Year 2023 in Biomedical Natural Language Processing: a Tribute to Large Language Models and Generative AI

Cyril Grouin[1], Natalia Grabar[2]
Contact: cyril.grouin@lisn.upsaclay.fr, natalia.grabar@univ-lille.fr

1 Université Paris Saclay, CNRS, LISN, 91400 Orsay, France
2 UMR8163 STL, CNRS, Université de Lille, Domaine du Pont-de-bois, 59653 Villeneuve-d'Ascq cedex, France

Summary

Objectives: This synopsis gives insights into scientific publications from 2023 in Natural Language Processing for the biomedical domain. We present the process we followed to identify candidates for NLP's best papers and the two best papers of this year. We also analyze the current trends in the 2023 publications.

Methods: We queried two bibliographic databases (Medline and the ACL anthology) and refined the outputs through automatic scoring. We then manually shortlisted publications to review and selected candidate papers through an adjudication process. External reviewers assessed the interest of the 13 selected candidates. At last, the section editors chose the best NLP papers.

Results: We collected 2,148 papers published in 2023, of which two were the best and selected as part of this NLP synopsis. Both address language models and propose solutions for data augmenta-tion, domain-specific model adaptation, and model distillation. Work is done on social media con-tent and electronic health records, using deep learning approaches such as ChatGPT and large lan-guage models.

Conclusion: Trends from 2023 cover classical NLP tasks (information extraction, text categoriza-tion, sentiment anal-ysis), existing topics from several years (medical education), mainstream applications (Chatbots, generative approaches), and specific issues (cancer, COVID-19, mental health). Specifically for COVID-19, current researches deal with post-COVID-19 conditions, and they explore the understanding of how this pandemic has been managed and welcomed by populations. In addition, due to language models, a few works have been done to process languages other than English, especially using language porta-bility approaches.

Keywords

Biomedical Natural Language Processing; Topics; Issues; Best Papers; 2023.

Figure 1. Query used on PubMed for collecting candidate publications for review.

liographic databases: Medline[1], specifically dedicated to the biomedical domain, and the ACL anthology[2], a database that brings together the major NLP conferences (ACL, COLING, EMNLP, LREC, NAACL, etc.) and journals, since some NLP studies concerning the biomedical domain are published in ACL conferences and journals that PubMed does not index.

The query we used (Figure 1) is voluntari-ly basic to bring back the maximum number of candidates. It consists of searching for all papers published in English in 2023, hav-ing an abstract, and indexed with 75 terms of clinical language processing, medical language processing, or natural language processing. As of 2023 January 17th, we col-lected 2,114 entries, much more than the two last years (1,204 entries in 2021 and 1,670 entries in 2022). We applied a similar query to the ACL anthology database and collected 34 additional entries. To refine the selection of the candidate's best papers, we defined an automatic scoring for each candidate based on five criteria (journal name, objectives, methods/corpus/resources used in the paper,

1. Introduction

Natural Language Processing (NLP) aims to provide methods, tools, and resources designed to mine textual and narrative doc-uments and enable access to the information they convey [1]. While human languages are complex (for example,learning a human language requires many years for a human to be fluent), the importance of using NLP approaches to mine documents produced by humans has been pointed out for a long time [2]. In this synopsis, we present the selection process applied this year and then analyze some publications' content.

2. Best papers selection process

2.1. Automatic extraction and scoring

2.1.1. Description

To identify the papers published during the year 2023 in the field of NLP for the biomedical domain, we queried two bib-

---

1 https://pubmed.ncbi.nlm.nih.gov/
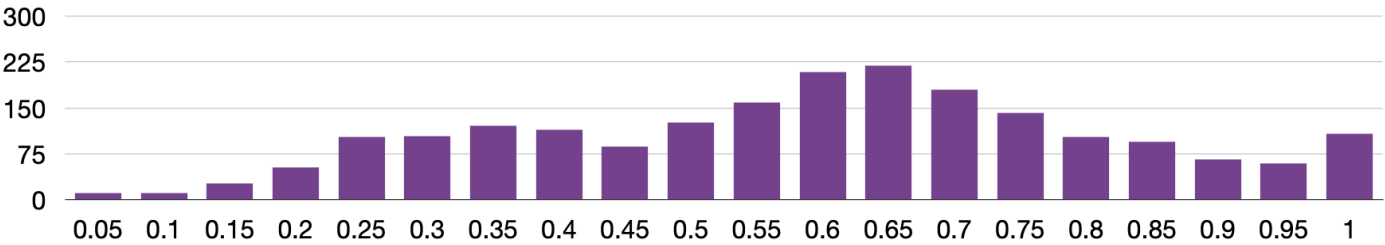2 https://www.aclweb.org/anthology/

Figure 2. Distribution of papers according to the filter scores.

evaluation/metrics used, particular concepts or key phrases used in the abstract). We aim to identify scientific papers valuable for the NLP community and discard medical papers that apply existing NLP methods without any valuable improvement. For example, the "language" concept may highlight the object study, which concerns NLP and computer science, or a cognitive aspect with language disabilities, which concerns neuroscience and psychiatry.

We also excluded yearbooks and survey papers since they do not fit the criteria for best paper candidates. To focus on original contributions, we gave a lower score to abstracts that specifically mention phrases like "using natural language processing" or "perform a natural language processing analysis" since the NLP dimension is not central in those submissions. Consequently, those papers are not good candidates for the best papers in NLP. For each of the 2,148 candidate papers, the final score ranged from 0 to 1 (Figure 2).

2.1.2. Analysis

We present in Figure 3 the evolution of the proportion of the first author's affiliation country from 2011 to 2023 in the PubMed export as provided by our query (Figure 1), limited to the ten countries best represented in the publications (U.S., Canada, several Western European countries, India, and Japan). Note that the y-axis is logarithmic to make visible countries with a low proportion of first authors.

Table 1 presents the number of published papers per year for each country of affiliation of the first author, based on the output of the PubMed query we defined previously (Figure 1).

