (Hiltmann et al., 2025). In their experiment, the zero-shot approach outperformed few-shot learning, likely due to the difficulty of providing high-quality historical examples. This also demonstrated that a high-quality instructional prompt was easier for this model to distinguish than overly complex instructions with many examples. When using LLMs for tasks such as NER, it is important to understand that they are designed for text generation rather than assigning labels to specific tokens (Wang et al., 2025). The prompt engineering is therefore one key parameter that influences the quality to ensure good performance with only limited training data. This underlined the need for a new methodological approach in this thesis, as a balance had to be found between clear task instructions, a small number of representative examples, and explicit guidance on typical errors to support the model in identifying geographic entities.

Therefore, this thesis proposes an LLM-based method for place-name extraction and the removal of geographic references, as these models integrate knowledge from implicit lexical resources and showcase broader contextual reasoning. The removal of geographic references first seems like a relatively simple task for the model, as the previous studies suggest. In the thesis, it is shown that GPT-5.1, with a more specific prompt including three examples performs better than pretrained SpaCy pipelines. In the workflow for writing the optimal prompt, common errors are incorporated to improve quality, as previously proposed. The problem is that all of these methods and models primarily focus on retrieving place references from nouns or doing NER in general. The previous experiments show that while the overall detection is good, for a more complex case of removing references in every word type, abbreviations, and distinguishing between national and international, the model reaches its limit. Further experiments showed that the quality differences in the outputs of the removal step have a strong dependence on the model configuration. The introduction of the reasoning models with the parameter reasoning effort in 2024 through the GPT-o series and further developed in the GPT-5 series enables access to the reasoning power of the models before an answer is created (OpenAI Developers, 2026). The results demonstrate that the word removal task is too complex for the model to perform effectively when the reasoning effort is set to *"minimal"*. This shows that good results not only depend on the chosen model alone, but also on prompt design and inference settings. Using the *"high"* reasoning effort setting can improve output quality significantly. Especially when working with long system prompts and complex inputs, the performance increases, although this comes at the cost of increased computational time and higher resource consumption.

These trade-offs must be carefully considered when designing experiments with GPT models. Nevertheless, the simpler processing using only 3-shot learning yields good results and is therefore a practical alternative. With this approach, the proposed methodology addresses challenges related to place-name ambiguity, as LLMs can capture contextual meaning and detect geographic references even when they appear in abbreviations or adjectival forms. Furthermore, the method allows precise specification of which geographic references should be removed.

6.2 Detection of linguistic traces

To answer, it can be said that it is possible for the model to detect the traces, however, the extent varies depending on model settings and article location. The data quality significantly affects prediction accuracy. The more effective the removal, the more difficult it is for the LLM to predict the location correctly. Nevertheless, even under optimal removal conditions, the model performs significantly better than random guessing, which indicates an ability to detect latent linguistic traces of the former border between East and West Germany. The accuracy

is consistently above 70%, and class-specific recalls are also above 50%. The lowest value is reported for East articles based on the *"high"* effort removal data with 51%. Given the large dataset size of 11,856 articles used in the experiment, this is still a significant result. The small experiment predicting with reasoning effort *"high"* on the ground-truth data shows an increase in performance similar to that described in Section 6.1. Taken together, these findings provide an answer to the first research question, suggesting that state-of-the-art LLMs can capture latent linguistic traces of phantom borders in German newspaper articles even when geographic references have been removed. Focusing on the extent to which the model can capture these traces, it needs to be said that the model does not achieve perfect and balanced classification results across East and West Germany. However, the consistent performance above random guessing demonstrates that regional linguistic patterns associated with the former inter-German border remain detectable in the texts. Furthermore, the extent of the ability to distinguish the latent linguistic traces can be increased with more reasoning effort.

Previous research on GenAIs, particularly LLMs, has similarly demonstrated their ability to capture geographic phenomena such as vague cognitive regions (VCRs), which lack objective ground truth and instead rely on human perception (Karimi et al., 2026, under review). The model appeared to be capable of reproducing patterns related to human spatial cognition and processing ambiguous geographic data. These results highlight the potential of LLMs for detecting spatial patterns in textual data, enabling researchers to investigate regional differences in large-scale datasets.

A further aspect worth discussing is the results in the context of Geo-alignment and responsible AI (Janowicz et al., 2025a). The capability of LLMs to detect latent linguistic traces of historical divisions can be seen as an indicator of spatial knowledge, but it also raises several concerns. In the case of East and West Germany, the division has given rise to numerous cultural stereotypes and has further divided Germany, laying the groundwork for persistent socioeconomic disparities and making the issue politically and socially sensitive. If LLMs reproduce or even amplify such a distinction, they could unintentionally increase the influence of this phantom border in the future, contributing to further division in Germany. This conflict raises the question of at what point the behavior of LLMs treating East and West as different is acceptable, and when it hinders social cohesion. For further research, the question is not whether LLMs can detect such regional differences, but under which circumstances their latent recognition becomes socially problematic. It may raise concerns if a user from former East Germany interacts with a widely used chat model, such as OpenAI's ChatGPT, and the system infers their regional background from linguistic cues. As already proposed by Janowicz (2023), AIs are far from neutral when representing geographical space. They function more like distorted mirrors when representing society and rely heavily on training data, incorporating biases. Such inference could lead the model to make implicit assumptions about demographic or socioeconomic conditions of the user, which might in turn influence the responses generated and risk reinforcing existing regional stereotypes. To address this issue, it is important to assess how outputs can avoid amplifying historical divides while preserving analytical validity.

But most striking is the fact that the prediction performance is inconsistent for East and West Germany. While out of scope for this thesis, this may indicate a systematic bias in the prediction performance. Research on geoparsing and geographic information extraction showed that for regions with less available data, the geographic knowledge of models is notably limited. In Liu et al. (2022), they showed that geoparsing models also favored highly populated places and that results were geographically skewed towards some regions, detecting the phenomenon of bias even in earlier methods before LLMs. Manvi et al. (2024) introduced and evaluated geographic bias in LLMs. They described the phenomenon as systematic errors in the models' geospatial predictions. Using geographic data in combination with socioeconomic factors enabled the identification of stereotypes by comparing them with ground-truth data on pop-

ulation density. Even in such objective tasks, a strong underestimation could be detected for India and African countries. Additionally, they examined ratings of residents' attractiveness, morality, intelligence, and work ethic across different countries. Here, the bias appeared even stronger with LLMs consistently providing lower ratings to people in Africa, the Middle East, and South Asia compared to those in North America and Europe. They could also show that these biases manifest at the local level across different areas of a city. Their study suggested that LLMs discriminate against regions with lower socioeconomic conditions. This could imply that existing economic differences between East and West Germany also influence the performance of the model for the proposed task of predicting the origin of newspaper articles. This is consistent with the results, as West is the region with a higher population and stronger economic indicators. But overall, East and West Germany are regions with relatively large amounts of data; it would be interesting to do further analysis with phantom borders lying in less-covered areas. This could show possible limitations of LLMs and their prediction capability, and adds to the broader field of bias analysis in GeoAI. The results of the justification experiment add to these findings, which are further discussed in Section 6.4.

6.3 Separation of articles through embeddings

Answering the second research question, it can be said that high-dimensional embeddings can show semantic differences in articles without reasoning models. This suggests that the LLM's misclassifications are not necessarily caused by the articles being semantically identical. The results are supported by the classifiers, which achieve an accuracy of 77% for the SVM, which is 5% higher than the results obtained using GPT-5.1 with the *"high"* effort geographic reference removal setting. Notably, the SVM performs equally well for both East and West articles. These results imply that the biased predictions towards West Germany are not grounded in inherent similarities within the articles themselves. Unlike LLMs, these classifiers do not infer contextual knowledge and instead make predictions based only on the numerical representations of the text. While the accuracy of LLMs is often expected to surpass embedding-based classifiers because they are able to process complex information, the data reveals a different trend. The data shows a possible distinction between East and West Germany, but the model does not exclusively rely on these cues. So how can it be that a simple classifier outperforms a GenAI model such as GPT-5.1? The SVM learns a distinct decision boundary while the LLM performs a generative inference task and appears to incorporate broader distributional or probabilistic patterns, prior knowledge, or context that favor West Germany's predictions. The embeddings show that phantom borders influence natural language data just like socio-demographic data, with historical spatial divisions continuing to shape contemporary media discourse in measurable ways. This implies that the embedding-based approaches provide a more transparent representation of the linguistic structure of the articles in the example. LLMs do not detect these differences clearly and rather internalize the differences arising from these borders themselves, providing potentially biased answers and predictions.

6.4 Inconsistency of model justifications

The third research question of whether models' justifications can provide insights into the reasoning process cannot be answered with complete certainty. The justification texts generally exhibit different semantic patterns when predicting East or West as the article origin. Similar to the article embeddings, the justifications can be partially separated using cluster distance measurements. However, the intra-cluster distance remains large, indicating that topic variation could influence the textual patterns of the justifications. Therefore, these results do not conclusively demonstrate that the justifications themselves are internally consistent or non-

contradictory. With the manual inspection, it can be confirmed that justifications are similar despite the predictions when looking at the article level and not at patterns across the article collection. As a result, it seems that the model can predict the article's location with an accuracy significantly above chance, but cannot provide stable or coherent explanations for its predictions. This means that the model outputs do not consistently cite the same cues for its decisions, making its justifications unreliable and difficult to interpret. Consequently, humans cannot trust these explanations and cannot learn which textual cues are relevant. The same words are seen as a cue in one justification and as universal, i.e., not distinguishing East and West, in another justification. Asking a model why it picked a specific answer can lead to inconsistent justifications. This effect was already shown by Madsen et al. (2024), answering the question if self-explanations provided by LLM are faithful and whether their justifications really output the model's reasoning process. The authors suggested that the provided explanations were more hallucinatory and could provide misleading cues. They tested the faithfulness of several models, for example, Llama2, Falcon, and Mistral, on different tasks and could show that the results vary between them. They concluded that the trustworthiness of self-explanations is highly model- and dataset-dependent and should not be trusted.

But in this case, the justifications still point to an interesting phenomenon in the provided texts of the model, perhaps providing insights into its encoding of East and West Germany. Every time it cannot find any cues in the article texts, it uses West Germany as a default, arguing with the population size and referring to West Germany as *"a cautious default guess"*. This matches the results from the previous experiments, where GPT showed a tendency to miss many East articles and predict West as the location over-proportionally often. The asymmetry between East and West predictions is grounded in the phenomenon that when the signal is not very strong, it assumes West. This pattern raises questions about how geographic knowledge is encoded in LLMs. For the LLM, the West seems to be more prominent. It raises the question if today's Germany, in the LLMs internal representation, is more similar to the former West than the East. Or is West overly dominant in the training data, and the model sees it as the more relevant location? Therefore, if no strong signal is in the article texts, it assumes West. When stronger regional signals are present in East German texts, the model shows high precision in identifying them, suggesting that East-related linguistic markers are reliable but less frequently detected. This behavior would match the results derived from the prediction experiments. The model favors the geographically more populated and socio-economically advanced region. This can be interpreted from a stochastic point of view. Selecting the most prominent example under uncertainty can be understood as a rational strategy to maximize the probability of a correct prediction. However, while this strategy may be statistically reasonable, it also highlights how underlying data distributions and historical imbalances can shape the model's decision-making process, as the most prominent can, for example, arise from data availability. This shows how stochastic reasoning combined with structural bias can influence model outputs and highlight the importance of critically examining how geographic knowledge is encoded in LLMs.

Geographic knowledge should include all places, not only large and economically advanced regions (Karimi et al., 2026, under review). A decision merely based on these factors questions the fairness of models. This is a term used across different disciplines and investigated by several researchers, which builds the ground for biased decisions (Mai et al., 2025). Prior work from Li et al. (2024) showed that when asking LLMs in different languages, the geographical and geopolitical understanding changed. An example was the very relevant topic, namely, the Crimea, a peninsula internationally recognized as part of Ukraine but annexed by Russia in 2014. When asking with a very simple prompt, which country owns the region, GPT-4 answered Russia when asked in Russian and Ukraine when asked in Ukrainian. The minor adjustments in the prompt, adding a person's perspective, for example, UN Peacekeeper, amplified or neutralized the biases. This strong dependence on the exact prompt and provided context was also re-

ferred to as the brittleness or prompt sensitivity of models. Larger models even reported higher bias and more brittle answers compared to less advanced models (Li et al., 2024). All LLMs that were tested in their study showed such inconsistencies, indicating that they did not learn facts and provide general knowledge but rather encoded different views of phenomena based on language and origin. This matters as LLMs are used for general knowledge tasks, relying on fact-based responses, but providing different answers depending on the user's language and cultural background can increase the polarization of the world's society. For the thesis, this implies that even subtle linguistic differences between East and West German texts could be detected and, as a consequence, could influence model outputs, along with an over-proportional influence of West Germany.

7 Future work and limitations

To the best of the author's knowledge, this thesis is the first study linking research on phantom borders with the latent geospatial representations by LLMs. As the results suggest, this encoded knowledge about linguistic differences in the context of a phantom border is complex and needs further research. It starts with suggestions about possible subsequent studies for the methodological part in Section 7.1, which built the background for all analyses. In Section 7.2, the limitations, shortcomings, and possible research gaps are proposed, suggesting future research in the field of phantom borders in LLMs.

7.1 Prospective research in named entity recognition

The study introduces workflows for using LLMs to execute sophisticated NER tasks, and in this case, especially the detection of geographic references in all word types, including abbreviations and individual names. As proposed in Section 6, the formulations in the prompts strongly influence the model's performance across the settings. In this study, results for 3-shot and 20-shot learning were compared. The research done by Hiltmann et al. (2025) indicated that providing more examples did not necessarily result in better results, as their findings demonstrated that up to a threshold of 16-shot learning, the quality did not increase. Further experiments with providing various combinations of examples should be done to investigate the model's performance. Additionally, in the case of this study, the articles provided as examples for both the 3-shot and 20-shot learning were picked manually, not using any standardized metrics. Using selection strategies as proposed by Cheng et al. (2024) could further increase the performance of the models even for few-shot learning. Not only the examples, but also the standardized system prompt text should be further investigated, as even small changes in wording have a substantial influence on model performances. The experiences of this study with words that are difficult to remove, phrases, and names could be used to develop an even more precise prompt. Further investigating which words were missed, especially with the 3-shot learning, could provide further insights into how the LLMs execute the task and help develop a prompt leading to even better results.

