

Reading the Current: Tracking Subsea Cable Status through LLM-Assisted News Analysis

Jonas Franken¹, Kasimir Romer², Timon Dörnfeld¹, Paula Meissner¹, Christian Reuter¹

Subsea Data Cables (SDCs) form the backbone of global digital infrastructure, handling over 99% of intercontinental data traffic. Still, they receive limited and dispersed media coverage, creating significant information gaps regarding their operational status. This paper aims to investigate the potential of Large Language Models (LLMs) in automating the extraction and analysis of SDC-related information from unstructured media sources. A comprehensive LLM-based information extraction module was developed and integrated into an existing SDC database. By systematically comparing different LLMs, including GPT-4o, Gemini 1.5 Flash, Claude 3.5 Sonnet, and Llama 3.1, this work identifies optimal model configurations for SDC news processing, considering accuracy, hallucination rate, and processing speed. The system implements Claude 3.5 Sonnet and GPT-4o as primary models, incorporating domain-specific prompt engineering and robust output validation mechanisms. A performance evaluation demonstrates substantial improvements over rule-based methods. While excelling at natural language processing tasks, the system revealed limitations in extracting more specific technical details such as construction costs and capacity measurements. The implementation provides a modular, adaptable framework for automated information extraction in specialised technical domains, including maritime ones. The results demonstrate that LLMs, when properly implemented with structured prompts and validation mechanisms, can significantly enhance the automated monitoring and analysis of SDC-related events.

KEY WORDS

- ~ Subsea data cables
- ~ Critical maritime infrastructure
- ~ Maritime domain awareness
- ~ Information extraction
- ~ Large language models
- ~ News analysis

¹ Science and Technology of Peace and Security - PEASEC, Technical University of Darmstadt, Darmstadt, Germany

² Technical University of Darmstadt, Department of Computer Science, Darmstadt, Germany

e-mail: franken@peasec.tu-darmstadt.de

doi: 10.7225/toms.v15.n02.w07

Received: 31 Mar 2026 / Revised: 16 Jun 2026 / Accepted: 16 Jun 2026 / Published: 20 Jun 2026

This work is licensed under



1. INTRODUCTION

Forming the physical backbone of intercontinental network traffic, Subsea Data Cables (SDCs) are among the most critical maritime infrastructures today (Bueger and Liebetrau, 2021; Kavanagh, Franken and Kulesza, 2026). This role makes them potential targets for sabotage and illicit interception of information, underscoring a significant vulnerability in global communications infrastructure (Bueger, Liebetrau and Franken, 2022). The urgency of addressing these threats is highlighted by recent geopolitical events. Since November 2023, the Houthi rebels in Yemen have launched significant maritime attacks against cargo vessels in the Red Sea, including explicit threats against submarine cable systems as part of their strategy to weaken the international economic stability (Johnson, 2024). This conflict saw the damage of at least three submarine cables in the Red Sea by February 2024. While these faults cannot be directly attributed to Houthi actions and likely resulted from anchor drag after a targeted vessel had been evacuated and drifted for several days, they highlight the vulnerability of networks relying on SDCs and the importance of swift repair (Stronge, 2024).

Despite their importance, SDCs tended to be a niche subject with limited presence in the public discourse. This may, in part, be explained with what Bueger and Liebetrau (2021) call “triple invisibility” of the subsea cable network: namely as infrastructure, as being buried, and as being under the sea. They note that infrastructures such as roads, buildings, and sewers often reside in a naturalised background, despite their fundamental role in our civilisation as the “connective tissues and circulatory systems of modernity”. Submarine cables, which are hidden on the seafloor and cause minimal disruption during construction, are among the least visible types of infrastructure. This invisibility is compounded by “sea blindness,” where the open sea is not a primary concern for most people, making the maritime infrastructure more hidden than terrestrial (Bueger and Edmunds, 2017).

However, the security of SDCs, especially in the Baltic Sea, has recently drawn more frequent media attention (Kavanagh, Franken and He, 2025). Multiple faults of seabed infrastructures caused by anchor drags between October 2023 and February 2025 have been increasingly reported by the news media and led to increased regulatory activity. Especially, the European Union’s efforts to an enhanced SDC awareness are tangible through the Cable Recommendation (European Commission, 2024), the Action Plan on Subsea Cables (European Parliament and European Council, 2025) and the most recent Cable Toolbox (European Commission, 2026). Still, media coverage on SDCs in general is sporadic at best, often relegated to specialised industry forums, private online groups, or smaller-scale, regularly updated publications like the *Almanac* by Clark and Humera (2024) or *SubCableNews* edited by Bruckschen (2025). At the same time, metaphor and analogy use is prevalent in articles from these sources (Franken, Schneider and Reuter, 2023), which complicates purely rule-based text mining approaches. This dispersed information landscape creates a significant gap in structured, up-to-date, and machine-readable data, making it difficult to fully assess parallel impacts on the SDC network. To overcome this, the authors propose investigating the usefulness and performance of LLMs in extracting and analysing information from various fragmented news sources regarding SDC. The advent of LLMs and other forms of generative artificial intelligence (AI) applications since the introduction of OpenAI’s ChatGPT in late 2022 has revolutionised numerous fields by enabling sophisticated text and visual analysis and parsing capabilities.

This paper will contribute to a more robust understanding of how to make use of the information landscape surrounding SDCs and develop a framework that could be applied to other critical maritime infrastructures and beyond. The primary objective is to develop a data processing pipeline that collects and analyses news from various sources, suggesting updates to an existing SDC database (Franken and Reuter, 2024). This system should be specifically designed to monitor SDC news but has a generalised architecture to facilitate easy adaptation for other subjects with minimal effort. The research will be guided by the research question

(RQ): How can Large Language Models be applied to extract and process unstructured news data on subsea data cables?

To address this research question, the work proceeds with an overview of related works (Sec. 2) in the fields of AI applications in the maritime domain and LLM-based Information Extraction (IE), followed by the methodology (Sec. 3) that explains the model-comparative approach taken in the subsequent model analysis (Sec. 4). Then, the information extraction pipeline as central artefact will be presented (Sec. 5) and evaluated via post-tests (Sec. 6). The work closes with a discussion (Sec. 7) and conclusion (Sec. 8).

2. RELATED WORK

This section provides an overview of current research in maritime applications of AI as well as Information Extraction (IE) approaches using LLMs.

2.1. Artificial Intelligence in Maritime Applications

Hadi *et al.* (2024) and Pahune and Chandrasekharan (2023) describe that LLMs are already being used at scale within the domains of finance, legal, health, code generation and many more. For the maritime domain, research has already explored AI to support in various operational contexts such as ports and autonomous shipping; however, most applications have focused more on operational optimisation (efficiency, safety, security) than on extracting structured information from fragmented reporting (Safuan and Syafira, 2024). For the subsea cables area, AI is often regarded as external factor driving demand for intercontinental data transmission that may require systems of petabit-per-second- scale in the coming decade (Palmer-Felgate, Perrier and Kawasaki, 2025). But, AI has also seen some applications for maritime domain awareness, for example as analytical tool for early warning and detection mechanisms as suggested by Lei *et al.* (2024) and ex-post visual analysis of cable damages as conceptualised by Karmakar (2023). With regard to the increasing implementation of various optical sensing technologies in SDCs, AI offers potential in seismic detection and general marine monitoring (Caruso *et al.*, 2026).

In maritime research, applications of text-based analytics using LLMs are rather rare. Cardenas *et al.* (2025) showed that natural language processing can improve the completeness of accident analysis when applied to diving and maritime operations linked to subsea cable installation – power and data cables alike. Their case study drew on a diverse document set, including incident records, regulations, standards, manuals, guidelines, specifications, installation plans, risk assessments, interviews, technical reports, and academic literature (Cardenas *et al.*, 2025). They conclude that this combined approach is especially useful for underreported, data-sparse safety contexts, which is a similarity shared by the information landscape on operational status of SDCs. Accident reports are also focused in the work of Soner (2025), where they showed that analysing unstructured maritime accident reports with topic modelling can reveal recurring and overlooked operational risk patterns. Based on 247 accident reports, they identified ten major topical clusters of safety incidents, including failures in emergency response, fires, grounding, towing risks, and communication problems. However, no studies could be identified that address the automated maintenance of datasets using LLMs, which makes it worthwhile to draw on the broader literature on information extraction (IE).

