

- Hill Museum and Manuscript Library (n. d.) *Reading Room*. Retrieved at <https://www.vhmml.org/readingRoom/view/139831> (accessed on 5. 3. 2024).
- Identità. (2024). *Public Registry*. Retrieved at <https://identita.gov.mt/public-registry-main-page/> (accessed on 3. 3. 2024).
- Malta Diocesan Archives. (2018). *Malta Parish Archives*. Retrieved at <https://www.maltaparisharchives.org/> (accessed on 5. 3. 2024).
- Romanova, E. (2019). Archival Science: Bridges between tradition and innovation. *Atlanti*, 29(1), 17–27.
- Serracino, C. (2022). *Ardet Amans. Essays in honour of Horatio Caesar Roger Vella*. Malta: Midsea Publishers.
- Tatò, G. (2019). Archives and the Society. In *Atlanti*, 29(2), 95–101.
- The Malta Independent. (5. 8. 2015). *National Archives of Malta announce completion of the digitisation of the magna curia castellanicae*. Retrieved at <https://www.independent.com.mt/articles/2015-08-05/local-news/National-Archives-of-Malta-announce-completion-of-the-digitisation-of-the-magna-curia-castellanicae-6736140033> (accessed on 3. 3. 2024).

Miroslav Milovanović<sup>1</sup>

## ZERO SHOT CLASSIFICATION FOR UNSTRUCTURED TEXT OF ARCHIVAL VALUE

### Abstract

**Purpose:** *The purpose of the article is to investigate if artificial intelligence and, subsequently, machine learning can provide any solutions to ease some of the archival tasks when dealing with classification of unstructured texts which have archival value. The research was aimed specifically on how to approach a specific archival task within content classification of unstructured texts.*

**Method/approach:** *In the research, the methods of content analysis and experiment were used. Different approaches to managing the classification of unstructured text with the use of machine learning were investigated, as well as the conduction of experiment testing of some of the most prominent technological solutions currently available.*

**Results:** *The research showed that the use of machine learning for the purpose of classification in managing unstructured text with archival value is achievable and effective.*

**Conclusion:** *The approach, with its method and technology, which was used in the research is mature, manageable, and available to carry out the archival task of classification of unstructured text where needed. Zero shot classification provides a suitable path to solve problems relating to the classification of unstructured texts of archival value where pre-labelled data for following the supervised approach to create the model for classification is not available.*

**Key words:** *Machine learning, unstructured text, classification, zero shot classification, description.*

---

<sup>1</sup> Miroslav Milovanović, PhD student of Archival Sciences at Alma Mater Europaea University, Slovenia, e-mail: miroslav.milovanovic1@almamater.si.

## INTRODUCTION

One of the biggest problems in dealing with unstructured texts which could have archival value is how to handle the ever-growing amount of individual unstructured records using archival practice. While certain technology and methods exist that could be used for executing some of the individual tasks when dealing with unstructured texts, it is still hard to find or develop an organized approach in the form of a model or a guideline as how to proceed to achieve completeness when doing certain archival tasks, such as arranging unstructured texts or providing archival description to individual unstructured texts. Novak (2019) exposes that using modern information technology solutions in archival professional work, requires an approach which uses many ad hoc skills that are currently hard to acquire through formally established educations and trainings.

Unstructured text has no predefined structure, such as format or data model. Unstructured text thus represents any form of record that can include any information. Because this type of data is not organized in a predetermined way, it is more difficult to process and analyse when using traditional methods (OpenText Corporation, 2024).

In recent years many organisations started with their digital transformation processes resulting in the creation of a large number of digital records. The exponential growth of the number of digital records thus exposes numerous problems when dealing with those records such as the aforementioned arrangement of unstructured texts or providing an archival description to an individual unstructured text. Burgener and Rydning (2022) have outlined that an ever-growing number of digital records will be dominated by unstructured data in the proportion of up to 90% which is created on a yearly basis, and the proportion will just keep increasing.

There are several approaches available as to how to proceed when handling unstructured data but, when proceeding with such approaches, certain challenges should be addressed and taken into consideration, such as (OpenText Corporation, 2024; Baig, 2023):

- Accessibility and usability of unstructured data: The rapid evolution of information technologies and diverse formats may impact the readability of data, posing a challenge in maintaining its usefulness for subsequent processing.