Furthermore, the influence of the parameter reasoning effort is only briefly investigated. As the new models allow for multiple settings, it needs to be shown which configuration provides the best trade-off between costs and performance. In this study, it seems that the setting change to the *"high"* reasoning effort for removal has a substantially larger impact on the performance than merely providing more examples for article texts. When using the complete reasoning power of the model, it enhances the model's ability to handle tasks that require long instructions and need to make difficult decisions. This is also indicated through the increase in performance of the prediction task when using the *"high"* reasoning effort. Research in this area remains limited because previous studies focus primarily on the total reasoning effort expended by a model when faced with substantially difficult tasks. There is a lack of analysis regarding how these performance differences manifest across specific applications (Estermann and Wattenhofer, 2025).

The trade-off between costs and efficiency remains a difficult topic when working with commercial models such as OpenAI's GPT-series. Future research should not only focus on the performance enhancement in general, but rather on finding the optimal configuration for each task, evaluating performance gain against computational time and costs. Strategies for optimal resource allocation, also in the context of AI sustainability, are important and will be even more relevant in the future (Janowicz, 2023).

7.2 Prospective research for phantom borders and LLMs

The results show a possible distinction between East and West Germany through linguistics, especially visible through the high accuracy provided by the simple classifiers. Further research could focus on extracting which words, phrases, and descriptions exactly create the differences captured by the embedding model. This could provide cues on how the differences, which are assumed to be caused by the phantom border between East and West Germany, are encoded in everyday language, showing which linguistic features exactly lead to the separation in embedding space. In the next step, the interest lies in the fact that LLMs seem to miss cues provided through the basic numerical embeddings or that these cues are marginalized by some other learned and encoded knowledge about the world. More specifically, in this case, this means knowledge about East and West Germany. This introduces prospective research on the differences in the success rates between the articles, depending on their location.

Another question is whether the prediction accuracy is also affected by other signals in the texts, probably more topic-related, which should be studied in the future. In this study, no differentiation is made between articles covering different topics. It would be interesting to see if, for example, sport-related articles are easier to classify than political articles, as they may encode more local emotions and understandings. Furthermore, the possible spread of topics in East and West Germany could provide more cues about the phantom border and its representation through LLMs. The differences between the two regions could be influenced not only through the language alone, but also through the thematic associations to either East or West Germany, which might be encoded in the LLM.

As newer models do not necessarily provide better results, further research should check if the differences between East and West German predictions could be more a result of increased brittleness in the newer versions of the models. Studies on the diversity of LLM outputs by Liu et al. (2025) demonstrated that older models actually provided more geographically diverse answers when asked to name specific geographic features. This further states the idea that the newer models do not necessarily perform better across different geographic phenomena and questions.

Due to the conflict already arising from the further separation between East and West Germany and also around the world in regions with such scars of history, it is very important to see the influence of these phenomena in the daily usage of LLMs. As mentioned in Section 6, LLMs do not always provide general answers. The phrasing not only influences their outputs in very complex tasks, but also, when asking questions that have some sort of statistical ground-truth, for example, asking about population sizes or country territories, the answers are highly dependent on the language used in the prompt. This gets even more controversial when asking a question not having any ground-truth, purely relying on human cognition. Further research on phantom borders in linguistic data should focus on how these differences influence possible LLM outputs when asking general questions about travel destinations, possible jobs, living places, and so on. The answers to these questions could be influenced by the assumption that a user is from East or West Germany, when stereotypes still affect the model's answers for people living or originating in these regions. As the justifications suggest, West Germany is picked due to its larger size and population. Designing experiments that would further underpin this assumption that a decision is often merely based on country size and population would be helpful to see if the explanations of the model really reflect the underlying reasoning process. Interestingly, previous research on geographic diversity by Liu et al. (2025) showed that when naming a specific country, models do not always prioritize the largest or most populous ones. This proves that LLMs do not select answers merely based on size. The amount of training data showed a substantial influence when it comes to providing answers as well. Hence, another possible explanation could be a larger amount of data from West Germany in LLM's training data. This would again point to the topic of biased decisions due to an unequal distribution of

data and knowledge around the world.

The entire analysis is performed on a broader geographical scale, only comparing East and West Germany. The newspaper articles reflect texts from across Germany, but no micro-level analysis is conducted. As previous research on phantom borders showed, effects decay with distance from the border. But also, the level at which influences of phantom borders are investigated plays a key role. It would be interesting to compare the differences on a smaller scale, to determine whether the semantic differences vary within, for example, East Germany, and also how the performance of the LLM is affected. The dataset would allow downscaling experiments to the NUTS regions, which could already provide more local differences.

The proposed analysis and experiments are all limited to newspaper articles. For these data sources, the writing style and topics differ from everyday conversations and situations. There are many other sources for local language, especially on social media platforms, which provide access to a large amount of data, potentially capturing linguistic differences between regions. Further analysis with LLMs and how they encode the phenomena of phantom borders could be done using different datasets. This would allow us to further investigate not only the phantom border itself, but also the LLM's behavior when confronted with local language from other datasets and backgrounds.

8 Conclusion

This thesis investigated the capability of a state-of-the-art LLM, GPT-5.1, in detecting latent linguistic traces of the inner German phantom border between East and West. Newspaper articles served as a source of local language to address the three research questions, and each question was answered through the course of this thesis. While the results for the first research question suggested that LLMs can detect latent linguistic traces of phantom borders in newspaper articles, the extent varies depending on the article's origin and specific model settings. Even after the removal of the geographic references, the model consistently performed above random guessing and reached accuracies above 70%. However, the results also revealed an asymmetry between regions, as West German articles were detected significantly more reliably than those from East Germany. The second research question was confirmed in the affirmative, which indicates that semantic differences were visible without the use of reasoning models. When combined with the first finding, this highlighted a potential bias in the LLMs' outputs due to their suboptimal performance compared to traditional embedding-based classifiers. The third research question could only be answered partially. The justifications differed between East and West, while several ambiguous cues appeared across different articles and within the justifications for the same article. Nevertheless, the provided explanations support the observation that West Germany may serve as the default category when clear signals are absent.

In conclusion, these findings indicate that a direct and transparent learning of linguistic cues in local contexts remains challenging. LLMs are not explicitly trained for such tasks, but are still able to capture important and more diffuse geographic phenomena such as phantom borders. The findings suggest LLMs could serve as useful analytical tools for identifying latent linguistic patterns in textual corpora that reflect the persistence of phantom border effects. This could allow researchers to explore hidden spatial structures, especially in unstructured data, complementing traditional approaches in geographic analysis. This is particularly relevant when using spatial data to avoid incorrect assumptions of spatial dependence and to properly assess spatial heterogeneity. Shifting from LLMs as tools to LLMs as study objects themselves, the thesis indicates that the phantom border effects might also influence the model performance and its spatial reasoning. As these LLMs are trained on large datasets, where phantom border effects are encoded, the persistence of phantom borders in language might be embedded in the models' internal representations. This indicates that even more vague concepts of spatial structure can be captured by LLMs. At the same time, this also highlights several risks associated with biased model predictions, which are identified as one of the most important topics in the GeoAI research. These probable biases need to be investigated in future work to improve transparency, to better understand how such biases emerge, and what their impacts on downstream tasks may be. Future work should focus on making potential risks more visible and developing best practices for applying LLMs in spatial and geographic research contexts. Overall, the thesis proposes LLMs not only as a new possibility to capture latent linguistic traces of phantom borders but also to see how phantom borders shape AI and, more specifically, GeoAI systems. The findings suggest that their implementation and interpretation of geographic space is influenced by phantom borders, probably affecting their performance in downstream spatial tasks.

Data availability statement

All scripts that were used to build the database, process the articles, and execute the experiments are available on my GitHub. The complete, balanced, and ground-truth datasets are also available on GitHub and can be downloaded for further use.

List of Tools – Use of AI Tools

A list of tools providing a transparent overview of how AI tools were used in this thesis.

AI Tool	**Purpose**	**Affected Sections**	**Remarks (Examples)**
Grammarly	Writing assistance	All chapters	Grammar and spelling corrections.
Notebook LM	Collecting relevant literature	Section 2 and 6	Only used for an overview of the literature, information in the thesis was derived directly from the papers.
OpenAI ChatGPT 4o	Generation of Python code	Experiments in method Section 4	The codes are provided on GitHub.
OpenAI ChatGPT 4o	Generation of LaTeX code	All chapters	Syntax to format tables, figures, and texts in Overleaf.

References

O. R. Abbasi, F. Welscher, G. Weinberger, and J. Scholz. The World As Large Language Models See It: Exploring the reliability of LLMs in representing geographical features, May 2025. URL http://arxiv.org/abs/2506.00203. arXiv:2506.00203 [cs].

B. Adams and K. Janowicz. On the Geo-Indicativeness of Non-Georeferenced Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 6(1):375–378, Aug. 2021. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v6i1.14309. URL https://ojs.aaai.org/index.php/ICWSM/article/view/14309.

M. A. Adler and A. Brayfield. East-West Differences in Attitudes about Employment and Family in Germany. *The Sociological Quarterly*, 37(2):245–260, 1996. URL http://www.jstor.org/stable/4120809.

L. Anselin. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2): 93–115, 1995. ISSN 1538-4632. doi: 10.1111/j.1538-4632.1995.tb00338.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.1995.tb00338.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1538-4632.1995.tb00338.x.

ApX. GPT-5 Mini: Model Specifications and Details, 2025. URL https://apxml.com/models/gpt-5-mini.

BeautifulSoup. Beautiful Soup Documentation — Beautiful Soup 4.4.0 documentation, 2026. URL https://beautiful-soup-4.readthedocs.io/en/latest/.

Bundeszentrale für politische Bildung. Deutsche Teilung - Deutsche Einheit, Mar. 2009. URL https://www.bpb.de/themen/deutsche-einheit/deutsche-teilung-deutsche-einheit/.

R. D. Caballar and C. Stryker. What Are Foundation Models? | IBM, Oct. 2024. URL https://www.ibm.com/think/topics/foundation-models.

Q. Cheng, L. Chen, Z. Hu, J. Tang, Q. Xu, and B. Ning. A novel prompting method for few-shot NER via LLMs. *Natural Language Processing Journal*, 8:100099, Sept. 2024. ISSN 29497191. doi: 10.1016/j.nlp.2024.100099. URL https://linkinghub.elsevier.com/retrieve/pii/S2949719124000475.

Chronik der Mauer. Bau und Fall der Berliner Mauer, 2016. URL https://www.chronik-der-mauer.de/.

Common Crawl. Common Crawl - Open Repository of Web Crawl Data, 2026. URL https://commoncrawl.org.

R. Egger. Text Representations and Word Embeddings: Vectorizing Textual Data. In R. Egger, editor, *Applied Data Science in Tourism*, pages 335–361. Springer International Publishing, Cham, 2022. ISBN 978-3-030-88388-1 978-3-030-88389-8. doi: 10.1007/978-3-030-88389-8_16. URL https://link.springer.com/10.1007/978-3-030-88389-8_16. Series Title: Tourism on the Verge.

B. Estermann and R. Wattenhofer. Reasoning Effort and Problem Complexity: A Scaling Analysis in LLMs, Mar. 2025. URL http://arxiv.org/abs/2503.15113. arXiv:2503.15113 [cs].

European Commission, editor. *Regions in the European Union: nomenclature of territorial units for statistics, NUTS 2016/EU-28: edition 2018*. Publications Office, Luxembourg, edition 2018 edition, 2018. ISBN 978-92-79-93675-3. doi: 10.2785/573744.

A. Fawzy Gad. Measuring Text Similarity Using the Levenshtein Distance | Paperspace Blog, Mar. 2020. URL https://blog.paperspace.com/measuring-text-similarity-using-levenshtein-distance/.

GeoNames. Geonames, 2026. URL https://www.geonames.org/.

P. Geyer. Phantom Borders: Division in East and West Germany, 2025. URL https://idrn.eu/phantom-borders-the-enduring-influence-of-division-in-east-and-west-germany/. Section: Economic Development.

J. Hills and S. Anadkat. Using logprobs. URL https://developers.openai.com/cookbook/examples/using_logprobs/.

T. Hiltmann, M. Dröge, N. Dresselhaus, T. Grallert, M. Althage, P. Bayer, S. Eckenstaler, K. Mendi, J. M. Schmitz, P. Schneider, W. Sczeponik, and A. Skibba. NER4all or Context is All You Need: Using LLMs for low-effort, high-performance NER on historical texts. A humanities informed approach, Feb. 2025. URL http://arxiv.org/abs/2502.04351. arXiv:2502.04351 [cs].

A. Holl. Temperature & Top-P Parameter in ChatGPT & OpenAI Playground, Mar. 2024. URL https://www.121watt.de/ki/parameter-temperature-top-p-chatgpt-openai-playground/. Section: Künstliche Intelligenz.

E. Holtmann. Deutschland 2020: unheilbar gespalten?: Anmerkungen zur Ost-West-Differenz im 30. Jahr der Wiedervereinigung. *Zeitschrift für Politikwissenschaft*, 30(3):493–499, Sept. 2020. ISSN 1430-6387, 2366-2638. doi: 10.1007/s41358-020-00223-6. URL https://link.springer.com/10.1007/s41358-020-00223-6.

X. Hu, Z. Zhou, Y. Sun, J. Kersten, F. Klan, H. Fan, and M. Wiegmann. GazPNE2: A General Place Name Extractor for Microblogs Fusing Gazetteers and Pretrained Transformer Models. *IEEE Internet of Things Journal*, 9(17):16259–16271, Sept. 2022. ISSN 2327-4662, 2372-2541. doi: 10.1109/JIOT.2022.3150967. URL https://ieeexplore.ieee.org/document/9711571/.

K. Janowicz. Philosophical Foundations of GeoAI: Exploring Sustainability, Diversity, and Bias in GeoAI and Spatial Data Science, Mar. 2023. URL http://arxiv.org/abs/2304.06508. arXiv:2304.06508 [cs].

K. Janowicz, Z. Liu, G. Mai, Z. Wang, I. Majic, A. Fortacz, G. McKenzie, and S. Gao. Whose Truth? Pluralistic Geo-Alignment for (Agentic) AI, Aug. 2025a. URL https://arxiv.org/abs/2508.05432v1.