2.2. Information Extraction through LLMs

While no previous paper has targeted the extraction of information from SDC-related news, several works focused on other source materials (Xu *et al.*, 2024). Three IE tasks are relevant for the application area.

2.2.1. Named Entity Recognition (NER)

In the past years, many works have analysed how LLMs can facilitate NER tasks. Evans *et al.* (2024) conducted a survey on how well an LLM can annotate entities in comparison to a human domain expert. Their research found that fine-tuning led to a significant improvement in LLM-performance. Similarly, Bogdanov *et al.* (Bogdanov *et al.*, 2024) used an off-the-shelf LLM to annotate NER data that were then used to train a smaller model in order to save costs. NuNER surpassed GPT-3.5 and possibly also GPT-4 after a few training rounds, making it a viable candidate for NER tasks. Likewise, Wang *et al.* (2023) transformed the NER task into a text generation task, solvable by LLMs. In order to avoid hallucination, they implemented a self-verification stage. Their results showed very good performance, also when training data was scarce. Zhang *et al.* (2024) proposed a framework called LinkNER that employed a two-step approach for NER. First, a specialised NER model identified entities in the text and assessed its own confidence in each prediction. Then, any entities the model was uncertain about were passed to a more general LLM for review and refinement. The final result combined the strengths of both models; the precision of the NER model and the broad knowledge base of the LLM.

2.2.2. Relation Extraction (RE)

Wan *et al.* (2023) introduced GPT-RE, a method to improve how LLMs perform RE tasks, which they have struggled with due to issues such as selecting irrelevant examples and misclassifying missing relationships. GPT-RE overcame these challenges by incorporating task-specific entity information and using reasoning logic derived from correct answers. Huguet Cabot and Navigli (2021) used the same approach as Zhang *et al.* (2024) and reformulated RE as a text generation problem LLMs can solve. Xu *et al.* (2023) investigated enhancing few-shot RE using GPT-3.5. The study explored two main strategies: In-context learning and data generation. Josifoski *et al.* (2022) evaluated their GenIE approach and discovered promising results for RE using a decoder-encoder model.

2.3.1 Event Extraction (EE)

Of particular interest is Nakshatri et al.'s (2023) use of LLMs to discover events from news streams. The framework first filtered news articles based on their temporal patterns to focus on periods of high activity and then clustered them to identify potential event candidates, offering great potential to understand the evolution of news narratives and identify key events within large news corpora.

Despite the extensive work done in the fields of SDCs, LLMs, and IE, a research gap remains unaddressed in the integration of current events into existing SDC documentation and visualisation approaches. Current maps, such as the Submarine Cable Map (TeleGeography, 2026) or the SDCS Map (Franken and Reuter, 2024) provide static representations of SDC networks. However, they do not reflect real-time changes or disruptions, such as cable damages, repairs, or new deployments. This paper offers an integrated approach that combines the LLM capabilities in processing natural language with real-time IE tailored to the SDC domain. These advancements can significantly improve global SDC networks' monitoring and analysis but also address the broader issue of IE of topical niches with low training data availability.

3. METHOD

For the analysis, four widely used industry LLMs were selected for comparison in order to reflect different model characteristics and usage contexts. The selection was intended to cover both leading commercial systems and an open-source alternative, while also accounting for strengths that appeared particularly relevant to the task of extracting information from news articles. ChatGPT-4o was included as one of the leading commercial LLMs and as a useful benchmark due to its broad adoption and strong general-purpose performance. Gemini 1.5 Flash was selected because it is designed to handle longer text inputs efficiently, which is especially relevant when processing lengthy or information-dense news articles. Claude 3.5 Sonnet was chosen on account of its documented strength in factual accuracy, a core consideration for a task in which reliable extraction of events, dates, or status is essential. Finally, Llama 3.1 was included as an open-source alternative, enabling a comparison between proprietary and openly available model families.

The evaluation dataset consisted of SDC-related articles ($n = 27$) selected through purposive sampling, see Appendix A. The sample was designed to cover different source types relevant to SDC-related information extraction. It includes English texts from general news outlets (BBC, Reuters, and The Guardian) and specialised sources (SubCableNews, SubTel Forum, and DatacenterDynamics). This decision is supported by the source evaluation shown in Figure 1, which illustrates substantial differences in the average delay between incidents and reporting across sources. General news outlets tend to report major incidents more quickly, whereas specialised sources show longer reporting delays, which reflects their more periodic or detailed publication formats. SubCableNews, for example, is a monthly newsletter but frequently offers more detailed information on incidents compared to general news outlets. Hence, the dataset intentionally combines timely general reporting with more domain-specific and technically detailed coverage.

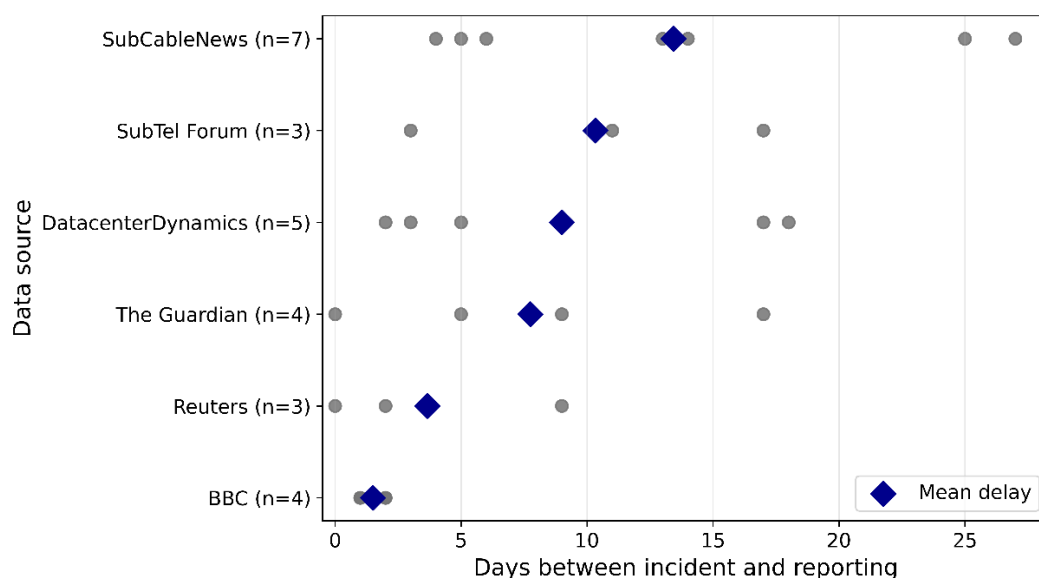


Figure 1: Delays between incidents and reports (grey dots) of SDC status changes.

The LLMs were then applied to this dataset to extract relevant data, performing IE tasks, including NER, RE, and EE. The performance evaluation was based on four measures: accuracy, depth, hallucination rate, and structural adherence.

First, accuracy measures how accurately an LLM extracts relevant details from news articles, including names, dates, and specific technical terms. The extracted information was compared against manually verified data to assess how well the model performs. Second, depth assesses how thoroughly the LLM extracts background information and context from articles, such as technical specifications, historical context, and industry impact. The completeness of the extracted information was rated against a benchmark of manually extracted data. Third, hallucination rate refers to the frequency with which the LLM introduces fabricated or incorrect information that is not present in the original source material. The results were reviewed to identify instances of hallucination. Fourth, structure adherence evaluates how well the LLM conforms to the specified JavaScript Object Notation (JSON) format, ensuring correct field names, proper nesting, and consistent formatting. Each output was fed to an online JSON validator to determine whether it is correctly structured and ready for direct use or requires modifications. Each aspect was rated on a scale from 1 (poor performance) to 5 (excellent performance), which enabled a systematic comparison of the strengths and weaknesses of the LLMs, see Appendix B.

