# Natural Language Processing in Advertising – A Systematic Literature Review

Vinh Truong
*RMIT University*
702 Nguyen Van Linh, Tan Phong, District 7, HCMC, Vietnam
vinh.truongnguyenxuan@rmit.edu.vn

*Abstract*—Computational or programmatic advertising is the new way to advertise products and services online and in real-time. In this emerging type of advertising, Natural language processing (NLP) is a powerful tool for intelligently targeting and placing advertisements at the right time and in the right place for the right audience in a very short period. This study systematically reviewed journal articles, book chapters, and conference proceedings for the last ten years to find out what are the uses, approaches, and challenges that the researchers have been recently facing in making use of natural language processing techniques in the domain of advertising. It is found that in the majority of studies, information extraction and sentiment analysis are still the main focus areas. Only a small number of advanced artificial intelligence (AI) techniques, such as deep learning and speech synthesis, are used. In addition, most of the studies are based on traditional forms of advertising (such as search engines, websites, and job listings), excluding the newer forms of mobile and app-based advertising. The ongoing challenge in the current literature is applying natural language processing to automatically target advertisements.

*Keywords—natural language processing, natural language, NPL, advertising, advertisement, ads, literature review, survey, systematic literature review*

## I. INTRODUCTION

Natural Language Processing (NLP) is a set of techniques used to extract grammatical structure and meaning from input to perform an actionable task. Theories and methods used in NLP enable effective communication in natural language between humans and computers. To achieve this goal, NLP uses techniques from many other domains, such as computer science, linguistics, and mathematics [2].

There have been four distinct phases in the development of NLP: the germination phase before 1956, the rapid development phase from 1957 to 1970, the low-speed development phase from 1971 to 1993, and the recovery phase from 1994 to the present day [1]. First, NLP was very useful in the fields of automatic answering systems and duplicate detection as well as computer-assisted tutors and database interfaces. Later, the applications of natural language processing have expanded to include a variety of fields of study such as machine translation and neural networks [3].

In the literature, related works to NLP can be divided into three primary categories: information extraction, semantic analysis, and deep learning [4]. Earlier research on information extraction methods concentrated on extracting text segments, which were then used for feature extraction and matching using common information retrieval metrics [5]. Semantic analyses are based on the idea that a description can be provided in a variety of ways, relying on implicit knowledge of how different terms relate to one another. The input data can be encoded in taxonomies, which map the relationships between different terms and can then be used when displaying outputs [6]. Deep learning methods rely heavily on annotated datasets to produce better results, in contrast to the data-agnostic methods, in which datasets were primarily used as expert evaluation sets to compare method effectiveness [1].

Nowadays, NLP has been used in a variety of fields due to its evolving methods and techniques. Natural language processing interests include speech recognition, coaching, applied mathematics, and machine learning. In terms of application, customer service and chatbots are likely to be the most commonly used because they are directly involved with human-computer interaction [7]. Besides, natural language processing techniques are widely used in market research and intelligence. The rise of ad networks recently (for example, Google Ads and Facebook Audience Network) has necessitated programmatic advertising research. On Facebook, 4.11M people are interested in "Natural language processing". In 2020, the average CPC for the natural language processing interest is $1.18, and the average CPM is $24.51 [8]. Companies use natural language processing techniques to match advertisement content and search queries to increase click-through rates. The fact that 70 per cent of advertisements are now delivered through automated systems demonstrates the significance of big data analyses in advertising [9].

The use of natural language in advertisements makes the NLP application an effective approach. Even though NLP is widely used in advertising in practice, it is not at that level of attention in theory [5]. It is intriguing to learn about the direction of advertising research that employs NLP methods and techniques as a new research area. The primary goal of this study, therefore, is to survey the most recent applications of NLP in advertising. To achieve these goals, the following research questions (RQ) were accordingly formulated:

RQ1: What are the uses of NLP in advertising?

RQ2: What NLP techniques are available in advertising?

RQ3: What are the challenges of employing NLP techniques in advertising?

The remainder of this paper's contents is organized as follows. Section 2 summarizes the related works – other systematic literature reviews. It includes those of natural language processing in related fields, such as user stories and radiology. It also includes those in other Artificial Intelligence areas, such as text mining and speech synthesis. Section 3 describes the study's methodology, including the sampling and data collection techniques. Section 4 is devoted to the data analysis results. The key findings of this study are discussed in Section 5. That section also discusses the study's implications and limitations. Section 6 summarized the paper's main points in its concluding remarks.

## II. RELATED WORKS

Advertising is a way of introducing a company's product/service to potential consumers to encourage them to purchase [10]. Advertising is therefore essential for the firm. A shift is inherent in the business, and one such change is the growth of advertising tactics. Advertising is now done with the assistance of a long-standing ally, Artificial Intelligence and Machine Learning [1]. The main issue for this new kind of advertising is ensuring that it is not utilized in such a way that the typical person is concerned about his or her privacy [11]. Advertising will become even more effective in the future as Artificial Intelligence and Machine Learning provide firms with greater control over their advertising campaigns [1]. According to many authors, that so-called programmatic advertising will be the pillar of automation and the only solution on which marketers will rely in the future [3]. However, the capabilities of programmatic advertising appear to have not yet reached their full potential, and we will see the channels and platforms for ad delivery evolve dramatically over the next few years [10]. As a result, it is critical to continue evaluating and accumulating knowledge about what makes new media advertising effective.