K. Janowicz, G. Mai, W. Huang, R. Zhu, N. Lao, and L. Cai. GeoFM: how will geo-foundation models reshape spatial data science and GeoAI? *International Journal of Geographical Information Science*, 39(9):1849–1865, Sept. 2025b. ISSN 1365-8816, 1362-3087. doi: 10.1080/13658816.2025.2543038. URL https://www.tandfonline.com/doi/full/10.1080/13658816.2025.2543038.

J. Jańczak. Phantom borders and electoral behaviour in Poland. Historical legacies, political culture and their influence on contemporary politics. *Erdkunde*, 69(2):125–137, June 2015. ISSN 00140015. doi: 10.3112/erdkunde.2015.02.03. URL https://www.erdkunde.uni-bonn.de/article/view/2764.

H. Jiang, H. Hong, Y. Chen, and V. Kulkarni. DialectGram: Detecting Dialectal Variation at Multiple Geographic Resolutions. 2019.

M. Kapuściński. OpenAI GPT-5.1: A Faster, Smarter, More Personal ChatGPT for Business, Nov. 2025. URL https://ttms.com/openai-gpt%e2%80%915-1-a-faster-smarter-more-personal-chatgpt-for-business/.

M. Karimi, K. Janowicz, Z. Liu, S. Wang, and A. Suess. Geo-alignment of vague cognitive regions: Representing uneven cognitive geographies of large language models. 2026. Under review.

V. Kolosov. Phantom borders: the role in territorial identity and the impact on society. *Belgeo*, 2, 2020. ISSN 1377-2368, 2294-9135. doi: 10.4000/belgeo.38812. URL https://journals.openedition.org/belgeo/38812.

L. Kriesch and S. Losacker. A geolocated dataset of German news articles. *Scientific Data*, 12(1):1128, July 2025. ISSN 2052-4463. doi: 10.1038/s41597-025-05422-w. URL https://www.nature.com/articles/s41597-025-05422-w.

D. Lapp, S. Ravada, Q. J. Xie, J. Sharma, H. A. S. Briseno, D. Esparza, and L. Jayapalan. Spatial Dependence, 2025. URL https://docs.oracle.com/en/cloud/paas/autonomous-database/serverless/cspai/spatial-dependence.html.

A. Leemann, C. Derungs, and S. Elspaß. Analyzing linguistic variation and change using gamification web apps: The case of German-speaking Europe. *PLOS ONE*, 14(12):e0225399, Dec. 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0225399. URL https://dx.plos.org/10.1371/journal.pone.0225399.

B. Li, S. Haider, and C. Callison-Burch. This Land is Your, My Land: Evaluating Geopolitical Bias in Language Models through Territorial Disputes. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3855–3871, Mexico City, Mexico, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.213. URL https://aclanthology.org/2024.naacl-long.213.

Z. Liu, K. Janowicz, L. Cai, R. Zhu, G. Mai, and M. Shi. Geoparsing: Solved or Biased? An Evaluation of Geographic Biases in Geoparsing. *AGILE: GIScience Series*, 3:1–13, June 2022. ISSN 2700-8150. doi: 10.5194/agile-giss-3-9-2022. URL https://agile-giss.copernicus.org/articles/3/9/2022/.

Z. Liu, K. Janowicz, I. Majic, M. Shi, A. Fortacz, M. Karimi, G. Mai, and K. Currier. Operationalizing Geographic Diversity for the Evaluation of AI -Generated Content. *Transactions in GIS*, 29(3):e70057, May 2025. ISSN 1361-1682, 1467-9671. doi: 10.1111/tgis.70057. URL https://onlinelibrary.wiley.com/doi/10.1111/tgis.70057.

A. Madsen, S. Chandar, and S. Reddy. Are self-explanations from Large Language Models faithful? In *Findings of the Association for Computational Linguistics ACL 2024*, pages

295–337, Bangkok, Thailand and virtual meeting, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.19. URL https://aclanthology.org/2024.findings-acl.19.

G. Mai, Y. Xie, X. Jia, N. Lao, J. Rao, Q. Zhu, Z. Liu, Y.-Y. Chiang, and J. Jiao. Towards the next generation of Geospatial Artificial Intelligence. *International Journal of Applied Earth Observation and Geoinformation*, 136:104368, Feb. 2025. ISSN 15698432. doi: 10.1016/j.jag.2025.104368. URL https://linkinghub.elsevier.com/retrieve/pii/S1569843225000159.

A. Mansourian and R. Oucheikh. ChatGeoAI: Enabling Geospatial Analysis for Public through Natural Language, with Large Language Models. *ISPRS International Journal of Geo-Information*, 13(10):348, Oct. 2024. ISSN 2220-9964. doi: 10.3390/ijgi13100348. URL https://www.mdpi.com/2220-9964/13/10/348.

R. Manvi, S. Khanna, M. Burke, D. Lobell, and S. Ermon. Large Language Models are Geographically Biased, Feb. 2024. URL https://arxiv.org/abs/2402.02680v2.

J. Mellon and C. Prosser. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3):2053168017720008, July 2017. ISSN 2053-1680, 2053-1680. doi: 10.1177/2053168017720008. URL https://journals.sagepub.com/doi/10.1177/2053168017720008.

H. Mollenkopf and R. Kaspar. Ageing in rural areas of East and West Germany: increasing similarities and remaining differences. *European Journal of Ageing*, 2(2):120–130, June 2005. ISSN 1613-9372, 1613-9380. doi: 10.1007/s10433-005-0029-2. URL http://link.springer.com/10.1007/s10433-005-0029-2.

P. Netrdová, M. Korčák, and V. Nosek. Measuring and mapping the existence of phantom borders at a local scale: example of Sudetenland in Czechia. *Journal of Maps*, 20(1):2300260, Dec. 2024. ISSN 1744-5647. doi: 10.1080/17445647.2023.2300260. URL https://www.tandfonline.com/doi/full/10.1080/17445647.2023.2300260.

Nominatim. Nominatim, 2026. URL https://nominatim.org/.

M. Obschonka, F. Wahl, M. Fritsch, M. Wyrwich, P. J. Rentfrow, J. Potter, and S. D. Gosling. Roma Eterna? Roman rule explains regional well-being divides in Germany. *Current Research in Ecological and Social Psychology*, 8:100214, 2025. ISSN 26666227. doi: 10.1016/j.cresp.2025.100214. URL https://linkinghub.elsevier.com/retrieve/pii/S2666622725000012.

J. Ogden, E. Summers, and S. Walker. Know(ing) Infrastructure: The Wayback Machine as object and instrument of digital research. *Convergence: The International Journal of Research into New Media Technologies*, 30(1):167–189, Feb. 2024. ISSN 1354-8565, 1748-7382. doi: 10.1177/13548565231164759. URL https://journals.sagepub.com/doi/10.1177/13548565231164759.

OpenAI. text-embedding-3-large Model | OpenAI API, 2026. URL https://developers.openai.com/api/docs/models/text-embedding-3-large.

OpenAI Developer Community. How do logprobs work for chat completion API (for GPT-4.1) - API, Sept. 2025. URL https://community.openai.com/t/how-do-logprobs-work-for-chat-completion-api-for-gpt-4-1/1357590/2. Section: API.

OpenAI Developers. Reasoning models | OpenAI API, 2026. URL https://developers.openai.com/api/docs/guides/reasoning/.

K. O'Sullivan, N. R. Schneider, and H. Samet. Metric Reasoning in Large Language Models. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, pages 501–504, Atlanta GA USA, Oct. 2024. ACM. ISBN 979-8-4007-1107-7. doi: 10.1145/3678717.3691226. URL https://dl.acm.org/doi/10.1145/3678717.3691226.

A. Plewnia and A. Rothe. Eine Sprach-Mauer in den Köpfen? Über aktuelle Spracheinstellungen in Ost und West pp. 2009.

N. Reimers and I. Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, Aug. 2019. URL http://arxiv.org/abs/1908.10084. arXiv:1908.10084 [cs].

L. Rensmann. Divided We Stand. *German Politics and Society*, 37(3):32–54, Sept. 2019. ISSN 1045-0300, 1558-5441. doi: 10.3167/gps.2019.370304. URL http://berghahnjournals.com/view/journals/gps/37/3/gps370304.xml.

SAP. How to work with Logprobs that come from the Open AI models in the SAP Generative AI Hub, June 2024. URL https://community.sap.com/t5/technology-blog-posts-by-sap/how-to-work-with-logprobs-that-come-from-the-open-ai-models-in-the-sap/ba-p/13720203. Section: Technology Blog Posts by SAP.

SBERT. SentenceTransformers Documentation — Sentence Transformers documentation, 2026. URL https://www.sbert.net/.

J. Schneider, C. Meske, and P. Kuss. Foundation Models: A New Paradigm for Artificial Intelligence. *Business & Information Systems Engineering*, 66(2):221–231, Apr. 2024. ISSN 2363-7005, 1867-0202. doi: 10.1007/s12599-024-00851-0. URL https://link.springer.com/10.1007/s12599-024-00851-0.

SpaCy. spaCy · Industrial-strength Natural Language Processing in Python, 2026. URL https://spacy.io/.

Statistikportal. Wahlen | Statistikportal.de. URL https://www.statistikportal.de/de/wahlen.

W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240, 1970. ISSN 00130095, 19448287. doi: 10.2307/143141. URL http://www.jstor.org/stable/143141.

A. Van Hoorn and R. Maseland. Cultural differences between East and West Germany after 1991: Communist values versus economic performance? *Journal of Economic Behavior & Organization*, 76(3):791–804, Dec. 2010. ISSN 01672681. doi: 10.1016/j.jebo.2010.10.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0167268110002118.

B. Von Hirschhausen, H. Grandits, C. Kraft, D. Müller, and T. Serrier. Phantom Borders in Eastern Europe: A New Concept for Regional Research. *Slavic Review*, 78(2):368–389, 2019. ISSN 0037-6779, 2325-7784. doi: 10.1017/slr.2019.93. URL https://www.cambridge.org/core/product/identifier/S0037677919000937/type/journal_article.

S. Von Löwis. Phantom borders in the political geography of East Central Europe: an introduction. *Erdkunde*, 69(2):99–106, June 2015. ISSN 00140015. doi: 10.3112/erdkunde.2015.02.01. URL https://www.erdkunde.uni-bonn.de/article/view/2762.

F. Wahl. Does European development have Roman roots? Evidence from the German Limes. *Journal of Economic Growth*, 22(3):313–349, Sept. 2017. ISSN 1381-4338, 1573-7020. doi: 10.1007/s10887-017-9144-0. URL http://link.springer.com/10.1007/s10887-017-9144-0.

S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, and G. Wang. GPT-NER: Named Entity Recognition via Large Language Models. 2025.

Y. Wu, Z. Zeng, K. Liu, Z. Xu, Y. Ye, S. Zhou, H. Yao, and S. Li. Text Geolocation Prediction via Self-Supervised Learning. *ISPRS International Journal of Geo-Information*, 14(4):170, Apr. 2025. ISSN 2220-9964. doi: 10.3390/ijgi14040170. URL https://www.mdpi.com/2220-9964/14/4/170.

Y. Yamada, Y. Bao, A. K. Lampinen, J. Kasai, and I. Yildirim. Evaluating Spatial Understanding of Large Language Models, Apr. 2024. URL http://arxiv.org/abs/2310.14540. arXiv:2310.14540 [cs].

A. Yang, C. Fu, Q. Jia, W. Dong, M. Ma, H. Chen, F. Yang, and H. Wu. Evaluating and enhancing spatial cognition abilities of large language models. *International Journal of Geographical Information Science*, 39(9):2009–2044, Sept. 2025. ISSN 1365-8816, 1362-3087. doi: 10.1080/13658816.2025.2490701. URL https://www.tandfonline.com/doi/full/10.1080/13658816.2025.2490701.

Zilliz Cloud. What is text-embedding-3-large?, 2025. URL https://milvus.io/ai-quick-reference/what-is-textembedding3large.

M. Šimon. Measuring phantom borders: the case of Czech/Czechoslovakian electoral geography. *Erdkunde*, 69(2):139–150, June 2015. ISSN 00140015. doi: 10.3112/erdkunde.2015.02.04. URL https://www.erdkunde.uni-bonn.de/article/view/2765.

A 3-shot removal examples

Example 1: Before removal: "Im Landkreis Kulmbach macht sich der Main breit Der starke Regen und das Sturmtief "Burglind" haben den Landkreis Kulmbach weitgehend verschont. Einige Orte sind aber vom Hochwasser betroffen . An den neuralgischen Punkten im Landkreis war für Autofahrer, sofern sie sich an die Verbotsschilder hielten, am Mittwoch kein Durchkommen mehr: Gesperrt waren die Zufahrt nach Ebersbach von der Straße zwischen Ködnitz und Kauerndorf aus sowie die Ortseinfahrt Langenstadt aus Richtung Kulmbach und die Straße bei Feuln. Über die Ufer trat der Weiße Main bereits bei der Forstlasmühle, wo das Wasser in der Nähe der dortigen Anwesen eine enorme Fließgeschwindigkeit erreichte. Auch im gesamten Tal bis nach Trebgast bildeten sich viele kleine und größere Seen."

After removal: Im Landkreis [...] macht sich der [...] breit Der starke Regen und das Sturmtief "Burglind" haben den Landkreis [...] weitgehend verschont. Einige Orte sind aber vom Hochwasser betroffen . An den neuralgischen Punkten im Landkreis war für Autofahrer, sofern sie sich an die Verbotsschilder hielten, am Mittwoch kein Durchkommen mehr: Gesperrt waren die Zufahrt nach [...] von der Straße zwischen [...] und [...] aus sowie die Ortseinfahrt [...] aus Richtung [...] und die Straße bei [...]. Über die Ufer trat der Weiße [...] bereits bei der Forstlasmühle, wo das Wasser in der Nähe der dortigen Anwesen eine enorme Fließgeschwindigkeit erreichte. Auch im gesamten Tal bis nach [...] bildeten sich viele kleine und größere Seen.

