Is the Extra Compute Worth It? Evaluating Rule Based and Language Model Approaches to Title Generation

Diego González-Suárez*[1][0009−0004−6388−5333], Noelia Rico[1][0000−0002−6169−4523], and Irene Díaz[1][0000−0002−3024−6605]

Department of Computer Science, University of Oviedo, Spain
{gonzalezsdiego,noeliarico,sirene}@uniovi.es

Abstract. This work presents a comparative study on automatic title generation for institutional communication, focusing on the trade-off between result quality and computational cost in rule based methods and large language models. The study is grounded in a real-world application developed for the communication team of the University of Oviedo, aimed at supporting the creation of titles for legal bulletin news items from regional, national, and European sources, including research calls, academic job vacancies, regulations, and other relevant announcements of interest to the university community.
A tool was designed and deployed to integrate three alternative title generation options: a rule based system based on manually defined transformations, and two generation pipelines based on large language models. This enables users to compare the generated titles, select their preferred option, or provide a custom title when none of the suggestions are suitable, facilitating the collection of interaction data and user preferences within their daily workflow.
The quality of the generated titles is evaluated in terms of expert preference, while computational cost is considered in terms of the economic cost associated with the hardware required to support each approach. The results highlight a trade-off between output quality and resource requirements, providing insights into whether the improvements achieved by language models justify their additional cost in practical settings.

Keywords: Automatic Title Generation · Pretrained Large Language Models · Human Evaluation · Cost–Quality Trade-off.

1 Introduction

In recent years, Large Language Models (LLMs) have demonstrated remarkable capabilities in performing high-quality tasks across a wide range of domains. As a result, these models have undergone rapid development and diversification, becoming increasingly accessible and widely used. These advances have fostered their adoption across different industries and application domains.

* Corresponding author: gonzalezsdiego@uniovi.es

In Spain, the adoption of this kind of LLMs is rapidly increasing across both public institutions and private companies. For example, according to a study carried out by ESADE in May 2025 [4], regarding the public sector, approximately 54% of employees report using AI-based tools at least occasionally, and 66% of these users perceive their impact as positive. These tools are commonly applied to tasks such as data analysis, translation, and processing large volumes of textual information. In document-intensive workflows, this often involves the need to identify and condense the most relevant information efficiently, making summarization a recurring and time-consuming activity. Automating this process, through the use of specialized LLMs, may have the potential to improve efficiency in workflows that rely on the management and dissemination of textual data.

In line with the growing range of tasks that can potentially be automated through LLM-based solutions, this work focuses on a specific problem arising in the daily activities of the communication office of the University of Oviedo. A daily announcement is distributed to the entire university community, comprising several sections, one of which includes news items extracted from official gazettes that are relevant to academic and research staff. For each of these items, the communication team is responsible for manually generating concise and informative titles. Although this task may appear minor, it constitutes a daily repetitive and time-consuming activity. When combined with other tasks that require greater attention and editorial effort, it contributes to an increased workload. This motivates the need to explore automated solutions that can assist in the generation of such titles, reducing manual effort while maintaining adequate quality. However, such automation is not without cost, and it becomes necessary to assess whether the use of computationally intensive approaches, such as large language models, is justified in comparison to lighter alternatives.

The paper is organized as follows. Section 2 describes the challenges associated with the current institutional workflow. Section 3 reviews the most relevant literature, with a particular focus on summarization and title generation. Section 4 presents the approach and experimental setup followed in this study. Section 5 exposes an analysis of the data obtained. Finally, Section 6 summarizes the findings and outlines directions for future research.

2 Problem Description

At the University of Oviedo, a small team of five communication professionals is responsible for the daily publication of news and events across websites, mailing lists, and other channels. This team reviews and curates communications submitted by members of the university community, deciding whether they are suitable for dissemination. The scale of this task is non-trivial, as the university comprises over 3,000 staff members, including academic, administrative, and technical personnel, as well as approximately 20,000 students. In addition to these responsibilities, the team is also involved in a wide range of other tasks typical of institutional communication services. As a result, they must manage

large volumes of information originating from a variety of sources. Among these activities, a particularly mechanical and recurring task consists of publishing information provided by the legal department, which is responsible for selecting relevant content from daily official gazettes based on the following sources:

- The Official Gazette of the Principality of Asturias (BOPA[1], *Boletín Oficial del Principado de Asturias*), corresponding to the autonomous community in which the University of Oviedo is located, where it serves as the only public university.
- The Spanish Official State Gazette (BOE[2], *Boletín Oficial del Estado*).
- The Official Journal of the European Union (DOUE[3], *Diario Oficial de la Unión Europea*)

The process followed to transform information from official gazettes into published news items involves several stages and coordination among departments, which are detailed in Table 1.

