

Year 2022 in Medical Natural Language Processing: Availability of Language Models as a Step in the Democratization of NLP in the Biomedical Area

Cyril Grouin¹, Natalia Grabar², Section Editors for the IMIA Yearbook Section on Natural Language Processing

¹ Université Paris Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400 Orsay, France

² UMR8163 STL, CNRS, Université de Lille, Domaine du Pont-de-bois, 59653 Villeneuve-d'Ascq cedex, France

Summary

Objectives: To analyse the content of publications within the medical Natural Language Processing (NLP) domain in 2022.

Methods: Automatic and manual preselection of publications to be reviewed, and selection of the best NLP papers of the year. Analysis of the important issues.

Results: Three best papers have been selected. We also propose an analysis of the content of the NLP publications in 2022, stressing on some of the topics.

Conclusion: The main trend in 2022 is certainly related to the availability of large language models, especially those based on Transformers, and to their use by non-NLP researchers. This leads to the democratization of the NLP methods. We also observe the renewal of interest to languages other than English, the continuation of research on information extraction and prediction, the massive use of data from social media, and the consideration of needs and interests of patients.

Keywords

Natural language processing; best paper; topics; issues; 2022

Yearb Med Inform 2023;244-52

<https://dx.doi.org/10.1055/s-0043-1768752>

1 Introduction

Natural Language Processing (NLP) aims at providing methods, tools and resources designed in order to mine textual and narrative documents, and to make it possible to access the information they convey [1]. While human languages are complex (as an example, learning a human language requires many years in order 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 first present the selection process applied this year and then we analyze the content of some publications. More particularly, we will focus on several important trends, as well as the originality of the research questions addressed in 2022.

2 The Selection Process

In order to identify all papers published during the year 2022 in the field of NLP for the biomedical domain, we queried two databases: MEDLINE¹, specifically dedicated to the biomedical domain, and the ACL anthology², a database that brings together the major NLP conferences (ACL, COLING, EMNLP, LREC, NAACL, etc.)

¹ <https://pubmed.ncbi.nlm.nih.gov/>

² <https://www.aclweb.org/anthology/>

and journals, since some NLP studies concerning the biomedical domain are published in conferences and journals which are not indexed by PubMed.

```
(English[LA]
AND journal article[PT] AND
2022[dp]
AND ((medical OR clinical OR
natural)
AND "language processing"))
```

Fig. 1 Query used for collecting candidate publications for review.

We applied on MEDLINE the basic query we defined in 2018 and we used since then (Figure 1): all journal papers published in English in 2022, having abstract, and composed of sequences “clinical language processing” or “medical language processing” or “natural language processing”. As of 2023, January 17th, we collected 1,670 entries, which is much more than last year (1,204 entries found in 2021). We applied a similar query on the ACL anthology database and collected 20 additional entries. In order to process those 1,690 papers, we automatically scored the papers. While we defined a set of rules in 2018 to filter out each year those candidates, we refined the automatic scoring this year: a score ranging from 0 to 1 is given to each candidate, for five criteria

(journal name, objectives, methods/corpus/resources used in the paper, evaluation/metrics used, special concepts/key phrases used in the abstract). The underlying idea is to focus on the NLP community's custom in terms of place where papers are published, keywords and metrics the community generally used, and phrases that highlight the use or the design of a method. Since "language" may refer either to the study object (i.e., language productions) or to a cognitive aspect, we considered positively papers published in the NLP-related conferences (*AMIA*, *MedInfo*, *MIE*, etc.) or some journals (*Journal of the American Medical Informatics Association*, *Journal of Biomedical Informatics*, *International Journal of Medical Informatics*, etc.) rather than cognitive journals (concerning brain, neuroscience, or psychiatrics research). We also excluded yearbooks and survey papers since they do not fit criterion for best paper candidates. In order 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, and consequently, those papers are not good candidates for NLP best papers.

For each of the 1,690 candidate papers, the final score ranked from 0.05 to 1 (Figure 2). This score has been used as a meta element during the manual selection of the top-13 papers. Indeed, the section editors did not fully rely on the scores but only used them as additional information. Hence, both section editors independently browsed the abstracts, keywords and automatic scores, and assigned the *Yes/Maybe/No* score to each paper. All papers having at least one *Yes* or *Maybe* score have been kept for the next step of the selection. At this stage, 116 candidate papers remained (i.e., a subset of 9.35% of the whole dataset). We then performed an adjudication process, in order to choose the final 13 candidates to be proofread by external reviewers. We paid attention to the topics addressed by the researchers so as to provide enough diversity. As result, out of the 13 papers, nine come from the USA, two from the UK, one from Canada and one from South Korea. The best papers are listed in Table 1 and the content summaries is in the Appendix.