The main findings of this evolution are the following:

- Each country generally publishes more papers each year, and those papers are much more indexed in PubMed (Table 1). This is not visible in the figure since the
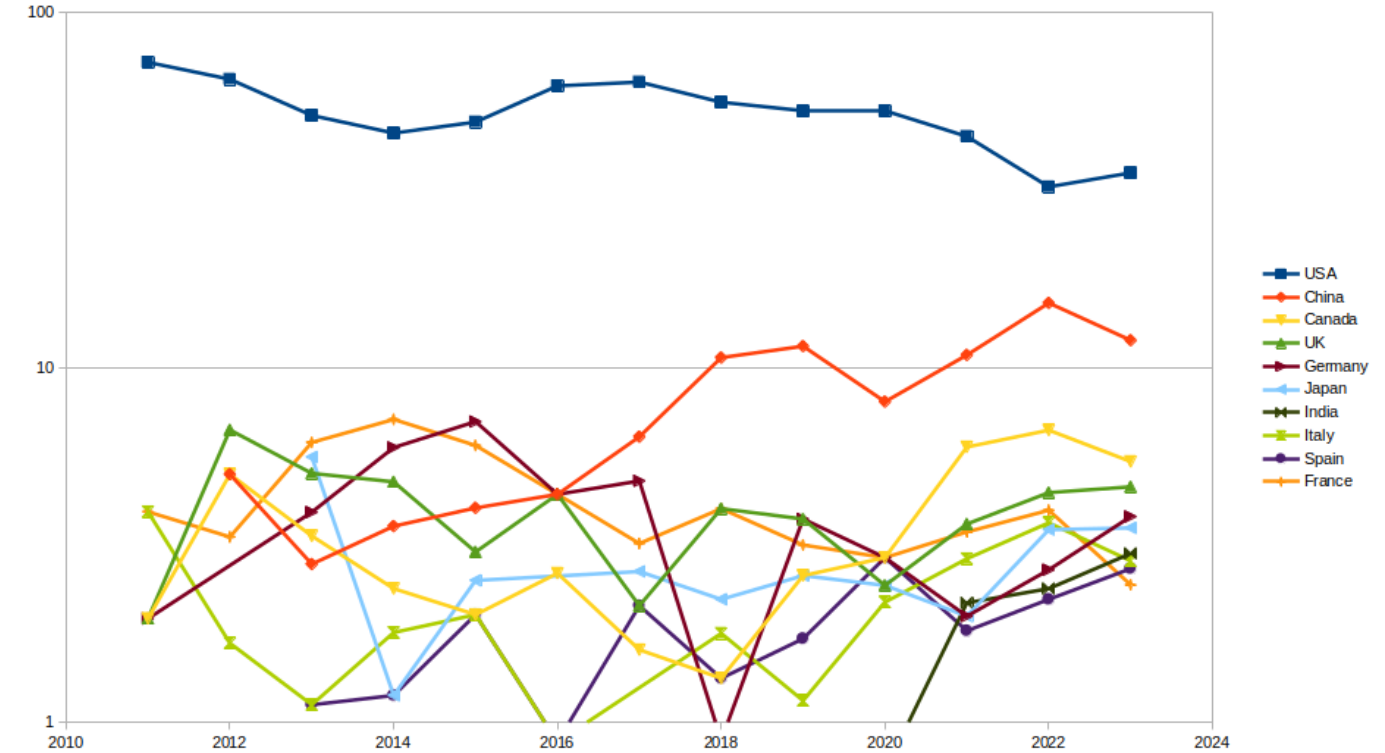

Figure 3. Top-10 first author's affiliation country from 2011 to 2023 in the PubMed extractions sorted by decreasing percentages of publications F(logarithmic scale).

Table 1. Number of papers collected in the PubMed database from 2013 to 2023 for each country based on the first author's affiliation country.

Country/Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Canada	6	4	4	3	3	3	9	12	33	42	40
China	5	6	8	5	12	24	40	34	60	96	88
France	11	12	12	5	6	9	11	12	19	25	18
Germany	7	10	14	5	9	2	13	12	11	17	28
India	0	0	0	0	0	1	2	3	12	15	22
Italy	2	3	4	1	0	4	4	9	16	23	21
Japan	10	2	5	0	5	5	9	10	11	22	26
Spain	2	2	4	1	4	3	6	12	10	14	20
U.K.	9	8	6	5	4	9	13	10	20	28	34
U.S.	92	77	98	71	120	126	184	218	251	206	261

y-axis shows how a country is represented by its first authors in terms of percentage and not the total number of publications of those first authors;

- Until 2020, the U.S. was concerned by more than half of the publications with a first author affiliated within an institution from the U.S. Yet, starting from 2021, the U.S. becomes represented by the first author in around a third of the published papers. The blue line in the figure is slowly decreasing, representing this lower representation. Conversely, the representation of China in the first authors of published papers is growing each year, and the orange line in the figure is thus increasing and closer to that of the U.S.;

- Since 2021 (post-COVID-19 pandemic), the ranking of the second country (China), third country (Canada), and fourth country (the U.K.) did not vary, while the ranking changed for countries beyond the fifth position.

2.2. Human Selection

2.2.1. Pre-selection

Based on titles, keywords, abstract contents, and the automatic scoring, used as additional information to make the process faster, each section editor rapidly reviewed those 2,148 entries. This year, we only processed papers with an automatic score higher than 0.5, allowing us to discard half of the entries that were generally unrelated

to NLP topics (see next Section). Each section editor independently classified the entries into three classes: Yes / Maybe / No. We thus collected 86 papers classified into Yes or Maybe classes by at least one section editor. We then performed an adjudication process to choose the final top 13 candidates to be proofread by external reviewers. We paid attention to the topics addressed by the researchers to provide enough diversity. As a result, out of the 13 papers, six come from the U.S., two from South Korea, and one from five other countries (China, Germany, Singapore, Spain, and the U.K.). Based on the external reviews and our reviews, we finally decided to keep only two papers as best for the NLP section.

2.2.2. Final best papers

Both best papers deal with language models, highlighting the difficulty of training language models on the clinical data, especially for large language models (LLMs), due to the difficulty to access those sensitive data. To tackle this issue, Tan *et al.* [3] propose to use data augmentation techniques to benefit from more available data and prompt engineering to fine-tune existing models for a classification task. While language models are generally heavy, Rohanian *et al.* [4] explore compacting biomedical models using continual learning of existing models on a PubMed dataset and distillation procedures. The first best paper proposes solutions regarding data and model adaptation to domain-specific data. In contrast, the second paper addresses the models to make them

lighter (see Appendix A for a more detailed summary of these two papers).

3. Current trends in biomedical NLP

As an introduction, we observed that language models (all BERT-related models) based on text content are currently also used to process sequences of characters that do not compose words but are helpful content to other disciplines. This is true for processing sequences of proteins [5, 6] or nucleotides in genomic sequences [7-9]. Second, those language models contributed to their democratization outside the NLP community, making them easily usable for non-specialists who published the results of their experiments in scientific papers. These considerations highlight the ability to transfer NLP techniques to computer-based approaches in biology and medicine. Another major trend we have observed for a few years now is how researchers published and valorized the results of their research. If making available the produced corpora is still complex, the availability of produced language models is effortless through the existing platforms (HuggingFace or GitHub repository).

