| | |
|---|---|
| **Titre:** Title: | NLP in SMEs for industry 4.0: opportunities and challenges |
| **Auteurs:** Authors: | Mathieu Bourdin, Thomas Paviot, Robert Pellerin, & Samir Lamouri |
| **Date:** | 2023 |
| **Type:** | Communication de conférence / Conference or Workshop Item |
| **Référence:** Citation: | Bourdin, M., Paviot, T., Pellerin, R., & Lamouri, S. (2023, November). NLP in SMEs for industry 4.0: opportunities and challenges [Paper]. 2023 International Conference on Enterprise Information Systems (CENTERIS 2023), Porto, Portugal. Published in Procedia Computer Science, 239. https://doi.org/10.1016/j.procs.2024.06.186 |



**Document publié chez l'éditeur officiel**
Document issued by the official publisher

| | |
|---|---|
| **Nom de la conférence:** Conference Name: | 2023 International Conference on Enterprise Information Systems (CENTERIS 2023) |
| **Date et lieu:** Date and Location: | 2023-11-10 - 2023-11-08, Porto, Portugal |
| **Maison d'édition:** Publisher: | Elsevier |
| **URL officiel:** Official URL: | https://doi.org/10.1016/j.procs.2024.06.186 |
| **Mention légale:** Legal notice: | © 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0) |

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 396–403

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

# NLP in SMEs for industry 4.0: opportunities and challenges

Mathieu Bourdin[a,b,*], Thomas Paviot[c], Robert Pellerin[d,e,f], Samir Lamouri[b,g]

[a]*Eurofins Analyses pour l'Environnement France, 5 rue d'Otterswiller, 67700 Saverne, France*
[b]*Arts et Métiers ParisTech, 121 boulevard de l'Hôpital, 75013 Paris, France*
[c]*ISAT/Université de Bourgogne, 49 rue Mademoiselle Bourgeois, BP 31, 58027 Nevers, France*
[d]*CIRRELT, 2920 chemin de la Tour, pavillon André Aisenstadt, bureau 3520, 11290 Montréal, Canada*
[e]*IVADO, 6666 rue Saint Urbain, 11290 Montréal, Canada*
[f]*Polytechnique Montréal, 2500 chemin de Polytechnique Montréal, 11290 Montréal, Canada*
[g]*LAMIH CNRS/Université Polytechnique Hauts-de-France, Campus Mont Houy, 59313 Valenciennes Cedex 9, France*

**Abstract**

Natural Language Processing is the field of Computer Science that focuses on analyzing and processing natural language, mainly human text or speech. Recent trends in Natural Language Processing have led to the development of Large Language Models (LLMs): huge models trained on high amounts of data that achieve unprecedented performances in many tasks, such as answering questions, summarizing texts, or coding. These new tools have a wide range of applications and are being developed by many companies. However, Small and Medium Enterprises (SMEs) struggle to implement these new technologies, mainly because of the lack of resources. This paper aims to show the opportunities and challenges related to NLP-based solutions in SMEs based on a literature review. The main result is that NLP-based solutions have a wide range of applications in various companies, including SMEs, and may lead to many changes. However, there are still many obstacles to developing these tools in SMEs: SMEs lack specialized know-how to develop these solutions and do not often have standardized data. Moreover, there exists nearly no support for SMEs in the scientific literature to develop these tools.



*Keywords:* Natural Language Processing; Industry 4.0; Machine Learning; Large Language Models; SMEs.

* Corresponding author.
*E-mail address:* mathieubourdin@hotmail.fr


10.1016/j.procs.2024.06.186

## 1. Introduction

Language Models (LM) are statistical models that can understand and generate human language based on statistical patterns. They can for example create text and analyze the sentiments expressed in texts [21]. These models are a fundamental component of Natural Language Processing (NLP), the branch of artificial intelligence that focuses on the interaction between computers and human language.

Language Models have recently attracted much attention in the media with the release of ChatGPT in November 2022 [40]. Since then, many other NLP models have emerged: Google's Bard, Meta's LLaMA, and OpenAI's GPT-4 are a few examples. These models and the related algorithms perform great on a wide variety of tasks: they can give human-like answers to various questions, provide creative ideas, or execute concrete tasks such as coding [24] or summarising documents [46].

