Faisal N Alhajri[1]

CURRENT TRENDS IN NATURAL LANGUAGE PROCESSING APPLICATION AND ITS APPLICATIONS IN IMPROVING THE QUALITY OF TOURISM COMMUNICATION

Abstract: This paper aimed to review the papers published since 2021 to evaluate the current trends in natural language processing. A search of Google Scholar with the time frame of 2021-2023 using a PRISMA methodology yielded 27 papers for review. The review revealed an evolutionary trend from simple NLP models to multitasking, word embedding, neural networks, sequence-to-sequence models, and attention mechanisms to the current trends of pre-trained NLP uses in different contexts. Many techniques and algorithms were used in this process. One of the latest developments was identifying false news about Covid-19 pandemic from social media posts, to reduce risks to the population. Healthcare seems to be a major beneficiary of these developments. Natural Language Processing and its applications to tourism communication present great potential. The consistent development of NLP technology is allowing for the use of automated translation services, personalized tourism messaging, and more natural interplays between humans and AI systems (e.g., through ChatBots). The innovative and powerful capabilities of NLP can potentially help to drive tourism-related services ahead of the curve. As the technology continues to innovate and evolve, the advantage it offers to the tourism industry cannot be understated. A few limitations of this review have been mentioned at the end. This research has implications for tourism communications.

Keywords: natural language processing, NLP, tourism communication, review

1. Introduction

"Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can." (IBM, 2023). In NLP, computational linguistics is combined with statistical, machine learning and deep learning models. NLP can help to convert texts or voices of human language to understand their full meaning. It can translate texts from one language to another. It can also respond to spoken commands. Many daily used applications like voice-operated GPS systems, digital assistants,

Corresponding author: Faisal N Alhajri
Email: f155@live.com

speech-to-text dictation software, and customer service chatbots use NLP. All of these can happen in real-time.

NLP helps in linguistic analysis through speech recognition, parts of speech tagging, word sense clarity, named entity recognition, co-reference resolution, sentimental analysis, and natural language generation. Some use cases of NLP are chatbots, spam detection, and social media sentiment analysis.

All the above mechanisms and uses of NLP were found through research. Therefore, it will be interesting to examine how research has contributed to different aspects of natural language processing. This paper aims to qualitatively review the most current trends by selecting papers published in 2021 and after.

2. Method & Results

The method of identifying and selecting the papers consisted of using the term "natural language processing'" to search Google Scholar setting the time in 2021 and beyond. A PRISMA methodology was used to shortlist papers as shown in Fig 1.This process yielded 27 papers for this review. These papers are discussed below.
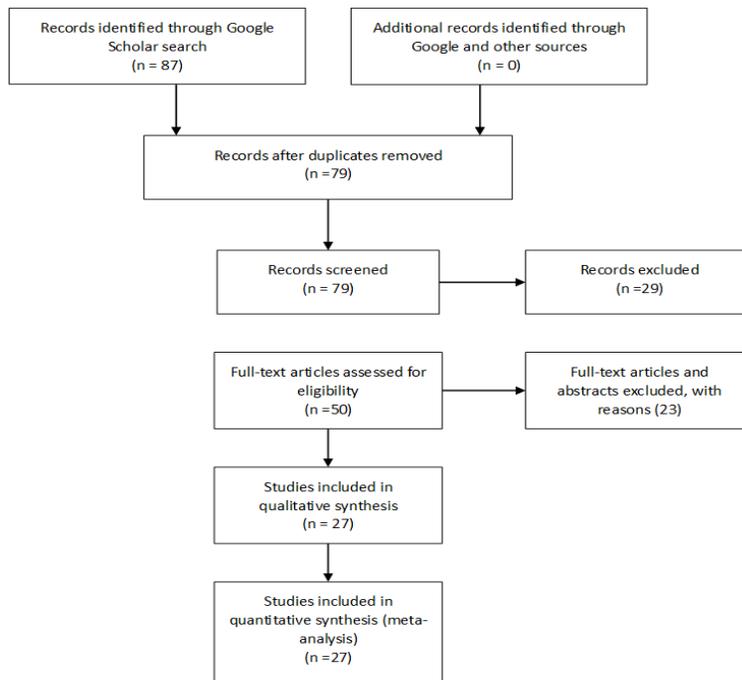