Finally, the LLMs' performances were optimised with prompt engineering. The prompts were refined by incorporating domain-specific knowledge and structured output requirements. The models with optimised prompts were then reevaluated using the same criteria system, allowing for a performance comparison between basic and engineered prompts and the quantification of the impact of prompt optimisation on extraction quality. The results were analysed to determine whether specialised prompt engineering could significantly improve the models' ability to extract information from SDC-related news compared to basic prompting approaches. The initial prompt and the optimised prompts can be found in Appendix C.

The size of the manually evaluated benchmark reflects the labour-intensive nature of the evaluation procedure. For each of the 27 articles, eight model outputs had to be assessed, covering four LLMs and two prompting variants. Outputs from the initial prompt were evaluated across three dimensions: accuracy, depth, and hallucination rate. Outputs from the optimised prompt were additionally assessed for structural adherence to the required JSON format. This resulted in 28 manual rating operations per article, or 756 individual assessment steps across the benchmark dataset. The dataset size therefore reflects a trade-off between source diversity, controlled model comparison, and the depth of manual verification required for each article.

4. MODEL ANALYSIS

This section reports the evaluation results with regard to the performance measures and the effect of prompt engineering that supported the structural adherence.

4.1. Evaluation of LLM Performance

The evaluation of the four chosen LLMs reveals distinct strengths and weaknesses across the key performance metrics: accuracy, depth, and hallucination rate. The average values with and without prompt optimization are shown in Figure 2. In terms of accuracy, Claude 3.5 Sonnet slightly outperforms the other models, indicating that it consistently extracts key facts more reliably than the others. ChatGPT-4o follows closely behind, also demonstrating strong performance in identifying correct details. Llama 3.1 and Gemini 1.5 Flash exhibit slightly lower accuracy, suggesting they may occasionally miss or misinterpret important information. Both Claude and ChatGPT excel in the depth of extraction and capture comprehensive background information, technical details, and contextual nuances from the articles. By contrast, Gemini and Llama can extract essential information but fail to provide the same level of depth and richness in their responses as the leading models. Finally, Gemini barely hallucinates, scoring 4.96 in hallucination rate. While Claude and Llama also achieve high scores, indicating that they are also very reliable in avoiding fabricated information, ChatGPT-4o shows a slightly lower score and occasionally introduces incorrect details.

4.2. Impact of Prompt Engineering

Following best practices of prompt engineering, a remodelled prompt included the instruction to output only valid JSON and not to hallucinate by limiting the results to information in the text (Subramanian, 2024). Additionally, it clearly defined the expected JSON structure and the collected data to include the cable name, the date of the event, and information related to the event, like the location, the impact, and additional details. Results of the improved prompt showed improvements in all three key performance metrics, see Figure 2. Accuracy improved slightly, though challenges in date extraction and fact-checking remained. Depth of extraction improved significantly, with some models providing more detailed context, though others continued to offer shallow responses. While reducing hallucinations was not the primary goal, most models showed better performance.

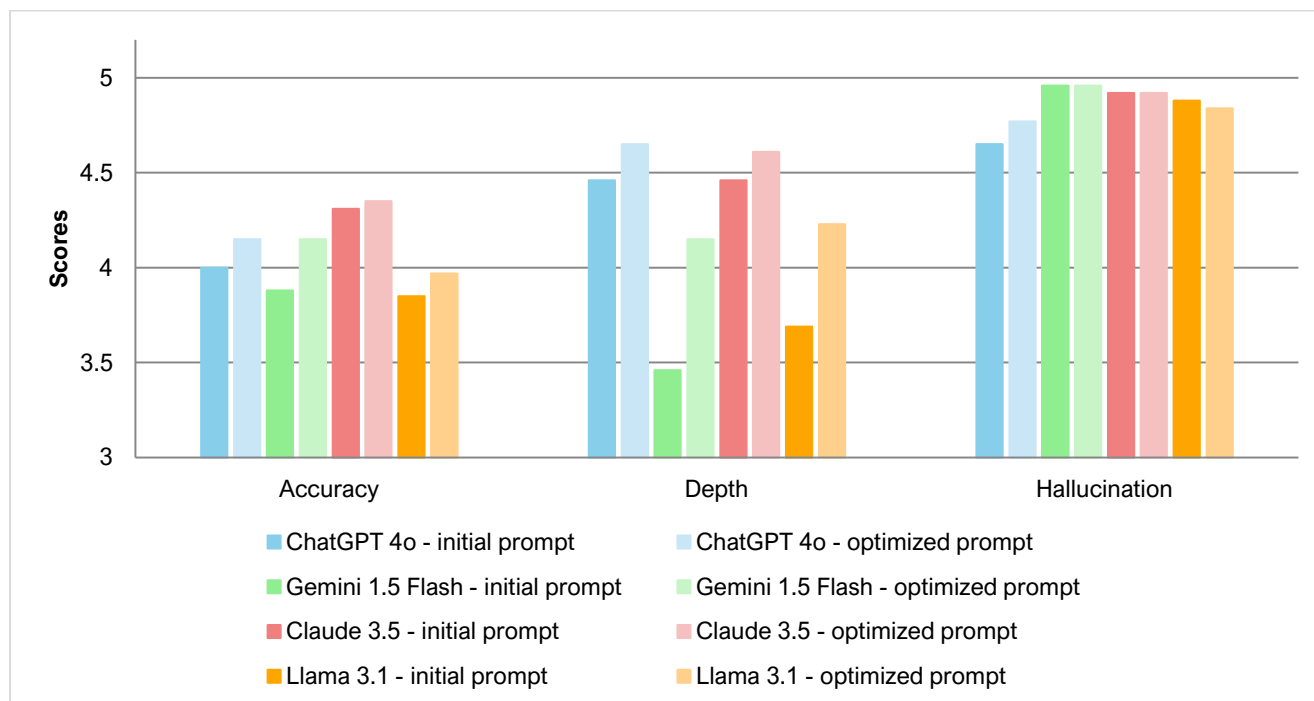


Figure 2. Performance comparison between original and optimised prompts.

4.3. Structural Adherence

Conformity of LLMs to the JSON output structure unexpectedly showed great variability, see Table 1. Claude was the only model to produce valid JSON consistently, while ChatGPT, Llama, and Gemini sometimes included comments within JSON, leading to invalid outputs. This underscores the need for stricter prompting, explicitly instructing models to exclude comments and follow structured formatting.

LLM	Structure Score	Violations of JSON
Claude 3.5 Sonnet	5.00	None
ChatGPT-4o	4.65	Minor structural deviations
Llama 3.1	4.56	Minor structural deviations
Gemini 1.5 Flash	4.38	Includes comments

Table 1. Structure Adherence scores for each LLM.

Overall, Claude 3.5 Sonnet emerges as the top performer, with high marks in accuracy, structural adherence, and depth, while maintaining a low hallucination rate. ChatGPT-4o also performs well across the board, particularly in depth, though it exhibits a slightly higher tendency to hallucinate. Therefore, both models, alongside their respective lightweight models Claude 3.5 Haiku and ChatGPT-4o mini, were chosen to be integrated into the data processing pipeline.

5. ARTEFACT GENERATION

This section details the development and implementation of the LLM-based IE system for SDC news as the main artefact. The goal was to create a seamless replacement for the existing rule-based system while significantly enhancing the platform's ability to extract and process information from SDC news articles.

5.1. Implementation of the LLM APIs

The implementation of Claude 3.5 Sonnet and GPT-4o follows a straightforward workflow: scraping SDC-relevant articles from the web or PDF-based newsletters and saving them as JSON file, loading the articles.json file, sending individual articles to the respective LLM for processing, converting the returned JSON responses (which might be more than one per article if multiple cables are mentioned) into a Pandas DataFrame, and finally generating the dataset update vectors, see Figure 3.