In what follows, we first analyze the keywords provided by the authors (Section 3.1). Second, we analyze the languages addressed (Section 3.2) and the sources of data exploited (Section 3.3). Then, we distinguish some typical tasks (Section 3.4) and some emerging topics (Section 3.5). Except for the keywords, the analysis is issued from the 200 top citations according to the scores computed automatically.

3.1. Main Keywords from Publications

Out of the 2,114 entries obtained while querying PubMed, the two main keywords used to index papers are Natural Language Processing/NLP (914 papers), part of our PubMed query (Figure 1), and Artificial Intelligence/AI (308 papers). The rapid development of language models leads to

a move towards AI-based approaches for several NLP tasks. We noticed that other keywords used by the authors to index their submissions refer to three main aspects of their research:

- First, the type of documents they process: Social Media/Twitter/Reddit (133 papers) or Electronic Health Records (EHR) (127 papers). Those two types are well balanced, highlighting that social media are an alternative solution when EHRs are unavailable.

- Second, the methods used in the experiments: Machine Learning (293 papers), Deep Learning (165 papers), ChatGPT (84 papers), Language Models/Transformers (52 papers), BERT (35 papers), LLMs (33 papers), and Long-Short-Term-Memory (11 papers). All methods currently used are statistical approaches, whatever the term refers to . If regular expressions are used, they are no longer expressed in the keywords list.

- Third, the topics or purpose addressed in the paper: Sentiment analysis/Text Mining/Text Classification, Chatbot, Medical Education, COVID-19, Cancer, Mental Health/Depression/Dementia. It should be no surprise that COVID-19 is still being studied in 2023. Research in this area is no longer focused on the search for treatment but on retrospective studies of the efficacy of vaccines and vaccination hesitancy [10-13] and post-COVID-19 conditions [14, 15], such as the impact on mental health [16]. More confidential research focuses on analyzing vaccine-related messages produced by specific laboratories and how these laboratories have managed their communications [17]. At last, research during COVID-19 is now helpful for re-use for more general medical applications, such as Chatbots [18,19].

3.2. Languages Addressed

As in previous years, we observed that Chinese is still the first other-than-English language addressed in NLP publications related to the biomedical domain in 2023. European languages from countries with vast populations and available resources are also well considered: French, German, and Spanish. We hypothesize that countries having few research labs are less represented in terms of processed languages (Dutch, Estonian, Greek, Italian, Norwegian, Swedish). Out of Europe, we found several papers concerning Korean and Japanese languages. The following list is not exhaustive and only reflects a few topics addressed in each language:

- Arabic: sentiment analysis on COVID-19 tweets [20, 21];

- Chinese: automatic text classification, named entity recognition, and information extraction tasks are well represented in papers, while traditional Chinese medicine is still a current research topic [22-33];

- Czech, Polish, and Slovak: information extraction in oncology records [34];

- Dutch: negation identification and language models for a specific clinical issue [35-37];

- Estonian: production of a BERT model for information extraction from Estonian texts [38];

- Finnish: analysis of social media during the pandemic issue [13];

- French: language portability of existing algorithms, adaptation of existing models, and information extraction from social media [17,39-42];

- German: production of corpora, corpus annotation, information extraction, and production of language models [43-48];

- Greek: the issue of COVID-19 long analyzed in social media [49];

- Italian: information extraction and opinion mining tasks, development of Chatbots, and vaccination hesitancy [50-53];

- Japanese: information extraction from social media and reports, use of ChatGPT to generate lists of diagnoses, and production of systems for traditional Kampo medicine [54-58];

- Korean: information extraction on adverse drug events (ADEs) [59];

- Moroccan dialect: named entity annotated dataset [60];

- Norwegian: production of information extraction systems and de-identification system based on related languages [61,62];

- Persian: sentiment analysis from cancer institute feedback forms [63];

- Portuguese: information extraction and sentiment analysis [17];

- Russian: production of a dataset from PubMed annotated into nested named entities [64];

- Spanish: production of resources (corpora, lexicon), information extraction, and langage portability methods to annotate corpora [17,65-70];

- Swedish: comparison of a BERT language model and human annotations in an adverse drug reaction triage task [71].

We observe that solutions are proposed for producing resources and language models, using existing resources from a close language (specifically between Danish, Norwegian, and Swedish, or between Arabic and dialects) or trying language portability, generally from English to another language.

3.3. Sources of the Data Processed

As in previous years, NLP researchers work with several sources of data, which have become emblematic of medical research: scientific literature, clinical documents and social media.

One notable difference in 2023 is that the use of clinical data increased importantly. The keywords analyzed also testify this fact. Indeed, several experiments with clinical data addressing a wide variety of tasks have been proposed, such as test of language models for long clinical texts [72], acquisition of lexicon for the family history [73], detection of negation and speculation in radiology reports [74], coreference resolution in clinical narratives [75], de-identification of radiology reports [76], rare disease identification from clinical notes [77], section identification and structuring in clinical narratives [68], adaptation of LLMs to clin-

ical data through prompting [78]. Let's also note that radiology reports become again attractive for researchers and are used for instance, for the extraction of entities and spatial relations [43,79], ICD coding [80,81], ADE extraction [82], patient similarity [40], or for inferring cancer disease responses from text and images [3]. In such works, clinical data can be provided from publicly available datasets, like MIMIC-III [83], or from own datasets with restricted access.

Social media remain another important source of information for researchers and health practitioners. Such data are typically used for the analysis of COVID-19-related issues [13,20,21,42,49], and for mental health [54,84-86].

As for the scientific literature, it is exploited in the context of randomized controlled trials (RCTs) and systematic reviews: extraction of evidence for the RCTs [87], detection of the outcome issues [88-90], automatic assistance for systematic reviews [91,90], detection of high-quality papers [92]. Besides, scientific literature can also be used in a variety of tasks, such as extraction of relations between suicide and drugs [93], simplification of scientific abstracts [94], or the creation of medical lexicon intended to complete the UMLS [66].

3.4. NLP Tasks performed

Among the general NLP tasks, we can mention traditional information extraction, categorization, and prediction. Besides, like in the general NLP area, the wave of generative LLMs has also entered the biomedical domain. Generative models are exploited in a lot of contexts. To provide a few examples, we can mention the extraction of clinical entities and relations [88,95], clinical data augmentation [3,41,76], generation of diagnosss lists [57,56], different public health issues [96], ADE signal extraction [97], and conversational agents [98].