According to previous reviews (e.g. [1]), AI has been used in advertising for a long time, but primarily to automate the process of bidding on ads. Publishers are the ones who make ad spaces available for bidding. On the demand side, advertisers are those who purchase them. AI was used in that process to automate all relevant bidding tasks. Previous AI studies in advertising appear to have focused solely on this process, discussing techniques such as automation, robotics, and machine learning [1]. AI, however, includes also machine vision and natural language processing techniques[12]. To the best of the researcher's knowledge, no systematic literature review on NLP in advertising has been conducted, even though there have been several studies on Natural Language Processing in other related fields. Having said that, although not fully explored in advertising, natural language processing techniques have been used widely in a variety of other areas, as evidenced by several systematic literature reviews [4, 13].

Natural language processing, for example, is gaining traction in management due to its capacity to automatically analyze and interpret human language [13]. The researchers [13] examined articles from the University of Texas at Dallas list of 24 leading business journals that utilize NLP as their primary analytical approach to demonstrate how textual data may be used to enhance management ideas across various disciplines. They discussed the different toolkits and processes for utilizing NLP as an analytical approach. To guide future studies, the authors of this systematic literature review pointed out several managerial and technological obstacles connected with the use of NLP in management to be further explored and overcome [13].

Similarly, the authors in the study [4] investigated how NLP benefits radiology by investigating the various applications for which it can be used. A comprehensive examination of the literature turned up 67 papers presenting NLP approaches that may be used in practical radiology applications. Individual studies were evaluated based on their tasks, NLP techniques and tools utilized, application goals, and performance outcomes. Natural language processing methods help convert text into a structured representation, allowing computers to infer meaning from human input. NLP methods, when applied to radiology records, allow for the automated identification and extraction of information [4]. This systematic literature review about NLP in radiology highlighted the importance of NLP in radiology and how NLP could even improve its accuracy and speed when reading medical records.

The authors of [14] discussed the changing landscape of marketing, particularly digital marketing. It then goes on to discuss the various sectors of digital marketing where Natural Language Processing has made a play. In the last decade, there has been rapid advancement in information technology and its applications. This has influenced how we perceive the world as well as how business is conducted. Today's emerging technologies, such as data analytics, mobile devices, automation, and Natural Language Processing (NLP), are fundamentally altering how society and individuals interact. Marketing is a microcosm of how new software solutions, machine learning, and big-data analytics are altering the game [14]. This systematic literature review about NLP in marketing has not only reported the current uses of NLPs in marketing but also their current challenges.

The authors of another study [5] identified existing clinical NLP systems that create structured information from unstructured free text using a systematic method based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. During two screening rounds, the researchers reviewed all records for relevance, and information regarding clinical NLP systems was gleaned from the final batch of articles. NLP systems were identified to be employed in a variety of critical clinical and scientific activities. Certain tasks, such as temporal information extraction or concept normalization to standard terminologies, are well addressed by current systems, whilst others, such as concept normalization to standard terminologies, are still open for future research. That systematic literature review identified several NLP systems capable of processing clinical free text which is valuable in prioritizing the development of new clinical NLP techniques [5].

Another thorough literature review was done as part of a project to capture the current state-of-the-art of NLP research on user stories [7]. They discovered 38 main studies that explore NLP methods in user stories after using the inclusion and exclusion criteria together with the snowballing technique. It was found that a majority of research used NLP approaches to

extract who, what, and why information from user stories. In general, NLP research in user stories aims to uncover flaws, create software artifacts, identify important abstractions of user stories, and trace relationships between models and user stories. Practically, NLP can assist system analysts in managing user stories. However, implementing NLP in user stories provides a plethora of potential as well as problems. Therefore, this systematic literature review pointed out that it is important to study NLP approaches as well as rigorous assessment procedures to achieve high-quality research. Understanding the context of a sentence, like NLP research in general, was found to be a continuing challenge to be addressed in the future [7].

All of the systematic literature reviews mentioned above survey NLP studies, but in different areas other than advertising. In the field of advertising, its NLP techniques have not been systematically surveyed and its NLP challenges have not been thoroughly discussed, which leaves a big gap for this study to fulfil.

## III. METHODOLOGY

The purpose of this study was to conduct a survey of NLP techniques in advertising literature. Accordingly, it decided on a search strategy together with inclusion and exclusion criteria. The study is producing a systematic literature review, which will be divided into three stages: planning, conducting, and reporting. As a guide, this SLR utilizes the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist, a common guideline for systematic literature reviewing [15].

### A. Uses of NLP for Advertisements

By selecting keywords, constructing a search string, and defining a database and search criteria, this study was able to acquire relevant publications. The keywords were picked with the aims and research topics in mind, especially the applications, techniques, and challenges of utilizing natural language processing (NLP) in advertising. Based on the research questions, it identified two major categories for determining keywords which are "natural language processing" and "advertising." To obtain comprehensive results, this study identified alternative spellings and synonyms too. The final set of keywords is listed in Table 1.