- Efficient handling of vast data volumes: Managing the exponential growth of unstructured data poses the challenge of processing and capturing information promptly to prevent potential losses.
- Complex indexing and classification: The diverse forms and unknown contents of records make indexing and classification a demanding process prone to errors, significantly affecting the quality of the obtained results.
- Security challenges: Safeguarding confidential data during processing becomes intricate, as this information can swiftly proliferate across diverse record formats and storage locations, leading to difficulties in identifying sensitive content.
- Support for diverse record formats: Unstructured data lacks predetermined standard record formats, complicating data processing by requiring versatile solutions to handle various types of formats effectively.
- Requirement for specialised resources and expertise: Unstructured data constitutes a majority of created material today, necessitating robust hardware for efficient processing and skilled personnel, often referred to as “data scientists,” capable of devising appropriate solutions for handling unstructured data.
- Considerable expense in establishing unstructured data processing systems: Beyond hardware and human resources, the cost of additional components, including specialised software, data storage equipment, and measures related to information security, must be considered when setting up systems for processing unstructured data.

## ARTIFICIAL INTELLIGENCE AND ARCHIVING

There are several definitions of the concept of artificial intelligence but, generally, artificial intelligence could be defined as “*a science whose goal is to make a machine that will do things which require human intelligence*” (Balič, 2004) and as a system that can design or execute independently without human intervention (Barredo et al., 2020). Klasinc (2023) defines the use of artificial intelligence in Archival science as collective solutions “*that help in generating and managing archival content, context and other relations established in archival material*”.

There are several advantages and disadvantages of using artificial intelligence.

Some of the advantages that can help in the field of archiving are “task automation” where faster task execution can provide a solution for mass data handling; no overload or stress when executing tasks; an ability to perform several tasks at the same time, low costs in relation to the work undertaken; possibility to discover relations, connections and patterns in previously unknown content etc, (Khanzode & Sarode, 2020; Bhosale, 2020). While there are some clear advantages to using artificial intelligence, there are also some disadvantages that need to be addressed when deciding if using such a system is appropriate, such as: a significantly higher inaccuracy given the potential errors when executing tasks; dependency and subjectivity with regards to the rules designed by the system architect (creativity and vision); the potential high cost in the development and implementation; dependence on specific technology; impact on a need for human resources; potential abuse and unethical use etc., (Khanzode & Sarode 2020; Bhosale, 2020).

## ARTIFICIAL INTELLIGENCE AND ETHICS

Ethics within the field of artificial intelligence primarily involves assessing the implications and potential outcomes associated with the development and deployment of artificial intelligence (Boddington, 2023). One significant consideration is how the advancement of artificial intelligence may impact the demand for human labour and influence the nature of work (Kumar et al., 2021). The archivist’s discretionary approach as someone who evaluates the archival material and subsequently carries out the archival tasks is one of the most important things for quality assurance for the long term preservation of archival material. Even if artificial intelligence is used for such tasks, the human factor (the archivist) is still needed to initially develop and later evaluate the execution of expert systems and results gathered.

There are several topics that should be taken into consideration when developing solutions, including artificial intelligence, with regard to the execution of archival tasks. Some of the topics can include how to handle or manage privacy, responsibility, trust, continuity and sustainability, dignity, solidarity, transparency and availability, freedom and autonomy to make decisions and the provision of guidelines to enforce harmless execution (Boddington, 2023).

## MACHINE LEARNING

Machine learning is a specific application of artificial intelligence. According to Murty and Avinash (2023), machine learning is considered a mature, dynamic and crucial field that has evolved over more than six decades. The accelerated expansion of machine learning can be attributed, in part, to the recent surge in the availability of machine-processable data and the enhancement and accessibility of hardware which is capable of efficiently processing substantial amounts of data.

At its core, machine learning relies on the processing of data to facilitate learning, making the format of the data a crucial factor. Data can be broadly categorised into structured, unstructured, semi-structured, and descriptive data or metadata. Furthermore, the quality and appropriateness of the input data play a vital role, influencing the approach and expected outcomes. Inadequate data or data containing extraneous information can significantly impact the expected results (see Caliskan et al., 2017).