Example 2: Before removal: "POL-PDNR: Verkehrsunfälle, Verkehrsunfallflucht Betzdorf, Kirchen, Schutzbach (ots) Am Samstag, den 09.12.2023 gegen 08:15 Uhr wurde der Polizei Betzdorf ein Verkehrsunfall in der Schützenstraße in Betzdorf gemeldet. Ein auf dem Gehweg geparkter Renault wurde auf der abschüssigen Straße nicht ordnungsgemäß gesichert und rollte auf der abschüssigen Straße gegen einen hinter ihm ebenfalls auf dem Gehweg geparkten BMW. AN dem BMW entstand Sachschaden in einer Höhe von ca. 1000EUR. Da beide Fahrzeuge jedoch zuvor verbotswidrig geparkt wurden, wurden beide Fahrer verkehrsrechtlich sanktioniert. Am Sonntag, den 10.12.2023 gegen 02:00 Uhr kam eine 18-jährige Fahrerin eines Peugeot auf der L 280 zwischen Alsdorf und Schutzbach auf regenasser Fahrbahn aufgrund nicht angepasster Geschwindigkeit ins Schleudern und überschlug sich. Die allein im Fahrzeug befindliche Fahrerin wurde nicht verletzt. Der PKW war allerdings nicht mehr fahrbereit. Am Samstag, den 09.12.2023 kam es gegen 14:50 Uhr zu einer Verkehrsunfallflucht in der Brückenstraße in Kirchen."

After removal: "[...]: Verkehrsunfälle, Verkehrsunfallflucht [...], [...], [...] ([...]) Am Samstag, den 09.12.2023 gegen 08:15 Uhr wurde der Polizei [...] ein Verkehrsunfall in der [...] in [...] gemeldet. Ein auf dem Gehweg geparkter Renault wurde auf der abschüssigen Straße nicht ordnungsgemäß gesichert und rollte auf der abschüssigen Straße gegen einen hinter ihm ebenfalls auf dem Gehweg geparkten [...]. AN dem [...] entstand Sachschaden in einer Höhe von ca. 1000EUR. Da beide Fahrzeuge jedoch zuvor verbotswidrig geparkt wurden, wurden beide Fahrer verkehrsrechtlich sanktioniert. Am Sonntag, den 10.12.2023 gegen 02:00 Uhr kam eine 18-jährige Fahrerin eines Peugeot auf der [...] zwischen [...] und [...] auf regenasser Fahrbahn aufgrund nicht angepasster Geschwindigkeit ins Schleudern und überschlug sich. Die allein im Fahrzeug befindliche Fahrerin wurde nicht verletzt. Der PKW war allerdings nicht mehr fahrbereit. Am Samstag, den 09.12.2023 kam es gegen 14:50 Uhr zu einer Verkehrsunfallflucht in der [...] in [...]."

Example 3: Before removal: " Dennis Diekmeier schießt erstes Profitor seiner Karriere Da gratuliert sogar der Ex-Verein: Nach einem Dutzend Jahren im deutschen Profifußball ist Den-

nis Diekmeier sein erstes Tor gelungen. Der Verteidiger schoss das Siegtor im Spiel seines SV Sandhausen gegen den SV Wehen Wiesbaden. Der HSV twitterte dazu, das habe sich ja "schon lange abgezeichnet", man freue sich mit ihm."

After removal: "Dennis Diekmeier schießt erstes Profitor seiner Karriere Da gratuliert sogar der Ex-Verein: Nach einem Dutzend Jahren im deutschen Profifußball ist Dennis Diekmeier sein erstes Tor gelungen. Der Verteidiger schoss das Siegtor im Spiel seines SV [...] gegen den SV [...] [...]. Der [...] twitterte dazu, das habe sich ja "schon lange abgezeichnet", man freue sich mit ihm."

B 20-shot removal examples

The following golddataset provides training data. Each dictionary contains the article before and after the desired removal. golddataset = [

"before1": "Netzverteiler symbolisch aktiviert Der Breitband-Ausbau im Sülzetal ist abgeschlossen. Er wurde ohne Fördermittel umgesetzt. 03.03.2019, 14:29 Osterweddingen l Mit einem öffentlichen, symbolischen Druck auf den roten Knopf an der Ecke Bahnhofstraße/Feldstraße in Osterweddingen nahmen kürzlich der Geschäftsführer der „Mitteldeutschen Gesellschaft für Kommunikation" (MDDSL), Andreas Riedel, der Referatsleiter aus dem Magdeburger Wirtschaftsministerium, Theodor Struhkampf, der Vorsitzende des Gemeinderats des Sülzetals und CDU-Landtagsabgeordnete, Guido Heuer, und der Wirtschafts- und Entwicklungsplaner der Gemeinde Sülzetal, Fred Fedder, den letzten Netzverteiler von MDDSL in der Gemeinde Sülzetal in Betrieb. Somit ist das Gebiet der Gemeinde mit „Breitband + 50 Mbit" ausgebaut. Gemeinsam haben die Telekom und die MDDSL, jeweils ohne Inanspruchnahme von Fördermitteln und kommunalen Haushaltsmitteln, die Wunschvorstellung der Bundesregierung für 2018 im Sülzetal gemeistert. „Die Einwohner und Unternehmer im Sülzetal können sich freuen und sollten diesen Erfolg wertschätzen", betonte der Wirtschaftsplaner Fedder bei dem Vor-Ort-Termin. „Wir haben den Anschluss mit allen uns zur Verfügung stehenden Mitteln forciert", erklärte Geschäftsführer Riedel, als der Netzverteiler leise vor sich hin brummte. „Wir danken der Gemeinde für ihre Unterstützung. Was wir hier in der Ortschaft Osterweddingen gebaut haben, ist eine absolut moderne Technik. Das war sehr, sehr aufwendig." Der zweite Teil Osterweddingens sei nun an das Giga-Glasnetz angebunden, sodass die MDDSL am Anfang 50 Mbit liefern könne, dann in zwei Wochen 100 Mbit und am Ende, in einem halben Jahr, sogar 200 Mbit.", "after1": "Netzverteiler symbolisch aktiviert Der Breitband-Ausbau im [...] ist abgeschlossen. Er wurde ohne Fördermittel umgesetzt. 03.03.2019, 14:29 [...] Mit einem öffentlichen, symbolischen Druck auf den roten Knopf an der Ecke [...]/[...] in [...] nahmen kürzlich der Geschäftsführer der „ [...] Gesellschaft für Kommunikation" ([...]), Andreas Riedel, der Referatsleiter aus dem [...] Wirtschaftsministerium, Theodor Struhkampf, der Vorsitzende des Gemeinderats des [...] und [...]-Landtagsabgeordnete, Guido Heuer, und der Wirtschafts- und Entwicklungsplaner der Gemeinde [...], Fred Fedder, den letzten Netzverteiler von MDDSL in der Gemeinde [...] in Betrieb. Somit ist das Gebiet der Gemeinde mit „Breitband + 50 Mbit" ausgebaut. Gemeinsam haben die [...] und die [...], jeweils ohne Inanspruchnahme von Fördermitteln und kommunalen Haushaltsmitteln, die Wunschvorstellung der Bundesregierung für 2018 im [...] gemeistert. „Die Einwohner und Unternehmer im [...] können sich freuen und sollten diesen Erfolg wertschätzen", betonte der Wirtschaftsplaner Fedder bei dem Vor-Ort-Termin. „Wir haben den Anschluss mit allen uns zur Verfügung stehenden Mitteln forciert", erklärte Geschäftsführer Riedel, als der Netzverteiler leise vor sich hin brummte. „Wir danken der Gemeinde für ihre Unterstützung. Was wir hier in der Ortschaft [...] gebaut haben, ist eine absolut moderne Technik. Das war sehr, sehr aufwendig." Der zweite Teil [...] sei nun an das

Giga-Glasnetz angebunden, sodass die MDDSL am Anfang 50 Mbit liefern könne, dann in zwei Wochen 100 Mbit und am Ende, in einem halben Jahr, sogar 200 Mbit.",

"before": "Landesklasse-Duell im Altmark-Strom-Pokal Im Altmark-Strom-Pokal der Stadtwerke Stendal geht es am Sonnabend bereits mit dem Achtelfinale weiter. 10.10.2020, 08:00 Stendal l Dabei sticht das Landesklassenduell zwischen dem Osterburger FC und Rot-Weiß Arneburg heraus. Alle sieben Begegnungen beginnen um 14 Uhr. Die Partie Spielvereinigung Havelberg/Kamern gegen den Möringer SV ist erst am 14. November um 13 Uhr . Im Duell der beiden Landesklassenvertreter aus Osterburg und Arneburg ist die Rolle des Favoriten schwer zu finden. Was für den OFC spricht, ist erst einmal das Heimrecht. Zudem gab es das Duell bereits in der Meisterschaft, wo die Biesestädter den Kontrahenten mit einem satten 8:0 auf deren Geläuf förmlich zerlegten. Doch danach hat Arneburg zugelegt, einige Partien in Folge gewonnen. Was den Elbestädtern fehlen könnte, ist das Flutlicht. RWA hat in den letzten Wochen meist am Freitagabend gespielt und gewonnen. In der Vorwoche hat die Dieckmann-Elf Moral bewiesen, aber auch etwas Glück gehabt. In Unterzahl schafften die Elbestädter in der regulären Spielzeit noch spät den Ausgleich bei der SG Bismark II/Kläden, hatten mit Marco Pawellek einen starken Aushilfstorwart, der für den mit Gelb-Rot vom Platz gestellten Tobias Maier zwischen die Pfosten ging. Ob Pawellek erneut ins Tor geht, bleibt abzuwarten, denn es gibt beim Gast da noch andere Personalien.", "after": "Landesklasse-Duell im Altmark-Strom-Pokal Im Altmark-Strom-Pokal der Stadtwerke [...] geht es am Sonnabend bereits mit dem Achtelfinale weiter. 10.10.2020, 08:00 [...] Dabei sticht das Landesklassenduell zwischen dem Osterburger FC und Rot-Weiß [...] heraus. Alle sieben Begegnungen beginnen um 14 Uhr. Die Partie Spielvereinigung [...]/ [...] gegen den [...] SV ist erst am 14. November um 13 Uhr Im Duell der beiden Landesklassenvertreter aus [...] und [...] ist die Rolle des Favoriten schwer zu finden. Was für den OFC spricht, ist erst einmal das Heimrecht. Zudem gab es das Duell bereits in der Meisterschaft, wo die [...] den Kontrahenten mit einem satten 8:0 auf deren Geläuf förmlich zerlegten. Doch danach hat Arneburg zugelegt, einige Partien in Folge gewonnen. Was den [...] fehlen könnte, ist das Flutlicht. RWA hat in den letzten Wochen meist am Freitagabend gespielt und gewonnen. In der Vorwoche hat die Dieckmann-Elf Moral bewiesen, aber auch etwas Glück gehabt. In Unterzahl schafften die [...] in der regulären Spielzeit noch spät den Ausgleich bei der SG [...] II/ [...], hatten mit Marco Pawellek einen starken Aushilfstorwart, der für den mit Gelb-Rot vom Platz gestellten Tobias Maier zwischen die Pfosten ging. Ob Pawellek erneut ins Tor geht, bleibt abzuwarten, denn es gibt beim Gast da noch andere Personalien.",

"before": "Bad Lobenstein (dpa/th) - Touristische Ziele in der Region Rennsteig und Thüringer Meer steuert ab diesem Samstag die neue Thüringer-Meer-und-Rennsteig-Linie ab Bad Lobenstein an. Jeweils samstags, sonntags und an Feiertagen fährt der Wander- und Rad-Bus nach Harra, Blankenstein, Pottiga, Birkenhügel, Frössen und Saaldorf, bevor er nach Bad Lobenstein zurückkehrt, wie der Tourismusverbundes Rennsteig-Saaleland am Freitag mitteilte. Die neue Verbindung ist ein Gemeinschaftsprojekt des Busunternehmens KomBus, der Kommunalen Arbeitsgemeinschaft Thüringer Meer und des Tourismusverbundes. Sie startet in diesem Jahr als Pilotprojekt und kann bei guter Auslastung im kommenden Jahr fest in den Fahrplan aufgenommen werden. Die Erprobungsphase dauert bis zum 31. Oktober.", "after": "[...] ([...]/[...]) - Touristische Ziele in der Region [...] und [...] steuert ab diesem Samstag die neue [...]-und-[-]-Linie ab [...] an. Jeweils samstags, sonntags und an Feiertagen fährt der Wander- und Rad-Bus nach [...], [...], [...], [...], [...] und [...], bevor er nach [...] zurückkehrt, wie der Tourismusverbundes [...] am Freitag mitteilte.Die neue Verbindung ist ein Gemeinschaftsprojekt des Busunternehmens [...], der Kommunalen Arbeitsgemeinschaft [...] und des Tourismusverbundes. Sie startet in diesem Jahr als Pilotprojekt und kann bei guter Auslastung

im kommenden Jahr fest in den Fahrplan aufgenommen werden. Die Erprobungsphase dauert bis zum 31. Oktober.",

"before": "LPI-J: Vier Verletzte nach Vorfahrtsunfall Weimar (ots) Am Donnerstagnachmittag kam es auf der B7, in der Ortslage Frankendorf, zu einem Verkehrsunfall. Ein 20jähriger kam mit seinem Opel aus Richtung Hammerstedt gefahren und wollte in Frankendorf die B7 überqueren. Hierbei missachtete er die Vorfahrt des auf der B7, in Richtung Weimar, fahrenden 72jährigen Toyota-Fahrers, wodurch es zum Zusammenstoß kam. Die jeweils beiden Fahrzeuginsassen der beteiligten Fahrzeuge wurden leicht verletzt, wobei drei davon in umliegende Krankenhäuser verbracht wurden. Beide Pkw waren nicht mehr fahrbereit und mussten abgeschleppt werden. Der Sachschaden beläuft sich auf insgesamt etwa 40 000 Euro. Während der Unfallaufnahme war die B7 zeitweise komplett gesperrt.", "after": " [...]: Vier Verletzte nach Vorfahrtsunfall [...] ([...]) Am Donnerstagnachmittag kam es auf der B [...], in der Ortslage [...], zu einem Verkehrsunfall. Ein 20jähriger kam mit seinem Opel aus Richtung [...] gefahren und wollte in [...] die B [...] überqueren. Hierbei missachtete er die Vorfahrt des auf der B [...], in Richtung [...], fahrenden 72jährigen Toyota-Fahrers, wodurch es zum Zusammenstoß kam. Die jeweils beiden Fahrzeuginsassen der beteiligten Fahrzeuge wurden leicht verletzt, wobei drei davon in umliegende Krankenhäuser verbracht wurden. Beide Pkw waren nicht mehr fahrbereit und mussten abgeschleppt werden. Der Sachschaden beläuft sich auf insgesamt etwa 40 000 Euro. Während der Unfallaufnahme war die B7 zeitweise komplett gesperrt.",