Time	**Stage**	**Description**
00:00 AM	**Bulletins published**	Official bulletins are published by the corresponding government authorities in official gazettes.
08:00 AM (45 min)	**Legal department filtering**	The university's legal department reviews official gazettes to retrieve relevant information based on institutional interests. This process is carried out manually by copying and pasting content from official sources into a text document, which is then shared with the communication team via email.
08:45 AM (15 min)	**Structure analysis**	Once the document is received, the communication team analyzes the content. This stage involves extracting key elements associated with each bulletin item, such as descriptions and links. Additionally, the diversity of file types, formats, and the complexity of the legal language further complicate this process.
09:00 AM (45 min)	**Title generation**	Based on the analysis, communication professionals manually generate concise titles for each bulletin item that summarize the most notable information. This stage requires an understanding of creativity and journalism techniques in order to create well-formed titles.
09:45 AM (5 min)	**Publication on communication channels**	Finally, news items, along with their corresponding titles, descriptions, and links, are published across different channels, including the university website and institutional mailing lists.

Table 1. Current workflow for processing institutional notifications from official gazettes. Times are indicative and may vary depending on the daily volume of news, either upward or downward.

The tasks outlined in Table 1 are carried out in an unstructured manner, and their outcomes often depend on contextual factors as well as the preferences of

[1] `https://miprincipado.asturias.es/bopa`
[2] `https://www.boe.es/`
[3] `https://eur-lex.europa.eu/oj/direct-access.html`

the individual responsible, which further complicates the overall process. Among these stages, the title generation phase is particularly challenging, as communication professionals must deal with complex legal texts. Consequently, this phase may represent an ideal candidate for assessing the integration of automated tools.

3 Related Work

Automatic title and headline generation has been previously explored in natural language processing (NLP). However, it still presents several open challenges and opportunities for further research. As with many NLP tasks, difficulty varies significantly between languages and is further shaped by contextual and cultural factors that influence how information is managed.

This task is closely related to summarization. In fact, it could be viewed as a particular case of summarization, where the goal is to condense the most relevant information into a very short text output, maintaining the original meaning [14]. In the context of summarization, early approaches were primarily extractive, relying on statistical and heuristic methods such as frequency-based techniques [3,7] and graph-based ranking algorithms such as TextRank [8]. These methods were later complemented by neural approaches based on sequence-to-sequence setups [15], which were further improved with the Transformer architecture and attention mechanisms [17] that selected the most relevant parts of the input data, improving the ability to generate shorter outputs. Summarization research has continued to advance, with approaches based on task-specific models such as BART [6], PEGASUS [19] and T5 [12]. More recently, general-purpose models such as LLaMA [16], GPT [10] and Qwen [11] have demonstrated strong capabilities across a wide range of NLP tasks, enabling effective text summarization in zero-shot and few-shot settings.

In parallel, a growing body of research has addressed title generation as a more specific task of summarization. Although some studies formulate title generation in relatively general terms, the task often depends on domain-specific conventions, vocabulary, sentence structures, and communication goals. For instance, Mishra et al. [9] propose a general Transformer-based pipeline, but validate it on scientific abstracts. Similarly, Khanna et al. [18] introduce a domain-controlled generation framework that constrains the output using domain-specific vocabulary, while Rehman et al. [13] explore the use of fine-tuned pretrained language models to generate titles from research abstracts.

4 Methodology

This study adopts a comparative approach to evaluate different methods for supporting and partially automating the title generation process in institutional communication, with the aim of improving the efficiency of communication professionals. To this end, we compare rule-based methods and LLM-assisted approaches to evaluate and analyze their relative effectiveness.

To carry out this comparative study, we have developed a dedicated web application that supports the title generation workflow. The platform allows users to upload documents and generate titles for each bulletin entry. It serves as an interface that displays the extracted information along with the generated title candidates. This setup enables a structured evaluation of the different approaches.

The following subsections describe the data collection process, the methods used for title generation, and the deployment of the experiment. Subsection 4.1 describes the data collection procedure, including the application interface and the interaction of users with the platform. Subsection 4.2 details the different approaches used to generate candidate titles. Finally, subsection 4.3 exposes technical aspects related to application deployment.