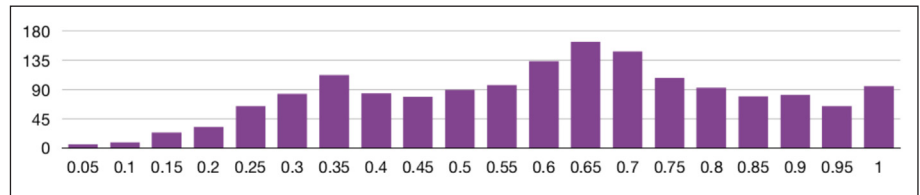


Fig. 2 Distribution of papers according to the filter scores.

Table 1 Selection of best papers for the 2023 IMIA Yearbook of Medical Informatics for the Natural Language Processing section. The articles are listed in alphabetical order by the first author's surname.

Section

Natural Language Processing

- Ahne A, Khetan V, Tannier X, Rizvi MdIH, Czernichow T, Orchard F, Bour C, Fano A, Faheerazzi G. Extraction of explicit and implicit cause-effect relationships in patient-reported diabetes-related tweets from 2017 to 2021: Deep learning approach. *JMIR Med Inform* 2022;10(7):e37201.
- Li Y, Wehbe RM, Ahmad FS, Wang H, Luo Y. A comparative study of pretrained language models for long clinical text. *J Am Med Inform Assoc* 2023 Jan 18;30(2):340-7.
- Phatak A, Savage DW, Ohle R, Smith J, Mago V. Medical text simplification using reinforcement learning (TESLEA): Deep learning-based text simplification approach. *JMIR Med Inform* 2022;10(11):e38095.

This year, we noticed that our process produces an unexpected selection of candidates mainly from English-speaking countries, with an over-representation of papers from the USA. We hypothesize that, in this post-COVID-19 period, researchers from the US maintained their level of submissions while European researchers decreased the number of submitted papers.

In the next sections, we present the main issues and approaches addressed in the pre-selected publications.

3 Current Trends in Biomedical NLP

To present the current trends in biomedical NLP observed over the last year, we propose an analysis of the top 200 citations according to the scores computed automatically. We focus on the topics studied by researchers. For this, we first analyze the keywords provided by the authors (Section 4.1). Second, we analyze the languages addressed (Section

4.2) in this year's publications. Then, we group our observations along two lines: main approaches used (Section 4.3) and some frequent topics investigated (Sections 4.4 to 4.6). Useless to say that many other topics are also investigated but we cannot mention them all in this synoptic paper.

3.1 Analysis of Keywords from Publications

We studied the distribution of keywords used to index the 200 top papers. This distribution provides a global overview of the keywords. By comparison with the last year, we have observed the following trends in 2022:

- among the key-words similar to those from last year, we can distinguish two series: source of data (social media, social network, Twitter, Twitter analysis, digital pharmacovigilance, reddit, Twitch) and methods (transformers, learning transfer, word2vec). Let's notice that the frequency of transformers is significantly increased in 2022;

- in a continuation of last year's observations, several studies address the pandemics, and especially the COVID-19 period. The researchers have now more global data and can provide more exhaustive analyses related to the pandemics, the vaccines, their safety, and global information on this period. Several keywords (Covid-19, covid-19 surveillance, Covid-19 vaccine, Sars-cov-2, vaccine hesitancy, vaccine safety, vaccine adverse events, vaccine misinformation, vaccine acceptance, patient safety, vaccination) confirm the situation. Besides, other keywords represent the methods used for the analysis of these data (sentiment analysis, emotion analysis, emotion detection, opinion mining, patient experience, patient reported outcome measures). Hence, we can say that the COVID-19 period revealed the concern of patients towards the vaccination, which may be explained by the technology used for the creation of vaccines and little knowledge about their efficiency and safety. The first need of population is related to the security, which is now studied through sentiment and emotion analysis. Another need is to check out the veracity of information, which is now studied through the detection of fake news and misinformation, and through fact-checking. Social media remain the main source of content created by patients;
- among the new trends, we can find indexing terms related to methods (reinforcement learning, explainable IA, knowledge graph, knowledge graph embeddings) and tasks (disinformation, fake news detection, misinformation, medical errors). We can also notice technological evolutions of AI methods towards the reinforcement learning [3, 4] and knowledge graphs [5, 6].