"before": "In vier Sorten Griebenschmalz könnten Verbraucher Reste von Metalldraht finden. Die Gläser werden bei Aldi-Nord verkauft. ©Aldi-Nord 28. November: Metalldraht in Aldi-Griebenschmalz Wegen einer möglichen Gesundheitsgefahr hat die Firma Thalheimer Bauernwurst Deuerlein Vertriebs GmbH mehrere Sorten Griebenschmalz im Bügelglas zurückgerufen. Die Warnung wurde am Mittwoch über das Portal lebensmittelwarnung.de verbreitet. Betroffen ist der Kräuter-, Klassik-, Bacon-, Apfel- und Zwiebelgriebenschmalz des Aldi Nord-Sortiments - jeweils in 100-Gramm-Packungen mit dem Mindesthaltbarkeitsdatum 20. März 2020. Als Grund für den Rückruf nannte das Unternehmen mit Sitz im bayerischen Gebertshofen, dass sich in einzelnen Packungen Metalldraht befinden könne. Der Artikel wurde nach Aldi-Angaben in einer Aktion am 18. November 2019 angeboten. Quelle: DPA 27. November: Kaufland ruft eigene Leberwurst zurück Die Neckarsulmer Supermarktkette Kaufland ruft die Leberwurst ihrer Hausmarke bundesweit zurück. In einer Mitteilung erklärte das Unternehmen, dass mehrere Sorten von "K-Favourites" Leberwurst, 150g verdorben sein könnten und nicht verzehrt werden sollten. Demnach geht es um alle Produkte mit einer Mindesthaltbarkeit 11.12. bzw. 18.12.2019. Betroffen seien folgende Geschmacksrichtungen: Apfel und Zwiebel (GTIN 4337185751202) Geröstete und gemahlene Kaffeebohnen (GTIN 4337185751226) Preiselbeeren (GTIN 4337185751189) Schwarze Trüffel (GTIN 433718575126) Möglicherweise wurde bei den betroffenen Produkten die Kühlkette unterbrochen. Deshalb könne man eine Gefährdung der Gesundheit nach dem Verzehr nicht ausschließen, so die Kaufland Warenhandel GmbH & Co.", "after": "In vier Sorten Griebenschmalz könnten Verbraucher Reste von Metalldraht finden. Die Gläser werden bei [...]-Nord verkauft. ©[...]?Nord 28. November: Metalldraht in [...]-Griebenschmalz Wegen einer möglichen Gesundheitsgefahr hat die Firma[...] Vertriebs GmbH mehrere Sorten Griebenschmalz im Bügelglas zurückgerufen. Die Warnung wurde am Mittwoch über das Portal lebensmittelwarnung.de verbreitet. Betroffen ist der Kräuter-, Klassik-, Bacon-, Apfel- und Zwiebelgriebenschmalz des [...]-Nord-Sortiments - jeweils in 100-Gramm-Packungen mit dem Mindesthaltbarkeitsdatum 20. März 2020. Als Grund für den Rückruf nannte das Unternehmen mit Sitz im [...] [...], dass sich in einzelnen Packungen Metalldraht befinden könne. Der Artikel wurde nach [...]-Angaben in einer Aktion

am 18. November 2019 angeboten. Quelle: [...] 27. November: [...] ruft eigene Leberwurst zurück Die [...] Supermarktkette [...] ruft die Leberwurst ihrer Hausmarke bundesweit zurück. In einer Mitteilung erklärte das Unternehmen, dass mehrere Sorten von "K-Favourites" Leberwurst, 150g verdorben sein könnten und nicht verzehrt werden sollten. Demnach geht es um alle Produkte mit einer Mindesthaltbarkeit 11.12. bzw. 18.12.2019. Betroffen seien folgende Geschmacksrichtungen: Apfel und Zwiebel (GTIN 4337185751202) Geröstete und gemahlene Kaffeebohnen (GTIN 4337185751226) Preiselbeeren (GTIN 4337185751189) Schwarze Trüffel (GTIN 433718575126) Möglicherweise wurde bei den betroffenen Produkten die Kühlkette unterbrochen. Deshalb könne man eine Gefährdung der Gesundheit nach dem Verzehr nicht ausschließen, so die [...] Warenhandel GmbH & Co.",

"before": "Hier haben Telekom, Vodafone und o2 ihre Netze ausgebaut - teltarif.de News Netzausbau 25.02.2022 18:48 Wie immer blicken wir vor dem Start ins Wochenende auf den Ausbau der Mobilfunknetze der Deutschen Telekom, von Vodafone und Telefónica. Vor allem die Telekom hat in den vergangenen Tagen über zahlreiche neue Basisstationen und Aufrüstungen an bestehenden Standorten berichtet. Bei Vodafone geht der 5G-Ausbau weiter, Telefónica forciert 5G neben 3600 MHz nun auch auf 700 und 1800 MHz. Baden-Württemberg Die Deutsche Telekom hat im Landkreis Tuttlingen in den vergangenen zwei Monaten einen Mobilfunk-Standort neu gebaut, einen mit LTE und vier mit 5G erweitert. Vier dieser Standorte befinden sich in Immendingen, dazu kommen Stationen in Trossingen und Tuttlingen. Vodafone hat in Waldkirch und Freiamt im Kreis Emmendingen neue 5G-Mobilfunkstationen in Betrieb genommen. An einem weiteren Vodafone-Standort im Kreis wird die 5G-Technologie bis Mitte 2022 eingebaut. Dieses 5G-Bauprojekt wird in Elzach realisiert. Für Simonswald ist eine neue LTE-Station geplant. In Esslingen (Kelterstraße) und Hemsbach (Gutenbergstraße) hat Telefónica je eine neue 5G-Station auf 3600 MHz in Betrieb genommen. 5G auf 700/1800 MHz gibt es neu in Bad Friedrichshall, Wuach und Radolfzell. In Schwendi hat o2 einen neuen LTE-Sender aufgeschaltet. Bayern Die Deutsche Funkturm GmbH baut neue Sendemasten für das Telekom-Mobilfunknetz in Harburg im Kreis Donau-Ries sowie in Hirschbach im Kreis Amberg-Sulzbach.", "after": "Hier haben [...], [...] und [...] ihre Netze ausgebaut - [...] News Netzausbau 25.02.2022 18:48 Wie immer blicken wir vor dem Start ins Wochenende auf den Ausbau der Mobilfunknetze der [...], von [...] und [...]. Vor allem die [...] hat in den vergangenen Tagen über zahlreiche neue Basisstationen und Aufrüstungen an bestehenden Standorten berichtet. Bei [...] geht der 5G-Ausbau weiter, [...] forciert 5G neben 3600 MHz nun auch auf 700 und 1800 MHz. [...] Die [...] hat im Landkreis [...] in den vergangenen zwei Monaten einen Mobilfunk-Standort neu gebaut, einen mit LTE und vier mit 5G erweitert. Vier dieser Standorte befinden sich in [...], dazu kommen Stationen in [...] und [...]. [...] hat in [...] und [...] im Kreis [...] neue 5G-Mobilfunkstationen in Betrieb genommen. An einem weiteren [...]-Standort im Kreis wird die 5G-Technologie bis Mitte 2022 eingebaut. Dieses 5G-Bauprojekt wird in [...] realisiert. Für [...] ist eine neue LTE-Station geplant. In [...] ([...]) und [...] ([...]) hat [...] je eine neue 5G-Station auf 3600 MHz in Betrieb genommen. 5G auf 700/1800 MHz gibt es neu in [...], [...] und [...]. In [...] hat o2 einen neuen LTE-Sender aufgeschaltet. [...] Die [...] baut neue Sendemasten für das [...]-Mobilfunknetz in [...] im Kreis [...] sowie in [...] im Kreis [...].",

"before": "News: Ravensburg - Baby mit Küchenpapier erstickt - 23-Jährige muss lebenslang in Haft — STERN.de Union und SPD wollen Klimaziel aufgeben+++ Brand in Trump-Tower +++ Angeklagter gesteht Anschlag auf BVB +++ Öltanker vor China droht zu explodieren +++ Die News des Tages im stern-Ticker. Die Meldungen im Kurz-Überblick: 2017 teuerstes US-Katastrophenjahr (20 Uhr) Baby mit Küchenpapier erstickt - 23-Jährige muss lebenslang in Haft (18.37 Uhr) 40-Jähriger stirbt in Polizeigewahrsam in NRW (14.55 Uhr) Feuer im Trump-Tower ausgebrochen (13.38 Uhr) Brennender Öltanker vor China droht zu explodieren (9.12

Uhr) Die Nachrichten des Tages im stern -Ticker: +++ 22.03 Uhr: Einjähriger Junge fällt in Bach und stirbt +++ Ein einjähriger Junge ist im sächsischen Königswalde in einen Bach gefallen und später im Krankenhaus gestorben. Am Montagnachmittag wurde die Polizei nach eigenen Angaben darüber informiert, dass die Mutter ihr Kleinkind vermisst. Daraufhin sei eine umfangreiche Suche von Landes- und Bundespolizei sowie mehrerer freiwilliger Feuerwehren angelaufen. Auch ein Polizeihubschrauber und ein Fährtensuchhund kamen zum Einsatz. Gegen 16.45 Uhr sei der Junge dann leblos in dem Bach entdeckt worden. Zwar kam er noch ins Krankenhaus, starb dort aber kurz darauf. +++ 21.39 Uhr: Fast tausend Menschen 2017 bei Polizeieinsätzen in den USA erschossen +++ In den USA sind im vergangenen Jahr laut einem Bericht 987 Menschen durch Polizeikugeln getötet worden.", "after": "News: [...] - Baby mit Küchenpapier erstickt - 23-Jährige muss lebenslang in Haft —[...] [...] und [...] wollen Klimaziel aufgeben+++ Brand in Trump-Tower +++ Angeklagter gesteht Anschlag auf BVB +++ Öltanker vor China droht zu explodieren +++ Die News des Tages im stern-Ticker. Die Meldungen im Kurz-Überblick: 2017 teuerstes US-Katastrophenjahr (20 Uhr) Baby mit Küchenpapier erstickt - 23-Jährige muss lebenslang in Haft (18.37 Uhr) 40-Jähriger stirbt in Polizeigewahrsam in [...] (14.55 Uhr) Feuer im Trump-Tower ausgebrochen (13.38 Uhr) Brennender Öltanker vor China droht zu explodieren (9.12 Uhr) Die Nachrichten des Tages im [...] -Ticker: +++ 22.03 Uhr: Einjähriger Junge fällt in Bach und stirbt +++ Ein einjähriger Junge ist im sächsischen [...] in einen Bach gefallen und später im Krankenhaus gestorben. Am Montagnachmittag wurde die Polizei nach eigenen Angaben darüber informiert, dass die Mutter ihr Kleinkind vermisst. Daraufhin sei eine umfangreiche Suche von Landes- und Bundespolizei sowie mehrerer freiwilliger Feuerwehren angelaufen. Auch ein Polizeihubschrauber und ein Fährtensuchhund kamen zum Einsatz. Gegen 16.45 Uhr sei der Junge dann leblos in dem Bach entdeckt worden. Zwar kam er noch ins Krankenhaus, starb dort aber kurz darauf. +++ 21.39 Uhr: Fast tausend Menschen 2017 bei Polizeieinsätzen in den USA erschossen +++ In den USA sind im vergangenen Jahr laut einem Bericht 987 Menschen durch Polizeikugeln getötet worden.",

"before": "Bayerisches Dorf feiert Wiedervereinigung Nach Schrecksekunde für Trumps Außenminister - Pompeo besucht Anschlagsort 11.11.19 Es war ein außergewöhnlicher Besuch: Am 7. November war US-Außenminister Mike Pompeo im bayerischen Mödlareuth - das Dorf feiert die Wiedervereinigung. Update vom 9. November, 17.17 Uhr: Am Donnerstag (7. November) besuchte US-Außenminister Mike Pompeo Mödlareuth. Das kleine Dorf an der bayerisch-thüringischen Grenze war bis zum Mauerfall geteilt. Heute gab es in dem 48-Einwohner-Dorf Feierlichkeiten anlässlich der Wiedervereinigung. „Keiner vermochte damals zu sagen, ob die Rufe nach Freiheit nicht wie die Studentenproteste auf dem Platz des Himmlischen Friedens in Peking mit Waffengewalt blutig erstickt würden", sagte Bayerns Innenminister Joachim Hermann bei seiner Rede. Danach durchbrach ein Korso aus historischen DDR-Fahrzeugen wie Trabants, Wartburgs und Barkas symbolisch eine aus Styropor aufgebaute Mauer. US-Außenminister in Deutschland: Mike Pompeo trifft sich mit Spitzenpolitikern Update vom 9. November, 10.40 Uhr: Bei seinem Besuch in Deutschland traf sich US-Außenminister Mike Pompeo mit zahlreichen deutschen Spitzen-Politikern. Darunter waren neben Finanzminister Olaf Scholz (SPD) auch Verteidigungsministerin Annegret Kramp-Karrenbauer (CDU) und Bundeskanzlerin Angela Merkel (CDU). Nach Schrecksekunde für Trumps Außenminister - Pompeo besucht Anschlagsort Update vom 8. November, 10.27 Uhr: Mike Pompeo ist auf Deutschland-Besuch. Der US-Außenminister besuchte am Donnerstag (7. November) das kleine Dorf Mödlareuth in Bayern.", "after": "[...] Dorf feiert Wiedervereinigung Nach Schrecksekunde für Trumps Außenminister - Pompeo besucht Anschlagsort 11.11.19 Es war ein außergewöhnlicher Besuch: Am 7. November war US-Außenminister Mike Pompeo im [...] [...] - das Dorf feiert die Wiedervereinigung. Update vom 9. November, 17.17 Uhr:

Am Donnerstag (7. November) besuchte US-Außenminister Mike Pompeo [...]. Das kleine Dorf an der [...]-[...] Grenze war bis zum Mauerfall geteilt. Heute gab es in dem 48-Einwohner-Dorf Feierlichkeiten anlässlich der Wiedervereinigung. „Keiner vermochte damals zu sagen, ob die Rufe nach Freiheit nicht wie die Studentenproteste auf dem Platz des Himmlischen Friedens in Peking mit Waffengewalt blutig erstickt würden“, sagte [...] Innenminister Joachim Hermann bei seiner Rede. Danach durchbrach ein Korso aus historischen [...]-Fahrzeugen wie Trabants, Wartburgs und Barkas symbolisch eine aus Styropor aufgebaute Mauer. US-Außenminister in [...]: Mike Pompeo trifft sich mit Spitzenpolitikern Update vom 9. November, 10.40 Uhr: Bei seinem Besuch in [...] traf sich US-Außenminister Mike Pompeo mit zahlreichen [...] Spitzen-Politikern. Darunter waren neben Finanzminister Olaf Scholz ([...]) auch Verteidigungsministerin Annegret Kramp-Karrenbauer ([...]) und Bundeskanzlerin Angela Merkel ([...]). Nach Schrecksekunde für Trumps Außenminister - Pompeo besucht Anschlagsort Update vom 8. November, 10.27 Uhr: Mike Pompeo ist auf [...]-Besuch. Der US-Außenminister besuchte am Donnerstag (7. November) das kleine Dorf [...] in [...].”,