There are many approaches to the use of machine learning, four of which are the most commonly used today (Sarker, 2021): supervised learning, which relies on data processing in respect to the relationship between input and output data, utilising pre-processed data which is readily available for learning and training; unsupervised learning which involves data processing without prior manipulation or human intervention; semi-supervised learning which combines elements of both supervised and unsupervised approaches, utilising both pre-processed and non-pre-processed data in processing; and reinforcement learning which is a machine learning approach facilitating the automatic assessment of optimal behaviour within a given context or environment to enhance performance.

## NATURAL LANGUAGE PROCESSING

One of the domains where machine learning is extensively used is “natural language processing” (NLP). Eisenstein (2018) draws a direct comparison between natural language processing and the term “computational linguistics.” Despite a significant overlap, a distinction persists between the two definitions. “Computational linguistics” primarily emphasises “linguistics,” wherein various forms of computer processing play a supportive role, while in the case of natural language processing, the focus is on the design and analysis of computer algorithms and approaches for natural human language processing.

The primary objective of natural language processing is to provide new computing capabilities related to human language, including tasks such as extracting information from texts, language translation, question answering and engaging in conversations etc. Khurana et al. (2022) define natural language processing (NLP) as a branch of artificial intelligence and linguistics dedicated to enabling computers to comprehend statements or words written in human languages. Natural language processing was developed to simplify user tasks and to fulfil the desire to communicate with computers in a natural language. Additionally, the authors categorise the field of natural language processing into the “understanding of natural language” and “natural language creation.”

Some of the most common uses of natural language processing using machine learning can include (McMullen, 2023): text summarisation, automated chat rooms, machine translation, classification of texts, answering questions, recognition of named entities, creation of natural language, discerning the meaning of words, sentiment analysis, speech recognition and connection of entities, etc.

## RESEARCH

Managing extensive digital content with unknown and unstructured data with regards to identifying potential archival value poses a significant challenge. To provide swift, efficient and accurate handling of content with archival tasks, such as providing descriptions or arranging records, becomes problematic, particularly when attempting to classify, edit or list such vast quantities of digital material and when not undertaken immediately may lead to a poor preservation process and an inferior quality in respect of the records retained (Popovici, 2022).

When dealing with unstructured and content-ambiguous records, there is a likelihood that numerous records may not be worth retaining or preserving (Moss and Gollins, 2017). Therefore, it is essential to devise an approach for identifying and preserving those records of archival value (Grigory, 2023) through content classification and evaluation. This involves ensuring the accessibility and usability of the retained material while distinguishing it from what should be discarded.

The research presented here outlines the “zero shot classification” (ZSC) approach to classifying unstructured texts with archival value using machine learning. The object of the research was also to determine how the ZSC compares to a more guid-

ed approach of building a decision model based on a learning process and if it is possible to create an implementation model for the processing of unstructured texts. An example that was directly covered in the research was “text classification”.

## **ZERO SHOT CLASSIFICATION**

Zero shot classification relates to an approach to achieve a classification through a prediction of a classes that were not part of the initial training of the model (Hugging Face, s. d. b). ZSC still uses a pre-trained model to achieve its goal, but the aim is to provide an approach which can achieve a task of classification where training data for supervised classification is scarce.

There is one problem in particular that can present a serious obstacle to building a dedicated classification model and that is the lack of quality labelled data set for training, especially in the case of supervised classification. This is even more highlighted when we are dealing with multilanguage content. Generally, the ZSC relies on learning to recognise the layer of semantic attributes while building the model which can be used to identify classes that were not visible in the training process of the pre-trained model (Alcoforado et al., 2022).

There are several practical cases where ZSC can be used, given that there are no required prerequisites apart from the pretrained model such as: categorisation or topic classification, identification of intent, sentiment analysis and even image classification. Some of the advantages that ZSC could bring to the aforementioned tasks are mostly related to the optimisation of the processes, such as the time required to achieve the tasks, flexibility with regards to the inclusion of new material and independence regarding the form of data etc.