Some works investigate specific issues, such as enriching LLMs with biomedical knowledge [99] or studying the bias [100].

Besides, as the use of LLMs is well established now, we have noticed several

literature reviews, which focus on specific research questions such as the use of ChatGPT in clinical practice [101], the use of transformers in biomedicine [102], as well as the use of NLP in emergency rooms [103], in oncology departments [104], on radiology reports [105], for the summarization [106], or on social media data [107,108].

3.5. NLP Research outside the Box

A few topics are addressed but not highlighted in the mainstream research. This is mainly true for literature discovery, where analysis of *known unknowns* reported in the scientific literature was done to produce an ignorance base [109]. On social media, research is generally done to identify hate and harmful speeches. In contrast, work has been done to explore the characteristics of *peace speech* and to provide an index of peace for each country [110]. Automatic speech recognition (ASR) systems produce text from speech, enabling the application of conventional NLP approaches to mine information. Nevertheless, information may be expressed by non-lexical conversational sounds such as *Mm-hm* and *Uh-uh* to express positive and negative answers, which have been used to refine ASR outputs from clinical conversations [111].

4.   Conclusion

This synopsis paper presents the process we followed to identify the best papers for the NLP section of the IMIA Yearbook. We selected two scientific papers published in biomedical journals. These papers deal with language models and the difficulty of training language models on clinical data. They propose solutions that address data augmentation, domain-specific model adaptation, and model distillation. Unsurprisingly, most biomedical NLP papers published in 2023 deal with deep learning and language models. Consequently, such papers are becoming increasingly numerous. They address classical NLP tasks (information extraction, text categorization, sentiment analysis), existing topics from several years (medical education), mainstream applications (Chatbots,

generative approaches), and specific issues (cancer, COVID-19, mental health). Specifically for COVID-19, current researches deal with post-COVID-19 conditions, and they explore the understanding of how this pandemic has been managed and welcomed by populations. Due to the democratization of language models, work is being done to cover more confidential languages (*e.g.*, Estonian) when producing models, making it possible to process close languages through language transfer methods (*e.g.*, Moroccan dialect from Arabic). Nevertheless, NLP publications do not cover African languages, and Asian languages are limited to a few (Chinese, Japanese, and Korean).

References

1.   PM Nadkarni, L Ohno-Machado, and WW Chapman. Natural Language Processing: an introduction. J Am Med Inform Assoc, 2011:18(5):544-51. DOI: 10.1136/amiajnl-2011-000464

2.   C Friedman and G Hripcsak. Natural language processing and its future in medicine. Academic Medicine, 1999:74(8):890-5. DOI: 10.1097/00001888-199908000-00012

3.   RSYC Tan, Q Lin, GH Low, R Lin, TC Goh, CCE Chang, et al. Inferring cancer disease response from radiology reports using large language models with data augmentation and prompting. J Am Med Inform Assoc, 2023:30(10):1657-64. DOI: 10.1093/jamia/ocad133

4.   O Rohanian, M Nouriborji, S Kouchaki, and DA Clifton. On the effectiveness of compact biomedical transformers. Bioinformatics, 2023:39(3):btad103. doi: 10.1093/bioinformatics/btad103

5.   C Tran, S Khadkikar, and A Porollo. Survey of Protein Sequence Embedding Models. Int J Mol Sci, 2023:24(4):3775. DOI: 10.3390/ijms24043775

6.   A Yoshimori and J Bajorath. Motif2Mol: Prediction of New Active Compounds Based on Sequence Motifs of Ligand Binding Sites in Proteins Using a Biochemical Language Model. Biomolecules, 2023:13(5):833. DOI: 10.3390/biom13050833

7.   MF Danilevicz, M Gill, CG Tay Fernandez, J Petereit, SR Upadhyaya, J Batley, et al. DNABERT-based explainable lncRNA identification in plant genome assemblies. Comput Struct Biotechnol J, 2023:21:5676-85. DOI: 10.1016/j.csbj.2023.11.025

8.   S Wang, Y Liu, Y Liu, Y Zhang, and X Zhu. BERT-5mC: an interpretable model for predicting 5-methylcytosine sites of DNA based on BERT. PeerJ, 2023:11:e16600. DOI: 10.7717/peerj.16600

9.   Y Ma, Y Pei, and C Li. Predictive Recognition of DNA-binding Proteins Based on Pre-trained

Language Model BERT. J Bioinform Comput Biol, 2023:21(6):2350028. DOI: 10.1142/S0219720023500282

10. JC Boucher, SY Kim, G Jessiman-Perreault, J Edwards, H Smith, N Frenette, et al. HPV vaccine narratives on Twitter during the COVID-19 pandemic: a social network, thematic, and sentiment analysis. BMC Public Health, 2023:23(1):694. DOI: 10.1186/s12889-023-15615-w

11. Z Zaidi, M Ye, F Samon, A Jama, B Gopalakrishnan, C Gu, et al. Topics in Antivax and Provax Discourse: Yearlong Synoptic Study of COVID-19 Vaccine Tweets. J Med Internet Res, 2023:25:e45069. DOI: 10.2196/45069

12. L Lösch, T Zuiderent-Jerak, F Kunneman, E Syurina, M Bongers, ML Stein, et al. Capturing Emerging Experiential Knowledge for Vaccination Guidelines Through Natural Language Processing: Proof-of-Concept Study. J Med Internet Res, 2023:25:e44461. DOI: 10.2196/44461

13. A Unlu, S Truong, T Tammi, and AL Lohiniva. Exploring Political Mistrust in Pandemic Risk Communication: Mixed-Method Study Using Social Media Data Analysis. J Med Internet Res, 2023:25:e50199. DOI: 10.2196/50199

14. H Ayadi, C Bour, A Fischer, M Ghoniem, and G Fagherazzi. The Long COVID experience from a patient's perspective: a clustering analysis of 27,216 Reddit posts. Front Public Health, 2023:11:1227807. DOI: 10.3389/fpubh.2023.1227807

15. E Dolatabadi, D Moyano, M Bales, S Spasojevic, R Bhambhoria, J Bhatti, et al. Using Social Media to Help Understand Patient-Reported Health Outcomes of Post-COVID-19 Condition: Natural Language Processing Approach. J Med Internet Res, 2023:25:e45767. DOI: 10.2196/45767