TABLE I. SEARCH KEYWORDS

| Category | Keywords |
|---|---|
| Natural Language Processing | Natural Language Processing, Natural Language, NPL |
| Advertising | Advertising, Advertisement, Ad |

The set of keywords was then linked using Boolean operators. This study restricted the publication period to the previous 10 years to get the most recent state-of-the-art ones. The relevant studies were chosen using the specified inclusion and exclusion criteria. To be included, a study must be peer-reviewed, in English, published between 2011 and 2022, and relevant to NLP in advertisements. A study is excluded if it is a summary article, patents, a book, a summary of conference keynotes, proposals, lecture notes, editorials, comments, tutorials, or review papers.

For the first screening, this study employed abstracts, titles, and keywords to evaluate publications based on the inclusion and exclusion criteria stated above. After that, the researcher retrieved the entire text of relevant papers to re-evaluate the inclusion and exclusion criteria. The researcher excluded the studies that did not match the requirements. The duplicate studies were removed. Finally, in this analysis, the snowballing approach was utilized to acquire more complete results and decrease the chance of missing important studies.

### B. Conducting the Review

This study used specified search phrases to search the following online libraries: SCOPUS, Google Scholar, SpringerLink, ScienceDirect, ACM Digital Library, and IEEE Xplore.
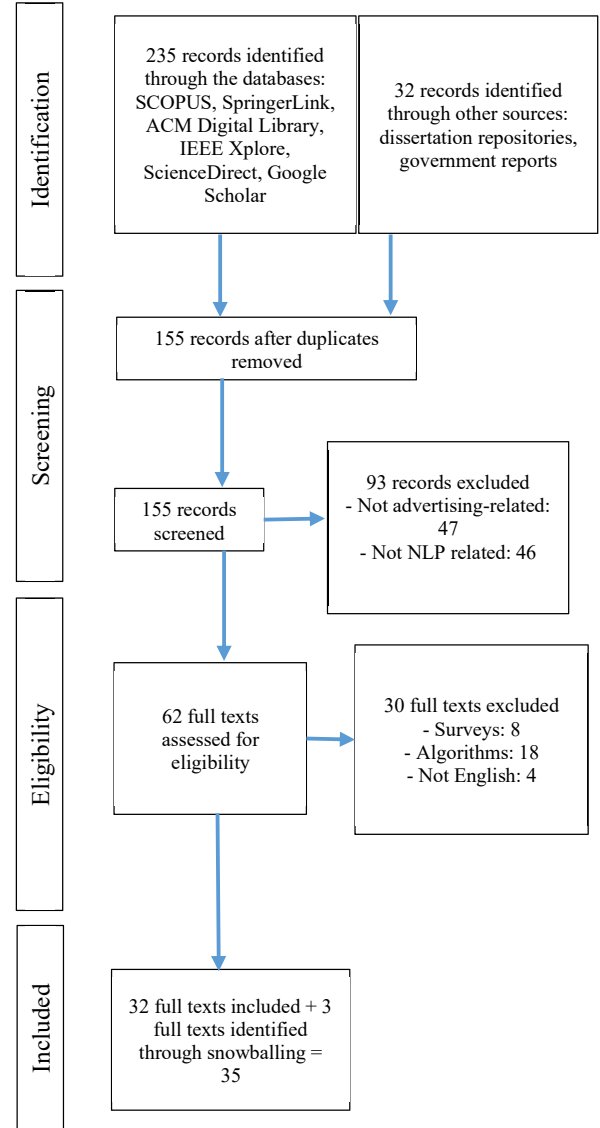


Fig. 1. PRIMA flow chart – Literature review process.

To optimize the efficacy of the search, the researcher ran it consecutively on electronic databases. It all began with

Authorized licensed use limited to: University of London: Online Library. Downloaded on June 10,2024 at 12:57:34 UTC from IEEE Xplore. Restrictions apply.

SCOPUS and Google Scholar searches. SCOPUS and Google Scholar searches were also included in the digital library category; hence, in the event of redundancy, the paper that appears may be quickly identified, minimizing efforts to handle redundant publications. The search results were sorted chronologically based on those discovered in the aforementioned databases. The screening was done on the titles, abstracts, and keywords. Relevant papers were highlighted on a spreadsheet before being downloaded and integrated into the Mendeley program.

Initially, this technique yielded a total of 235 relevant studies. For eligibility, the complete text of the research was examined. This was achieved by going over the inclusion and exclusion criteria again and ensuring that the item was suitable for the review.

After identifying the additional studies, the backward and forward snowballing strategies were utilized. That means it shifted through a reference list to locate studies that were related to backward snowballing. To assess whether the studies were relevant, the title of the reference was first examined. The whole text of the research candidates is then retrieved to further analyze its content.

The study search and selection process is depicted in Fig. 1.