Figure 1. PRISMA flowchart

In a review, Maulud, Zeebaree, Jacksi, Sadeeq, and Sharif (2021) compared seven NLP algorithms for accuracy and area under the curve. LSTM performed best, followed by CNN. Continuing the most recent developments identified by the authors were semantic analysis, topic modelling, tokenisation, named entry recognition, generalised searching, natural language recognition, automated extraction, supervised machine learning, and ontology. The accuracy of most of them was 85 to 95%.

The Arabic language can be categorised into three forms: classical Arabic (CA), Modern Standard Arabic (MSA), and Arabic Dialect

(AD). MSA and AD could be written either in Arabic or in Roman script (Arabizi), which corresponds to Arabic written with Latin letters, numerals, and punctuation. CA is used in literary texts and in the Quran. Guellil, Saâdane, Azouaou, Gueni, and Nouvel (2021) identified works on NLP of all three forms of Arabic. In these works, the NLP tools used for Arabic were either adaptations of the tools used for English or developed specifically for Arabic.

Five sources of bias in NLP identified by Hovy and Prabhumoye (2021) were the data, the annotation process, the input representations, the models, and the research design. The authors provided a diagram of these biases and offered countermeasures for each bias.

As new NLP tasks emerge, research needs more datasets. In this respect, Lhoest, et al. (2021) announced the availability of a library of 650 datasets contributed by over 650 researchers. The dataset was named Hugging Face. Each dataset is tagged with its type, construction, and usage. Hugging Face Datasets is an open-source, community-driven library that standardizes the processing, distribution, and documentation of NLP datasets. This dataset is in continual development and more datasets will be added over time. Details of how to access and use the datasets in Hugging Face have been provided. To illustrate some of its uses, case studies on N-task pretraining benchmarks, reproducible shared tasks, and robustness evaluation have been described.

There are substantial gains on many NLP tasks by pretraining large neural models of NLP like BERT. Most pretraining works are done on the general domain corpora like the web. The underlying assumption here is that general domain language models can be extended to domain-specific pretraining. This assumption was challenged by Gu, et al. (2021). They showed that for domains which have many unlabelled texts, pretraining language models from scratch is better than

continual pretraining of general domain language models (Fig 2).

A comprehensive biomedical NLP benchmark from published datasets was compiled. Experiments with these datasets showed that the domain-specific pretraining laid a strong foundation for many biomedical NLP tasks, leading to new advanced results for different applications. A comprehensive benchmark, BLURB, was created for this purpose. The tested tasks were named entity recognition, relation extraction, document classification, and question answering. Their NLP model was named PubMed BERT.
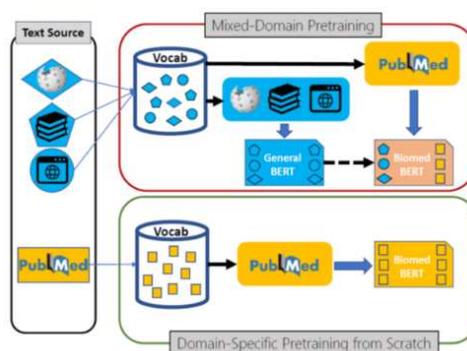

Figure 2. The difference between the classical assumption and the current proposition (Gu, et al., 2021).

The recent research works on transformer-based pretrained language models were reviewed by Kalyan, Rajasekharan, and Sangeetha (2021). The topics covered were different pretraining methods, pretraining tasks, embeddings, downstream adaptation methods, intrinsic and extrinsic benchmarks, and useful libraries to work with T-PTLMs. A new taxonomy to categorize various T-PTLMs was also proposed.