16. J Zhu, N Yalamanchi, R Jin, DR Kenne, and NH Phan. Investigating COVID-19's Impact on Mental Health: Trend and Thematic Analysis of Reddit Users' Discourse. J Med Internet Res, 2023:25:e46867. DOI: 10.2196/46867

17. D Catalan-Matamoros, I Prieto-Sanchez, and A Langbecker. Crisis Communication during COVID-19: English, French, Portuguese, and Spanish Discourse of AstraZeneca Vaccine and Omicron Variant on Social Media. Vaccines (Basel), 2023:11(6):1100. DOI: 10.3390/vaccines11061100

18. H Chin, G Lima, M Shin, A Zhunis, C Cha, J Choi, et al. User-Chatbot Conversations During the COVID-19 Pandemic: Study Based on Topic Modeling and Sentiment Analysis. J Med Internet Res, 2023:25:e40922. DOI: 10.2196/40922

19. F Moutsana Tapolin, J Liaskos, E Zoulias, and J Mantas. A Conversational Web-Based Chatbot to Disseminate COVID-19 Advisory Information. Stud Health Technol Inform, 2023:305:483-86. DOI: 10.3233/SHTI230538

20. MJ Althobaiti. An open-source dataset for arabic fine-grained emotion recognition of online content amid COVID-19 pandemic. Data Brief, 2023:51:109745. doi: 10.1016/j.dib.2023.109745

21. S Alhumoud, A Al Wazrah, L Alhussain, L Alrushud, A Aldosari, RN Altammami, et al. ASAVACT: Arabic sentiment analysis for

vaccine-related COVID-19 tweets using deep learning. PeerJ Comput Sci, 2023:9:e1507. DOI: 10.7717/peerj-cs.1507

22. E Zhu, Q Sheng, H Yang, Y Liu, T Cai, and J Li. A unified framework of medical information annotation and extraction for Chinese clinical text. Artif Intell Med, 2023:142:102573. DOI: 10.1016/j.artmed.2023.102573

23. J Wei, T Hu, J Dai, Z Wang, P Han, and W Huang. Research on named entity recognition of adverse drug reactions based on NLP and deep learning. Front Pharmacol, 2023:14:1121796. DOI: 10.3389/fphar.2023.1121796

24. ZY Feng, XH Wu, JL Ma, M Li, GF He, DS Cao, et al. DKADE: a novel framework based on deep learning and knowledge graph for identifying adverse drug events and related medications. Brief Bioinform, 2023:24(4):bbad228. DOI: 10.1093/bib/bbad228

25. M Li, C Gao, K Zhang, H Zhou, and J Ying. A weakly supervised method for named entity recognition of Chinese electronic medical records. Med Biol Eng Comput, 2023:61(10):2733-43. DOI: 10.1007/s11517-023-02871-6

26. J Cai, S Chen, S Guo, S Wang, L Li, X Liu, et al. RegEMR: a natural language processing system to automatically identify premature ovarian decline from Chinese electronic medical records. BMC Med Inform Decis Mak, 2023:23(1):126. DOI: 10.1186/s12911-023-02239-8

27. H Peng, Z Zhang, D Liu, and X Qin. Chinese medical entity recognition based on the dual-branch TENER model. BMC Med Inform Decis Mak, 2023:23(1):136. DOI: 10.1186/s12911-023-02243-y

28. X Xu, Y Chang, J An, and Y Du. Chinese text classification by combining Chinese-BER-Tology-wwm and GCN. PeerJ Comput Sci, 2023:9:e1544. DOI: 10.7717/peerj-cs.1544

29. L Fu, Z Weng, J Zhang, H Xie, and Y Cao. MMBERT: a unified framework for biomedical named entity recognition. Med Biol Eng Comput, 2023:62(1):327-41. doi: 10.1007/s11517-023-02934-8

30. MW Ma, XS Gao, ZY Zhang, SY Shang, L Jin, PL Liu, et al. Extracting laboratory test information from paper-based reports. BMC Med Inform Decis Mak, 2023:23(1):251. doi: 10.1186/s12911-023-02346-6

31. H Sun, K Zhang, W Lan, Q Gu, G Jiang, X Yang, et al. An AI Dietitian for Type 2 Diabetes Mellitus Management Based on Large Language and Image Recognition Models: Preclinical Concept Validation Study. J Med Internet Res, 2023:25:e51300. DOI: 10.2196/51300

32. J Liu and T Jiang. Methods for Analyzing Unknown Health Risk Based on Nature Language Process (NLP). Stud Health Technol Inform, 2023:308:633-9. DOI: 10.3233/SHTI230894

33. Z Cui, K Yu, Z Yuan, X Dong, and W Luo. Language inference-based learning for Low-Resource Chinese clinical named entity recognition using langage model. J Biomed Inform, 2023:149:104559. DOI: 10.1016/j.jbi.2023.104559

34. K Anetta. Understanding Health Records in West Slavic Languages: Available Resources, Case Study in Oncology. Stud Health Technol Inform,

2023:305:97-101. DOI: 10.3233/SHTI230433

35. B van Es, LC Reteig, SC Tan, M Schraagen, MM Hemker, SRS Arends, et al. Negation detection in Dutch clinical texts: an evaluation of rule-based and machine learning methods. BMC Bioinformatics, 2023:24(1):10. DOI: 10.1186/s12859-022-05130-x

36. TM Seinen, JA Kors, EM van Mulligen, E Fridgeirsson, and PR Rijnbeek. The added value of text from Dutch general practitioner notes in predictive modeling. J Am Med Inform Assoc, 2023:30(12):1973-84. DOI: 10.1093/jamia/ocad160

37. M Homburg, E Meijer, M Berends, T Kupers, T Olde Hartman, J Muris, et al. A Natural Language Processing Model for COVID-19 Detection Based on Dutch General Practice Electronic Health Records by Using Bidirectional Encoder Representations From Transformers: Development and Validation Study. J Med Internet Res, 2023:25:e49944. DOI: 10.2196/49944

38. H Šuvalov, S Laur, and R Kolde. Information Extraction from Medical Texts with BERT Using Human-in-the-Loop Labeling. Stud Health Technol Inform, 2023:302:831-2. DOI: 10.3233/SHTI230281

39. T Fabacher, EA Sauleau, N Leclerc Du Sablon, H Bergier, JE Gottenberg, et al. Evaluating the Portability of Rheumatoid Arthritis Phenotyping Algorithms: A Case Study on French EHRs. Stud Health Technol Inform, 2023:302:768-72. DOI: 10.3233/SHTI230263