*C. Reporting the Review*

The review's results were grouped and presented according to their RQs as in Section 4. Furthermore, the characteristics of all selected studies are also summarized and reported in Appendix. By strictly following the PRISMA framework, this review has identified all related studies and accordingly selected the qualified ones. The outcomes from reading and examining those manuscripts truly reflect the current context of NLP uses, techniques, and challenges in advertising, opening the door for future research. Those will be accordingly reported in Section 4 and discussed in Section 5.

## IV. RESULTS

This study identified 35 primary studies using the mentioned literature review methodology. Twelve (34.3%) of the studies were published in journals, twenty-one (60.0%) in conferences, and two (5.7%) as book chapters. The studies were distributed both in journal and conference venues, indicating that the authors did not favour a single source.

Preliminary studies made up nearly half of the total. Fourteen studies (40.0 percent) expressed ideas and provided proof of concept in the form of experimentation or case studies. Twenty-one studies (60.0 percent) conducted their research in an in-lab setting. There were no studies conducted in industrial settings. Several researchers, on the other hand, used real industry datasets in their research. As shown in the Appendix, the number of publications per year is steadily increasing. There has been a significant increase in the number of publications since 2019.

*A. Uses of NLP for advertisements*

The review found a variety of natural language processing and natural language generating usages in advertising, which include:

- reading ads with natural language processing (NLP) techniques
- creating ads with natural language generating (NLG) techniques
- targeting ads with deep learning (DL) techniques

These are the three major problems that drew the attention of the majority of the researchers. These topics have been researched on an ongoing basis since 2011. Because researchers are still attempting to comprehend a novel and distinct feature of advertisements, the topic of reading advertisements was chosen as the primary focus in the early stages [6]. Because it can improve advertising effectiveness, the topic of ad targeting is always a challenge in advertising research [2]. Less attention is paid, however, to the creation of advertisements, which should be the subject of future research [16]. NLP techniques seem to do a good job of reading content but not writing content, according to the result of this survey.

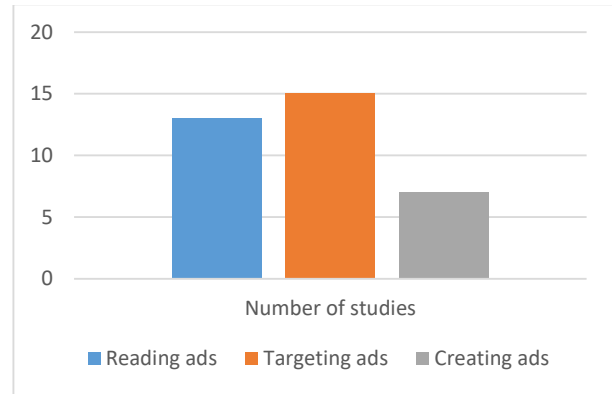Figure 2 depicts the number of studies conducted in each application.



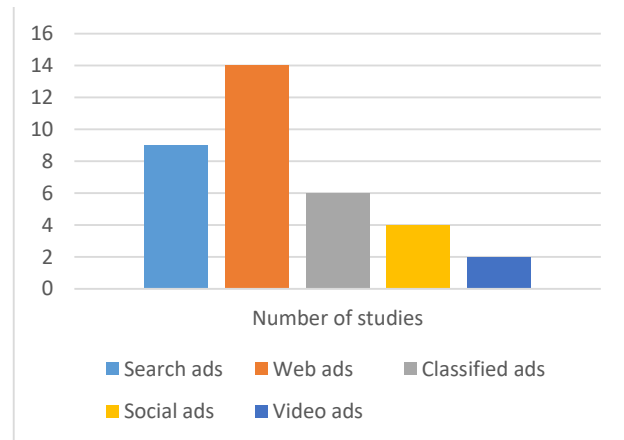Fig. 2. NLP uses in advertising studies.



Figure 3. NLP applications in current advertising studies

Furthermore, NLP techniques are widely used in search, social, web, and classified ads, as shown in Figure 3. Those kinds of advertisements are quite traditional [10]. NLP research is scarce in emerging types like video, mobile, and apps. It

revealed that mobile in-app advertising now accounts for nearly 70% of all spending. More research should be conducted on such novel advertising formats.

### B. NLP Approaches in Advertising

To address RQ2, this study separated the methodologies available in research on NLP in advertising into two categories: NLP techniques and NLP tools. This study looked at the different NLP approaches that have been reported in primary research, such as information extraction, semantic analysis, and deep learning [17].

In terms of NLP tools, there is a bulk of research utilizing SpaCy or Stanford NER to conduct NLP. Some claimed to have utilized WordNet and the Stanford POS tagger, while others claimed not to have any of those. In several research, more than one tool was utilized [16].