The models' performance strongly depends on their size: the bigger the model, the fewer mistakes the model does. Language models have therefore increased their sizes from billions of parameters a few years ago to hundreds of trillions of parameters today [47]. With this increase in the models' sizes, these models, now called Large Language Models (LLM), have reached an unprecedented level of performance in NLP.

However, this improvement has a cost: a huge model implies a lot of computational resources to train the model and a high development cost. To try and reduce the costs, companies have developed Foundation Models: models that are trained on a vast amount of data and can be later adapted to different tasks. But even Foundation Models remain too expensive for many companies.

With the recent works made to reduce the sizes of LLMs [50] and the development of open-source language models [48], these models will likely become financially affordable to all companies in the near future, and even for Small and Medium Enterprises (SMEs). Other obstacles to developing these innovations in SMEs include small companies' lack of know-how and unstandardized data [49].

This paper aims to give an overview of the opportunities and challenges related to adopting NLP-based algorithms in SMEs based on a literature review. To this end, two questions will be answered:

1) In which domains will NLP have the most significant potential impact in the future? And
2) What obstacles deter the adoption of these tools in SMEs?

The structure of this paper is the following: section 2 draws up an overview of the different NLP use cases in companies of all sizes based on recent scientific literature. Section 3 then deals with why SMEs still struggle to adopt these new tools and the potential solutions to the encountered obstacles. Detailed research methodologies are presented at the beginning of each section.

## 2. NLP Use Cases for Industry 4.0

### *2.1. Methodology*

The following list is a non-exhaustive list based on research made with the keywords "NLP", "LLM", and "Foundation Models" in Science Direct. After filtering review articles from 2019 and later in Science Direct, 47 results remained, then reduced to 33 after analyzing titles and abstracts. The same research was made on Google Scholar with results from 2019 and later to complete this listdone. For each keyword, the first two pages were analyzed and deepened if needed. The total 93 articles found have then been sorted out and grouped by similar activity areas, which are presented in section 2.2. Systematic reviews have been removed since they are not use cases. Duplicate papers have also been removed.

This section aims to give an overview of the different applications of NLP algorithms in companies of all sizes. This list is not exhaustive but aims to show the wide range of applications of NLP solutions.

### *2.2. NLP use cases*

#### *2.2.1. Healthcare*

Healthcare is one of the most documented domains for using NLP: 21 of the 33 use cases found in ScienceDirect were related to healthcare. NLP algorithms can indeed help in both physical and mental Healthcare. They can:

- Assist diagnosis-making [1];

- Detect pathologies that were unidentified in patients' medical reports [2];
- Inform patients on potential illnesses or treatment [3];
- Assist with image analysis (for example, radiology) [4];
- Detect mental illnesses [5] or loneliness [6] through texts or conversations; and
- Provide 24/7 medical support for individuals [7].

*2.2.2. Training*

Language Models can also be useful in the domain of training, where retrieving information and understanding complex concepts are keys. They can:

- Answer students' questions thanks to chatbots [8];
- Provide personalized teaching to each student [9];
- Explain by generating analogies [10];
- Generate questions for students' assessment [11];
- Evaluate students' work, including essays [12]; and
- Help with grammar correction, including foreign languages [13].

*2.2.3. Legal*

The use of NLP algorithms in the legal domain is a vast topic. The adoption in this field is high due to the significant number of documents and consequent time required when looking for information. LM is used in legal to:

- Look for previous similar cases [14];
- Predict judgment results for a case [15];
- Answer legal questions [16];
- Summarize information on legal documents [17];
- Detect unfair clauses [18]; and
- Detect crimes on social media [19].

*2.2.4. Finance*

Finance is another economy's key domain. Thanks to the high amount of available data in this field (companies' financial reports or public posts on social media), NLP algorithms can be used in finance to:

- Detect fraud in companies' annual reports [20];
- Predict stock return based on social media posts [21];
- Extract risk sentences from companies' reports [22]; and
- Assess companies' ESG based on media coverage [23].