There are also some disadvantages which could impact the decision to use ZSC as an approach to undertake archival tasks such as the quality of the pre-trained model and connected class descriptions, extreme variation between the content and classes which were used for pre-training the model data and data which is intended for zero shot classification. One of the biggest drawbacks in using ZSC is also how to provide a tangible evaluation process with the intention of measuring the performance of the zero-shot learning process, as there are no preexisting labels that could provide any type of quantification such as in the supervised classification approaches (Xian et al., 2020).



Since ZSC is not a supervised approach there are also concerns in using such an approach in respect of tasks which may include ethical considerations. Given that the general guideline of the ZSC model relies heavily on the pre-trained model, certain points need to be taken into consideration like “content subjectivity” with an unsuitable class description process, which does not correctly capture the relationship between model decisions and the intended architectural layout, “biased decisions” for the same reason given the unsuitable class description, “misclassification” with regards to not understanding the context and “privacy concerns” etc. (Van Otten, 2023).

An inherent limitation of utilising the ZSC approach can also be in their narrow focus when declaring the hypothesis to implement the classification. If these approaches fail to detect the specific semantics they’re trained to identify initially, it’s unlikely they’ll retrieve it any time later. This outlines the importance of maintaining consistency and clear instructions in the criteria creating the hypothesis. Any alterations from its concise meaning must be carefully considered to avoid disrupting the coherence of understanding as to what it needs to search for and how it should understand the similarities between classes. Utilising machine learning and ZSC can provide the basis for providing autonomously established rules for assessing content relevance, given the use of a suitable pre-trained model. This facilitates the possibility of identification and segregation of material with potential archival significance.

## EXPERIMENT

The purpose of the research was to assess the efficiency of ZSC on individual text records when compared to the more controlled and supervised approach to text classification.

The data which was used to execute the ZSC were publicly available unstructured news articles in the English language, aggregated in the machine-readable textual form.

The following architecture was used to test the zero shot classification approach and compared with the variant of the supervised classification approach:

- 1000 individual records, each record containing one news article (Guardian Media Group, 2023); all individual records were pre-labelled under three sections or topics which were named: “government”, “business” and “sports”.

- For Zero Shot Classification, the pre-trained model “Roberta-large-mnli” (Hugging Face, s. d. a) was used.
- For the comparison classification using the supervised approach pre-trained model, the “Bert-base-cased” (Hugging Face, s. d.) model was used.

The workflow for executing the classification was divided in two parts. The first one being the Zero Shot Classification approach, which directly executed predictions, and the second one with the supervised approach including the additional “learning step” and, subsequently, the creation of a decision model to make predictions. Since the ZSC method does not include the “learning” step, the evaluation of efficiency would prove to be difficult as such so for valuation purposes, the data was pre-labelled with three topics.

The Zero Shot Classifier parametrisation, with the use of a pretrained model, was as follows:

- 1000 individual records as input for executing the Zero Shot Classification.
- A classifier was executed on each individual record.
- Content of each individual record was in its original, unstructured form.
- Candidate labels of expected classifying classes were “government”, “business” and “sports”.
- Custom hypothesis was used with the following narrative “The topic of this content is {}”
- Batch size: 10.
- 1000 individual records as output with assigned predictions on individual records in respect of the pre-labelled data
- Analysis and quantified distances (efficiency) between pre-labelled classes and predicted classes.

The supervised classification approach, using pre-trained model parametrisation, was as follows:

- 1000 individual records as input for executing the learning process and the creation of a decision model for classification.
- A classifier was executed on each individual record.
- Content of each individual record was in its original unstructured form.

- The classification learner included data (individual records) which were pre-labelled with one of the following classes (labels): “government”, “business” and “sports”.
- Maximum sequence length: 512.
- Number of epochs: 8.
- Batch size:32.
- Validation batch size:20.
- Optimiser: Adam; Learning rate: 0.001.
- 1000 individual records as output with assigned predictions on individual records in respect of the pre-labelled data
- Analysis and degree of difference (efficiency) between pre-labelled classes and predicted classes.

## RESULTS

### Zero Shot Classification

The Zero Shot Classification result was as follows:

- 71% accuracy in respect of the pre-labelled data.
- 0.57 Cohen’s Kappa.