NLP techniques to identify the creation and impact of new technologies in the US patents were developed by Arts, Hou, and Gomez (2021). The method was validated by comparing it with traditional methods using two case-control studies. Patents linked to the Nobel Prize and National Inventor Hall of Fame were collected first. These patents consist of novel technologies which had

subsequent impacts on technological progress and patenting. A second set of patents granted by the US patent office and rejected by the EU or Japan. These consisted of weak technologies with poor subsequent impacts.

The NLP used for this purpose was described in detail. Overall, both case-control studies validated the use of NLP to measure the technical novelty and impact of patents. The results also supported the improvement over traditional approaches based on patent classification and citations. However, the new NLP was correlated weakly with the traditional measures. New keyword combinations and their reuse by later patents showed high discriminatory power to classify patents related to both prestigious awards and rejected patents. The performance of the new NLP was superior to other text-based metrics and the traditional measures based on patent classification and citations.

NLP-supported requirements engineering (RE) applies NLP techniques, tools and resources to different requirements documents or artefacts to support various linguistic analysis tasks performed at various RE phases. These tasks include the identification of language issues, and key domain concepts, and the establishment of traceability links between requirements. A systematic mapping approach was used by Zhao, et al. (2021) to review literature on five key aspects: the state of the literature, the state of empirical research, the research focus, the state of the practice, and the NLP technologies.

Much research literature, mostly concerned with solutions for evaluation by laboratory experiments or using an example, most studies focus on analytical aspects for detection as the main task by the processing of requirements specification documents, 130 new tools for various NLP tasks (without adequate evidence of their long-term adoption or industrial applications), and 140 NLP techniques, 66 NLP tools, 25 NLP

resources were identified from the review. However, most of the techniques, tools and resources, particularly novel and specialised ones, were used rarely. On the other hand, syntactic analysis techniques, general-purpose tools, and generic language lexicons were used very frequently. The huge gap between research and practice due to inadequate industrial validation, lack of evidence of industrial adoption of the proposed tools, insufficiently shared RE-specific language resources, and the lack of NLP expertise in requirements research to advise industries on the choice of NLP technologies, were indicated by this review. Mapping diagrams have been provided for all five aspects of this review.

Trankit, a lightweight Transformer-based Toolkit for multilingual NLP, was presented by Van Nguyen, Lai, Veyseh, and Nguyen (2021). It provides a trainable pipeline for basic NLP tasks for about 100 languages and 90 pre-trained pipelines for 56 languages. It was built on a state-of-the-art pre-trained language model. Trankit outperformed some previous multilingual NLP pipelines on sentence segmentation, part-of-speech tagging, morphological feature tagging, and dependency parsing. It performed similarly to other multilingual NLPs on tokenization, multi-word token expansion, and lemmatization over 90 Universal Dependencies treebanks.

Identification of misinformation about covid pandemic in social media is essential to prevent risks associated with them. The use of BERT (Bidirectional Encoder Representations from Transformers) has been successful in detecting such misinformation. Ayoub, Yang, and Zho (2021) proposed an explainable NLP model based on DistilBERT and SHAP (Shapley Additive exPlanations) to fight against misinformation about COVID-19. These NLP models are highly efficient and effective in detecting misinformation. The authors tested a dataset of 984 claims about COVID-19 for fact-checking. Then, the sample size was doubled using back-

translation. To this data, the DistilBERT model was applied. This resulted in obtaining an accuracy of 0.972 and areas under the curve of 0.993) in identifying misinformation about COVID-19.

This NLP model was used on a larger dataset for AAAI2021-COVID-19 Fake News Detection Shared Task. This test led to an accuracy of 0.938 and areas under the curve of 0.985. Both results were better than the traditional models. To enhance public trust in model prediction, SHAP was used to enhance model explainability. Then, it was evaluated using a between-subjects experiment with three conditions: text (T), text + SHAP explanation (TSE), and text + SHAP explanation + source and evidence (TSESE). The participants were more likely to trust and share information related to COVID-19 in the TSE and TSESE conditions than text alone. However, the results tables of the paper show that BERT (0.993, 0.999), and Aug-BERT (0.994, 0.999) were superior to DistilBERT (0.938, 0.985) or Aug-DistilBERT (0.972, 0.993).