“before”: “Vorbereitungen für Sanierung auf A9 und A14 - WELT Halle (dpa/sa) - Auf den Autobahnen im Süden Sachsen-Anhalts müssen Autofahrer ab Montag wegen neuer Bauarbeiten mit Einschränkungen rechnen. Auf der A9 zwischen der Anschlussstelle Leipzig-West und dem Autobahnkreuz Rippachtal sowie auf der A14 zwischen Löbejün und Halle-Trotha werden nach Angaben des Verkehrsministeriums Sanierungsarbeiten vorbereitet. Insgesamt wird auf einer Länge von rund sieben Kilometern gebaut. Die Fahrbahnsanierung kostet etwa 13 Millionen Euro und soll Ende des Jahres abgeschlossen sein. Auf der A9 wird die Fahrbahn in Richtung Berlin auf fünf Kilometern gesperrt. Der Verkehr wird über die Gegenfahrbahn geführt. Das gilt auch für die A14, hier wird die Fahrbahn in Richtung Dresden gesperrt.”, “after”: “Vorbereitungen für Sanierung auf A[...] und A[...] - [...] [...] ([...]/[...]) - Auf den Autobahnen im Süden [...] müssen Autofahrer ab Montag wegen neuer Bauarbeiten mit Einschränkungen rechnen. Auf der A[...] zwischen der Anschlussstelle [...] und dem Autobahnkreuz [...] sowie auf der A[...] zwischen [...] und [...] werden nach Angaben des Verkehrsministeriums Sanierungsarbeiten vorbereitet. Insgesamt wird auf einer Länge von rund sieben Kilometern gebaut. Die Fahrbahnsanierung kostet etwa 13 Millionen Euro und soll Ende des Jahres abgeschlossen sein. Auf der A[...] wird die Fahrbahn in Richtung [...] auf fünf Kilometern gesperrt. Der Verkehr wird über die Gegenfahrbahn geführt. Das gilt auch für die A[...], hier wird die Fahrbahn in Richtung [...] gesperrt.”,

“before”: “Ein blühendes Mohnfeld ziert das Titelblatt des Casekower Kalenders. Auch für 2021 wird es wieder einen Kalender vom Dorfverein Casekow geben. Er ist gerade erschienen und wird ab nächster Woche erhältlich sein. Seit vielen Jahren gibt der Verein des Jahresweiser mit schönen Ansichten des Dorfes heraus. Der aktuelle ist allerdings etwas anders. „Wir haben erstmals auch alle anderen Ortsteile der Gemeinde Casekow mit ins Boot geholt“, erklärt Thorsten Bolle, der den Hauptanteil der Arbeit geleistet hat. Er nahm die Kontakte mit den Vereinen und Ortsvertretern auf und steuerte auch die Mehrzahl der Fotos, passend zu den Jahreszeiten, bei. „Wir haben versucht, jeden Ortsteil zu erwähnen, leider hat uns das entsprechende Bildmaterial etwas gefehlt“, sagt er. Vielleicht klappt das ja im nächsten Jahr besser. Die Ortsteile Casekow, Wartin, Blumberg, Luckow, Petershagen und Woltersdorf sind mit einigen Sehenswürdigkeiten und zumindest einem Foto davon auf einer Karte vertreten. Im Kalendarium sind auch die Feiertage und Ferien aufgeführt. Künftig auch Veranstaltungsplaner Den Herausgebern schwebt vor, den Kalender künftig auch als Veranstaltungsplaner zu nutzen und die Vorhaben und Termin aus den Ortsteilen mit aufzunehmen. „Der Mühlentag am 24. Mai und der Weihnachtsmarkt am 27. November in Casekow stehen diesmal schon drin“, so Thorsten Bolle. Durch mehr Abnehmer hat sich auch die Auflage des Kalenders

erhöht.", "after": "Ein blühendes Mohnfeld ziert das Titelblatt des [...] Kalenders. Auch für 2021 wird es wieder einen Kalender vom Dorfverein [...] geben. Er ist gerade erschienen und wird ab nächster Woche erhältlich sein. Seit vielen Jahren gibt der Verein des Jahresweiser mit schönen Ansichten des Dorfes heraus. Der aktuelle ist allerdings etwas anders. „Wir haben erstmals auch alle anderen Ortsteile der Gemeinde [...] mit ins Boot geholt", erklärt Thorsten Bolle, der den Hauptanteil der Arbeit geleistet hat. Er nahm die Kontakte mit den Vereinen und Ortsvertretern auf und steuerte auch die Mehrzahl der Fotos, passend zu den Jahreszeiten, bei. „Wir haben versucht, jeden Ortsteil zu erwähnen, leider hat uns das entsprechende Bildmaterial etwas gefehlt", sagt er. Vielleicht klappt das ja im nächsten Jahr besser. Die Ortsteile [...], [...], [...], [...], [...] und [...] sind mit einigen Sehenswürdigkeiten und zumindest einem Foto davon auf einer Karte vertreten. Im Kalendarium sind auch die Feiertage und Ferien aufgeführt. Künftig auch Veranstaltungsplaner Den Herausgebern schwebt vor, den Kalender künftig auch als Veranstaltungsplaner zu nutzen und die Vorhaben und Termin aus den Ortsteilen mit aufzunehmen. „Der Mühlentag am 24. Mai und der Weihnachtsmarkt am 27. November in [...] stehen diesmal schon drin", so Thorsten Bolle. Durch mehr Abnehmer hat sich auch die Auflage des Kalenders erhöht.",

"before": "Orkan, Starkregen, Unwetterwarnungen — vergangene Woche hielt eine Abfolge von Sturmtiefs Bayern fest im Griff. BR24 hat die Daten ausgewertet: Welche Regionen waren am stärksten betroffen, welche Schäden entstanden und wie gut haben die Vorhersagen funktioniert? Die Sturmserie begann am Dienstag mit Tief "Xaver", das vor allem im Norden Deutschlands starke Böen brachte. Bayern blieb zunächst verschont, doch bereits am Mittwoch erreichte Tief "Ylva" den Freistaat. Meteorologen registrierten in den Alpen und im Alpenvorland Böen von bis zu 120 Kilometern pro Stunde. In Teilen Oberbayerns und Schwaben wurden Bäume entwurzelt, Stromleitungen beschädigt und Straßen gesperrt. Besonders betroffen waren die Landkreise Weilheim-Schongau und Garmisch-Partenkirchen. Parallel dazu brachte ein weiteres Tiefschaufel-System ergiebigen Regen. In Mittelfranken fiel binnen 24 Stunden die Monatsnorm, es kam zu lokalen Überschwemmungen, vollgelaufenen Keller und Hangrutschungen. Insgesamt wurden mehrere hundert wetterbedingte Einsätze für Feuerwehren und Rettungsdienste gemeldet; glücklicherweise gab es nur wenige Verletzte und keine Todesopfer. BR24 hat für die Auswertung Daten von Wetterstationen, Notrufen und Verkehrsinformationen zusammengeführt. Die Analyse zeigt, dass die höchsten Windspitzen in bergigen Lagen auftraten, während die urbanen Zentren durch Starkregen und daraus resultierende Verkehrsbeeinträchtigungen litten. Die Bahn verzeichnete zahlreiche Verspätungen und Ausfälle; regionaler Zugverkehr musste stellenweise eingestellt werden. Wie gut waren die Vorhersagen? Laut Deutschem Wetterdienst (DWD) lagen die stündlichen Prognosen für Windspitzen und Niederschläge im Großen und Ganzen richtig, doch lokal traten stärkere Böen auf als prognostiziert — insbesondere an Hanglagen und in engen Tälern, wo Kanalisierungseffekte Wind verstärken kann.", "after": "Orkan, Starkregen, Unwetterwarnungen — vergangene Woche hielt eine Abfolge von Sturmtiefs [...] fest im Griff. [...] hat die Daten ausgewertet: Welche Regionen waren am stärksten betroffen, welche Schäden entstanden und wie gut haben die Vorhersagen funktioniert? Die Sturmserie begann am Dienstag mit Tief "Xaver", das vor allem im Norden [...] starke Böen brachte. [...] blieb zunächst verschont, doch bereits am Mittwoch erreichte Tief "Ylva" den Freistaat. Meteorologen registrierten in den [...] und im [...] Böen von bis zu 120 Kilometern pro Stunde. In Teilen [...] und [...] wurden Bäume entwurzelt, Stromleitungen beschädigt und Straßen gesperrt. Besonders betroffen waren die Landkreise [...] und [...]. Parallel dazu brachte ein weiteres Tiefschaufel-System ergiebigen Regen. In [...] fiel binnen 24 Stunden die Monatsnorm, es kam zu lokalen Überschwemmungen, vollgelaufenen Keller und Hangrutschungen. Insgesamt wurden mehrere hundert wetterbedingte Einsätze für Feuerwehren und Rettungsdienste gemeldet; glücklicherweise gab es nur wenige Verletzte

und keine Todesopfer. [...] hat für die Auswertung Daten von Wetterstationen, Notrufen und Verkehrsinformationen zusammengeführt. Die Analyse zeigt, dass die höchsten Windspitzen in bergigen Lagen auftraten, während die urbanen Zentren durch Starkregen und daraus resultierende Verkehrsbeeinträchtigungen litten. Die Bahn verzeichnete zahlreiche Verspätungen und Ausfälle; regionaler Zugverkehr musste stellenweise eingestellt werden. Wie gut waren die Vorhersagen? Laut [...] lagen die stündlichen Prognosen für Windspitzen und Niederschläge im Großen und Ganzen richtig, doch lokal traten stärkere Böen auf als prognostiziert — insbesondere an Hanglagen und in engen Tälern, wo Kanalisierungseffekte Wind verstärken kann.",

"before": "Der Pollertshof ist Geschichte Lübbecke/Preußisch Oldendorf (WB). Seit 1977 ist der Pollertshof in Preußisch Oldendorf das Jugendfreizeit- und Seminarhaus des Kirchenkreises Lübbecke. In wenigen Wochen werden die Türen geschlossen. Der Kirchenkreis hat sich zu diesem Schritt entschieden, da die Einrichtung finanziell als nicht mehr tragbar gilt. Die Zukunft des Hofes ist ungewiss. Montag, 25.11.2019, 06:00 Uhr 25.11.2019, 06:01 Uhr Etwas zu grell leuchten die orange-gelben Wände im Kontrast zum gesprenkelten Fliesenboden, die Vorhänge sind an manchen Stellen bereits vergilbt. Und doch könnte es für Ann-Sophie Sandrock keinen schöneren Ort geben, wenn sie die Treppen zu den Schlafräumen im ersten Stock des Pollertshofs hochgeht: »Es ist wie nach Hause kommen. Der Pollertshof ist meine gesamte Kindheit«, sagt die 17-Jährige aus Wehdem, die viele Stunden im Pollertshof verbracht hat und eine berufliche Laufbahn als Jugendreferentin anstrebt. »Der Ort hat vielen Menschen ein Zuhause gegeben« Ann-Sophie Sandrock war eine von vielen Weggefährten des Pollertshofes (»Poho«), die am Samstag zur offiziellen Abschiedsfeier kamen. Beinahe jeder der Anwesenden war durch seine eigene Geschichte mit dem geschichtsträchtigen Haus verbunden. So wie der ehemalige Jugendpfarrer Helmut Schlingheide, der den Pollertshof 1976 zum Jugendfreizeitheim umfunktionierte und bis 1999 die Leitung innehatte: »Es ist nicht bloß ein Jugendfreizeitheim, es ist ›unser‹ Pollertshof.", "after": "Der [...] ist Geschichte [...]/[...] ([...]). Seit 1977 ist der [...] in [...] das Jugendfreizeit- und Seminarhaus des Kirchenkreises [...]. In wenigen Wochen werden die Türen geschlossen. Der Kirchenkreis hat sich zu diesem Schritt entschieden, da die Einrichtung finanziell als nicht mehr tragbar gilt. Die Zukunft des Hofes ist ungewiss. Montag, 25.11.2019, 06:00 Uhr 25.11.2019, 06:01 Uhr Etwas zu grell leuchten die orange-gelben Wände im Kontrast zum gesprenkelten Fliesenboden, die Vorhänge sind an manchen Stellen bereits vergilbt. Und doch könnte es für Ann-Sophie Sandrock keinen schöneren Ort geben, wenn sie die Treppen zu den Schlafräumen im ersten Stock des [...] hochgeht: »Es ist wie nach Hause kommen. Der [...] ist meine gesamte Kindheit«, sagt die 17-Jährige aus [...], die viele Stunden im [...] verbracht hat und eine berufliche Laufbahn als Jugendreferentin anstrebt. »Der Ort hat vielen Menschen ein Zuhause gegeben« Ann-Sophie Sandrock war eine von vielen Weggefährten des [...] (»Poho«), die am Samstag zur offiziellen Abschiedsfeier kamen. Beinahe jeder der Anwesenden war durch seine eigene Geschichte mit dem geschichtsträchtigen Haus verbunden. So wie der ehemalige Jugendpfarrer Helmut Schlingheide, der den [...] 1976 zum Jugendfreizeitheim umfunktionierte und bis 1999 die Leitung innehatte: »Es ist nicht bloß ein Jugendfreizeitheim, es ist ›unser‹ [...].",

"before": "Nach missglücktem Überholmanöver: Zwei E-Bike-Fahrer schwer verletzt Bei einem Überholmanöver kam es zu einem schlimmen Unfall Bayerisch Gmain / Schaffelpoint – Schlimmer Radlunfall an der österreichischen Grenze. Am Freitagnachmittag gegen 14.20 Uhr stürzten auf dem Radweg neben der B20 am unteren Hallthurmer Berg gegenüber von Hohenfried (Schaffelpoint) zwei einheimische E-Bike-Fahrer. Nach erster Einschätzung verletzten sich dabei beide schwer. Unfallursache: Ein 73-Jähriger wollte offenbar einen vorausfahrenden 85-Jährigen überholen und berührte ihn dabei mit der Schulter. Ersthelfer kümmerten sich um die Männer und setzten einen Notruf ab. Die Leitstelle Traunstein schickte das Reichenhaller