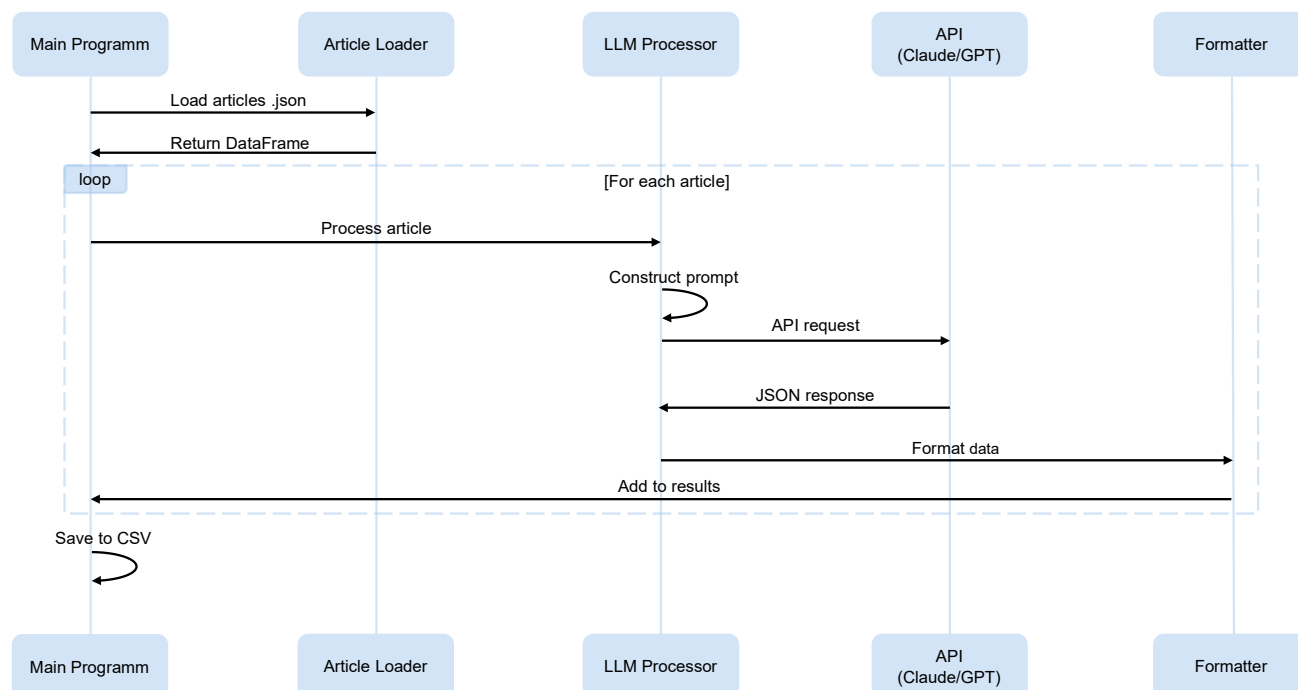


Figure 3. Sequence diagram of the complete IE process.

5.2. Multi-Model Support

The implementation allows for model selection at runtime, either through command-line parameters or default configuration. The core IE logic remains constant, with only API-specific handling differing. This is achieved through two different functions for the actual API call, one for Anthropic and one for OpenAI. They both return output of the same structure, so the post-processing is executed by other, model-independent functions.

5.3. Token Counting and Cost Management

The LLM processing is accompanied by a monitoring unit to track performance indicators and costs. To manage processing costs, a token counting function tracks both input and output tokens, as costs are based on the total tokens used. Different models have varying token limits, such as Claude 3.5 Sonnet (200,000 tokens) and GPT-4o (128,000 tokens). The implementation includes real-time cost estimation using current pricing data, though actual costs may vary due to provider optimizations. The full system architecture is displayed in Figure 4.

6. EVALUATION AND POST-TESTS

The evaluation covers two main areas: performance analysis of the implemented solution and a comparison with a rule-based approach.

6.1. Performance Analysis

The performance analysis revealed distinct variations in processing speed and cost efficiency across different extraction approaches, see Figure 5. The rule-based approach, serving as our baseline, demonstrated consistent processing speeds of 17 seconds for 20 articles. When evaluating LLM-based solutions, including the newly integrated Claude 3.5 Haiku model, both Claude variants showed comparable processing patterns, though Haiku exhibited slightly longer processing times for large datasets while offering significant cost advantages - processing 176 articles cost \$0.4704 with Haiku versus \$1.395 with Sonnet. OpenAI's models displayed notable performance characteristics: GPT-4o performed well with small datasets but experienced slower processing times with larger ones due to API rate limitations. In contrast, GPT-4o-mini proved more efficient, processing the complete dataset in 5 minutes compared to GPT-4o's 11 minutes, while maintaining the lowest cost at \$0.0714 for 176 articles. The time complexity analysis showed linear scaling for the rule-based approach, while LLM processing exhibited a flattened curve due to their ability to skip articles lacking relevant SDC information, unlike the rule-based system which processes all articles uniformly.

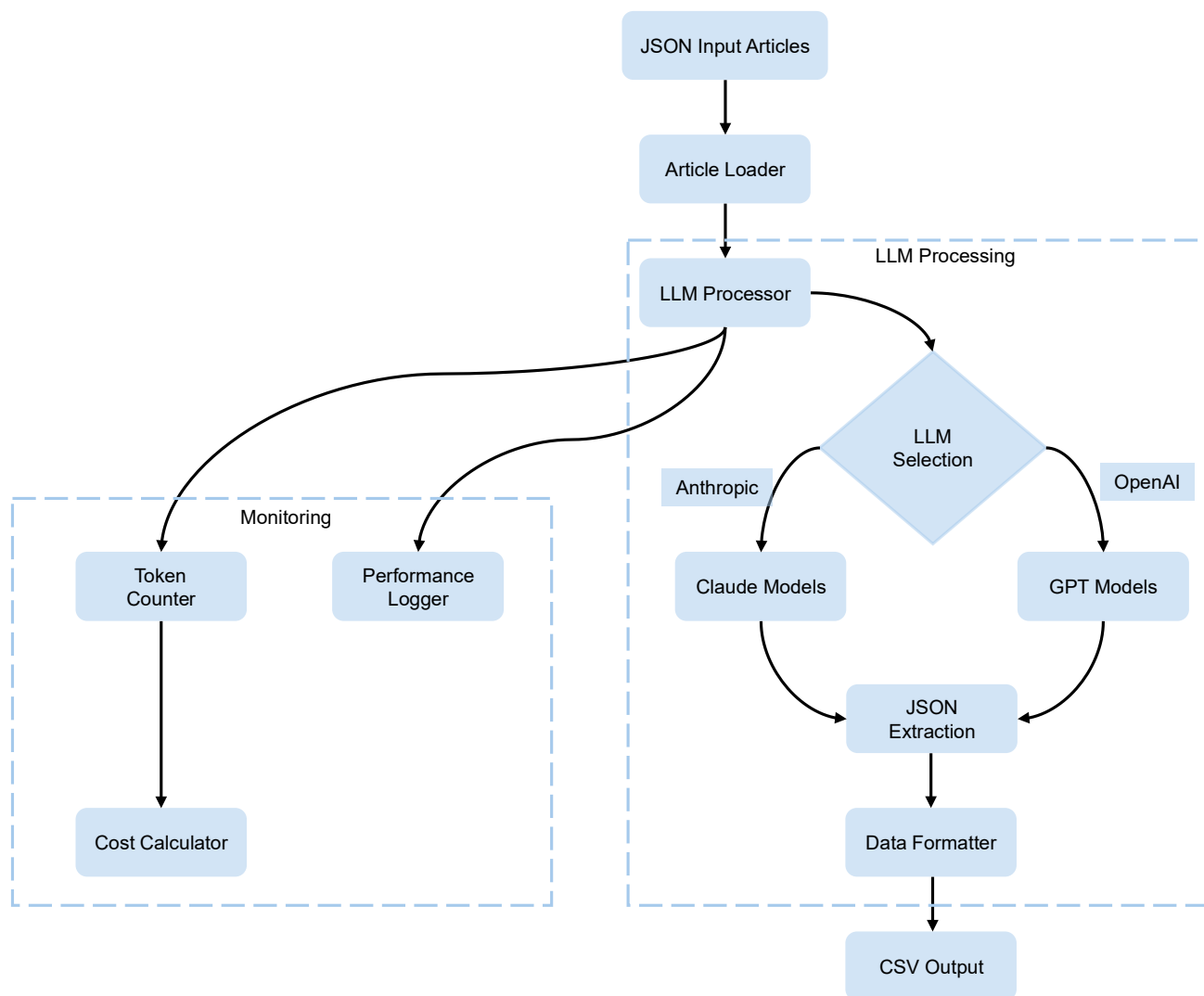


Figure 4. System architecture of the LLM-based information extraction pipeline.