*2.2.5. Robotics and computer science*

In Computer Science, LM can be used to generate code from human instruction [24], to complete the missing part of a code [25], to test software [26], or even to generate comments [27]. It is also possible to control robots with human language or pictures [28]. Robots can "see" thanks to NLP [29] and detect emotions [30]. Eventually, NLP is even used to design other robots [31].

*2.2.6. Business Process Management (BPM)*

In Business Process Management, LM can be used to retrieve information and classify documents [32], to detect whether a maintenance issue will block production [33], to recommend activities [34], to schedule tasks [35], or to help developing new products [36].

*2.2.7. Energy*

Eventually, NLP can be used in the domain of energy production, for example, to answer specific questions on the nuclear domain [37] or to classify research articles related to this domain [38]. There are also examples of optimizing energy storage systems associated with renewable energies [39].

*2.2.8. Other use cases and general discussion*

NLP use cases are vast. To give a few other examples of situations where NLP can be used, one can mention: entertainment (creation of fictional stories [40]), journalism (detection of fake news [41]), translation [42], marketing (analyzing trends [43]), construction (accident analyzing in a construction site [44]) or transports (acquire data from different sources for traffic management [45]). Table 1 summarises the different application areas of NLP use cases.

NLP-based solutions have many advantages compared to the human workforce, depending on the use case: they are faster, can deal with higher amounts of data, and are available 24/7. However, they also have drawbacks: mainly the lack of

in the outputs but also the need to acquire many data to train the models. The advantages and drawbacks in each use case are summarised in Table 1.

Table 1. Summary of NLP use cases (non-exhaustive)

| Area | Useful NLP function | Comparison with human workforce | | Potential impact |
|---|---|---|---|---|
| | | Advantage | Drawback | |
| Healthcare | Provide information<br>Analyse text and image<br>Mimic human behaviour | Available 24/7<br>Remote access<br>Cheaper | Higher risk of error | Provide universal access to medical support |
| Training | Provide information<br>Analyse text (essays)<br>Create content (questions, tests) | Available 24/7<br>Remote access | Higher risk of error | Provide universal access to education |
| Legal | Provide information | Faster | Higher risk of error | Provide universal access to legal information |
| Finance | Analyse text (financial reports, media)<br>Analyse sentiment | Faster<br>Can handle high amounts of data | - | - |
| Robotics and computer science | Generate code<br>Analysing data from multiple sources<br>Analyze image and sentiment | Faster | Higher risk of error | Lead to the creation of autonomous robots |
| Business Process Management | Analysing data from multiple sources | Can handle high amounts of data | Require many data | Assist decisions makers in their everyday life |
| Energy | Provide information<br>Mathematical reasoning | Available 24/7<br>Faster | Higher risk of error | Reduce energy consumption worldwide |

## 3. Obstacles to future development in SMEs

Even though NLP applications are numerous, SMEs still struggle to implement them. This section aims to understand why and to give some solutions to the encountered issue.

The methodology applied in this section is similar to the one in the previous section. A first literature review has been made in ScienceDirect with the same keywords as precedent ("NLP", "LLM", and "Foundation Models"), but the keyword “SME” was added (and its variant orthographs). The research only gave 4 results, none of which were relevant to this study. This first observation is a finding in itself: very few documents about using NLP in SMEs.

The same keywords have been searched in Google Scholar to complete this research. The first 5 pages of results have been analyzed and filtered for each keyword based on titles and abstracts. This allowed us to find nine relevant articles dealing, which are summarised in this section. This shows that despite the many applications presented in Section 2, there is still very little research on implementing NLP solutions in SMEs. Results are grouped by encountered issues.

*3.1. Financial obstacles*

As mentioned in the introduction, state-of-the-art performances in NLP are obtained with Large Language Models (LLMs). These models require a lot of computational resources for both the training and inference phases.