In terms of NLP techniques, Figure 4 shows that the sentiment analysis approach was used in the majority of the studies. That is, advertisements are not only a source of information but also a place to analyze opinions and sentiments. Some studies have also looked into advanced text mining and deep learning techniques. Their goal is to create advertisements rather than simply read them. Creating ads required advanced NLP techniques of deep learning, and that seems to be less focus by researchers at this point as shown previously.
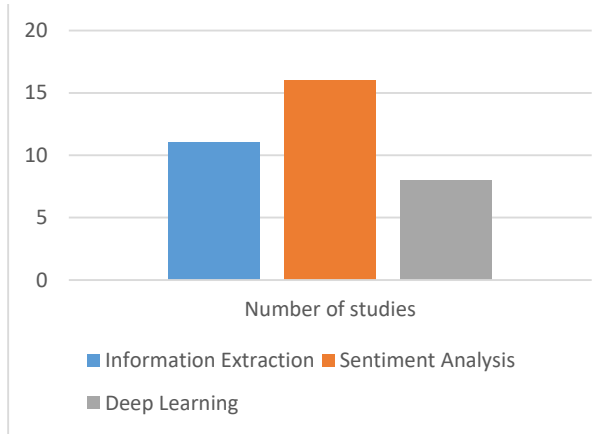


Fig. 4. NLP approaches in current advertising studies.

### C. Challenges of Using NLP in Advertising Research

Several difficulties were reported in the primary studies. Some were related to recall and precision improvement, dataset, pragmatic analysis, and multi-channel as shown in Figure 5.

Even though the recall and precision findings were within the predicted ranges, the researchers noted that the accuracy results were not as expected throughout the examination. A study [18] found persistent recall levels of more than 68%, although the average accuracy value remained in the 60s. They even got extremely low precision values [6]. The researchers concur that getting a high-precision number remains difficult [19].

Datasets confront several problems, including heterogeneity, a small volume of data, and human data labelling. Due to the limited number of freely available advertising datasets,

obtaining significant volumes of advertising data is challenging. The data's minimal heterogeneity is an issue [16]. For their objectives, researchers generally gather advertising datasets on their own. When evaluating advertising using machine learning algorithms, the problem of heterogeneity and a lack of data has become a concern. Another difficulty is collecting reliable and large data. To achieve general findings, a very big dataset is necessary [20]. Most research continues to concentrate on specific data without attempting to reproduce it in complicated systems or real-world applications [21].

The pragmatic analysis is a step in the extraction of information from text. It is the stage that focuses on taking a set of text structures and determining what the actual meaning was. It comes from the field of linguistics, where the context of the text is taken into account. Understanding proper sentence interpretation remains difficult. Some of the difficulties include difficult-to-identify compounds, verbs that are difficult to link to the appropriate object, and conjunctions [21].

In NLP research on advertising, there is still an ongoing challenge to interpreting text, image, and even video simultaneously. Many studies found it difficult to apply NLP together with other processing techniques [16]. This kind of advertisement seems to get more popular these days when users are consuming different types of media in their daily internet use. NLP is just one domain of AI, and it needs to coordinate with other domains to deliver a better advertising experience to its users [12]. An integrated approach is needed in these cases.
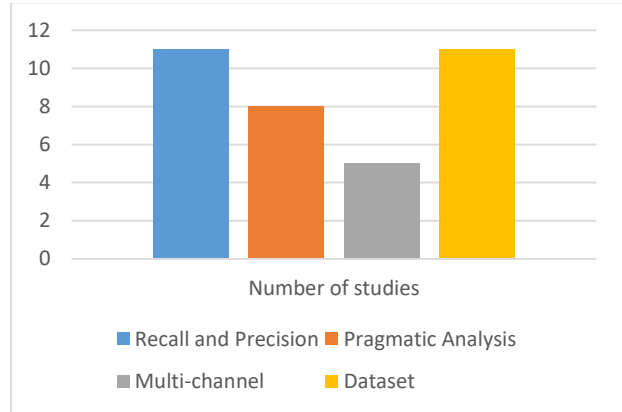


Fig. 5. NLP challenges in current advertising studies.

### V. DISCUSSION

Several conclusions may be drawn from the results provided above. This section explained the significance of the findings, their ramifications, and their limitations.

### A. Key Findings

According to the review, current literature still focuses on advertisements themselves, rather than advertising as a process. The majority of NLP research on advertisements continues to focus on information extraction and sentiment analysis. The recognition of verbs and nouns is the foundation for reading advertisements, but not for targeting and creating them. Future research should look more into the areas of using NLP for programming ads.

Information extraction and semantic analysis are still the dominant techniques in current advertising studies. Meanwhile, to discover the context in which to target advertisements, deep learning techniques should be required to achieve its goals. As a result, the generation of advertisements from free text has received little attention. Deep learning techniques could also help the advertiser to work with difficult-to-understand free text features and a language structure that requires more in-depth analysis. In the future, advertisers can use deep learning to create more personalized experiences, target the right audience, and identify the right thought leaders and influencers in a broader context.

The review also discovered that the majority of the literature continues to focus on search and web advertisements while ignoring emerging formats such as video, mobile, and apps. Mobile advertising is a kind of internet advertising that specifically targets users of mobile phones and other mobile devices. This market is projected to develop shortly. In 2019, global mobile ad expenditure is anticipated to reach about 234 billion US dollars. This number is anticipated to grow to 495 billion US dollars by 2024. With a market value of over 150 million US dollars projected in 2023, mobile in-app advertising is the leading channel for businesses [9]. Future research should concentrate on these novel forms of advertising.