Even if those keywords give a global view of work done in these papers, they only reflect a small and general part of the work that has been done by the authors. Indeed, keywords are essentially used for indexing papers while we used them to draw an overview of the research done in 2022. Nevertheless, the trends observed within this set of keywords are also reflected in the analysis provided in the following sections.

3.2 Languages Addressed

Contrary to previous COVID-19 and post-COVID-19 years, we finally witness another wave of research on non-English-language data. Indeed, several languages are addressed in the publications in 2022, which we indicate in what follows together with the topics addressed:

- *Arabic*: sentiment analysis in COVID-19 tweets [7], assessment of web page credibility [8];
- *Bengali*: detection of depression severity in social media posts [9];
- *Chinese*: ICD-coding [10], classification of patient-doctor dialogues [11], question-answering [12], clinical named entity recognition [13, 14, 15], prescription recommendation of traditional Chinese medicine [16];
- *Danish*: extraction of adverse drug reactions (ADRs) from clinical narratives [17];
- *Dutch*: predicting COVID-19 symptoms from clinical narratives [18], detecting changes in help seeker conversations on a suicide prevention during the COVID-19 pandemic [19];
- *French*: analysis of tweets about COVID-19 vaccines [20], coding of ADRs from patient reports [21], extraction of explicit and implicit cause-effect relationships from tweets [22], ICD-10 coding [23], construction of cohorts of similar patients [24], processing of electronic medical records [25], understanding of patient's answers in a French medical chatbot [26];
- *German*: evaluation of Transformers on clinical notes [27];
- *Greek*: improving the performance of localized healthcare virtual assistants [28];
- *Hindi*: classification of COVID-19 texts [29], chatbot for information sexual and reproductive health for young people [30];
- *Italian*: analysis of social media for quality of life in Parkinson's patients [31], sentiment analysis of opinion on COVID-19 vaccines [32,33], estimation of the incidence of infectious disease cases [34];
- *Japanese*: understanding psychiatric illness [35], detection of adverse events from narrative clinical documents [36];
- *Korean*: BERT model for processing med-

ical documents [37], sentiment analysis of tweets about COVID-19 vaccines [38];

- *Malayalam*: generation of synoptic clinical reports [39];
- *Polish*: prediction of cardiovascular diseases in electronic health records [40];
- *(Brazilian) Portuguese*: description of an annotated clinical corpus [41], ICD-10 coding [42];
- *Serbian*: sentiment analysis in COVID-19 tweets [43];
- *Spanish*: ICD-coding [10, 44], negation and uncertainty detection in clinical narratives [45], training and evaluation of word embeddings for the clinical domain [46];
- *Swedish*: ICD-10 coding [44];
- *Turkish*: sentiment analysis of tweets about COVID-19 vaccines [47].

Aside from English, which is addressed in a huge number of publications, the most frequently processed languages are Chinese, Japanese, Spanish, Arabic, and French. We can also notice new and rare languages like Bengali, Malayalam, Hindi, Greek, or Serbian. Another interesting fact is that some publications address multilingual data or data in several languages: ICD-10 coding in English, Spanish and Swedish [44], analysis of social media for quality of life in Parkinson's patients and their caregivers in English, French, Italian, Spanish, and German [31], term normalization using the UMLS [48].

The research work in English takes undoubtedly advantage of the existing datasets annotated within various challenges (I2B2, N2C2, KDD, etc.) and institutions (like MIMIC-III), as well as data from social media, hospitals, bibliographical datasets, clinical trials, etc. The research in other languages is possible mainly thanks to the availability of data from social media [7, 9, 19, 20, 22, 38, 43, 47] and documents from local hospitals [10, 13, 14, 17, 18, 23, 25, 27, 36, 37, 40, 42]. Besides, this set of works in languages other than English relies on the dedicated language models, which cover a great variety of languages by now. We expect that this trend will continue. We also hope that large annotated datasets in languages other than English will become available for the research.