In their studies, Liu, Shin, and Burns (2021) used big data of 3.78 million tweets from the top 15 luxury brands with the highest number of Twitter followers. NLP was applied to quantify the unstructured big data of the downloaded tweets. The procedure adopted in applying NLP to this data has been described in detail in the paper. Entertainment, trendiness, and interactions had significant effects on customer engagement in the case of tweets on luxury goods. Thus, if a luxury firm invests in enhancing entertainment, interaction, and trendiness, it will commercially benefit the company by increasing customer engagement with the brand-related social media content.

A detailed description of LexNLP was provided by Bommarito II, Katz, and Detterman (2021). LexNLP is an open-source Python package. It has NLP and machine learning capabilities for legal and regulatory texts. In the package, there is functionality for segmenting documents, identifying key text such as titles and section headings, extracting over eighteen types of structured information like distances and dates, extracting named entities such as companies and geopolitical entities, transforming text into features for model training, and building unsupervised and supervised models like word embedding or tagging models. LexNLP also carries pre-trained models based on several unit tests drawn from real documents sourced from the SEC EDGAR database and various judicial and regulatory proceedings. LexNLP can be used in both academic research and industrial applications.

In the field of the rapidly developing area of Neural Architecture Search (NAS), training neural networks requires high levels of computational power. This makes NAS unreachable for most researchers. The recently introduced benchmarks with precomputed neural architecture performances are only for the computer vision domain and are built from image datasets and convolution-derived architectures. To solve these problems, Klyuchnikov, et al. (2022) proposed a NAS by leveraging the language modelling task, outside the computer vision domain as the core of natural language processing (NLP). The proposed benchmark was provided with search space of recurrent neural networks on the text datasets and trained 14k architectures within it. Both intrinsic and extrinsic evaluations of the trained models using datasets for semantic relatedness and language understanding evaluation and testing of several NAS algorithms were done. The results demonstrated how the pre-computed results can be utilized.

NLP has been used for predictive uses so far. An emerging multidisciplinary research area is changing this to the classic reason for scientific research, namely, finding causal relationships. Feder, et al. (2022) reviewed various aspects of using NLP to find causal relationships. However, certain confounders, and lack of robustness of the data, might

mask the causal relationships between two variables in the text. Certain assumptions of causal relationships in NLP are the independence of statistical treatments from counterfactual outcomes, the probability of receiving treatment being bounded away from 0 and 1 for all values of the confounders X and uniform outcomes for successive treatment levels. Challenges of estimating causal effects from texts were discussed. These challenges arise from confounders and assumptions made. For future works, heterogeneity of textual contents and interpretations, difficulty in extraction of low dimensional text features, absence of benchmarks, and the need to use control the generation of texts of unwanted significance are some challenges. Causal reasoning to solve traditional NLP tasks is also possible. Robust predictions are one area.

The use of counterfactual data both in training and testing sets can improve the robustness of observed causal relationships. Instead of data augmentation, new algorithms can also be used for direct operation on the data. Fairness and absence of bias are two other aspects of robust causal relationships. Interpretations predictions from NLP models for causal relationships need to be bias-free for the detection of errors in the relationships identified. A comparison of predictions for each sample with its counterfactual can be adopted for this purpose. The use of manually done counterfactuals helps to observe information flows and identify the site encoded in the model.

In an extensive review Khurana, Koli, Khatter, and Singh (2023) noted that in the 1940s, although machine translation had started, the concept of NLP was unknown. Despite many research works done in England and Russia, machine translation almost stopped by 1966. However, other works related to NLP were progressing. Towards the end of the 20th century, rapid developments in research on building various linguistic capabilities of computers established NLP. The recent developments in NLP were discussed by the authors using a diagram (Fig 3). The objectives of NLP are the interpretation, analysis, and manipulation of natural language data for a defined purpose using various algorithms, tools, and methods. In 2001, neural language modelling was introduced to estimate the probability of occurrence of the next word (token) given a certain number of previous words. The concept of a feed-forward neural network and lookup table were used for this purpose. The application of multitask learning was introduced in 2008. In this application, two convolutional models with maximum pooling were used to tag parts of speech and named entity recognitions. In 2013, a word embedding process was developed for dense vector representation of text. Neural networks were introduced in the same year, in which, variable length input is taken for further processing.