Row ID	<input type="checkbox"/> government	<input type="checkbox"/> business	<input type="checkbox"/> sports
government	319	21	4
business	169	157	16
sports	75	2	237

**Figure 1: Confusion matrix ZSC (Knime, 2024)**

### Supervised classification approach

The supervised classification result was as follows:

- 97% accuracy in respect of the pre-labelled data.
- 0.96 Cohen’s Kappa.

Row ID	<input type="checkbox"/> government	<input type="checkbox"/> business	<input type="checkbox"/> sports
government	103	3	0
business	1	99	1
sports	2	1	90

**Figure 2: Confusion matrix supervised classification (Knime, 2024)**

## DISCUSSION AND CONCLUSIONS

### Discussion

Results for the ZSC approach show that in the case of classification, where no explicit learning process was involved and based on a generic pre-trained model, an accuracy of 71% was recorded. This in turn represents a good result given that the only “fine tuning” that was undertaken was where hypotheses and expected labels (classes) were provided without an actual pre-understanding of the data that was intended for classification. Because the ZSC approach is generally aimed at classifying data that can’t be used for a supervised learning process of classification, it does provide a good way to solve problems classifying unstructured data with potential archival value where content is unknown.

The confusion matrix (Figure 1), in turn, shows where discrepancies were identified. While the “sports” class showed certain deviations towards being misclassified as “government” it was the “business” class that showed the biggest deviation with regards to the “government” class, as it was in most cases classified as “government” instead of, “business,” as expected. This may, in a way, be explained by the fact that the content in individual records, which were pre-labelled with those classes mentioned, is very similar between two classes and, as such, provided the biggest challenge to make a suitable distinction. This deviation could also be attributed to the process of using different weights on the side of the pre-labelling process where “pre-classification” has already been undertaken for the evaluation purposes where the classifier could use a slightly different approach thus producing the contrasting results seen using the ZSC approach.

While the ZSC classifier went through the process of classification without the additional learning activity provided for input data, this was not the case in the second approach creating a decision model, based on additional activity of learning on the actual content that was intended for the classification prior to the final classification task. Because the second approach included the learning phase on the input data provided, accuracy was measured at 97%. This certainly represents an excellent result, but this could only be achieved while having the pre-labelled data available for training purposes. When dealing with unstructured texts with potential archival value, such availability or un-availability of pre-labelled data may prove to be one of the biggest obstacles for using this approach.

The confusion matrix (Figure 2) in the second approach shows that the learning process was crucial for eliminating confusion regarding content similarity. It shows that understanding content and its attributed values before actual classification does provide for a more coherent way to accept decisions.

## Conclusions

Zero Shot Classification does provide a useful approach when dealing with unknown content which has potential archival value and with no accessible means to train the decision model based on pre-labelled data. It does exhibit certain disadvantages when compared to the supervised classification approach, but it also shows certain advantages in specific real-life scenarios. There are many archives and many creators which could benefit from the use of ZSC for providing usable content in accessing archival records, such as a virtual archive reading room (Sabadin, 2023) or any other digital platform accessible to the public.

With the proliferation of digital content creation and usage, the need for effective management of mass, unstructured data of potential archival value has become paramount. However, alongside this challenge comes the dilemma of determining what content to capture, when to capture it, and how to do it, raising questions about the necessity of capturing all digital material or just a selection. Given the vast volume of digital content being generated, it's imperative to establish criteria for evaluating material, identifying what warrants long-term retention or preservation as archival material. Adaptations and enhancements to evaluation approaches are essential to accommodate the sheer volume of digital content, ensuring that management processes maintain their quality without compromising efficiency. To address these challenges, it is crucial to exert control over new methods and technologies used for evaluating the value of potential archival content and ensuring transparency throughout the process which uses such technologies. Additionally, clear procedures and mechanisms must be established to ensure compliance with archival regulations and standards, fostering a clear and comprehensive environment conducive to effective management of unstructured text when using machine learning approaches to classify records.