### B. Contributions

Theoretically, this study provided a thorough review of NLP in the advertising literature. In comparison to other areas, NLP in advertising has received little attention from researchers. This study is one of the first attempts to dig into this promising area. Practically, this review identified several key areas for future research. This paper summarizes some common techniques and tools that may be useful for practitioners to continue researching in related fields.

### C. Limitations

Some articles might have gone unnoticed, resulting in an incomplete set of results. This study followed a predetermined methodology, did a comprehensive search, and accessed numerous databases to decrease this risk. To get a thorough main study, the research search and selection procedure additionally employed forward and backward snowballing. Even though the risk has been very much mitigated, the chance of missing articles is still there.

This study simply examined each database based on the title, abstract, and keywords, which may have influenced publications on the list that were irrelevant, relevant, or unrelated. This study covered the phases of full-text publications examined for eligibility to avoid irrelevant studies. The danger of untracked relevant publications was acknowledged in this review since the fundamental context of the work should be present in the title, abstract, and keywords, except for those whose title, abstract and keywords do not follow the rule.

## VI. Conclusions

This study conducted a review of NLP implementation in advertising. The first search generated 235 studies after applying the inclusion and exclusion criteria, of which 32 were chosen as primary studies. After utilizing forward and backward snowballing, this study augmented this count with several additional studies. The research then evaluated those studies' full texts. The whole process strictly follows the PRISMA framework.

According to the review, the number of current studies on Natural Language Processing in advertising is low. It is found that in the majority of studies, information extraction and sentiment analysis are still the main focus areas. Only a small number of advanced artificial intelligence (AI) techniques, such as deep learning and speech synthesis, are used. In addition, most of the studies are based on traditional forms of advertising (such as search engines, websites, and job listings), excluding the newer forms of mobile and app-based advertising. The ongoing challenge in the current literature is how to apply natural language processing to targeting advertisements programmatically and creating them personally.

According to the review, this research field is still in its infancy and requires further investigation. This study is one of the first to dig into this area. The findings from such studies could apply to other areas, and bring benefits to many stakeholders.

### References

[1] Shah, N.; Engineer, S.; Bhagat, N.; Chauhan, H.; Shah, M. Research Trends on the Usage of Machine Learning and Artificial Intelligence in Advertising. Augmented Human Research 2020, 5.

[2] Qiu, X.P.; Sun, T.X.; Xu, Y.G.; Shao, Y.F.; Dai, N.; Huang, X.J. Pre-Trained Models for Natural Language Processing: A Survey. Science China Technological Sciences 2020, 63, 1872–1897.

[3] Qiu, X.P.; Sun, T.X.; Xu, Y.G.; Shao, Y.F.; Dai, N.; Huang, X.J. Pre-Trained Models for Natural Language Processing: A Survey. Science China Technological Sciences 2020, 63, 1872–1897.

[4] E., P.; LM., B.; MG., H.; JA., K. Natural Language Processing in Radiology: A Systematic Review. Radiology 2016, 279, 329–343.

[5] Kreimeyer, K.; Foster, M.; Pandey, A.; Arya, N.; Halford, G.; Jones, S.F.; Forshee, R.; Walderhaug, M.; Botsis, T. Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review. Journal of Biomedical Informatics 2017, 73, 14–29.

[6] Thun, A. Matching Job Applicants to Free Text Job Ads Using Semantic Networks and Natural Language Inference 2020.

[7] Raharjana, I.K.; Siahaan, D.; Fatichah, C. User Stories and Natural Language Processing: A Systematic Literature Review. IEEE Access 2021, 9, 53811–53826.

[8] Adcostlly Natural Language Processing Facebook Ad Cost Report Available online: https://adcostly.com/facebook-ads-cost/natural-language-processing.

[9] Statista Programmatic Advertising Worldwide - Statistics & Facts Available online: https://www.statista.com/topics/2498/programmatic-advertising/.

[10] Shelly, R.; Esther, T. Digital Advertising: Theory and Research; Taylor and Francis, 2017;

[11] Fang, Y.; Liu, J. A Novel Prior-Based Real-Time Click through Rate Prediction Model. International Journal of Machine Learning and Cybernetics 2014, 5, 887–895, doi:10.1007/s13042-014-0231-7.

[12] IBM How AI Is Changing Advertising Available online: https://www.ibm.com/watson-advertising/thought-leadership/how-ai-is-changing-advertising.

[13] Kang, Y.; Cai, Z.; Tan, C.W.; Huang, Q.; Liu, H. Natural Language Processing (NLP) in Management Research: A Literature Review. Journal of Management Analytics 2020, 7, 139–172.

[14] Verma, S.; Sharma, R.; Deb, S.; Maitra, D. Artificial Intelligence in Marketing: Systematic Review and Future Research Direction.

International Journal of Information Management Data Insights 2021, 1, 100002.

[15] Tricco AC; Lillie E; Zarin w; O'Brian KK; Colquhoun H; Levac D; Moher D PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine 2018, 169, 467–473.

[16] Hernández-Álvarez, M.; Granizo, S.L. Detection of Human Trafficking Ads in Twitter Using Natural Language Processing and Image Processing. Advances in Intelligent Systems and Computing 2021, 1213 AISC, 77–83.