4.1 Data Collection

Data collection was carried out through the designed application during its use by communication professionals. As users interact with the application, they upload documents provided by the legal department. After processing the documents with the news chosen by the legal department, the system presents multiple candidate titles for each bulletin item. In this step, users may select one of the proposed titles or manually edit or create a new one based on their expertise and preferences. To avoid bias in this step, users are not informed about the generation method associated with each candidate title. The final selected title is stored as the reference for publication, while all generated candidate titles are also retained. This process is repeated for each bulletin entry detected in the document. It makes possible to take advantage of their expertise, allowing for a comparative analysis between human-selected titles and those produced by different generation methods.

As shown in Figure 1, the interface is organized into a split-view layout designed to facilitate data verification. On the left-hand side, the application displays the document crafted by the legal department with the items selected from each official gazette. This allows users to directly inspect the source content and compare it with the candidate titles. The right-hand side works as the primary control panel; it features a selector to navigate among different detected bulletin entries, showing for the one selected the full list of generated candidate titles, the extracted description, and the links identified by the system. Users can choose a title via radio buttons, use quick-copy shortcuts for each text, and click submit to finalize.

4.2 Title Generation Methods

This subsection describes the title generation methods evaluated in this study. The considered approaches are as follows:

- **Rule-based.** These methods are suitable for headline and title generation when concise and stylistically controlled outputs are needed, as shown in

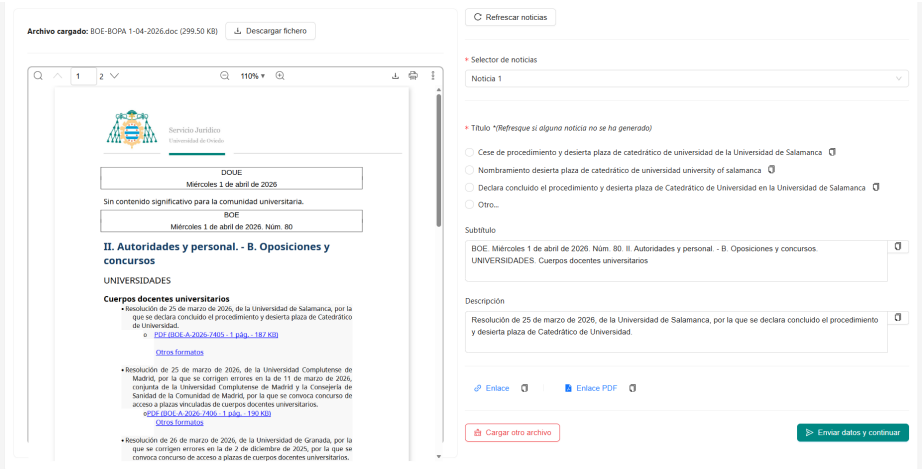


Fig. 1. Screenshot of the developed application interface.

early linguistic-rule-based work [2] and in more recent title generation approaches that partially incorporate linguistic rules [1]. Our approach implements a lightweight deterministic pipeline tailored to institutional legal bulletins. The method applies manually defined transformations derived from practices of communication professionals. The characteristics of this method are summarized as follows:

- **Pipeline:** The input description is first cleaned by enclosing quotation marks, trailing punctuation, and redundant whitespace. Formal legal prefixes containing document type and date, such as *Resolución de 9 de enero de 2026*, are then removed. The system subsequently applies pattern-based rewritings to common legal structures, such as *de/de la <entity>, por la que se <action>*.
- **Domain-specific nominalization:** Recurrent legal actions are rewritten into concise nominal expressions commonly used in institutional communication, for example *publica la convocatoria de* → *Convocatoria de*, *nombra* → *Nombramiento de*, and *aprueba* → *Aprobación de*.
- **Deterministic and efficient:** Identical input always produces identical output. The method does not require external models or GPU resources.
- **Conservative generation:** The method is designed to not introduce information that is not present in the original description. It only removes, reorders, or normalizes existing textual segments.