3.3. Availability of Large Language Models as a Step towards the Democratization of NLP

Language models group together a set of methods based on word embeddings. Such models often encode several levels of linguistic knowledge, hence their efficiency and self-sufficiency. In the past few years, large language models based on Transformers have been created stating with the BERT (Bidirectional Encoder Representations from Transformers) model [49]. Hence, BERT has been conjugated within the biomedical domain through models like SciBERT [50], Clinical BERT [51], BioBERT [52], PubMedBERT [53], or BioM [54]. At the same time, the BERT model has been adapted to languages other than English: French [55, 56], Spanish [57], Dutch [58], Finish [59], Italian [60, 61], Portuguese [62], Japanese [63], etc. In some languages, medicine-specific BERT models have also been proposed, such as French DrBERT [64] or Japanese clinical BERT [65]. This set of Transformer models is successfully exploited for several tasks such as categorization, POS-tagging, semantic similarity, named entity and relation extraction.

It is important to notice that these language models are freely available for research purpose, which opens new possibilities for those interested in their testing and use within various clinical and health-related tasks. We can even tell that the availability of such models leads to the democratization of this NLP approach. Indeed, without the help of NLP researchers, biologists, radiologists, pharmacists and other clinicians can now use these models within the clinical context and we can find several such experiments, such as: analysis of tweets for user opinions and side effects on COVID-19 vaccines [66, 33], fact-checking of posts on COVID-19 vaccines [67], identification of social determinants of health in EHRs [68], analysis of literature for drug-induced liver injury [69], labeling of diagnosis in cardiovascular Magnetic resonance imaging (MRI) [70], analysis of social media on the quality of life in Parkinson's patients [31], extraction of biomedical relations from the scientific literature [71].

More recently, the GPT (Generative Pre-trained Transformer) models [72] have been proposed for various generation tasks. These

models are also coming into the clinical domain but with only few works published in 2022: creation of BioGPT (Generative Pre-trained Transformer for biomedical text generation and mining) [73], prediction and suggestion of medical text in dental medical notes [74], challenges for GPT-3 in ophthalmology [75]. This generative model proposes text on the basis of the training corpora. Even if huge training corpora are used, the models do not cover the whole language and, more specifically, all situations. For instance, such models cannot describe a given patient, his lab or MRI results. This limits seriously the use of GPTs within the clinical domain, in which reliable and realistic data are required. Yet, the GPT-generated text can be used as suggestions which must be verified and approved by human experts. For instance, the application related to the automatic text completion can show its utility since the text is approved by human users.

3.4 Social Media as the Preferred Source of Information

Social media prove to remain the preferred source of information for researchers in several situations:

- when other sources of information are not freely available, such as in languages other than English [7, 9, 19, 20, 22, 38, 43, 47];
- when researchers investigate questions related to patients and population, while these questions are not discussed with medical doctors or require large population samples. We can mention for instance sentiment analysis on medication and vaccines [7, 20, 33, 38, 43, 47], and adverse drug effects [76, 77];
- when mental health of patients is concerned in cases like depression [9, 4, 78], eating disorders [79], suicide detection and prevention [19, 80], quality of life of patients [31, 81], and drug misuse [82-84].

We expect that, for these topics at least, social media continue to occupy an important place in the research community.

3.5 Typical NLP tasks: IE and Prediction

Information extraction (IE) has the purpose to localize within narrative documents exact pieces of information (drug, disorder, age of patients...) and to extract them for further processing. Prediction is related to the classification of texts or text spans into a given class. For several years now, these two tasks are well processed and provide reliable results for a given corpus, class, hospital, etc. Several clinical questions are addressed, such as: extraction of clinical information [15, 17, 85, 86], ICD-10 coding of medical records [10, 23, 42, 44], prediction of diseases [18, 40, 87, 88, 89], mortality [90], risks [91], and patient outcomes [89, 92, 93]. IE and prediction tasks will certainly continue to perform on clinical and health data in the next years. Providing reliable and robust models, which may work on data from different sources, should be one of the objectives for future research.

3.6 Patients as Actors of the Healthcare Process

Recently, patients became one of the true actors within the healthcare process. For instance, their opinions matter in the decision-making process, therapeutic choice, procedures, etc. Besides, medical doctors are interested in discovering the everyday life of patients, their quality of life, their opinion on the healthcare-related aspects, etc. Publications from 2022 represent these research questions. We have already mentioned various issues on the COVID-19 vaccine perception and hesitancy [7, 20, 33, 66]. Other observations on patients are also important, such as classification of self-harm behaviors in EHRs [94], analysis of unnoticed and unresolved safety incidents in patients [95], analysis of web-based reviews of sanctioned physicians [96], and patient experience and satisfaction in online reviews [97].