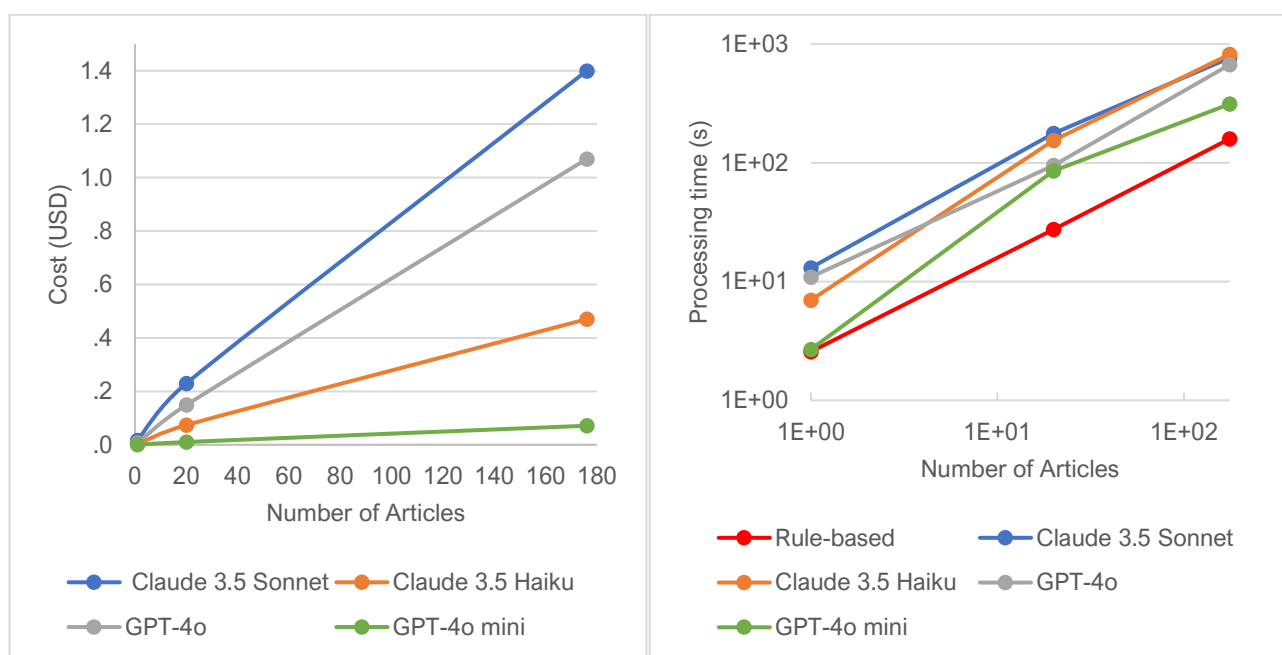


Figure 5. Cost comparison (left) and performance comparison (right, log scaled).

6.2. Comparative Evaluation

The comparison used three standard metrics: precision (the proportion of correctly extracted information among all extractions), recall (the proportion of relevant information successfully extracted) and the F1 score (the harmonic mean of precision and recall, providing a balanced measure of precision) which are standard metrics in statistical LLM analysis. These metrics were calculated by comparing the extracted data with a manually annotated ground truth dataset. The results, as presented in Table 2, demonstrate a substantial performance gap between the LLM-based and rule-based approaches. Claude 3.5 Sonnet achieved the highest overall F1 score of 0.268, followed by GPT-4o (0.233) and GPT-4o mini (0.208), while the rule-based system significantly underperformed with an F1 score of 0.016. This order-of-magnitude difference highlights the limitations of rigid pattern matching and predefined rules in handling the complexity and variability of natural language in news articles. At the same time, even those F1 scores are relatively low and point at the need for further refinement of the extraction systems.

Model	Precision	Recall	F1 Score	Processing Time
Claude 3.5 Sonnet	0.230	0.329	0.268	2m 47s
GPT-4o	0.228	0.263	0.233	1m 40s
GPT-4o mini	0.176	0.266	0.208	1m 01s
Rule-based	0.018	0.016	0.016	20s

Table 2. Performance comparison of IE approaches

Analysis of specific extraction categories, as shown in Table 3, reveals distinct patterns in performance across different types of IE tasks. All LLM variants demonstrated the strongest performance in identifying cable deployment status (F1 scores ranging from 0.459 to 0.561) and operational status information, such as cable breaks. (F1 scores between 0.400 and 0.536). This success can be attributed to the explicit and standardised way in which these events are typically reported in news articles. The rule-based system struggled significantly with these tasks, achieving negligible F1 scores in the two categories it was optimised for detection.

The detection of cable breaks, a critical aspect of SDC monitoring, demonstrates well the advantages of LLM-based approaches. As shown in Table 4, Claude 3.5 Sonnet achieved the highest performance in break detection. GPT-4o showed similar precision but lower recall, while the rule-based system did not correctly identify any cable breaks, producing only false positives and missing all actual events. This performance variation reveals the LLMs' ability to understand context and variations in how cable breaks are described in news articles.

Category	Claude 3.5	GPT-4o	GPT-4o mini	Rule-based
pot_new_cable	0.540	0.561	0.459	0.100
broken	0.536	0.468	0.400	N/A
planned	0.542	0.478	0.340	N/A
out_of_service	0.302	0.200	0.196	0.048

Table 3. Comparison of F1 scores across key extraction categories.

The performance advantage of LLMs comes with a trade-off in processing time. While the rule-based system completed the extraction task for the test data set of 20 articles in just 20 seconds, the LLM approaches required significantly more time. Claude 3.5 Sonnet took 2 minutes and 47 seconds, GPT-4o needed 1 minute and 40 seconds, and GPT-4o mini took 1 minute and 1 second. However, given the substantial improvement in extraction accuracy, this time investment appears justified for applications where precision is paramount. Still, even advanced LLMs face challenges when dealing with highly specific technical details, although they outperformed rule-based methods in this evaluation.

Model	Precision	Recall	F1 Score	TP/FP/FN
Claude 3.5 Sonnet	0.469	0.625	0.536	15/17/9
GPT-4o	0.478	0.458	0.468	11/12/13
GPT-4o mini	0.370	0.435	0.400	10/17/13

Table 4. Detailed analysis of cable break detection performance.
True Positives (TP), False Positives (FP), False Negatives (FN).

7. DISCUSSION

The analysis and the successful implementation of a data extractor demonstrate that LLMs can be effectively applied through a structured, multi-stage approach that leverages their natural language understanding capabilities while accounting for domain-specific requirements. In direct response to the RQ, this work demonstrates that LLMs can be effectively applied to extract and process SDC news data through (1) careful model selection based on empirical performance evaluation for the specific use case, (2) prompt optimization tailored to specific extraction tasks, and (3) robust implementation of structured data models and validation mechanisms.