40. X Chen, C Faviez, M Vincent, S Saunier, N Garcelon, and A Burgun. Improving Patient Similarity Using Different Modalities of Phenotypes Extracted from Clinical Narratives. Stud Health Technol Inform, 2023:302:1037-41. DOI: 10.3233/SHTI230342

41. TD Le, R Noumeir, J Rambaud, G Sans, and P Jouvet. Adaptation of Autoencoder for Sparsity Reduction From Clinical Notes Representation Learning. IEEE J Transl Eng Health Med, 2023:11:469-78. DOI: 10.1109/JTEHM.2023.3241635

42. JSM Gable, R Sauvayre, and C Chauvière. Fight Against the Mandatory COVID-19 Immunity Passport on Twitter: Natural Language Processing Study. J Med Internet Res, 2023:25:e49435. DOI: 10.2196/49435

43. M Jantscher, F Gunzer, R Kern, E Hassler, S Tschauner, and G Reishofer. Information extraction from German radiological reports for general clinical text and language understanding. Sci Rep, 2023:13(1):2353. DOI: 10.1038/s41598-023-29323-3

44. J Frei and F Kramer. German Medical Named Entity Recognition Model and Data Set Creation Using Machine Translation and Word Alignment: Algorithm Development and Validation. JMIR Form Res, 2023:7:e39077. DOI: 10.2196/39077

45. S Nowak, D Biesner, YC Layer, M Theis, H Schneider, W Block, et al. Transformer-based structuring of free-text radiology report databases. Eur Radiol, 2023:33(6):4228-36. DOI: 10.1007/s00330-023-09526-y

46. F Meineke, L Modersohn, M Loeffler, and M Boeker. Announcement of the German Medical Text Corpus Project (GeMTeX). Stud Health

Technol Inform, 2023:302:835-6. DOI: 10.3233/SHTI230283

47. J Frei and F Kramer. Annotated dataset creation through large language models for non-english medical NLP. J Biomed Inform, 2023:145:104478. DOI: 10.1016/j.jbi.2023.104478

48. J Frei, L Frei-Stuber, and F Kramer. GER-NERMED++: Semantic annotation in german medical NLP through transfer-learning, translation and word alignment. J Biomed Inform, 2023:147:104513. DOI: 10.1016/j.jbi.2023.104513

49. A Katika, E Zoulias, V Koufi, and F Malamateniou. Mining Greek Tweets on Long COVID Using Sentiment Analysis and Topic Modeling. Stud Health Technol Inform, 2023:305:545-8. DOI: 10.3233/SHTI230554

50. R Catelli, S Pelosi, C Comito, C Pizzuti, and M Esposito. Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy. Comput Biol Med, 2023:158:106876. DOI: 10.1016/j.compbiomed.2023.106876

51. J Franceschi, L Pareschi, E Bellodi, M Gavanelli, and M Bresadola. Modeling opinion polarization on social media: Application to Covid-19 vaccination hesitancy in Italy. PLoS One, 2023:18(10):e0291993. DOI: 10.1371/journal.pone.0291993

52. A Cappello, S Mora, DR Giacobbe, M Bassetti, and M Giacomini. Defining a Preprocessing Pipeline for the MULTI-SITA Project and General Medical Italian Natural Language Data. Stud Health Technol Inform, 2023. 309:48-52. DOI: 10.3233/SHTI230737

53. C Crema, TM Buonocore, S Fostinelli, E Parimbelli, F Verde, C Fundarò, et al. Advancing Italian biomedical information extraction with transformers-based models: Methodological insights and multicenter practical application. J Biomed Inform, 2023:148:104557. DOI: 10.1016/j.jbi.2023.104557

54. S Wang, H Ning, X Huang, Y Xiao, M Zhang, EF Yang, et al. Public Surveillance of Social Media for Suicide Using Advanced Deep Learning Models in Japan: Time Series Study From 2012 to 2022. J Med Internet Res, 2023:25:e47225. DOI: 10.2196/47225

55. A Maeda-Minami, T Yoshino, T Yumoto, K Sato, A Sagara, K Inaba, et al. Development of a novel drug information provision system for Kampo medicine using natural language processing technology. BMC Med Inform Decis Mak, 2023:23(1):119. DOI: 10.1186/s12911-023-02230-3

56. T Kuroiwa, A Sarcon, T Ibara, E Yamada, A Yamamoto, K Tsukamoto, et al. The Potential of ChatGPT as a Self-Diagnostic Tool in Common Orthopedic Diseases: Exploratory Study. J Med Internet Res, 2023:25:e47621. DOI: 10.2196/47621

57. T Hirosawa, R Kawamura, Y Harada, K Mizuta, K Tokumasu, Y Kaji, et al. ChatGPT-Generated Differential Diagnosis Lists for Complex Case-Derived Clinical Vignettes: Diagnostic Accuracy Evaluation. JMIR Med Inform, 2023:11:e48808. DOI: 10.2196/48808

58. K Sugimoto, S Wada, S Konishi, K Okada, S Manabe, Y Matsumura, et al. Extracting Clinical Information From Japanese Radiology Reports Using a 2-Stage Deep Learning Approach: Algorithm Development and Validation. JMIR Med Inform, 2023:11:e49041. DOI: 10.2196/49041

59. S Kim, T Kang, Tae Kyu Chung, Y Choi, YS Hong, K Jung, et al. Automatic Extraction of Comprehensive Drug Safety Information from Adverse Drug Event Narratives in the Korea Adverse Event Reporting System Using Natural Language Processing Techniques. Drug Saf, 2023:46(8):781-95. DOI: 10.1007/s40264-023-01323-2

60. HN Moussa and A Mourhir. DarNERcorp: An annotated named entity recognition dataset in the Moroccan dialect. Data Brief, 2023:48:109234. DOI: 10.1016/j.dib.2023.109234

61. GT Berge, OC Granmo, TO Tveit, BE Munkvold, AL Ruthjersen, and J Sharma. Machine learning-driven clinical decision support system for concept-based searching: a field trial in a Norwegian hospital. BMC Med Inform Decis Mak, 2023:23(1):5. DOI: 10.1186/s12911-023-02101-x

62. A Lamproudis, S Mora, TO Svenning, T Torsvik, T Chomutare, PD Ngo, et al. De-identifying Norwegian Clinical Text using Resources from Swedish and Danish. AMIA Annu Symp Proc, 2024:456-64. [Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/pmc10785939/]

63. A Yazdani, M Shamloo, M Khaki, and A Nahvijou. Use of sentiment analysis for capturing hospitalized cancer patients' experience from free-text comments in the Persian language. BMC Med Inform Decis Mak, 2023:23(1):275. DOI: 10.1186/s12911-023-02358-2