In 2014, a general framework for sequence-to-sequence was developed for using encoder and decoder networks to map from sequence to vector and back respectively. Much research and use of neural networks have been done over the years. Neural networks in different forms were developed. These include convolutional neural networks (CNN) to classify visual images for further analysis. The use of CNN was extended to sentence classification, sentiment analysis, text classification and summarisation, machine translation and answer relations. Recurrent Neural Networks (RNN) were introduced to perform the same function on every data occurring recurrently. These can be texts, speech, time series, financial data, audio, or video. A popular modification of RNN is the Long Short-term Memory (LSTM), useful to retain only the important information desired for a long time.

A simpler RNN, Gated Recurrent Guru (GRU), was developed as an improvement of LSTM to provide better results. Attention mechanisms can be used to suggest a network to learn to pay attention to components of current hidden data and

annotate together using transformers. There had been many developments of transformers applicable to fixed or variable text lengths. These include Transformer XL, Compressive Transformer, Deep learning, and BERT (Bidirectional Encoder Representations from Transformers). Tag sets have been developed for Indo-European languages but are less researched in the case of Asian and Middle Eastern languages. Some of the most recent developments are tag Chunking, named entity recognition, emotion detection, semantic role labelling, and event discovery in social media feeds.
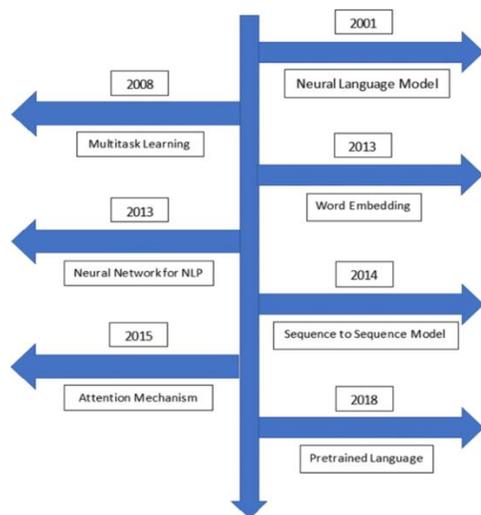

Figure 3. Recent developments in NLP (Khurana, Koli, Khatter, & Singh, 2023)

A systematic review of prompting methods was conducted by Liu, et al. (2023). Traditional supervised learning trains a model to predict an output as p (y/x), using an input. In the case of prompt-based learning, the probability of text is modelled directly. In fully supervised learning, a task-specific model is trained solely on a dataset of input-output examples for the target task. As this method was highly insufficient, feature engineering was used in early NLP works. In this method, they used their domain knowledge to define and extract salient features from raw data. To learn from

this limited data, and provide models, appropriate inductive biases were used. When neural network models were introduced for NLP, learning of salient features occurred jointly with the training of the model itself. Thus, architectural engineering came into existence. In these models, inductive bias was provided through the design of suitable network architecture for learning such features. Since 2017, this paradigm of fully supervised learning shifted to the pre-train and fine-tune paradigm. In this system, a model with a fixed architecture is pre-trained as a language model (LM). This is used for predicting the probability of observed textual data.

As large volumes of raw textual data are readily available to train LMs, such large datasets can be used for training. The process of learning robust general-purpose features of the language evolves from this modelling. This pre-trained LM can be adapted to different downstream tasks for which fine-tuning, and introduction of additional parameters can be done using task-specific objective functions. This paradigm is known as objective engineering, in which, the training objectives are used at both the pre-training and fine-tuning stages. This pre-train, fine-tune concept has been replaced by the pre-train, prompt and predict concept now. Here, instead of adapting pre-trained LMs for downstream tasks through objective engineering, downstream tasks are reformulated to appear like those solved during the original LM training using a textual prompt.