– **Model.** In particular, this approach employs the *Qwen2.5-7B-Instruct* [11] model which is prompted to produce concise and informative titles based on the extracted description and associated metadata. The generation process is configured with the following characteristics:

  - **Prompting:** Targeted prompt engineering that encodes domain constraints (length, forbidden tokens, keyword style mappings, mandatory inclusion of institutional names and uppercase rules).
  - **Temperature:** Low model temperature (0.2) to increase determinism.
  - **Post-processing:** Deterministic cleanup (escaped quotes, trailing punctuation, whitespace and capitalization).
- **Model Few-shot.** On the basis of the model approach, this method adopts a few-shot setting in which the model is provided with examples of input-output pairs. These examples illustrate the desired style of the titles, allowing the model to better adapt to the domain-specific conventions of institutional communication.
  - **Example:**

    Source: *"Resolución de 5 de enero de 2026, de la Universidad de Oviedo, por la que se nombran Catedráticas y Catedráticos de Universidad."*
    Few-shot Output: *"Nombramientos de catedráticas y catedráticos de la Universidad de Oviedo."*
- **Custom.** This alternative relies on communication professionals to manually create titles. This method is typically used when previous automated approaches do not meet specific requirements.

4.3 Deployment and Infrastructure

The developed application is deployed in separate containers that are co-located on a virtual machine within the Google Cloud Platform (GCP).[4] To balance high-performance requirements with cost-efficiency, the infrastructure is managed through scheduled availability, being open for three hours daily. This ensures that the GPU-accelerated instance is operational only during defined time slots (see Table 1).

5 Results

This section presents the results obtained from the dataset collected through the application. We describe the data obtained and compare the performance of the different title generation approaches detailed in Section 4.2. Each entry of the dataset corresponds to a processed bulletin item and is related to the original uploaded document. In particular, the dataset stores the human-selected title for publication and the titles generated by the rule-based and LLM-based methods. Beyond titles, the system also captures descriptions, subtitles, and timestamps for each bulletin entry.

[4] The virtual machine is a *n1-standard-2* instance equipped with 2 vCPUs, 7.5 GB RAM, x86-64 architecture and with an NVIDIA T4 GPU.

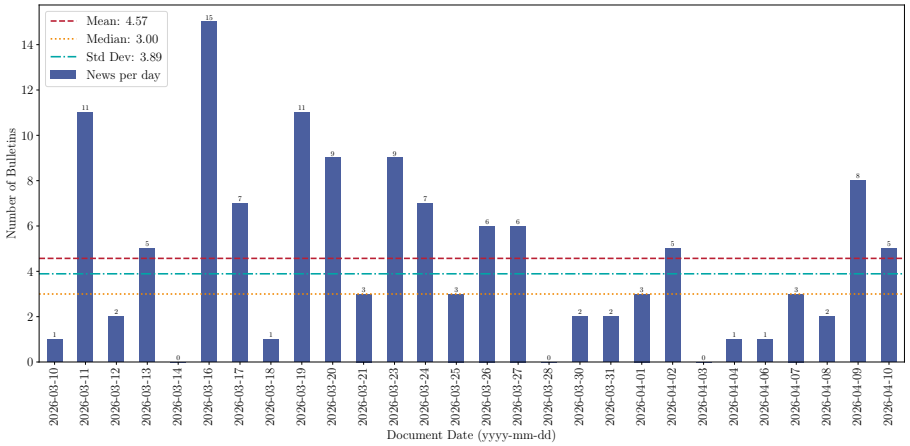


Fig. 2. Number of bulletin items per day *(March 10–April 10, 2026).*

Figure 2 presents the number of bulletin items processed per day during the data collection period (March 10 – April 10), with a total of 128 items collected. On average, approximately 4.6 items are processed daily. The standard deviation (around 4) also reveals variability across days, with occasional peaks reflecting higher workload periods.

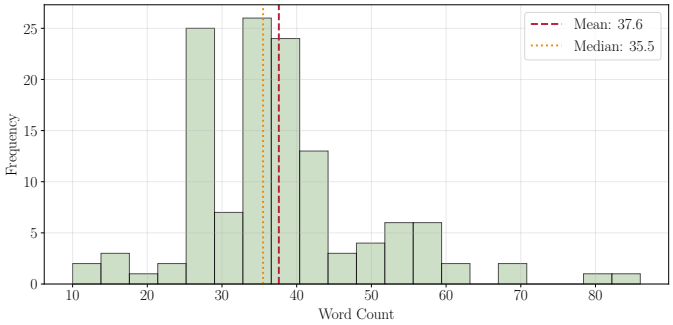


Fig. 3. Word count of descriptions used to generate titles *(March 10–April 10, 2026).*

For each bulletin entry, the candidate titles are derived from the original descriptions. Figure 3 presents the distribution of word counts in these descriptions. On average, descriptions contain 37.6 words. Despite their moderate length, these descriptions often encapsulate formal actions and legal vocabulary, requiring precise identification of the most relevant information for effective title generation.