[17] Hu, Y.; Mao, H.; McKenzie, G. A Natural Language Processing and Geospatial Clustering Framework for Harvesting Local Place Names from Geotagged Housing Advertisements. International Journal of Geographical Information Science 2019, 33, 714–738, doi:10.1080/13658816.2018.1458986.

[18] della Volpe, M.; Esposito, F. The Data Scientist on LinkedIn: Job Advertisement Corpus Processing with NooJ. Communications in Computer and Information Science 2020, 1153 CCIS, 79–90.

[19] Zhang, Y.; Ono, J.; Ogata, T. An Advertising Rhetorical Mechanism for Single Event Combined with Conceptual Dictionary in Narrative Generation System. In Proceedings of the NLP-KE 2011 - Proceedings of the 7th International Conference on Natural Language Processing and Knowledge Engineering; 2011; pp. 340–343.

[20] Katsumata, S.; Motohashi, E.; Nishimoto, A.; Toyosawa, E. The Contents-Based Website Classification for the Internet Advertising Planning: An Empirical Application of the Natural Language Analysis. The Review of Socionetwork Strategies 2017, 11, 129–142, doi:10.1007/s12626-017-0007-0.

[21] Karakatsanis, I.; AlKhader, W.; MacCrory, F.; Alibasic, A.; Omar, M.A.; Aung, Z.; Woon, W.L. Data Mining Approach to Monitoring the Requirements of the Job Market: A Case Study. Information Systems 2017, 65, 1–6., 1989.

APPENDIX

| No | Authors | Title | Year | Methodology | Use | Application | Technique | Tool | Challenge |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Hu Y., Mao H., and McKenzie G. | A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements | 2019 | modellingntal | Reading | Classified | Information Extraction | spaCy, Stanford NER | Recall and Precision |
| 2 | García-Sánchez F., Colomo-Palacios R., Valencia-García R. | A social-semantic recommender system for advertisements | 2020 | Experimental | Targeting | Social | Sentiment Analysis | Recommender systems | Dataset |
| 3 | Lee D., Hosanagar K., Nair H.S. | Advertising content and consumer engagement on social media: Evidence from Facebook | 2018 | Experimental | Targeting | Social | Information Extraction | AMT | Dataset |
| 4 | Zhang Y., Ono J., and Ogata T. | An advertising Rhetorical mechanism for a single event combined with a conceptual dictionary in a narrative generation system | 2011 | Conceptual | Creating | Web | Information Extraction | Goi-Taikei, WordNet | Recall and Precision |
| 5 | Amador-Domínguez E., Serrano E., Mateos-Nobre J.D., et al. | An intelligent and auto-adaptive system of virtual identities based on deep learning for the analysis of online advertising networks | 2019 | Conceptual | Targeting | Web | Deep Learning | VI | Dataset |
| 6 | Liu P., Azimi J., Zhang R. | Automatic Keywords generation for contextual advertising | 2014 | Experimental | Targeting | Web | Information Extraction | LDA | Recall and Precision |
| 7 | Esfahani modellingfarella M.J., Pouyan M.B., et al. | Context-specific language modeling for human trafficking detection from online advertisements | 2020 | Experimental | Reading | Social | Information Extraction | BERT | Recall and Precision |
| 8 | Karakatsanis I., AlKhader W., Ma Crory F., et al. | Data mining approach to monitoring the requirements of the job market: a case study | 2017 | Conceptual | Reading | Classified | Deep Learning | BeautifulSoup, NLTK, Gensim | Dataset |

| # | Authors | Title | Year | | | | | | |
|---|---------|-------|------|---|---|---|---|---|---|
| 9 | Di S. | Deep Interest Network for Taobao advertising data Click-Through Rate Prediction | 2021 | Experimental | Targeting | Search | Deep Learning | DIN | Dataset |
| 10 | Hernández-Álvarez M., Granizo S. L. | Detection of human trafficking ads in Twitter using natural language processing and image processing | 2021 | Experimental | Reading | Social | Deep Learning | Lexical normalization, TweetNorm_ES | Multimedia |
| 11 | González-Eras A., Aguilar J. | Determination of Professional Competencies Using an Alignment Algorithm of Academic Profiles and Job Advertisements, Based on Competence Thesauri and Similarity Measures | 2019 | Conceptual | Reading | Classified | Information Extraction | - | Recall and Precision |
| 12 | Su S., He S., Sun C., Zhang H., et al. | Do landscape amenities impact private housing rental prices? A hierarchical hedonic modeling approach based on semantic and sentimental analysis of online housing advertisements across five Chinese megacities | 2021 | Experimental | Reading | Classified | Sentiment Analysis | LDA | Dataset |
| 13 | Guerini M., Strapparava C., and Stock O. | Evaluation Metrics for Persuasive NLP with Google AdWords | 2010 | Conceptual | Targeting | Search | Sentiment Analysis | Evaluation Metrics (Google AdWords) | Dataset |
| 14 | Chakeri A., Lowe M. | From geolocation-based only to semantically-aware digital advertising: A neural embedding approach | 2019 | Experimental | Targeting | Web | Sentiment Analysis | SLL | Pragmatic analysis |
| 15 | Atıcı B., İlhan Omurca S. | Generating Classified Ad Product Image Titles with Image Captioning | 2021 | Experimental | Reading | Web | Sentiment Analysis | | Multimedia |
| 16 | Poornachandran P., Hrudya P., Reghunadh J. | Human vs. machine: Analyzing the robustness of advertisement based CAPTCHAs | 2014 | Conceptual | Creating | Web | Information Extraction | CAPTCHA | Recall and Precision |
| 17 | Burbano D., Hernandez-Alvarez M. | Illicit, Hidden Advertisements on Twitter | 2018 | Experimental | Reading | Social | Information Extraction | | Recall and Precision |
| 18 | Pryzant R., Basu S., and Sone K. | Interpretable Neural Architectures for Attributing an Ad's Performance to its Writing Style | 2019 | Conceptual | Creating | Search | Deep Learning | Selectors | Pragmatic analysis |
| 19 | Thun A. | Matching job applicants to free text job ads using | 2020 | Experimental | Targeting | Classified | Sentiment Analysis | NLI, Random | Recall and Precision |