Since patients are part of the healthcare process, they can access their medical records. In relation with this, one work proposes to study stigmatizing language in the EHRs [98] in order to protect patients. More generally, patients should also be able

to understand the content of medical and clinical documents. The main difficulty is that training and knowledge of patients do not provide the necessary basis for such understanding. One solution is to make clinical content more patient-friendly thanks to the text simplification [99, 3, 100, 101, 102]. This research question is addressed more frequently now. Even if these works are done on English-language documents, the interest in this issue is important. We expect that, in future, needs and requirements of patients will continue to attract researchers.

4 Conclusion

In 2022, we can observe the evolution of the research in several directions. The most interesting direction is certainly related to the democratization of the NLP methods. Indeed, large language models built with Transformers, which encode several levels of linguistic knowledge, are now freely available for research. Moreover, several of such models are adapted or fine-tuned to the data from the medical domain. This means that clinicians, often without the help of NLP researchers, can use the available language models for different tasks, such as IE, categorization and prediction, coding, or computing of text similarity. This trend does not really contribute to the NLP research, yet it permits to test the available models in different contexts and on different datasets, and to reveal the current limitations of these models.

The technological evolution of AI methods towards the reinforcement learning and knowledge graphs is yet another new trend, which may develop in the following years.

Another evolution is related to the renewal of interest to languages other than English. First, the language models are also adapted to or created for these languages. Second, while standard annotated datasets are rarely available for languages other than English, researchers exploit data from social media and from local hospitals instead. In continuation of the past trends, IE and prediction tasks are widely addressed by researchers and clinicians on various datasets. These methods are quite efficient in precise

contexts and on some categories. Let's also notice that patients continue to occupy an important role within the healthcare process and that several works are dedicated to the needs of patients.