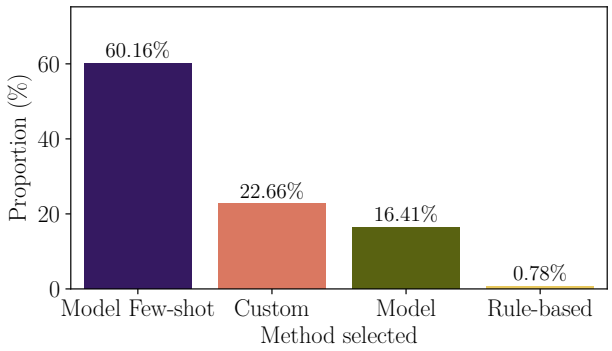


Fig. 4. Distribution of methods selected for title generation *(March 10–April 10, 2026)*.

The frequency of the methods selected for title generation is shown in Figure 4. The results indicate an important preference for LLM-based approaches, particularly in the few-shot setting, which accounts for 60.16% of the selected titles. This suggests that providing examples significantly improves the alignment of generated titles with user expectations. The standard LLM-based approach is selected in 16.41% of cases, showing a lower but still relevant contribution. In contrast to LLM-based approaches, the rule-based method is rarely selected (0.78%), indicating that its transformations are generally insufficient to meet the requirements of communication professionals. Manually created titles (referred to as *Custom*) account for 22.66% of the selections. Although this indicates that human intervention is still required in a subset of cases, it is worth to notice that the majority of selections correspond to automatically generated titles, demonstrating that these approaches are highly effective in most scenarios.

In the context of the environmental cost of LLM-based methods, a single execution of the system in our setup leads to an estimated footprint of 12.90 gCO_2 per day,[5] assuming that active GPU and CPU usage is limited to 25% of the three-hour operational window. This impact is relatively modest at a small scale, and the scheduled hours prevents it to become meaningful since the system is not deployed in high-demand or large-scale environments. In terms of economic cost, the deployment incurred a total expense of EUR 53.29 over the data collection period (March 10 – April 10),[6] corresponding to an average cost of approximately EUR 2.32 per business day. This cost is significantly lower than the operational costs associated with manual title generation due to the time invested by the workers of the communication office prior to the existence of the application.

[5] Calculated using green-algorithms.org v3.0 [5].

[6] Estimated from Google Cloud billing for a single virtual machine instance used to deploy the application during the data collection period.

6 Conclusions and Future Work

This work has explored the integration of automated title generation methods within an institutional communication workflow. By comparing rule-based and LLM-based approaches in a real-world setting, the results show a clear preference for LLM-generated titles, particularly when guided through few-shot prompting. Despite rule-based approaches' lower effectiveness, such methods remain lightweight and may still be relevant in resource-constrained environments. This study shows that LLM-assisted solutions, representing approximately 77% of selected titles, can effectively support communication professionals.

From a practical perspective, the deployment of LLM-based systems introduces additional computational and economic costs associated with maintaining GPU-enabled infrastructure. However, these costs are offset by the expense of human labor, as automation substantially reduces the manual effort required for title generation. This suggests a favorable trade-off in terms of overall operational efficiency. Moreover, automating this task allows communication professionals to focus on higher-value activities, notably reducing the burden of repetitive and time-consuming work. Regarding the environmental footprint of LLM-based methods, limiting their usage to specific time windows helps keep this impact within manageable levels.

This work constitutes a preliminary study in which the requirements of the problem are analyzed, a supporting tool is designed, and the system is deployed to the communication team to enable data collection in a real-world setting. The data collected will be used to analyze user preferences and editing patterns, with the aim of refining the behavior of the language model.

In particular, future work will explore the possibility of fine-tuning smaller and more efficient models that can achieve the desired performance while reducing computational cost. This direction is motivated by the proportion of manually created titles, which indicates that current approaches do not fully capture certain domain-specific nuances. Additionally, further improvements are planned in the information extraction stage from official gazettes. This includes incorporating a new user role within the application to allow the legal department, responsible for selecting relevant items, to perform a preliminary filtering step. This integration is expected to streamline the information flow within the system and mitigate issues related to external data formats.

Acknowledgments. The authors gratefully acknowledge the financial support of the Asturian Agency for Science, Business Competitiveness and Innovation (SEKUENS) under Grant Agreement No. SEK-25-GRU-GIC-24-018. Noelia Rico and Irene Díaz also acknowledge support from the European project SCIMIN-CRM (Grant Agreement No. 101177746) and the funding from the Spanish Ministry of Science and Innovation through project PID2022-139886NB-I00.