| No | Author | Title | Year | Type | Stage | Platform | Method | Tool | Evaluation |
|---|---|---|---|---|---|---|---|---|---|
| | | semantic networks and natural language inference | | | | | | suitability classifier | |
| 20 | Vedula N., Sun W., Lee H., et al. | Multimodal content analysis for effective advertisements on Youtube | 2017 | Conceptual | Creating | Video | Deep Learning | | Multimedia |
| 21 | Admov A., Adali E. | Opinion Mining and Sentiment Analysis for Contextual Online Advertisement | 2017 | Conceptual | Reading | Web | Sentiment Analysis | - | Pragmatic analysis |
| 22 | Adamov A.Z., Adali E. | Opinion mining and Sentiment Analysis for contextual online-advertisement | 2017 | Experimental | Reading | Web | Sentiment Analysis | | Pragmatic analysis |
| 23 | Gauba H., Kumar P., Roy P.P., et al. | Prediction of advertisement prefer by fusing EEG response and sentiment analysis | 2017 | Experimental | Creating | Video | Sentiment Analysis | EEG | Multimedia |
| 24 | Qi W., Gong Y., Yan Y. et al. | ProphetNet-Ads: A Looking Ahead Strategy for Generative Retrieval Models in Sponsored Search Engine | 2020 | Experimental | Targeting | Search | Sentiment Analysis | ProphetNet-Ads | Pragmatic analysis |
| 25 | Singh M., Lamba R. | Proposing Contextually Relevant Advertisements for Online Videos | 2020 | Conceptual | Creating | Web | Deep Learning | - | Pragmatic analysis |
| 26 | Kabaso S., Ade-Ibijola A. | Sell-Bot: An Intelligent Tool for Advertisement Synthesis on Social Media | 2020 | Experimental | Creating | Social | Sentiment Analysis | Sell-Bot | Pragmatic analysis |
| 27 | Curtis J.R., Chen L., Higginbotham P., et al. | Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising | 2017 | Experimental | Reading | Social | Information Extraction | Treato | Recall and Precision |
| 28 | Jalal A.A. | Text mining: Design of interactive search engine based regular expressions of online automobile advertisements | 2020 | Conceptual | Targeting | Classified | Deep Learning | | Dataset |
| 29 | Katsumata S., Motohashi E., Nishimoto A., et al. | The contents-based website classification for the internet advertising planning - an empirical application of the natural language analysis | 2017 | Experimental | Targeting | Web | Sentiment Analysis | LDA, MCMC, MeCab | Dataset |
| 30 | della Volpe M., and Esposito F. | The Data Scientist on LinkedIn: Job Advertisement Corpus Processing with NooJ | 2020 | Experimental | Reading | Social | Deep Learning | LG, NooJ | Dataset |
| 31 | Yu T., Yang X., Jiang Y., et al. | TIRA in Baidu image advertising | 2021 | Experimental | Targeting | Web | Deep Learning | TIRA, ResNet, BERT | Multimedia |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 32 | Wu L., Dodoo N. A., Wen T. J., et al. | Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing | 2021 | Experimental | Readings | Social | Sentiment Analysis | LDA, Mallet, VADER | Recall and Precision |
| 33 | Konapure R.C., Lobo L.M.R.J. | Video Content-Based Advertisement Recommendation System using Classification Technique of Machine Learning | 2021 | Conceptual | Targeting | Web | Deep Learning | - | Pragmatic analysis |
| 34 | Guo L., Ye H., Su W., Liu H., et al. | Visualizing and understanding deep neural networks in CTR prediction | 2018 | Conceptual | Targeting | Web | Deep Learning | | Dataset |
| 35 | Katsumata S., Motohashi E., Nishimoto A., Toyosawa E. | Website Classification Using Latent Dirichlet Allocation and Its Application for Internet Advertising | 2016 | Conceptual | Targeting | Web | Information Extraction | LDA | Recall and Precision |