References

1. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural Language Processing: an introduction. *J Am Med Inform Assoc* 2011 Sep-Oct;18(5):544-51. doi: 10.1136/amia-jnl-2011-000464.
2. Friedman C and Hripesak G. Natural language processing and its future in medicine. *Acad Med* 1999 Aug;74(8):890-5. doi: 10.1097/00001888-199908000-00012.
3. Phatak A, Savage DW, Ohle R, Smith J, Mago V. Medical text simplification using reinforcement learning (TESLEA): Deep learning-based text simplification approach. *JMIR Med Inform* 2022 Nov 18;10(11):e38095. doi: 10.2196/38095.
4. Cui B, Wang J, Lin H, Zhang Y, Yang L, Xu B. Emotion-based reinforcement attention network for depression detection on social media: Algorithm development and validation. *JMIR Med Inform* 2022 Aug 9;10(8):e37818. doi: 10.2196/37818.
5. Sima AC, Mendes de Farias R, Anisimova M, Dessimoz C, Robinson-Rechavi M, Zbinden E, et al. Bio-SODA UX: enabling natural language question answering over knowledge graphs with user disambiguation. *Distrib Parallel Databases* 2022;40(2-3):409-40. doi: 10.1007/s10619-022-07414-w.
6. Kaur M, Costello J, Willis E, Kelm K, Reformat MZ, Bolduc FV. Deciphering the diversity of mental models in neurodevelopmental disorders: Knowledge graph representation of public data using natural language processing. *J Med Internet Res* 2022 Aug 5;24(8):e39888. doi: 10.2196/39888.
7. Albahli S. Twitter sentiment analysis: An Arabic text mining approach based on COVID-19. *Front Public Health* 2022 Oct 10;10:966779. doi: 10.3389/fpubh.2022.966779.
8. Alasmari A, Alhothali A, Allinjaw A. Hybrid machine learning approach for Arabic medical web page credibility assessment. *Health Informatics J* 2022 Jan-Mar;28(1):14604582211070998. doi: 10.1177/14604582211070998.
9. Kabir MK, Islam M, Kabir ANB, Haque A, Rhaman MdK. Detection of depression severity using Bengali social media posts on mental health: Study using natural language processing techniques. *JMIR Form Res* 2022 Sep 28;6(9):e36118. doi: 10.2196/36118.
10. Shuai Z, Xiaolin D, Jing Y, Yanni H, Meng C, Yuxin W, et al. Comparison of different feature extraction methods for applicable automated ICD coding. *BMC Med Inform Decis Mak* 2022 Jan 12;22(1):11. doi: 10.1186/s12911-022-01753-5.
11. Sun Y, Gao D, Shen X, Li M, Nan J, Zhang W. Multilabel classification in patient-doctor dialogues with the RoBERTa-WWM-ext + CNN (robustly optimized bidirectional encoder representations from transformers pretraining approach with whole word masking extended combining a convolutional neural network) model: Named entity study. *JMIR Med Inform* 2022 Apr 21;10(4):e35606. doi: 10.2196/35606.
12. Zhang L, Yang X, Li S, Liao T, Pan G. Answering medical questions in Chinese using automatically mined knowledge and deep neural networks: an end-to-end solution. *BMC Bioinformatics* 2022 Apr 15;23(1):136. doi: 10.1186/s12859-022-04658-2.
13. An Y, Xia X, Chen X, Wu FX, Wang J. Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF. *Artif Intell Med* 2022 May;127:102282. doi: 10.1016/j.artmed.2022.102282.
14. Li C, Ma K. Entity recognition of Chinese medical text based on multi-head self-attention combined with BiLSTM-CRF. *Math Biosci Eng* 2022;19(3):2206-18.
15. Guo S, Yang W, Han L, Song X, Wang G. A multi-layer soft lattice based model for Chinese clinical named entity recognition. *BMC Med Inform Decis Mak* 2022 Jul 30;22(1):201. doi: 10.1186/s12911-022-01924-4.
16. Zhang H, Zhang J, Ni W, Jiang Y, Liu K, Sun D, et al. Transformer- and generative adversarial network-based inpatient traditional Chinese medicine prescription recommendation: Development study. *JMIR Med Inform* 2022 May 31;10(5):e35239. doi: 10.2196/35239.
17. Kaas-Hansen BS, Placido D, Rodríguez CL, Thorsen-Meyer HC, Gentile S, Nielsen AP, et al. Language-agnostic pharmacovigilant text mining to elicit side effects from clinical notes and hospital medication records. *Basic Clin Pharmacol Toxicol* 2022 Oct;131(4):282-93. doi: 10.1111/bcpt.13773.
18. Van Olmen J, Van Nooten J, Philips H, Sollie A, Daelemans W. Predicting COVID-19 symptoms from free text in medical records using artificial intelligence: Feasibility study. *JMIR Med Inform* 2022 Apr 27;10(4):e37771. doi: 10.2196/37771.
19. Salmi S, Mérelle S, Gilissen R, van der Mei R, Bhulai S. Detecting changes in help seeker conversations on a suicide prevention helpline during the COVID-19 pandemic: in-depth analysis using encoder representations from transformers. *BMC Public Health* 2022 Mar 18;22(1):530. doi: 10.1186/s12889-022-12926-2.
20. Sauvayre R, Vernier J, Chauvière C. An analysis of French-language tweets about COVID-19 vaccines: Supervised learning approach. *JMIR Med Inform* 2022 May 17;10(5):e37831. doi: 10.2196/37831.
21. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Saf* 2022 May;45(5):535-48. doi: 10.1007/s40264-022-01153-8.
22. Ahne A, Khetan V, Tannier X, Rizvi MdIH, Czernichow T, et al. Extraction of explicit and implicit cause-effect relationships in patient-reported diabetes-related tweets from 2017 to 2021: Deep learning approach. *JMIR Med Inform* 2022 Jul