The work revealed three critical success factors: balancing accuracy with efficiency in model selection, implementing robust error handling mechanisms, and maintaining a modular architecture adaptable to emerging models and requirements. It demonstrated that LLMs can effectively transform unstructured textual information from various sources into structured data, though their capabilities vary across information types. The significant performance improvements over rule-based approaches, combined with the system's inherent flexibility, establish LLMs as a viable solution for the technical domain IE. Despite limitations in processing highly specific technical details, the successful implementation establishes a framework adaptable to various technical domains requiring automated text processing. This approach provides a foundation for future applications where structured data must be extracted from unstructured text sources.

7.1. Limitations

The evaluation is based on a dataset of 27 articles, which limits the statistical generalisability of the model comparison, particularly in assessing model performance patterns and comparative analyses. Similarly, the evaluation of the implementation, comparing the new LLM-based extractor with the rule-based approach, relies on a relatively small ground truth, which may affect the robustness of the findings. Future work should expand the manually annotated benchmark, include additional sources and languages, and further validate the extraction results against a larger ground-truth dataset. While the implemented pipeline was tested on a larger corpus of 176 articles to assess processing time, cost, and practical feasibility, this larger corpus was not manually annotated to the same level of detail as the benchmark dataset. Consequently, the larger test demonstrates the operational scalability of the pipeline, but it does not provide the same level of evidence for extraction accuracy. Therefore, future evaluations may combine larger-scale processing tests with more extensive manual annotation or semi-automated validation procedures.

Additionally, the rapid advancement of LLMs presents a challenge, as newer models could outperform those evaluated here, potentially altering the comparative results over time. Another current limitation is the lack of advanced Event Coreference Resolution (ECR) mechanisms, as while the system successfully identifies individual events, it does not yet effectively track and correlate related events across multiple sources (Lu and Ng, 2018). Implementing ECR in future research could offer greater analytical value and improve event monitoring capabilities. Those interested to further advance this research are invited to contact the first author for access to the public parts of the code, project documentation, and prompts.

7.2. Future Work

The current system, developed for SDC news, demonstrates the potential for broader applications across various infrastructure domains. It could be particularly valuable for analysing other ICT infrastructures, helping to organise and synthesise fragmented information landscapes. More broadly, the system could be adapted for any domain characterised by infrastructure invisibility – whether physical (underground/undersea), institutional, or perceptual – where there is a need to aggregate and analyse dispersed information. This could include critical infrastructures such as underground power grids, pipeline networks, or distributed communication systems. The methodology's ability to extract and synthesise information

from diverse sources could help improve visibility and understanding of these often overlooked yet crucial systems, supporting better decision-making and risk assessment in infrastructure management.

The potential of fine-tuning LLMs specifically for SDC news analysis represents a particularly compelling direction. Fine-tuning is the process of further training a pre-trained LLM on a specific dataset to adapt and optimise its performance for particular tasks or domains, for example, in the domain of code generation as shown by Weyssow et al. (2024), using Parameter-Efficient Fine-Tuning (PEFT) techniques. Hu et al. (2023) introduce the concept of adapters to enhance PEFT. Parthasarathy et al. (2024) describe a seven-step framework for fine-tuning that includes hyperparameter optimisation. Although the current implementation achieves promising results through systematic prompt engineering, a fine-tuned model could potentially offer enhanced performance, particularly in processing technical specifications and industry-specific terminology. Fine-tuning can be implemented on both commercially available models and proprietary hardware. However, this approach necessitates a substantial, annotated dataset within the target domain - specifically, a more comprehensive collection of annotated news articles about SDCs - to achieve optimal results.

Additionally, future work could use quantitative hallucination measurement to examine the quality of model outputs. Widely used frameworks, like FactScore (Min *et al.*, 2023) for atomic factuality checking and RARR (Gao *et al.*, 2023) for attribution, produce clear indicators of models' propensity for hallucination, and could be used to complement qualitative Likert scores.

8. CONCLUSION

This paper investigated the application of LLMs for extracting and processing news data in the SDC network domain. The research addressed the automated monitoring and analysis of SDC events, demonstrating that LLMs can effectively replace rule-based approaches while providing substantially improved performance. Most notably, the research demonstrated that carefully prompted LLMs can achieve significantly higher accuracy in IE compared to rule-based methods. A key contribution of this work is the development of a modular, adaptable system architecture that enables the seamless integration of different LLM providers. The implementation demonstrated that LLMs can be used effectively for specialised technical domains while maintaining output reliability and consistency. The use of structured prompts and robust output validation mechanisms proved essential for achieving consistent performance across various IE tasks. The research highlights limitations in LLM applications for technical domains. While being effective in reading between the lines of natural language news articles, the models struggled with specific details like costs and capacities. Here, a hybrid approach may be optimal. Additionally, analysis of processing times, token usage, and costs can inform future deployments. Though LLMs are slower than rule-based methods, their accuracy often justifies trade-offs, and emerging smaller models may soon reduce this gap.

The implications of this research extend beyond the domain of SDCs. The demonstrated success in applying LLMs to extract domain-specific information suggests potential applications in other specialised technical fields. The modular architecture and prompt engineering approaches developed in this work provide a framework that could be adapted for similar information extraction tasks in different domains.

ACKNOWLEDGEMENTS

This research work was funded by The German Federal Ministry of Education and Research and the Hessian Ministry of Science and Research, Arts and Culture within their joint support of the National Research Center for Applied Cybersecurity ATHENE as well as the LOEWE initiative (Hesse, Germany) within the emergenCITY centre. Funding code: LOEWE/1/12/519/03/05.001(0016)/72. The funding sources were not involved in the contents of the study. The authors would like to thank Karen Beuck, Lydia Ebbinghaus, Thies Lukas, and Lennart Oestreich for their programming contributions to the rule-based data processing pipeline and the ground truth annotation, as well as SubCableNews for the permission to analyse SDC-related articles from their database.

CONFLICT OF INTEREST

Authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

NOTICE

The present paper has been invited to be published as an extended version of the authors' previous work on this subject. It is an extension of the conference paper 'Between the Lines: Leveraging LLMs for Information Extraction of Subsea Data Cable News' presented at the 5th European Workshop on Maritime Systems, Resilience and Security 2025 (MARESEC

2025), Rostock, Germany (doi: <https://doi.org/10.5281/zenodo.17119935>). The structure of the original paper has been modified to meet the requirements of this journal. The content of this paper is based on the previously published version, but it has been revised and supplemented with several new aspects.