Training is the first phase of developing an LLM: the model uses a high amount of data to adjust its parameters and learn from the text, which includes deep relations between words. This training is costly. For example, the BLOOM model, a 176B-parameter model, required 3,5 months of training on 384 NVIDIA A100 80GB GPUs [51]. Each NVIDIA A100 costs around $10k, which means a hardware cost of at least $3M for the training of BLOOM, which is out of the range of most SMEs.

Two solutions can be envisaged to face this:

1) The use of cloud computing; and
2) The use of open-source models.

Cloud computing is a solution to reduce the development cost of LLM in SMEs. Instead of acquiring the hardware required to train a model, an SME can rent it and use it only for the time required to train itneeded [52]. This solution is not ideal because even if it is less expensive than acquiring the hardware, it still represents a significant cost for the company. Moreover, there is a security risk with the data being transferred to an external company.

An alternative to this solution is the use of open-source models. Some effective open-source models have recently been released, such as LLaMA, Alpaca, GPT4All, or BLOOM. Even if some of these are restricted to research purposes, others, such as BLOOM [51], are open to all, even SMEs. These recent open-source models allow to develop their own NLP-based algorithm on standard hardware such as a single computer.

### *3.2. Human-related obstacles*

In [49], Bauer indicates that SMEs usually do not have dedicated employees for data science because of their size. This lack of knowledge is the main challenge to adopting Machine Learning (ML) in SMEs. Even though NLP and Machine Learning are different, one can assume that the obstacles that deter SMEs from adopting ML are the same as those that prevent them from adopting NLP. The lack of know-how in SMEs would therefore be an obstacle to developing NLP in SMEs.

Still in [49], Bauer gives three solutions to this issue: 1°) to exchange with other companies that are already more mature on these topics 2°) to resort to external companies (consultants or service providers), or 3°) to cooperate with universities or research faculties.

### *3.3. Data-related obstacles*

#### *3.3.1. Model's output*

Even though NLP has significantly improved over the past years, it still suffers from the issue of inconsistent outputs. This is a significant obstacle to the use of NLP in decision-making. Even the best state-of-the-art algorithms sometimes provide wrong or inexact information. This issue, called hallucination [53], makes NLP hard to use in many tasks, such as predicting the result of a legal judgment. Hopefully, exact outputs are not required in all tasks, for example, when creating a fictional story or speech [40], or in tasks where mistakes have few consequences, such as assisting humans in retrieving information [14]. However, this lack of consistency in NLP methods is one of the highest obstacles to their use in SMEs.

One solution to this issue is to systematically cite the references used to give an answer. This solution was developed, for example, by OpenAI in their WebGPT [54]. When answering a question, WebGPT not only answers but also always cites the different sources used to provide the answer. This allows the user to check the answer before using it and therefore lightens the drawback of giving inaccurate answers.

#### *3.3.2. Outdated data*

Another drawback of NLP-based solutions is their time-limited knowledge. Indeed, language models are initially trained on a given data set but then stop learning as long as they are used; the model does not acquire any new information. This is a major obstacle for many SMEs that need updated information.

One of the solutions that can be envisaged is the previously mentioned WebGPT [54]. This model cites its sources and retrieves information directly on the Web. This shows the feasibility of connecting a Language Model to the Web. No solution has been found in the literature for other kinds of updates (for example, information unavailable on the

Internet).

*3.3.3. Non-standardised data*

One big difference between SMEs and large companies is their generally lower maturity in data management [49]. SMEs have usually fewer data or unstandardized data. This is an issue when building a language model, requiring much standardized data. A few solutions exist.

*3.3.3.1. Adapting the model to the unstandardized data*

A first approach to this issue would be to adapt the model to several data sources. Some works have already been made to this end, such as ToolFormer [57], a model that can retrieve information from different data sources by itself (for example, a question-and-answer database, a search engine, or a calendar). This work suggests that future NLP models could be taught to call by themselves different SME databases, such as their Enterprise Resource Planning (ERP), potential SQL databases, user's mailboxes, or any other company's tool. However, these solutions are still at an early stage of development and have not been proven effective yet on so many different applications.