- 19;10(7):e37201. doi: 10.2196/37201.
23. Falissard L, Morgand C, Ghosn W, Imbaud C, Bounebach K, Rey G. Neural translation and automated recognition of ICD-10 medical entities from natural language: Model development and performance assessment. *JMIR Med Inform* 2022 Apr 11;10(4):e26353. doi: 10.2196/26353.
24. Gérardin C, Mageau A, Mékinian A, Tannier X, Carrat F. Construction of cohorts of similar patients from automatic extraction of medical concepts: Phenotype extraction study. *JMIR Med Inform* 2022 Dec 19;10(12):e42379. doi: 10.2196/42379.
25. Schiappa R, Contu S, Culie D, Thamphyha B, Chateau Y, Gal J, et al. RUBY: Natural language processing of french electronic medical records for breast cancer research. *JCO Clin Cancer Inform* 2022 Jul;6:e2100199. doi: 10.1200/CCI.21.00199.
26. Blanc C, Bailly A, Francis E, Guillotin T, Jamal F, Wakim B, et al. FlauBERT vs. CamembERT: Understanding patient's answers by a French medical chatbot. *Artif Intell Med* 2022 May;127:102264. doi: 10.1016/j.artmed.2022.102264.
27. Lentzen M, Madan S, Lage-Rupprecht V, Kühnel L, Fluck J, Jacobs M, et al. Critical assessment of transformer-based AI models for German clinical notes. *JAMIA Open* 2022 Nov 15;5(4):oac087. doi: 10.1093/jamiaopen/ooac087.
28. Malamas N, Papangelou K, Symeonidis AL. Upon improving the performance of localized healthcare virtual assistants. *Healthcare (Basel)* 2022 Jan 4;10(1):99. doi: 10.3390/healthcare10010099.
29. Jain V, Kashyap KL. Ensemble hybrid model for Hindi COVID-19 text classification with metaheuristic optimization algorithm. *Multimed Tools Appl*. 2023;82(11):16839-16859. doi: 10.1007/s11042-022-13937-2.
30. Wang H, Gupta S, Singhal A, Muttreja P, Singh S, Sharma P, et al. An artificial intelligence chatbot for young people's sexual and reproductive health in India (SnehAI): Instrumental case study. *J Med Internet Res* 2022 Jan 3;24(1):e29969. doi: 10.2196/29969.
31. Damier P, Henderson EJ, Romero-Imbroda J, Galimam L, Kronfeld N, Warnecke T. Impact of off-time on quality of life in Parkinson's patients and their caregivers: Insights from social media. *Parkinsons Dis* 2022 Dec 3;2022:1800567. doi: 10.1155/2022/1800567.
32. Stracqualursi L, Agati P. Covid-19 vaccines in Italian public opinion: Identifying key issues using twitter and natural language processing. *PLoS One* 2022 Nov 17;17(11):e0277394. doi: 10.1371/journal.pone.0277394.
33. Semeraro A, Vilella S, Ruffo G, Stella M. Emotional profiling and cognitive networks unravel how mainstream and alternative press framed AstraZeneca, Pfizer and COVID-19 vaccination campaigns. *Sci Rep* 2022 Aug 24;12(1):14445. doi: 10.1038/s41598-022-18472-6.
34. Lanera C, Baldi I, Francavilla A, Barbieri E, Tramontan L, Scamarcia A, et al. A deep learning approach to estimate the incidence of infectious disease cases for routinely collected ambulatory records: The example of Varicella-Zoster. *Int J Environ Res Public Health* 2022 May 13;19(10):5959. doi: 10.3390/ijerph19105959.
35. Kishimoto T, Nakamura H, Kano Y, Eguchi Y, Kitazawa M, Liang-KC, et al. Understanding psychiatric illness through natural language processing (UNDERPIN): Rationale, design, and methodology. *Front Psychiatry* 2022 Dec 1;13:954703. doi: 10.3389/fpsy.2022.954703.
36. Mashima Y, Tamura T, Kunikata J, Tada S, Yamada A, Tanigawa M, et al. Using natural language processing techniques to detect adverse events from progress notes due to chemotherapy. *Cancer Inform* 2022 Mar 22;21:11769351221085064. doi: 10.1177/11769351221085064.
37. Kim Y, Kim JH, Lee JM, Jang MJ, Yum YJ, Kim, S et al. *Sci Rep* 2022 Aug 16;12(1):13847. doi: 10.1038/s41598-022-17806-8.
38. Eom G, Yun S, Byeon H. Predicting the sentiment of South Korean Twitter users toward vaccination after the emergence of COVID-19 Omicron variant using deep learning-based natural language processing. *Front Med (Lausanne)* 2022 Sep 14;9:948917. doi: 10.3389/fmed.2022.948917.
39. Tan WM, Teoh KH, Ganggayah MD, Taib NA, Zaini HS, Dhillon SK. Automated generation of synoptic reports from narrative pathology reports in university Malaya medical center using natural language processing. *Diagnostics (Basel)* 2022 Apr 1;12(4):879. doi: 10.3390/diagnostics12040879.