REFERENCES

- Bogdanov, S. et al. (2024) 'NuNER: Entity Recognition Encoder Pre-training via LLM-Annotated Data', in Y. Al-Onaizan, M. Bansal, and Y.-N. Chen (eds) Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Miami, Florida, USA: Association for Computational Linguistics, pp. 11829–11841. Available at: <https://doi.org/10.18653/v1/2024.emnlp-main.660>.
- Bruckschen, E. (2025) SubCableNews. Available at: <https://subcablenews.com/> (Accessed: 3 March 2026).
- Bueger, C. and Edmunds, T. (2017) 'Beyond seablindness: a new agenda for maritime security studies', *International Affairs*, 93(6), pp. 1293–1311. Available at: <https://doi.org/10.1093/ia/iix174>.
- Bueger, C. and Liebetrau, T. (2021) 'Protecting Hidden Infrastructure: The Security Politics of the Global Submarine Data Cable Network', *Contemporary Security Policy*, 42(3), pp. 391–413. Available at: <https://doi.org/10.1080/13523260.2021.1907129>.
- Bueger, C., Liebetrau, T. and Franken, J. (2022) Security threats to undersea communications cables and infrastructure – consequences for the EU. In-Depth Analysis Requested by the SEDE sub-committee PE702557. European Parliament. Available at: [https://www.europarl.europa.eu/RegData/etudes/IDAN/2022/702557/EXPO_IDA\(2022\)702557_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2022/702557/EXPO_IDA(2022)702557_EN.pdf) (Accessed: 3 March 2026).
- Cardenas, I.C. et al. (2025) 'Accident analysis reinforced by natural language processing: the case of interactions between maritime and diving operations', *Journal of Marine Engineering & Technology*, pp. 1–14. Available at: <https://doi.org/10.1080/20464177.2025.2546226>.
- Caruso, M. et al. (2026) 'Seismic detection using submarine cable polarization signals with machine learning', *Communications Earth & Environment* [Preprint]. Available at: <https://doi.org/10.1038/s43247-026-03434-x>.
- Clark, K. and Humera, S. (2024) Submarine Cable Almanac - Issue 50. Edited by W. Nielsen. 50. Submarine Telecoms Forum, Inc. Available at: https://issuu.com/subtelforum/docs/almanac_issue_50 (Accessed: 3 March 2026).
- European Commission (2024) Recommendation on the security and resilience of submarine cable infrastructures. Commission Recommendation C(2024) 1181 final. Brussels, pp. 1–15. Available at: <https://digital-strategy.ec.europa.eu/en/library/recommendation-security-and-resilience-submarine-cable-infrastructures> (Accessed: 3 March 2026).
- European Commission (2026) Security and Resilience of EU Submarine Cable Infrastructures: Submarine Cable Security Toolbox and Cable Projects of European Interest. Brussels: Submarine Cable Infrastructures informal Expert Group. Available at: <https://digital-strategy.ec.europa.eu/en/library/submarine-cable-security-toolbox-and-cable-projects-european-interest> (Accessed: 30 March 2026).
- European Parliament and European Council (2025) EU Action Plan on Cable Security. JOIN(2025) 9. Brussels: High Representative of the Union for Foreign Affairs and Security Policy. Available at: <https://digital-strategy.ec.europa.eu/en/library/joint-communication-strengthen-security-and-resilience-submarine-cables> (Accessed: 3 March 2026).
- Evans, J., Sadruddin, S. and D'Souza, J. (2024) Astro-NER – Astronomy Named Entity Recognition: Is GPT a Good Domain Expert Annotator? Available at: <http://arxiv.org/abs/2405.02602> (Accessed: 19 June 2024).
- Franken, J. and Reuter, C. (2024) 'The Subsea Data Cable Security Map – Fusing Public Information for Enhanced Critical Maritime Infrastructure Security', Proceedings of the MARESEC 2024 / Transactions on Maritime Science. Bremerhaven, Germany. Available at: <https://doi.org/10.5281/zenodo.14216269>.
- Franken, J., Schneider, F. and Reuter, C. (2023) 'The Internet's Plumbing Consists of Garden Hoses: A Critical Analysis of the Advantages and Pitfalls of Metaphors Use for Critical Maritime Infrastructures', in H. Schilling (ed.) Dreizack 23. Kiel: The Kiel Seapower Series, pp. 1–8. Available at: https://www.kielseapowerseries.com/files/ispk/content/workshops/Dreizack/Dreizack23_Konferenzband.pdf (Accessed: 3 March 2026).
- Gao, L. et al. (2023) 'RARR: Researching and revising what language models say, using language models', Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, pp. 16477–16508. Available at: <https://doi.org/10.18653/v1/2023.acl-long.910>.
- Hadi, M.U. et al. (2024) 'Large Language Models: A Comprehensive Survey of Its Applications, Challenges, Limitations, and Future Prospects'. Techrxiv. Available at: <https://doi.org/10.36227/techrxiv.23589741.v6>.
- Hu, Z. et al. (2023) 'LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models', Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp. 5254–5276. Available at: <https://doi.org/10.18653/v1/2023.emnlp-main.319>.
- Huguet Cabot, P.-L. and Navigli, R. (2021) 'REBEL: Relation Extraction By End-to-end Language Generation', in M.-F. Moens et al. (eds) Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, pp. 2370–2381. Available at: <https://doi.org/10.18653/v1/2021.findings-emnlp.204>.
- Johnson, K. (2024) The Houthis' Next Target May Be Underwater. Foreign Policy. Available at: <https://foreignpolicy.com/2024/02/07/houthi-red-sea-attacks-submarine-cables/> (Accessed: 3 March 2026).
- Josifoski, M. et al. (2022) 'GenIE: Generative Information Extraction', Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, pp. 4626–4643. Available at: <https://doi.org/10.18653/v1/2022.naacl-main.342>.

- Karmakar, S. (2023) 'Modeling, navigation and artificial intelligence (AI) based fault detection of subsea cable network', 2023 IEEE Green Technologies Conference (GreenTech), pp. 244–249. Available at: <https://doi.org/10.1109/GreenTech56823.2023.10173836>.
- Kavanagh, C., Franken, J. and He, W. (2025) Achieving Depth: Subsea Telecommunications Cables as Critical Infrastructure. Geneva: United Nations Institute for Disarmament Research (UNIDIR). Available at: <https://unidir.org/publication/achieving-depth-subsea-telecommunications-cables-as-critical-infrastructure/> (Accessed: 3 March 2026).
- Kavanagh, C., Franken, J. and Kulesza, J. (2026) 'In Oceans and in Orbits: The Infrastructure Essential to the Public Core of the Internet', in D.W.J. Brouders and Arun M. Sakumar (eds) *Protecting the Internet's Core* (forthcoming). Cambridge, MA: MIT Press.
- Lei, H. et al. (2024) 'Design and Application of New Technology for Intelligent Inspection and Early Warning of Submarine Cables against External Breakage', 2024 4th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), pp. 696–701. Available at: <https://doi.org/10.1109/ACCTCS61748.2024.00129>.
- Lu, J. and Ng, V. (2018) 'Event Coreference Resolution: A Survey of Two Decades of Research', *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization*, pp. 5479–5486. Available at: <https://doi.org/10.24963/ijcai.2018/773>.
- Min, S. et al. (2023) 'Factscore: Fine-grained atomic evaluation of factual precision in long form text generation', *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 12076–12100. Available at: <https://doi.org/10.18653/v1/2023.emnlp-main.741>.
- Nakshatri, N. et al. (2023) 'Using LLM for Improving Key Event Discovery: Temporal-Guided News Stream Clustering with Event Summaries', in H. Bouamor, J. Pino, and K. Bali (eds) *Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics*, pp. 4162–4173. Available at: <https://doi.org/10.18653/v1/2023.findings-emnlp.274>.
- Pahune, S. and Chandrasekharan, M. (2023) 'Several Categories of Large Language Models (LLMs): A Short Survey', *International Journal for Research in Applied Science and Engineering Technology*, 11(7), pp. 615–633. Available at: <https://doi.org/10.22214/ijraset.2023.54677>.
- Palmer-Felgate, A., Perrier, P.A. and Kawasaki, T. (2025) 'Directions for future subsea cable systems', in J. Chesnoy and J.-C. Antona (eds) *Undersea Fiber Communication Systems (Third Edition)*. Third Edition. Academic Press, pp. 301–324. Available at: <https://doi.org/10.1016/B978-0-443-29886-8.00006-2>.
- Parthasarathy, V.B. et al. (2024) *The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities*. Available at: <http://arxiv.org/abs/2408.13296> (Accessed: 3 March 2026).
- Safuan and Syafira, A. (2024) 'Artificial Intelligence in Indonesian Ports: Opportunities and Challenges', *Transactions on Maritime Science*, 13(2). Available at: <https://doi.org/10.7225/toms.v13.n02.w07>.
- Soner, O. (2025) 'Discovering operational risk patterns in maritime accident reports through latent theme extraction', *Australian Journal of Maritime & Ocean Affairs*, pp. 1–21. Available at: <https://doi.org/10.1080/18366503.2025.2534261>.
- Stronge, T. (2024) What We Know (And Don't) About Multiple Cable Faults in the Red Sea. Available at: <https://blog.telegeography.com/what-we-know-and-dont-about-multiple-cable-faults-in-the-red-sea> (Accessed: 3 March 2026).
- Subramanian, A. (2024) An Introduction to Prompting for LLMs. Medium. Available at: <https://towardsdatascience.com/an-introduction-to-prompting-for-llms-61d36aec2048> (Accessed: 3 March 2026).
- TeleGeography (2026) Submarine Cable Map. Available at: <https://www.submarinecablemap.com/> (Accessed: 3 March 2026).
- Wan, Z. et al. (2023) 'GPT-RE: In-context Learning for Relation Extraction Using Large Language Models', *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, pp. 3534–3547. Available at: <https://doi.org/10.18653/v1/2023.emnlp-main.214>.
- Wang, S. et al. (2023) 'GPT-NER: Named Entity Recognition via Large Language Models'. Arxiv. Available at: <http://arxiv.org/abs/2304.10428> (Accessed: 3 March 2026).
- Weyssow, M. et al. (2024) 'Exploring Parameter-Efficient Fine-Tuning Techniques for Code Generation with Large Language Models'. Arxiv. Available at: <http://arxiv.org/abs/2308.10462> (Accessed: 3 March 2026).
- Xu, D. et al. (2024) 'Large Language Models for Generative Information Extraction: A Survey', *Frontiers of Computer Science*, 18(6), p. 186357. Available at: <https://doi.org/10.1007/s11704-024-40555-y>.
- Xu, X. et al. (2023) 'How to Unleash the Power of Large Language Models for Few-Shot Relation Extraction?', *Proceedings of the Fourth Workshop on Simple and Efficient Natural Language Processing (SustainLP)*, pp. 190–200. Available at: <https://doi.org/10.18653/v1/2023.sustainlp-1.13>.
- Zhang, Z. et al. (2024) 'LinkNER: Linking Local Named Entity Recognition Models to Large Language Models Using Uncertainty', *Proceedings of the ACM Web Conference 2024, ACM*, pp. 4047–4058. Available at: <https://doi.org/10.1145/3589334.3645414>.