*3.3.3.2. Adapting the data to the model*

Since adopting the model to various data sources seems complicated, an easier approach to the data issue could be to standardize the data.

Unfortunately, no support has been found in the literature to help in this direction, that seems easier than training a model to access different databases. The data gathering and preparing processes are very little documented. The authors of some papers explain how they did themselves to prepare their models [1][55][56]. Still, there is no general methodology in the literature to help with the topic of data preparation: how to collect data? How much data are required? In which format? How to update them? What pre-treatment has to be applied? Etc.

This lack of methodological framework in the literature is a major obstacle preventing SMEs from developing their own NLP-based solutions. This highlights the need for a methodological framework to help SMEs with data collection and preparation.

## 4. Conclusion

This paper gave an overview of the different fields where NLP is currently being developed and could become a major asset in the future for SMEs: healthcare, training, legal, finance, robotics, and more. Even if the potential of NLP is high, as shown by the numerous different use cases, SMEs still struggle to adopt these technologies. The main obstacles to this are SMEs' lack of know-how and support to help them implement these technologies.

The use of free, open-source language models and the collaboration with other entities such as universities can help SMEs in the development of these new technologies, which are faster than humans and can handle higher amounts of data. However, these tools still lack of exactness and are therefore not commonly used in decision-making. To face this issue, it is important for SMEs to keep updated and standardized data to train or update their models. Moreover, the model itself should be small enough to run on low-cost hardware and should be able to cite its data sources to mitigate the lack of precision.

As a future work, a valuable contribution to support SMEs would be creating a theoretical framework to help developing language tools, particularly to indicate how how to identify, collect, prepare, and maintain the data the data. This would be a major addition to the scientific knowledge and greatly assist SMEs.

## References

[1] Yunxiang, L., Zihan, L., Kai, Z., Ruilong, D., & You, Z. (2023). Chatdoctor: A medical chat model fine-tuned on llama model using medical domain knowledge. arXiv preprint arXiv:2303.14070.

[2] Siddiqui, Zohair, et al. "Rules-based natural language processing to extract features of large vessel occlusion and cerebral edema from radiology reports in stroke patients." Neuroscience Informatics (2023): 100129.

[3] Ayanouz, S., Abdelhakim, B. A., & Benhmed, M. (2020, March). A smart chatbot architecture based NLP and machine learning for health care assistance. In Proceedings of the 3rd international conference on networking, information systems & security (pp. 1-6).

[4] Pandey, Babita, et al. "A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions." Journal of King Saud University-Computer and Information Sciences 34.8 (2022): 5083-5099.

[5] Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. NPJ digital medicine, 5(1), 46.

[6] Badal, V. D., Lee, E. E., Jeste, D. V., & Kim, H. C. Leveraging AI for Psychiatric Research: Natural Language Processing to Assess Loneliness in Older Adults.

[7] Fadhil, A. (2018). Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals. arXiv preprint arXiv:1803.06000.

[8] Sreelakshmi, A. S., Abhinaya, S. B., Nair, A., & Nirmala, S. J. (2019, November). A question answering and quiz generation chatbot for education. In 2019 Grace Hopper Celebration India (GHCI) (pp. 1-6). IEEE.

[9] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. Ieee Access, 8, 75264-75278.

[10] Bhavya, B., Xiong, J., & Zhai, C. (2022). Analogy Generation by Prompting Large Language Models: A Case Study of InstructGPT. arXiv preprint arXiv:2210.04186.

[11] Lu, O. H., Huang, A. Y., Tsai, D. C., & Yang, S. J. (2021). Expert-authored and machine-generated short-answer questions for assessing students learning performance. Educational Technology & Society, 24(3), 159-173.

[12] Ormerod, C. M., Malhotra, A., & Jafari, A. (2021). Automated essay scoring using efficient transformer-based language models. arXiv preprint arXiv:2102.13136.

[13] Solyman, Aiman, et al. "Synthetic data with neural machine translation for automatic correction in Arabic grammar." Egyptian Informatics Journal 22.3 (2021): 303-315.

[14] Tran, V., Nguyen, M. L., & Satoh, K. (2019, June). Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law (pp. 275-282).