## REFERENCES

- Alcoforado, A. & Ferraz, T. P. & Gerber, R. & Bustos, E. & Oliveira, A. S. & Veloso, B. M. & Siqueira, F. L. & Costa, A. H. R. (2022). ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling. In P. Gamallo, R. Amaro, C. Scarton, F. Batista, D. Silva, C. Magro & H. Pinto (eds), *PROPOR 2022: Computational Processing of the Portuguese Language* (pp. 125-136). Fortaleza, Brazil: Springer.
- Baig, J. (2023). *Unstructured Data Challenges for 2023 and their Solutions*. Retrieved at <https://www.astera.com/type/blog/unstructured-data-challenges/> (accessed on 10.03.2024).
- Balič, J. (2004). *Inteligentni obdelovalni sistemi*. Maribor: Faculty of mechanical engineering, University of Maribor.
- Barredo, A. A. & Díaz-Rodríguez, N. & Del Ser, J. & Bennetot, A. & Tabik, S. & Barbado, A. & Garcia, S. & Gil-Lopez, S. & Molina, D. & Benjamins, R. & Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Bhosale, S. & Pujari, V. & Multani, Z. (2020). Advantages and Disadvantages of Artificial Intelligence. *Aayushi International Interdisciplinary Research Journal*, 77, 227–230.
- Blei, M. D. & Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3(2003), 993–1022. Retrieved at <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf> (accessed on 10.03.2024).
- Boddington, P. (2023). *AI Ethics*. Singapore: Springer Nature Singapore Pte Ltd.
- Burgener, E. & Rydning, J. (2022). High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises. *IDC White paper*. Retrieved at <https://www.delltechnologies.com/asset/en-my/products/storage/industry-market/h19267-wp-idc-storage-reqs-digital-enterprise.pdf> (accessed on 10.03.2024).
- Caliskan, A. & Bryson, J. J. & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.

- Eisenstein, J. (2018). *Natural language processing*. MIT press. Retrieved at <https://cseweb.ucsd.edu/~nnakashole/teaching/eisenstein-nov18.pdf> (accessed on 10.03.2024).
- Grigory, L. N. (2023). Archival science in the postindustrial society. *Atlanti+*, 33(1), 38–44. Retrieved at <https://journal.almamater.si/index.php/atlantiplus/issue/view/41> (accessed on 10.03.2024).
- Guardian Media Group. (2023). *The Guardian Open Platform*. Retrieved at <https://open-platform.theguardian.com/> (accessed on 10.03.2024).
- Hugging Face (s. d.). *Bert-base-cased*. Retrieved at <https://huggingface.co/google-bert/bert-base-cased> (accessed on 12.03.2024).
- Hugging Face. (s. d. a). *Roberta-large-mnli*. Retrieved at <https://huggingface.co/FacebookAI/roberta-large-mnli> (accessed on 12.03.2024).
- Hugging Face. (s. d. b). *Zero shot classification*. Retrieved at <https://huggingface.co/tasks/zero-shot-classification> (accessed on 10.03.2024).
- Khanzode, C. A. & Sarode, R. D. (2020). Advantages and disadvantages of artificial intelligence and machine learning: a literature review. *International Journal of Library & Information Science (IJLIS)*, 9(1), 30–36.
- Khurana, D. & Koli, A. & Khatter, K. & Singh, S. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications*, 82(6). Retrieved at [https://www.researchgate.net/publication/319164243\\_Natural\\_Language\\_Processing\\_State\\_of\\_The\\_Art\\_Current\\_Trends\\_and\\_Challenges](https://www.researchgate.net/publication/319164243_Natural_Language_Processing_State_of_The_Art_Current_Trends_and_Challenges) (accessed on 10.03.2024).
- Klasinc, P. P. (2023). Archivistics, Archival science and Artificial intelligence. *Atlanti+*, 33(2), 25–36. Retrieved at <https://journal.almamater.si/index.php/atlantiplus/issue/view/42/31> (accessed on 22.05.2024).
- Knime. (2024). *Knime*. Retrieved at <https://www.knime.com/> (accessed on 12.03.2024).
- Kumar, P. & Jain, V. K. & Kumar, D. (2021). *Artificial Intelligence and Global Society*. Boca Raton: Taylor & Francis Group, LLC.
- McMullen, M. (11. 5. 2023). 11 NLP Use Cases: Putting the Language Comprehension Tech to Work. *Readwrite*. Retrieved at <https://readwrite.com/11-nlp-use-cases-putting-the-language-comprehension-tech-to-work/> (accessed on 10.03.2024).