Moreover, the authors would like to express their sincere gratitude to the communication team of the University of Oviedo for their continuous collaboration, careful evaluation, and consistently constructive feedback. Their active involvement and commitment have been essential to making this work possible.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Callegari, E., Vajdecka, P., Xhura, D., Ingason, A.K.: Enhancing Academic Title Generation Using SciBERT and Linguistic Rules. In: Proceedings of the Second Workshop on Information Extraction from Scientific Publications. pp. 74–83. Association for Computational Linguistics, Bali, Indonesia (2023), https://aclanthology.org/2023.wiesp-1.10
2. Dorr, B., Zajic, D., Schwartz, R.: Hedge Trimmer: a parse-and-trim approach to headline generation. In: Proceedings of the HLT-NAACL 03 on Text summarization workshop. vol. 5, pp. 1–8. Association for Computational Linguistics (2003), http://portal.acm.org/citation.cfm?doid=1119467.1119468
3. Edmundson, H.P.: New Methods in Automatic Extracting. Journal of the ACM **16**(2), 264–285 (Apr 1969). https://doi.org/10.1145/321510.321519
4. Galindo, J., Pérez, M.H., Victoria, C., Santos, J.M., Fernández, A., Raigada, T.: El impacto de la inteligencia artificial en el sector público español: Avances para el acceso a servicios básicos. Tech. rep., EsadeEcPol - Center for Economic Policy (May 2025), https://www.esade.edu/ecpol/wp-content/uploads/2025/05/AAFF_EsadeEcPolxGoogle_AIbenefits__ES-final.pdf
5. Lannelongue, L., Grealey, J., Inouye, M.: Green Algorithms: Quantifying the Carbon Footprint of Computation. Advanced Science **8**(12), 2100707 (Jun 2021). https://doi.org/10.1002/advs.202100707
6. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension (2019). https://doi.org/10.48550/ARXIV.1910.13461, version Number: 1
7. Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development **2**(2), 159–165 (Apr 1958). https://doi.org/10.1147/rd.22.0159
8. Mihalcea, R., Tarau, P.: TextRank: Bringing Order into Text. In: Lin, D., Wu, D. (eds.) Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. pp. 404–411. Association for Computational Linguistics, Barcelona, Spain (Jul 2004), https://aclanthology.org/W04-3252/
9. Mishra, P., Diwan, C., Srinivasa, S., Srinivasaraghavan, G.: Automatic Title Generation for Text with Pre-trained Transformer Language Model. In: 2021 IEEE 15th International Conference on Semantic Computing (ICSC). pp. 17–24. IEEE, Laguna Hills, CA, USA (Jan 2021). https://doi.org/10.1109/ICSC50631.2021.00009
10. OpenAI: GPT-4 Technical Report (2023). https://doi.org/10.48550/ARXIV.2303.08774
11. Qwen: Qwen2.5 Technical Report (2024). https://doi.org/10.48550/ARXIV.2412.15115
12. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (2019). https://doi.org/10.48550/ARXIV.1910.10683, version Number: 4

13. Rehman, T., Sanyal, D.K., Chattopadhyay, S.: Can pre-trained language models generate titles for research papers? (2024). https://doi.org/10.48550/ARXIV.2409.14602, version Number: 2
14. Rush, A.M., Chopra, S., Weston, J.: A neural attention model for abstractive sentence summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing. pp. 379–389 (2015), https://arxiv.org/abs/1509.00685
15. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Advances in neural information processing systems **27** (2014), https://doi.org/10.48550/arXiv.1409.3215
16. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G.: LLaMA: Open and Efficient Foundation Language Models (2023). https://doi.org/10.48550/ARXIV.2302.13971, version Number: 1
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need (2017). https://doi.org/10.48550/ARXIV.1706.03762, version Number: 7
18. Waheed, A., Goyal, M., Mittal, N., Gupta, D.: Domain-Controlled Title Generation with Human Evaluation. In: Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds.) International Conference on Innovative Computing and Communications, vol. 1388, pp. 461–474. Springer Singapore, Singapore (2022). https://doi.org/10.1007/978-981-16-2597-8_39, series Title: Advances in Intelligent Systems and Computing
19. Zhang, J., Zhao, Y., Saleh, M., Liu, P.J.: PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization (2019). https://doi.org/10.48550/ARXIV.1912.08777, version Number: 3