64. N Loukachevitch, S Manandhar, E Baral, I Rozhkov, P Braslavski, V Ivanov, et al. NEREL-BIO: a dataset of biomedical abstracts annotated with nested named entities. Bioinformatics, 2023:39(4):btad161. DOI: 10.1093/bioinformatics/btad161

65. M Chizhikova, P López-Úbeda, J Collado-Montañez, T Martín-Noguerol, MC Díaz-Galiano, A Luna, et al. CARES: A Corpus for classification of Spanish Radiological reports. Comput Biol Med, 2023:154:106581. DOI: 10.1016/j.compbiomed.2023.106581

66. L Campillos-Llanos. MedLexSp - a medical lexicon for Spanish medical natural language processing. J Biomed Semantics, 2023:14(1):2. DOI: 10.1186/s13326-022-00281-5

67. I Goenaga, E Andres, K Gojenola, and A Atutxa. Advances in monolingual and crosslingual automatic disability annotation in Spanish. BMC Bioinformatics, 2023:24(1):265. DOI: 10.1186/s12859-023-05372-3

68. I de la Iglesia, M Vivó, P Chocrón, G de Maeztu, K Gojenola, and A Atutxa. An open source corpus and automatic tool for section identification in Spanish health records. J Biomed Inform, 2023:145:104461. DOI: 10.1016/j.jbi.2023.104461

69. DM Mendoza-Urbano, J Felipe Garcia, JS Moreno, JC Bravo-Ocaña, AJ Riascos, A Zambrano Harvey, et al. Automated extraction of information from free text of Spanish

oncology pathology reports. Colomb Med (Cali), 2023:54(1):e2035300. DOI: 10.25100/cm.v54i1.5300

70. O Solarte-Pabón, O Montenegro, A García-Barragán, M Torrente, M Provencio, E Menasalvas, et al. Transformers for extracting breast cancer information from Spanish clinical narratives. Artif Intell Med, 2023:143:102625. DOI: 10.1016/j.artmed.2023.102625

71. E Bergman, L Dürlich, V Arthurson, A Sundström, M Larsson, S Bhuiyan, et al. BERT based natural language processing for triage of adverse drug reaction reports shows close to human-level performance. PLOS Digit Health, 2023:2(12):e0000409. DOI: 10.1371/journal.pdig.0000409

72. Y Li, RM Wehbe, FS Ahmad, H Wang, and Y Luo. A comparative study of pretrained language models for long clinical text. J Am Med Inform Assoc, 2022:30(2):340-7. DOI: 10.1093/jamia/ocac225

73. L Wang, H He, A Wen, S Moon, S Fu, KJ Peterson, et al. Acquisition of a Lexicon for Family History Information: Bidirectional Encoder Representations From Transformers-Assisted Sublanguage Analysis. JMIR Med Inform, 2023:11:e48072. DOI: 10.2196/48072

74. KH Weng, CF Liu, and CJ Chen. Deep Learning Approach for Negation and Speculation Detection for Automated Important Finding Flagging and Extraction in Radiology Report: Internal Validation and Technique Comparison Study. JMIR Med Inform, 2023:11:e46348. DOI: 10.2196/46348

75. Y Liao, H Liu, and I Spasić. Fine-tuning coreference resolution for different styles of clinical narratives. J Biomed Inform, 2023:149:104578. DOI: 10.1016/j.jbi.2023.104578

76. PJ Chambon, C Wu, JM Steinkamp, J Adleberg, TS Cook, and CP Langlotz. Automated deidentification of radiology reports combining transformer and "hide in plain sight" rule-based methods. J Am Med Inform Assoc, 2022:30(2):318-28. DOI: 10.1093/jamia/ocac219

77. H Dong, V Suárez-Paniagua, H Zhang, M Wang, A Casey, E Davidson, et al. Ontology-driven and weakly supervised rare disease identification from clinical notes. BMC Med Inform Decis Mak, 2023:23(1):86. DOI: 10.1186/s12911-023-02181-9

78. S Sivarajkumar and Y Wang. HealthPrompt: A Zero-shot Learning Paradigm for Clinical Natural Language Processing. AMIA Annu Symp Proc, 2023:972-981. [Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/pmc10148337/]

79. S Datta and K Roberts. Weakly supervised spatial relation extraction from radiology reports. JAMIA Open, 2023:6(2):ooad027. DOI: 10.1093/jamiaopen/ooad027

80. HA Xu, B Maccari, H Guillain, J Herzen, F Agri, and JL Raisaro. An End-to-End Natural Language Processing Application for Prediction of Medical Case Coding Complexity: Algorithm Development and Validation. JMIR Med Inform, 2023:11:e38150. DOI: 10.2196/38150

81. MJ Kane, C King, D Esserman, NK Latham, EJ Greene, and DA Ganz. A compressed large language model embedding dataset

Grouin et al.

of ICD 10 CM descriptions. BMC Bioinformatics, 2023:2023.04.24.23289046. doi: 10.1101/2023.04.24.23289046

82. MM Zitu, S Zhang, DH Owen, C Chiang, and L Li. Generalizability of machine learning methods in detecting adverse drug events from clinical narratives in electronic medical records. Front Pharmacol, 2023:14:1218679. DOI: 10.3389/fphar.2023.1218679

83. AE Johnson, TJ Pollard, L Shen, LW Lehman, M Feng, M Ghassemi, et al. MIMIC-III, a freely accessible critical care database. Sci Data, 2016:3:160035. DOI: 10.1038/sdata.2016.35

84. S Kim, J Cha, D Kim, and E Park. Understanding Mental Health Issues in Different Subdomains of Social Networking Services: Computational Analysis of Text-Based Reddit Posts. J Med Internet Res, 2023:25:e49074. DOI: 10.2196/49074

85. M Afshar, S Adelaine, F Resnik, MP Mundt, J Long, M Leaf, et al. Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults. JMIR Med Inform, 2023:11:e44977. DOI: 10.2196/44977

86. SG Weiner, YC Lo, AD Carroll, L Zhou, A Ngo, DB Hathaway, et al. The Incidence and Disparities in Use of Stigmatizing Language in Clinical Notes for Patients With Substance Use Disorder. J Addict Med, 2023:17(4):424-30. DOI: 10.1097/ADM.0000000000001145