Thus, by selecting the appropriate prompts, it is possible to manipulate the behaviour of the model so that the pre-trained LM itself is used to predict the desired output, even without any additional task-specific training in some cases. But to find suitable prompts, prompt engineering is required.

An evaluation of COVID-Twitter-BERT (CT-BERT), a transformer-based model that pre-trained on a large corpus of COVID-19-related Twitter messages, was done by

Müller, Salathé, and Kummervold (2023). CT-BERT was designed specifically for use on COVID-19 content in social media. It can be used for NLP tasks like classification, question-answering, and chatbots. For the evaluation, five different classification datasets were used. This included one target domain. The performance of the model was compared with its base model, BERT-LARGE, to measure marginal improvement. CT-BERT performed better than BERT-LARGE with a marginal improvement of 10-30% on all five classification datasets. The highest improvement was in the case of the target domain. This type of analysis can be used to monitor public sentiment and to develop chatbots to provide precise covid-related information to the public.

Natural language processing has been used to enhance tourism communication as demonstrated by Filieri, D'Amico, Destefanis, Paolucci, and Raguseo (2021). They found that AI and NLP was being used extensively to enhance tourism communication in Europe. A number of authors have also found that AI and NLP based ChatBots are being used to enhance tourism communication in various contexts (e.g., Pillai & Sivathanu (2020), Hasan et al. (2021), Borghi & Mariani (2021), Tuo, Ning, & Zhu (2021), Calvaresi et al. (2021), Orden-Mejia & Huertas (2022), Calvaresi et al. (2023), Mich & Garigliano (2023).

3. Conclusion

As Khurana et al. (2023) observed, the evolutionary process of NLP started with simple models through multitasking, word embedding, neural networks, sequence-to-sequence models, and attention mechanisms to the current trends of pre-trained NLP uses in different contexts. Many techniques and algorithms were used in this process. One of the latest developments was identifying false news about covid pandemic from social media posts, to reduce risks to the population. Healthcare seems to be a major beneficiary of these developments.

Natural Language Processing and its applications to tourism communication present great potential. The consistent development of NLP technology allows for the use of automated translation services, personalized tourism messaging, and more natural interplays between humans and AI systems. The innovative and powerful capabilities of NLP can potentially help to drive tourism-related services ahead of the curve. As technology continues to innovate and evolve, the advantage it offers to the tourism industry cannot be understated.

The focus of this paper was narrowed down to developments since 2021. This narrow focus would have missed some important contributions to NLP made earlier. This narrow focus also limited the number of papers selected for this review.

References:

Arts, S., Hou, J., & Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy, 50*(2), 104144. doi:10.1016/j.respol.2020.104144

Ayoub, J., Yang, X. J., & Zho, F. (2021). Combat COVID-19 infodemic using explainable natural language processing models. *Information processing & management, 58*(4), 102569. doi:10.1016/j.ipm.2021.102569

Bommarito II, M. J., Katz, D. M., & Detterman, E. M. (2021). Natural language processing and information extraction for legal and regulatory texts. In R. Vogl (Ed.), *Research Handbook on Big Data Law* (pp. 216-227). Edward Elgar Publishing. doi:10.4337/9781788972826.00017

Borghi, M., & Mariani, M. (2021). Service robots in online reviews: an empirical study of online robotic discourse. *Annals of Tourism Research, 87*.

Calvaresi, D., Ibrahim, A., Calbimonte, J. P., Schegg, R., Fragniere, E., & Schumacher, M. (2021). The evolution of chatbots in tourism: A systematic literature review. In *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference, January 19–22, 2021* (pp. 3-16). Springer International Publishing.

Calvaresi, D., Ibrahim, A., Calbimonte, J. P., Fragniere, E., Schegg, R., & Schumacher, M. I. (2023). Leveraging inter-tourists interactions via chatbots to bridge academia, tourism industries and future societies. *Journal of Tourism Futures, 9*(3), 311-337.