Rote Kreuz mit beiden Rettungswagen und der Notärztin zum Unfallort. Ärztin und Notfallsanitäter versorgten die Patienten und brachten sie dann ins Salzburger Landeskrankenhaus und in die Kreisklinik Bad Reichenhall. Beamte der Reichenhaller Polizei nahmen die Ermittlungen zum genauen Hergang auf und regelten den restlichen Verkehr, der die Kolonne aus Einsatzfahrzeugen wechselseitig passieren konnte. Zur Klärung der genauen Unfallursache forderte die Staatsanwaltschaft Traunstein einen Gutachter an. Während des Verkehrsunfalls musste das Teisendorfer Rote Kreuz zu einem weiteren internistischen Notfall im Bayerisch Gmain ausrücken. Bereits gegen 11.45 Uhr mussten das Berchtesgadener Rote Kreuz, der Berchtesgadener Notarzt und die Besatzung des Salzburger Notarzthubschraubers „Christophorus 6“ auf die Alte Reichenhaller Straße in der Ramsau ausrücken, wo sich eine 25-jährige Urlauberin aus der Oberpfalz bei einem Radsturz schwer verletzt hatte.", "after": "Nach missglücktem Überholmanöver: Zwei E-Bike-Fahrer schwer verletzt Bei einem Überholmanöver kam es zu einem schlimmen Unfall [...] / [...] – Schlimmer Radlunfall an der österreichischen Grenze. Am Freitagnachmittag gegen 14.20 Uhr stürzten auf dem Radweg neben der B[...] am unteren [...] gegenüber von [...] ([...]) zwei einheimische E-Bike-Fahrer. Nach erster Einschätzung verletzten sich dabei beide schwer. Unfallursache: Ein 73-Jähriger wollte offenbar einen vorausfahrenden 85-Jährigen überholen und berührte ihn dabei mit der Schulter. Ersthelfer kümmerten sich um die Männer und setzten einen Notruf ab. Die Leitstelle [...] schickte das [...] [...] mit beiden Rettungswagen und der Notärztin zum Unfallort. Ärztin und Notfallsanitäter versorgten die Patienten und brachten sie dann ins Salzburger Landeskrankenhaus und in die Kreisklinik [...]. Beamte der [...] Polizei nahmen die Ermittlungen zum genauen Hergang auf und regelten den restlichen Verkehr, der die Kolonne aus Einsatzfahrzeugen wechselseitig passieren konnte. Zur Klärung der genauen Unfallursache forderte die Staatsanwaltschaft [...] einen Gutachter an. Während des Verkehrsunfalls musste das [...] [...] zu einem weiteren internistischen Notfall im [...] ausrücken. Bereits gegen 11.45 Uhr mussten das [...] Rote Kreuz, der [...] Notarzt und die Besatzung des Salzburger Notarzthubschraubers „Christophorus 6“ auf die Alte [...] Straße in der [...] ausrücken, wo sich eine 25-jährige Urlauberin aus der [...] bei einem Radsturz schwer verletzt hatte.",

"before": "Oberliga Westfalen: Bövinghausen neuer Spitzenreiter, Rhynern mit Big Points Auf dem Weg in Richtung Aufstieg in die Regionalliga ist der TuS Bövinghausen am Sonntag beim TuS Ennepetal gestolpert. Gegen den Abstiegskandidaten kamen die Dortmunder nicht über ein 2:2 hinaus. In dem umkämpften Duell ging Ennepetal durch Marius Müller (28.) in Führung, ehe Migel Max-Schmeling (31.) und Cedric Golbig (53.) die Partie drehten. Doch Müller schlug erneut zu (61.) und sicherte dem Gastgeber einen Zähler. Bövinghausen übernimmt dank des Punktgewinns vorerst die Tabellenführung von Preußen Münsters U23, deren Heimspiel gegen Siegen ausgefallen ist. Die Münsteraner dürfen ohnehin nicht aufsteigen - anders als die U21 des SC Paderborn. Der SCP hatte beim 0:0 in Lotte tags zuvor Punkte liegen lassen. Das nutzte der Tabellenvierte Westfalia Rhynern und rückte bis auf einen Zähler heran. Der SV gewann sein Spiel am Sonntag mit 3:2 gegen den Delbrücker SC. Mann des Tages war Winter-Neuzugang Jonas Telschow, der alle drei Tore erzielte (26., 44., 52.). In der Schlussphase musste Rhynern zittern, denn Jannik Tödtmann (52.) und Patrice Heisinger brachten Delbrück noch einmal heran (79.). Der DSC bleibt nach der unbelohnten Aufholjagd das Schlusslicht der Tabelle. Der FC Gütersloh hätte Schritt halten können mit dem vor dem Spieltag punktgleichen Rhynern.", "after": "Oberliga [...]: [...] neuer Spitzenreiter, [...] mit Big Points Auf dem Weg in Richtung Aufstieg in die Regionalliga ist der TuS [...] am Sonntag beim TuS [...] gestolpert. Gegen den Abstiegskandidaten kamen die [...] nicht über ein 2:2 hinaus. In dem umkämpften Duell ging [...] durch Marius Müller (28.) in Führung, ehe Migel Max-Schmeling (31.) und Cedric Golbig (53.) die Partie drehten. Doch Müller schlug erneut zu (61.) und sicherte dem Gastgeber einen Zähler. [...] übernimmt dank des Punktgewinns

vorerst die Tabellenführung von [...] [...] U23, deren Heimspiel gegen [...] ausgefallen ist. Die [...] dürfen ohnehin nicht aufsteigen - anders als die U21 des SC [...]. Der SCP hatte beim 0:0 in [...] tags zuvor Punkte liegen lassen. Das nutzte der Tabellenvierte [...] [...] und rückte bis auf einen Zähler heran. Der SV gewann sein Spiel am Sonntag mit 3:2 gegen den [...] SC. Mann des Tages war Winter-Neuzugang Jonas Telschow, der alle drei Tore erzielte (26., 44., 52.). In der Schlussphase musste [...] zittern, denn Jannik Tödtmann (52.) und Patrice Heisinger brachten [...] noch einmal heran (79.). Der DSC bleibt nach der unbelohnten Aufholjagd das Schlusslicht der Tabelle. Der FC [...] hätte Schritt halten können mit dem vor dem Spieltag punktgleichen [...].",

"before": "POL-BN: Zülpich: Verdacht eines versuchten Tötungsdeliktes - Polizei und Staatsanwaltschaft ermitteln Zülpich (ots) Die Bonner Polizei und die Staatsanwaltschaft Bonn haben am Dienstagmorgen (15.08.2023) die Ermittlungen wegen eines versuchten Tötungsdeliktes in Zülpich aufgenommen. Gegen 05:00 Uhr ging bei der Polizei Euskirchen ein Notruf ein, der Polizei und Rettungsdienst der Feuerwehr zu einer größeren Hofanlage im Zülpicher Stadtteil Füssenich führte. In dem Objekt fanden die Einsatzkräfte die Bewohner, eine 77-jährige Frau und einen 76-jährigen Mann, schwerverletzt vor. Das Ehepaar wurden umgehend in Krankenhäuser gebracht. Beiden Geschädigten wurden nach Auskunft der Ärzte lebensgefährliche Verletzungen beigebracht. Aufgrund der Gesamtumstände übernahm eine Mordkommission der Bonner Polizei unter Leitung von Kriminalhauptkommissar Messerschmidt in enger Abstimmung mit Staatsanwalt Martin Kriebisch die weiteren Ermittlungen. Die Hintergründe der Tat sind derzeit unklar. Neben der Spurensicherung vor Ort, den Ermittlungen zum Geschehensablauf und den Hintergründen suchen die Ermittler Zeugen, die Angaben zu verdächtigen Wahrnehmungen am und um den Tatort in der Nacht zu Dienstag machen können. Hinweise nimmt die Polizei unter der Rufnummer entgegen.", "after": "[...]: [...]: Verdacht eines versuchten Tötungsdeliktes - Polizei und Staatsanwaltschaft ermitteln [...] ([...]) Die [...] Polizei und die Staatsanwaltschaft [...] haben am Dienstagmorgen (15.08.2023) die Ermittlungen wegen eines versuchten Tötungsdelites in [...] aufgenommen. Gegen 05:00 Uhr ging bei der Polizei [...] ein Notruf ein, der Polizei und Rettungsdienst der Feuerwehr zu einer größeren Hofanlage im [...] Stadtteil [...] führte. In dem Objekt fanden die Einsatzkräfte die Bewohner, eine 77-jährige Frau und einen 76-jährigen Mann, schwerverletzt vor. Das Ehepaar wurden umgehend in Krankenhäuser gebracht. Beiden Geschädigten wurden nach Auskunft der Ärzte lebensgefährliche Verletzungen beigebracht. Aufgrund der Gesamtumstände übernahm eine Mordkommission der Bonner Polizei unter Leitung von Kriminalhauptkommissar Messerschmidt in enger Abstimmung mit Staatsanwalt [...] die weiteren Ermittlungen. Die Hintergründe der Tat sind derzeit unklar. Neben der Spurensicherung vor Ort, den Ermittlungen zum Geschehensablauf und den Hintergründen suchen die Ermittler Zeugen, die Angaben zu verdächtigen Wahrnehmungen am und um den Tatort in der Nacht zu Dienstag machen können. Hinweise nimmt die Polizei unter der Rufnummer entgegen.",

"before": "mt-news.de / Münsterländische Tageszeitung - Garrel Garrel Sonntag, 15.12.2019 Lokalnachrichten - Garrel Weihnachtsmarkt wirkt noch Tage nach Der „Garreler Weihnachtszauber“ lockte am Wochenende zahlreiche Besucher auf den Platz vor der Mensa der Oberschule Garrel. ... Freitag, 13.12.2019 Lokalnachrichten - Garrel 30-Jähriger auf der B72 schwer verletzt Ein 30-Jähriger ist am Freitagmorgen mit seinem Auto auf der B72 bei Garrel nahezu ungebremst auf das Heck eines Kleinbusses gefahren. Er wurde schwer verletzt. ... Mittwoch, 11.12.2019 Lokalnachrichten - Garrel Mit dem Bürgermeister auf Du und Du Seinen Kollegen hat Garrels neuer Bürgermeister gleich zu Beginn das „Du“ angeboten und ihnen versichert, dass Bürotür und Ohren immer offen sind für ein vertrauliches Gespräch. ... Dienstag, 10.12.2019 Lokalnachrichten - Garrel Polar-Express bringt Musiker ins Schwitzen Über

Wochen haben sich die Musiker auf diesen Abend vorbereitet und mit frenetischem Applaus wurde ihr Übungsfleiß belohnt. Jetzt kann das Jubiläum kommen. ... Montag, 09.12.2019 Lokalnachrichten - Garrel Kollision auf Kreuzung Bei Blechschaden ist es am Montag glücklicherweise geblieben, als zwei Autos auf der Kreuzung B 72/Garreler Straße in Varrelbusch miteinander kollidierten. ... Donnerstag, 05.12.2019 Lokalnachrichten - Garrel Musikverein: Verkauf der Jubiläums-CD startet Nicht nur, dass sich die Garreler Musiker und Musikerinnen auf ihr alljährlich stattfindendes Adventskonzert freuen, sie starten auch mit dem Verkauf ihrer vereinseigenen Jubiläums-CD anlässlich ihres 100-jährigen Bestehens.", "after": "[...] / [...] Tageszeitung - [...] [...] Sonntag, 15.12.2019 Lokalnachrichten - [...] Weihnachtsmarkt wirkt noch Tage nach Der „[...] Weihnachtszauber" lockte am Wochenende zahlreiche Besucher auf den Platz vor der Mensa der Oberschule [...]. ... Freitag, 13.12.2019 Lokalnachrichten - [...] 30-Jähriger auf der B[...] schwer verletzt Ein 30-Jähriger ist am Freitagmorgen mit seinem Auto auf der B[...] bei [...] nahezu ungebremst auf das Heck eines Kleinbusses gefahren. Er wurde schwer verletzt. ... Mittwoch, 11.12.2019 Lokalnachrichten - [...] Mit dem Bürgermeister auf Du und Du Seinen Kollegen hat [...]s neuer Bürgermeister gleich zu Beginn das „Du" angeboten und ihnen versichert, dass Bürotür und Ohren immer offen sind für ein vertrauliches Gespräch. ... Dienstag, 10.12.2019 Lokalnachrichten - [...] Polar-Express bringt Musiker ins Schwitzen Über Wochen haben sich die Musiker auf diesen Abend vorbereitet und mit frenetischem Applaus wurde ihr Übungsfleiß belohnt. Jetzt kann das Jubiläum kommen. ... Montag, 09.12.2019 Lokalnachrichten - [...] Kollision auf Kreuzung Bei Blechschaden ist es am Montag glücklicherweise geblieben, als zwei Autos auf der Kreuzung B [...]/[...] in [...] miteinander kollidierten. ... Donnerstag, 05.12.2019 Lokalnachrichten - [...] Musikverein: Verkauf der Jubiläums-CD startet Nicht nur, dass sich die [...] Musiker und Musikerinnen auf ihr alljährlich stattfindendes Adventskonzert freuen, sie starten auch mit dem Verkauf ihrer vereinseigenen Jubiläums-CD anlässlich ihres 100-jährigen Bestehens.",