87. T Kang, Y Sun, JH Kim, C Ta, A Perotte, K Schiffer, et al. EvidenceMap: a three-level knowledge representation for medical evidence computation and comprehension. J Am Med Inform Assoc, 2023:30(6):1022-31. DOI: 10.1093/jamia/ocad036

88. Y Hu, VK Keloth, K Raja, Y Chen, and H Xu. Towards precise PICO extraction from abstracts of randomized controlled trials using a section-specific learning approach. Bioinformatics, 2023:39(9):btad542. DOI: 10.1093/bioinformatics/btad542

89. A Newbury, H Liu, B Idnay, and C Weng. The suitability of UMLS and SNOMED-CT for encoding outcome concepts. J Am Med Inform Assoc, 2023:30(12):1895-903. DOI: 10.1093/jamia/ocad161

90. A Dhrangadhariya and H Müller. Not so weak PICO: leveraging weak supervision for participants, interventions, and outcomes recognition for systematic review automation. JAMIA Open, 2023:6(1):ooac107. DOI: 10.1093/jamiaopen/ooac107

91. E Orel, I Ciglenecki, A Thiabaud, A Temerev, A Calmy, O Keiser, et al. An Automated Literature Review Tool (LiteRev) for Streamlining and Accelerating Research Using Natural Language Processing and Machine Learning: Descriptive Performance Evaluation Study. J Med Internet Res, 2023:25:e39736. DOI: 10.2196/39736

92. Y Lin, J Li, H Xiao, L Zheng, Y Xiao, H Song, et al. Automatic literature screening using the PAJO deep-learning model for clinical practice guidelines. BMC Med Inform Decis Mak, 2023:23(1):247. DOI: 10.1186/s12911-023-02328-8

93. K Karapetian, SM Jeon, JW Kwon, and YK Suh. Supervised Relation Extraction Between Suicide-Related Entities and Drugs: Development and Usability Study of an Annotated PubMed Corpus. J Med Internet Res, 2023:25:e41100. DOI: 10.2196/41100

94. Y Guo, W Qiu, G Leroy, S Wang, and T Cohen. Retrieval augmentation of large language models for lay language generation. J Biomed Inform, 2024:149:104580. DOI: 10.1016/j.jbi.2023.104580

95. C Peng, X Yang, Z Yu, J Bian, WR Hogan, and Y Wu. Clinical concept and relation extraction using prompt-based machine reading comprehension. J Am Med Inform Assoc, 2023:30(9):1486-93. DOI: 10.1093/jamia/ocad107

96. M Pillai, AC Griffin, CA Kronk, and T McCall. Toward Community-Based Natural Language Processing (CBNLP): Cocreating With Communities. J Med Internet Res, 2023:25:e48498. DOI: 10.2196/48498

97. G Dong, A Bate, F Haguinet, G Westman, L Dürlich, A Hviid, and M Sessa. Optimizing Signal Management in a Vaccine Adverse Event Reporting System: A Proof-of-Concept with COVID-19 Vaccines Using Signs, Symptoms, and Natural Language Processing. Drug Saf, 2023:47(2):173-82. DOI: 10.1007/s40264-023-01381-6

98. A Brown, AT Kumar, O Melamed, I Ahmed, YH Wang, A Deza, et al. A Motivational Interviewing Chatbot With Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study. JMIR Ment Health, 2023:10:e49132. DOI: 10.2196/49132

99. TM Lai, CX Zhai, and H Ji. KEBLM: Knowledge-Enhanced Biomedical Language Models. J Biomed Inform, 2023:143:104392. DOI: 10.1016/j.jbi.2023.104392

100. Y Jin, Y Xiong, D Shi, Y Lin, L He, Y Zhang, et al. Learning from undercoded clinical records for automated International Classification of Diseases (ICD) coding. J Am Med Inform Assoc, 2022 :30(3):438-46. DOI: 10.1093/jamia/ocac230

101. J Liu, C Wang, and S Liu. Utility of ChatGPT in Clinical Practice. J Med Internet Res, 2023:25:e48568. DOI: 10.2196/48568

102. S Zhang, R Fan, Y Liu, S Chen, Q Liu, and W Zeng. Applications of transformer-based language models in bioinformatics: a survey. Bioinform Adv, 2023:3(1):vbad001. DOI: 10.1093/bioadv/vbad001

103. J Stewart, J Lu, A Goudie, G Arendts, SA Meka, S Freeman, et al. Applications of natural language processing at emergency department triage: A narrative review. PLoS One, 2023:18(12):e0279953. DOI: 10.1371/journal.pone.0279953

104. M Gholipour, R Khajouei, P Amiri, S Hajesmaeel Gohari, and L Ahmadian. Extracting cancer concepts from clinical notes using natural language processing: a systematic review. BMC Bioinformatics, 2023:24(1):405. DOI: 10.1186/s12859-023-05480-0

105. D Reichenpfader, H Müller, and K Denecke. Large language model-based information extraction from free-text radiology reports: a scoping review protocol. BMJ Open, 2023:13(12):e076865. DOI: 10.1136/bmjopen-2023-076865

106. D Keszthelyi, C Gaudet-Blavignac, M Bjelogrlic, and C Lovis. Patient Information Summarization in Clinical Settings: Scoping Review. JMIR Med Inform, 2023:11:e44639. DOI: 10.2196/44639

107. JM Lane, D Habib, and B Curtis. Linguistic Methodologies to Surveil the Leading Causes of Mortality: Scoping Review of Twitter for Public Health Data. J Med Internet Res, 2023:25:e39484. DOI: 10.2196/39484

108. Y Chi and HY Chen. Investigating Substance Use via Reddit: Systematic Scoping Review. J Med Internet Res, 2023:25:e48905. DOI: 10.2196/48905

109. MR Boguslav, NM Salem, EK White, KJ Sullivan, M Bada, TL Hernandez, et al. Creating an ignorance-base: Exploring known unknowns in the scientific literature. J Biomed Inform, 2023:143:104405. DOI: 10.1016/j.jbi.2023.104405

110. LS Liebovitch, W Powers, L Shi, A Chen-Carrel, P Loustaunau, and PT Coleman. Word differences in news media of lower and higher peace countries revealed by natural language processing and machine learning. PLoS One, 2023:18(11):e0292604. DOI: 10.1371/journal.pone.0292604

111. BD Tran, K Latif, TL Reynolds, J Park, J Elston Lafata, M Tai-Seale, et al. "Mm-hm," "Uh-uh": are non-lexical conversational sounds deal breakers for the ambient clinical documentation technology? J Am Med Inform Assoc, 2023:30(4):703-11. DOI: 10.1093/jamia/ocad001

Copyright