Feder, A., Keith, K. A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., & Eisenstein, J. e. (2022). Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics, 10*, 1138-1158. doi:10.1162/tacl_a_00511

Filieri, R., D'Amico, E., Destefanis, A., Paolucci, E., & Raguseo, E. (2021). Artificial intelligence (AI) for tourism: an European-based study on successful AI tourism start-ups. *International Journal of Contemporary Hospitality Management, 33*(11), 4099-4125.

Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., . . . Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH), 3*(1), 1-23. doi:10.1145/3458754

Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. (2021). Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences, 33*(5), 497-507. doi:10.1016/j.jksuci.2019.02.006

Hasan, R., Koles, B., Zaman, M., & Paul, J. (2021). The potential of chatbots in travel and tourism services in the context of social distancing. *International Journal of Technology Intelligence and Planning, 13*(1), 63-83.

Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass, 15*(8), e12432. doi:10.1111/lnc3.12432

IBM. (2023). *What is natural language processing?* Retrieved June 24, 2023, from IBM: https://www.ibm.com/topics/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can.

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). Ammus: A survey of transformer-based pretrained models in natural language processing. *arXiv, 2108*(August), 05542. doi:10.48550/arXiv.2108.05542

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications, 82*(3), 3713-3744. doi:10.1007/s11042-022-13428-4

Klyuchnikov, N., Trofimov, I., Artemova, E., Salnikov, M., Fedorov, M., Filippov, A., . . . Evgeny. (2022). Nas-bench-nlp: neural architecture search benchmark for natural language processing. *IEEE Access, 10*, 45736-45747. doi:10.1109/ACCESS.2022.3169897

Lhoest, Q., del Moral, A. V., Jernite, Y., Thakur, A., von Platen, P., Patil, S., & Chaumond, J. e. (2021). Datasets: A community library for natural language processing. *arXiv preprint, 2109*(September), 02846. doi:10.48550/arXiv.2109.02846

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys, 55*(9), Art 195. doi:10.1145/3560815

Liu, X., Shin, H., & Burns, A. C. (2021). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business research, 125*(March), 815-826. doi:10.1016/j.jbusres.2019.04.042

Maulud, D. H., Zeebaree, S. R., Jacksi, K., Sadeeq, M. A., & Sharif, K. H. (2021). State of art for semantic analysis of natural language processing. *Qubahan Academic Journal, 1*(2), 21-28. doi:10.48161/qaj.v1n2a44

Mich, L., & Garigliano, R. (2023). ChatGPT for e-Tourism: a technological perspective. *Information Technology & Tourism*, 1-12.

Müller, M., Salathé, M., & Kummervold, P. E. (2023). Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. *Frontiers in Artificial Intelligence, 6*, 102381. doi:10.3389/frai.2023.1023281

Orden-Mejia, M., & Huertas, A. (2022). Analysis of the attributes of smart tourism technologies in destination chatbots that influence tourist satisfaction. *Current Issues in Tourism, 25*(17), 2854-2869.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. *International Journal of Contemporary Hospitality Management, 32*(10), 3199-3226.

Tuo, Y., Ning, L., & Zhu, A. (2021). How artificial intelligence will change the future of tourism industry: The practice in China. In *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference, January 19–22, 2021* (pp. 83-94). Springer International Publishing.

Van Nguyen, M., Lai, V. D., Veyseh, A. P., & Nguyen, T. H. (2021). Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. *arXiv, 2101*(September), 0329. doi:10.48550/arXiv.2101.03289

Zhao, L., Alhoshan, W., Ferrari, A., Letsholo, K. J., Ajagbe, M. A., Chioasca, E.-V., & Batista-Navarro, R. T. (2021). Natural language processing for requirements engineering: A systematic mapping study. *ACM Computing Surveys, 54*(3), 1-41. doi:10.1145/3444689

Faisal N Alhajri
Deakin University, Australia
f155@live.com
ORCID 0009-0001-0339-5824