[15] Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal Judgment Prediction via Topological Learning. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3540–3549, Brussels, Belgium. Association for Computational Linguistics

[16] Duan, X., Wang, B., Wang, Z., Ma, W., Cui, Y., Wu, D., ... & Liu, Z. (2019). Cjrc: A reliable human-annotated benchmark dataset for chinese judicial reading comprehension. In Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18 (pp. 439-451). Springer International Publishing.

[17] Kanapala, A., Pal, S., & Pamula, R. (2019). Text summarization from legal documents: a survey. Artificial Intelligence Review, 51, 371-402.

[18] Lippi, M., Pałka, P., Contissa, G., Lagioia, F., Micklitz, H. W., Sartor, G., & Torroni, P. (2019). CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. Artificial Intelligence and Law, 27, 117-139.

[19] Drury, Brett, et al. "A social network of crime: A review of the use of social networks for crime and the detection of crime." Online Social Networks and Media 30 (2022): 100211.

[20] Chen, Y. J., Wu, C. H., Chen, Y. M., Li, H. Y., & Chen, H. K. (2017). Enhancement of fraud detection for narratives in annual reports. International Journal of Accounting Information Systems, 26, 32-45.

[21] Hiew, J. Z. G., Huang, X., Mou, H., Li, D., Wu, Q., & Xu, Y. (2019). BERT-based financial sentiment index and LSTM-based stock return predictability. arXiv preprint arXiv:1906.09024.

[22] Yu-Wen Liu, Liang-Chih Liu, Chuan-Ju Wang, and Ming-Feng Tsai. 2018. RiskFinder: A Sentence-level Risk Detector for Financial Reports. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pages 81–85, New Orleans, Louisiana. Association for Computational Linguistics.

[23] Fischbach, J., Adam, M., Dzhagatspanyan, V., Mendez, D., Frattini, J., Kosenkov, O., & Elahidoost, P. (2022). Automatic ESG Assessment of Companies by Mining and Evaluating Media Coverage Data: NLP Approach and Tool. arXiv preprint arXiv:2212.06540.

[24] Karampatsis, R. M., Babii, H., Robbes, R., Sutton, C., & Janes, A. (2020, June). Big code!= big vocabulary: Open-vocabulary models for source code. In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering (pp. 1073-1085).

[25] Alon, U., Sadaka, R., Levy, O., & Yahav, E. (2020, November). Structural language models of code. In International conference on machine learning (pp. 245-256). PMLR.

[26] Garousi, Vahid, Sara Bauer, and Michael Felderer. "NLP-assisted software testing: A systematic mapping of the literature." Information and Software Technology 126 (2020): 106321.

[27] Movshovitz-Attias, D., & Cohen, W. (2013, August). Natural language models for predicting programming comments. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 35-40).

[28] Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Ichter, B., ... & Florence, P. (2023). Palm-e: An embodied multimodal language model. arXiv preprint arXiv:2303.03378.

[29] El-Komy, A., Shahin, O. R., Abd El-Aziz, R. M., & Taloba, A. I. (2022). Integration of computer vision and natural language processing in multimedia robotics application. Inf. Sci, 7(6).

[30] Graterol, W., Diaz-Amado, J., Cardinale, Y., Dongo, I., Lopes-Silva, E., & Santos-Libarino, C. (2021). Emotion detection for social robots based on NLP transformers and an emotion ontology. Sensors, 21(4), 1322.

[31] Stella, F., Della Santina, C., & Hughes, J. (2023). Can Large Language Models design a Robot?. arXiv preprint arXiv:2303.15324.

[32] Wu, Chengke, et al. "Natural language processing for smart construction: Current status and future directions." Automation in Construction 134 (2022): 104059.

[33] Usuga-Cadavid, J. P., Grabot, B., Lamouri, S., & Fortin, A. (2021). Artificial data generation with language models for imbalanced classification in maintenance. In Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA LATIN AMERICA 2021 (pp. 57-68). Springer International Publishing.