"before": "Porta Nigra: Geschichte und Infos zur Top-Sehenswürdigkeit in Trier Erbaut als Stadttor ist die Porta Nigra heute das Wahrzeichen von Trier. Trier Zahllose Touristen haben in Trier bereits die Porta Nigra bewundert. Das Bauwerk, das die antiken Römer hinterlassen haben, wurde in seiner Geschichte sogar schon für mehrere Hundert Jahre zur Kirche umgebaut. Sie ist das Wahrzeichen von Trier, ja vielleicht sogar das von ganz Rheinland-Pfalz: die Porta Nigra. Sie ist die Top-Sehenswürdigkeit der Stadt und eines der beliebtesten Ausflugziele an der Mosel. Im Jahr 2017 zierte die Porta Nigra sogar eine 2-Euro-Sondermünze im Rahmen der Bundesländer-Serie. Das ist kein Wunder. Die Römer sind ein wichtiger Bestandteil deutscher Geschichte links des Rheins. Die Porta Nigra wurde von ihnen erbaut, der Grundstein vermutlich im Jahr 170 gelegt. Die Porta Nigra ist das am besten erhaltene römische Stadttor in Deutschland. Seit 1986 ist sie UNESCO-Weltkulturerbe. Wer hat die Porta Nigra gebaut? Erbaut wurde die Porta Nigra um das Jahr 170 von den Römern. Sie waren es auch, die die Stadt Trier, die damals Augusta Treverorum hieß, gegründet haben. Wie alle ihre Städte umgaben die Römer auch Augusta Treverorum mit einer Stadtmauer, um sich vor Feinden zu schützen. Die Porta Nigra war dabei Bestandteil der nördlichen Stadtmauer und damit das nördliche Stadttor.", "after": "[...]: Geschichte und Infos zur Top-Sehenswürdigkeit in [...] Erbaut als Stadttor ist die [...] heute das Wahrzeichen von [...]. [...] Zahllose Touristen haben in [...] bereits die [...] bewundert. Das Bauwerk, das die antiken Römer hinterlassen haben, wurde in seiner Geschichte sogar schon für mehrere Hundert Jahre zur Kirche umgebaut. Sie ist das Wahrzeichen von [...], ja vielleicht sogar das von ganz [...]: die [...]. Sie ist die Top-Sehenswürdigkeit der Stadt und eines der beliebtesten Ausflugziele an der [...]. Im Jahr 2017 zierte die[...] sogar eine 2-Euro-Sondermünze im Rahmen der [...]-Serie. Das ist kein Wunder. Die Römer sind ein wichtiger Bestandteil [...] Geschichte links des [...]. Die[...] wurde von ihnen erbaut, der Grundstein vermutlich im Jahr 170 gelegt. Die[...] ist das am besten

erhaltene römische Stadttor in [...]. Seit 1986 ist sie UNESCO-Weltkulturerbe. Wer hat die [...] gebaut? Erbaut wurde die [...] um das Jahr 170 von den Römern. Sie waren es auch, die die Stadt [...], die damals[...] hieß, gegründet haben. Wie alle ihre Städte umgaben die Römer auch [...] mit einer Stadtmauer, um sich vor Feinden zu schützen. Die [...] war dabei Bestandteil der nördlichen Stadtmauer und damit das nördliche Stadttor.",

"before": "Landtagswahl Saarland 2022 im Wahlkreis Saarlouis: Prognose, Hochrechnung und Ergebnis Das sind die Spitzenkandidaten für die Landtagswahl 2022 im Saarland Liveblog Welche Parteien schaffen es in den Landtag und welche nicht? Und wer regiert das Saarland künftig? Die Entscheidung fällt an diesem Sonntag bei der Landtagswahl. Im Wahlkreis Saarlouis treten gleich zwei saarlandweite Spitzenkandidatinnen an. Alle Infos hier im Live-Ticker. Wird Anke Rehlinger die nächste Ministerpräsidentin oder kann Tobias Hans entgegen der Umfragen doch noch im Amt bleiben? Das haben die Wähler im Saarland an diesem Sonntag, 27. März, bei der Landtagswahl in der Hand. Mehr wissen wir, wenn die Wahllokale schließen und die erste Prognose um 18 Uhr erscheint. Danach beginnt die Auszählung, die wir hier live für Sie begleiten, stets mit der aktuellsten Hochrechnung. Im Wahlkreis Saarlouis mit dem Kreis Saarlouis und Merzig-Wadern treten mit Anke Rehlinger und Angelika Hießerich-Peter gleich zwei Spitzenkandidatinnen an. Wer zieht aus dem Wahlkreis in den Landtag ein? Das dürfte erst am späten Abend feststehen. Das sind die Spitzenkandidaten der größten Parteien bei der Landtagswahl im Wahlkreis Saarlouis: CDU: Raphael Schäfer SPD: Anke Rehlinger Linke: Dagmar Ensch-Engel AfD: Carsten Becker Grüne: Kiymet Göktas FDP: Angelika Hießerich-Peter Zum Wahlkreis Saarlouis gehören: Beckingen, Bous, Dillingen, Ensdorf, Lebach, Losheim am See, Merzig, Mettlach, Nalbach, Perl, Rehlingen-Siersburg, Saarlouis, Saarwellingen, Schmelz, Schwalbach, Überherrn, Wadern, Wadgassen, Wallerfangen und Weiskirchen Die zwei weiteren Wahlkreise im Saarland sind: Saarbrücken Neunkirchen Und hier ist der Saarland-Ticker zur Landtagswahl.", "after": "Landtagswahl [...] 2022 im Wahlkreis [...]: Prognose, Hochrechnung und Ergebnis Das sind die Spitzenkandidaten für die Landtagswahl 2022 im [...] Liveblog Welche Parteien schaffen es in den Landtag und welche nicht? Und wer regiert das [...] künftig? Die Entscheidung fällt an diesem Sonntag bei der Landtagswahl. Im Wahlkreis [...] treten gleich zwei [...]weitige Spitzenkandidatinnen an. Alle Infos hier im Live-Ticker. Wird Anke Rehlinger die nächste Ministerpräsidentin oder kann Tobias Hans entgegen der Umfragen doch noch im Amt bleiben? Das haben die Wähler im [...] an diesem Sonntag, 27. März, bei der Landtagswahl in der Hand. Mehr wissen wir, wenn die Wahllokale schließen und die erste Prognose um 18 Uhr erscheint. Danach beginnt die Auszählung, die wir hier live für Sie begleiten, stets mit der aktuellsten Hochrechnung. Im Wahlkreis [...] mit dem Kreis [...] und [...] treten mit Anke Rehlinger und Angelika Hießerich-Peter gleich zwei Spitzenkandidatinnen an. Wer zieht aus dem Wahlkreis in den Landtag ein? Das dürfte erst am späten Abend feststehen. Das sind die Spitzenkandidaten der größten Parteien bei der Landtagswahl im Wahlkreis [...]: [...]: Raphael Schäfer [...]: Anke Rehlinger [...]: Dagmar Ensch-Engel [...]: Carsten Becker [...]: Kiymet Göktas [...]: Angelika Hießerich-Peter Zum Wahlkreis [...] gehören: [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...], [...] Die zwei weiteren Wahlkreise im [...] sind: [...] [...] Und hier ist der [...]-Ticker zur Landtagswahl.",

"before": "Landesweiter Probealarm: Hier in Franken heulen heute die Sirenen Die Regierung von Mittelfranken informiert über den landesweiten Warntag am Donnerstag (9. März 2023). Welche Regionen betroffen sind und worauf sich die Bürger einstellen müssen. Mit einem Heulton von einer Minute Dauer wird am Donnerstag, den 09. März 2023, um 11.00 Uhr in weiten Teilen Bayerns die Auslösung des Warnsystems zur Warnung der Bevölkerung geprobt, das berichtet die Regierung von Mittelfranken in einer Pressemitteilung. Neben den vorhandenen Sirenen der Gemeinden und Kreisverwaltungsbehörden werden auch andere Warnmit-

tel, z. B. Warn-Apps, eingesetzt. Der Probealarm dient dazu, die Funktionsfähigkeit der vorhandenen Warnsysteme zu überprüfen und die Bevölkerung auf die Bedeutung des Sirenensignals hinzuweisen. Probealarm in Bayern: Hier heulen am Donnerstag die Sirenen Der Heulton soll die Bevölkerung bei schwerwiegenden Gefahren für die öffentliche Sicherheit veranlassen, ihre Rundfunkgeräte einzuschalten und auf Durchsagen zu achten. Der Probealarm wird in Mittelfranken in folgenden kreisfreien Städten und Landkreisen ausgelöst: Stadt Ansbach Stadt Erlangen Stadt Fürth Stadt Nürnberg Stadt Schwabach Landkreis Erlangen-Höchstadt: Gemeinde Adelsdorf (Feuerwehrgerätehaus und OT Aisch, Neuhaus, Lauf, Weppersdorf) Markt Eckental (Feuerwehrgerätehaus Forth) Gemeinde Hemhofen (Feuerwehrgerätehaus Hemhofen-Zeckern) Markt Heroldsberg (Feuerwehrgeräte-/Rathaus) Stadt Herzogenaurach (Stadtgebiet und Feuerwehrgerätehaus Höfen-Zweifelsheim, OT Niederndorf) Gemeinde Heßdorf (Feuerwehrgerätehaus) Stadt Höchstadt a.d. Aisch (Feuerwehrgerätehaus, Fortuna Kulturfabrik, OT Saltendorf-Bösenbechhofen, Förtschwind-Greuth, Medbach-Kieferndorf) Landkreis Fürth: Markt Cadolzburg (nur Kernort und Ortsteile Egersdorf, Greimersdorf, Roßendorf, Seckendorf, Steinbach, Wachendorf) Gemeinde Obermichelbach Gemeinde Puschendorf Gemeinde Tuchenbach Landkreis Nürnberger Land: Gemeinde Engelthal Gemeinde Happurg Gemeinde Henfenfeld Stadt Hersbruck Gemeinde Pommelsbrunn Stadt Röthenbach a.d.", "after": "Landesweiter Probealarm: Hier in [...] heulen heute die Sirenen Die Regierung von [...] informiert über den landesweiten Warntag am Donnerstag (9. März 2023). Welche Regionen betroffen sind und worauf sich die Bürger einstellen müssen. Mit einem Heulton von einer Minute Dauer wird am Donnerstag, den 09. März 2023, um 11.00 Uhr in weiten Teilen [...] die Auslösung des Warnsystems zur Warnung der Bevölkerung geprobt, das berichtet die Regierung von [...] in einer Pressemitteilung. Neben den vorhandenen Sirenen der Gemeinden und Kreisverwaltungsbehörden werden auch andere Warnmittel, z. B. Warn-Apps, eingesetzt. Der Probealarm dient dazu, die Funktionsfähigkeit der vorhandenen Warnsysteme zu überprüfen und die Bevölkerung auf die Bedeutung des Sirenensignals hinzuweisen. Probealarm in [...]: Hier heulen am Donnerstag die Sirenen Der Heulton soll die Bevölkerung bei schwerwiegenden Gefahren für die öffentliche Sicherheit veranlassen, ihre Rundfunkgeräte einzuschalten und auf Durchsagen zu achten. Der Probealarm wird in [...] in folgenden kreisfreien Städten und Landkreisen ausgelöst: Stadt [...] Stadt [...] Stadt [...] Stadt [...] Stadt [...] Landkreis [...]: Gemeinde [...] (Feuerwehrgerätehaus und OT [...], [...], [...], [...]) Markt [...] (Feuerwehrgerätehaus Forth) Gemeinde [...] (Feuerwehrgerätehaus [...]-Zeckern) Markt [...] (Feuerwehrgeräte-/Rathaus) Stadt [...] (Stadtgebiet und Feuerwehrgerätehaus [...]-[...], OT [...]) Gemeinde [...] (Feuerwehrgerätehaus) Stadt [...] a.d. [...] (Feuerwehrgerätehaus, [...] Kulturfabrik, OT [...]-[...], [...]-[...], [...]-[...]) Landkreis [...]: Markt [...] (nur Kernort und Ortsteile [...], [...], [...], [...], [...], [...]) Gemeinde [...] Gemeinde [...] Gemeinde [...] Landkreis [...]: Gemeinde [...] Gemeinde [...] Gemeinde [...] Stadt [...] Gemeinde [...] Stadt [...] a.d. [...]",

"before": "Öl Bayernoil-Produktion nach Explosion halbiert Durch die Explosion und den Großbrand in der Bayernoil-Raffinerie im oberbayerischen Vohburg an der Donau ist die Produktion des Unternehmens vorläufig etwa halbiert. „Die Vohburger Raffinerie steht still“, sagte Geschäftsführer Michael Raue am Dienstag. Wann dort wieder produziert werden könne, sei nicht absehbar. Am zweiten Standort im niederbayerischen Neustadt an der Donau laufe die Produktion hingehen normal weiter. Vohburg an der Donau. Trotz der Einschränkungen gibt es laut Raue keinen Grund, einen Teil der knapp 800 Mitarbeiter in Kurzarbeit zu schicken. „Wir sind mit Aufräumen beschäftigt“, sagte er angesichts der großen Schäden am Standort Vohburg. Normalerweise verarbeitet Bayernoil in den beiden Raffinerien pro Jahr mehr als zehn Millionen Tonnen Rohöl. Bei dem Feuer war ein noch nicht genau festgestellter Schaden im Millionenbereich entstanden, 16 Menschen wurden verletzt. Die Kripo aus Ingolstadt ermittelt derzeit gemeinsam mit einem Experten des Landeskriminalamtes zur Unglücksursache.

Die Polizei geht davon aus, dass diese Untersuchung noch erhebliche Zeit brauchen wird. Bei der Explosion waren auch mehr als 100 Gebäude in der Umgebung des Betriebs beschädigt worden.", "after": "Öl [...]-Produktion nach Explosion halbiert Durch die Explosion und den Großbrand in der [...]-Raffinerie im [...] [...] an der Donau ist die Produktion des Unternehmens vorläufig etwa halbiert. „Die [...] Raffinerie steht still", sagte Geschäftsführer Michael Raue am Dienstag. Wann dort wieder produziert werden könne, sei nicht absehbar. Am zweiten Standort im [...] [...] an der Donau laufe die Produktion hingehen normal weiter. [...]. Trotz der Einschränkungen gibt es laut Raue keinen Grund, einen Teil der knapp 800 Mitarbeiter in Kurzarbeit zu schicken. „Wir sind mit Aufräumen beschäftigt", sagte er angesichts der großen Schäden am Standort [...]. Normalerweise verarbeitet Bayernoil in den beiden Raffinerien pro Jahr mehr als zehn Millionen Tonnen Rohöl. Bei dem Feuer war ein noch nicht genau festgestellter Schaden im Millionenbereich entstanden, 16 Menschen wurden verletzt. Die Kripo aus [...] ermittelt derzeit gemeinsam mit einem Experten des Landeskriminalamtes zur Unglücksursache. Die Polizei geht davon aus, dass diese Untersuchung noch erhebliche Zeit brauchen wird. Bei der Explosion waren auch mehr als 100 Gebäude in der Umgebung des Betriebs beschädigt worden."