APPENDIX

The appendix includes the articles used for the evaluation (Table 5), comparison metrics across different LLM models with scores on the 1-5 scale for various criteria (Table 6), and the various prompts used (C.1-C.3).

Analysed Articles

ID	URL
BBC-1	https://www.bbc.com/news/articles/cjr77zj8lj4o
BBC-2	https://www.bbc.com/news/world-asia-60021759
BBC-3	https://www.bbc.com/news/articles/cprg0yn8q81o
BBC-4	https://www.bbc.com/news/world-africa-68569022
Reuters-1	https://www.reuters.com/world/asia-pacific/three-vietnams-five-undersea-internet-cables-are-down-2024-06-17/
Reuters-2	https://www.reuters.com/world/africa/internet-disruption-hits-west-central-africa-netblocks-cloudflare-data-shows-2024-03-14/
Reuters-3	https://www.reuters.com/world/europe/sweden-says-telecom-cable-with-estonia-damaged-2023-10-17/
Guardian-1	https://www.theguardian.com/technology/2024/mar/14/much-of-west-and-central-africa-without-internet-after-undersea-cable-failures
Guardian-2	https://www.theguardian.com/world/2023/oct/17/telecoms-cable-between-sweden-and-estonia-damaged-sweden-says
Guardian-3	https://www.theguardian.com/world/article/2024/jul/16/parts-of-tonga-without-internet-after-cables-damaged-and-starlink-ordered-to-cease-operations
Guardian-4	https://www.theguardian.com/uk-news/2022/oct/20/shetland-loses-telephone-internet-services-subsea-cable-damaged
SCN-1	BlueMed landing in Bastia (SCN No. 258)
SCN-2	The India-Europe-Xpress cable lands in Djibouti (SCN No. 258)
SCN-3	New Egypt-Jordan submarine cable to power AI and data centers demands (SCN No. 260)
SCN-4	KLI submarine cable system inaugurated (SCN No. 259)
SCN-5	Bayobab lands 2Africa subsea cable in Ghana and Nigeria to strengthen internet connectivity across West Africa (SCN No. 260)
SCN-6	Disruption to international submarine cables off the West African Coast (WACS, SAT-3, MainOne, ACE) (SCN No. 261)
SCN-7	Seacom confirms cable outage in Red Sea (SCN No. 261)
STF-1	https://subtelforum.com/another-internet-cable-fault-hits-south-africa/
STF-2	https://subtelforum.com/deep-blue-one-cable-goes-live-in-suriname/
STF-3	https://subtelforum.com/vavau-and-haapai-suffer-third-cable-blackout/
DCD-1	https://www.datacenterdynamics.com/en/news/sea-h2x-subsea-cable-lands-in-the-philippines/
DCD-2	https://www.datacenterdynamics.com/en/news/peace-subsea-cable-goes-live/
DCD-3	https://www.datacenterdynamics.com/en/news/seamewe-6-lands-in-the-maldives/
DCD-4	https://www.datacenterdynamics.com/en/news/deep-blue-one-cable-goes-live-in-suriname/
DCD-5	https://www.datacenterdynamics.com/en/news/digicels-deep-blue-one-cable-goes-live-in-the-caribbean-and-south-america/

Table 5. Sources and URLs of analysed Articles

Scoring Criteria

Score	Accuracy measures how precisely the LLM extracts relevant details.	Depth evaluates the thoroughness of contextual and background information.	Hallucination Rate tracks the frequency of fabricated or incorrect information.	Structure assesses adherence to the specified JSON format.
1	Significant errors or omissions.	Only basic information extracted.	Frequent hallucinations.	Invalid JSON with major structural errors.
2	Frequent minor errors.	Limited detail, missing significant context.	Several hallucinations.	Valid JSON but poor structure adherence.
3	Some inaccuracies but generally correct.	Moderate depth, some background provided.	Occasional hallucinations.	Correct basic structure with minor deviations.
4	Mostly accurate with very few mistakes.	Detailed extraction, with most relevant context.	Rare hallucinations.	Well-structured with minimal deviations.
5	Perfect or near-perfect accuracy.	Comprehensive and thorough extraction.	No hallucinations detected.	Perfect structure adherence.

Table 6. Scoring criteria for LLM evaluation

Prompts

Initial prompt

Extract all relevant details from the following news article about submarine data cables in a structured format. The details should include:

- 1) *Event Type*
- 2) *Submarine Cable Name*
- 3) *Date of Event*
- 4) *Location*
- 5) *Stakeholders/Companies Involved*
- 6) *Impact/Consequence*

Provide the extracted information as JSON. News article:
[...]

Optimised prompt

Extract all relevant details from the following news article about SDCs in a structured JSON format. Only extract details that are explicitly mentioned in the article. If any of the requested information is not available, leave the respective field blank or null. Output only a JSON in the following form:

```
{
  "event_type": "", // Specify if it is a "cable deployment", "cable maintenance", "cable outage", or other significant event.
  "cable_name": "", // Provide the name of the submarine cable affected. If more are affected, return an array of strings. If unknown, output the string "unknown".
  "event_date": "", // Mention the date the event occurred. Format: YYYY-MM-DD
  "location": "", // Specify the location where the event occurred. Output as one string.
  "stakeholders": [], // List the companies or entities involved.
  "technical_specs": {}, // Include detailed technical specifications, such as cable capacity, length, and technology (if available).
  "impact": "", // Describe any impact or consequence mentioned in the article (e.g. network disruption, economic effects).
  "historical_context": "", // Provide historical or strategic background mentioned in the article.
  "financial_info": "", // Extract any financial data, such as investment amount or costs.
```

```
"additional_details": "" // Extract any further explanations, analysis, or indirect impacts mentioned.  
}
```

Ensure the output is properly formatted in JSON. Please provide detailed and comprehensive information, avoiding overly brief or summary responses. News article: [...]

System prompt

"You are an expert analyst specialising in submarine cable systems and telecommunications infrastructure. Your role is to extract precise, factual information from news articles about SDCs. You should focus on technical details, stakeholder information, and event specifics. Avoid speculation and only report information explicitly stated in the given text. Your responses should be structured, concise, and directly relevant to SDC systems. If information is not provided or unclear, indicate this rather than making assumptions. Your expertise in telecommunications terminology and industry trends should guide your analysis, ensuring accurate and relevant data extraction."