[34] Sola, D., van der Aa, H., Meilicke, C., & Stuckenschmidt, H. (2023). Activity Recommendation for Business Process Modeling with Pre-trained Language Models. ESWC. Springer.

[35] Prieto, S. A., Mengiste, E. T., & García de Soto, B. (2023). Investigating the use of ChatGPT for the scheduling of construction projects. Buildings, 13(4), 857.

[36] Burggräf, Peter, Johannes Wagner, and Tim Weißer. "Knowledge-based problem solving in physical product development—A methodological review." Expert Systems with Applications: X 5 (2020): 100025.

[37] Jain, A., Meenachi, D. N., & Venkatraman, D. B. (2020). NukeBERT: A pre-trained language model for low resource nuclear domain. arXiv preprint arXiv:2003.13821.

[38] Burke, L., Pazdernik, K., Fortin, D., Wilson, B., Goychayev, R., & Mattingly, J. (2021). NukeLM: Pre-Trained and Fine-Tuned Language Models for the Nuclear and Energy Domains. arXiv preprint arXiv:2105.12192.

[39] Das, B., & Kumar, A. (2018). A NLP approach to optimally size an energy storage system for proper utilization of renewable energy sources. Procedia Computer Science, 125, 483-491.

[40] Dale, R. (2021). GPT-3: What's it good for?. Natural Language Engineering, 27(1), 113-118.

[41] Ahsan, Mohammad, Madhu Kumari, and T. P. Sharma. "Rumors detection, verification and controlling mechanisms in online social networks: A survey." Online Social Networks and Media 14 (2019): 100050.

[42] Yulianto, A., & Supriatnaningsih, R. (2021). Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). AJELP: Asian Journal of English Language and Pedagogy, 9(2), 109-127.

[43] Dash, G., Sharma, C., & Sharma, S. (2023). Sustainable Marketing and the Role of Social Media: An Experimental Study Using Natural Language Processing (NLP). Sustainability, 15(6), 5443.

[44] Zhang, Fan, et al. "Construction site accident analysis using text mining and natural language processing techniques." Automation in Construction 99 (2019): 238-248.

[45] Zheng, O., Abdel-Aty, M., Wang, D., Wang, Z., & Ding, S. (2023). ChatGPT is on the horizon: Could a large language model be all we need for Intelligent Transportation?. arXiv preprint arXiv:2303.05382.

[46] Goyal, T., Li, J. J., & Durrett, G. (2022). News summarization and evaluation in the era of gpt-3. arXiv preprint arXiv:2209.12356.

[47] Koubaa, A. (2023). GPT-4 vs. GPT-3.5: A concise showdown.

[48] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

[49] Bauer, M., van Dinther, C., & Kiefer, D. (2020). Machine learning in SME: an empirical study on enablers and success factors.

[50] Kurtic, E., Frantar, E., & Alistarh, D. (2023). ZipLM: Hardware-Aware Structured Pruning of Language Models. arXiv preprint arXiv:2302.04089.

[51] Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... & Manica, M. (2022). Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100.

[52] Kaymakci, C., Wenninger, S., Pelger, P., & Sauer, A. (2022). A Systematic Selection Process of Machine Learning Cloud Services for Manufacturing SMEs. Computers, 11(1), 14.

[53] Beutel, G., Geerits, E., & Kielstein, J. T. (2023). Artificial hallucination: GPT on LSD?. Critical Care, 27(1), 1-3.

[54] Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., ... & Schulman, J. (2021). Webgpt: Browser-assisted question-answering with human feedback. arXiv preprint arXiv:2112.09332.

[55] Maguire Herriman, , Elana Meer, , Roy Rosin, MBA, Vivian Lee, MD, PhD, MBA, Vindell. Washington, MD, Kevin G. Volpp, MD, PhD. (2020) Asked and Answered: Building a Chatbot to Address Covid-19-Related Concerns. DOI: 10.1056/CAT.20.0230

[56] Kapočiūtė-Dzikienė, J. (2020). A Domain-Specific Generative Chatbot Trained from Little Data. Applied Sciences, 10(7), 2221.

[57] Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., ... & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. arXiv preprint arXiv:2302.04761.