40. Anetta K, Horak A, Wojakowski W, Wita K, Jadczyk T. Deep learning analysis of Polish electronic health records for diagnosis prediction in patients with cardiovascular diseases. *J Pers Med* 2022 May 25;12(6):869. doi: 10.3390/jpm12060869.
41. Oliveira LESE, Peters AC, Pucca da Silva AM, Gebelua CP, Gumiel YB, Cintho LMM, et al. SemClinBr - a multi-institutional and multi-specialty semantically annotated corpus for portuguese clinical NLP tasks. *J Biomed Semantics* 2022 May 8;13(1):13. doi: 10.1186/s13326-022-00269-1.
42. Coutinho I, Martins B. Transformer-based models for ICD-10 coding of death certificates with Portuguese text. *J Biomed Inform* 2022 Dec;136:104232. doi: 10.1016/j.jbi.2022.104232.
43. Ljajić A, Prodanović N, Medvečki D, Bašaragin B, Mitrović J. Uncovering the reasons behind COVID-19 vaccine hesitancy in Serbia: Sentiment-based topic modeling. *J Med Internet Res* 2022 Nov 17;24(11):e42261. doi: 10.2196/42261.
44. Blanco A, Remmer S, Pérez A, Dalianis H, Casillas A. Implementation of specialised attention mechanisms: ICD-10 classification of gastrointestinal discharge summaries in English, Spanish and Swedish. *J Biomed Inform* 2022 Jun;130:104050. doi: 10.1016/j.jbi.2022.104050.
45. Pabón OS, Montenegro O, Torrente M, González AR, Provencio M, Menasalvas E. Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach. *PeerJ Comput Sci* 2022 Mar 7;8:e913. doi: 10.7717/peerj-cs.913.
46. Chiu C, Villena F, Martin K, Núñez F, Besa C, Dunstan J. Training and intrinsic evaluation of lightweight word embeddings for the clinical domain in Spanish. *Front Artif Intell* 2022 Sep 21;5:970517. doi: 10.3389/frai.2022.970517.
47. Mermer G, Özsezer G. Discussions about COVID-19 vaccination on Twitter in Turkey: Sentiment analysis. *Disaster Med Public Health Prep* 2022 Oct 13;17:e266. doi: 10.1017/dmp.2022.229.
48. Yuan Z, Zhao Z, Sun H, Li J, Wang F, Yu S. CODER: Knowledge-infused cross-lingual medical term embedding for term normalization. *J Biomed Inform* 2022 Feb;126:103983. doi: 10.1016/j.jbi.2021.103983.
49. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proc of NAACL-HLT 2019, Minneapolis, Minnesota. Association for Computational Linguistics*; 2019. p. 4171-86.
50. Beltagy I, Lo K, Cohan A. SciBERT: A pre-trained language model for scientific text. In: *Proc of the 2019 Conf on Empirical Methods in Natural Language Processing and the 9th Inter Joint Conf on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, November 2019. Association for Computational Linguistics; 2019. p. 3615-20.
51. Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. In: *Proc of the 2nd Clinical Natural Language Processing Workshop*, 2019. p. 72-8.
52. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020 Feb 15;36(4):1234-40. doi: 10.1093/bioinformatics/btz682.
53. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans Comput Healthc* 2020;3(1):1-23.
54. Alrowili S, Shanker V. BioM-transformers: Building large biomedical language models with BERT, ALBERT and ELECTRA. In: *Proc of the 20th Workshop on Biomedical Language Processing*, Online, June 2021. Association for Computational Linguistics; 2021. p. 221-7.
55. Le H, Vial L, Frej J, Segonne V, Coavoux M, Lecouteux B, et al. FlauBERT: Unsupervised language model pre-training for French. In: *Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 2020. European Language Resources Association*; 2020. p. 2479-90.
56. Martin L, Muller B, Suárez PJO, Dupont Y, Romary L, Villemonte de La Clergerie E, et al. CamembERT: a tasty French language model. In: *ACL 2020 - 58th Annual Meeting of the Association for Computational Linguistics, Seattle / Virtual, United States*, 2020.
57. Cañete J, Chaperon G, Fuentes R, Ho JH, Kang H, and Pérez J. Spanish pre-trained bert model and evaluation data. In: *PML4DC at ICLR 2020*; 2020.
58. de Vries W, van Cranenburgh A, Bisazza A, Caselli T, van Noord G, and Nissim M. Bertje

- MN: A dutch bert model. ArXiv, December 2019.
59. Virtanen A, Kanerva J, Ilo R, Luoma J, Luotolahti J, Salakoski T, et al. Multilingual is not enough: BERT for Finnish. ArXiv, 2019.
 60. Polignano M, Basile P, de Gemmis M, Semeraro G, Basile V. AlBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In: Proceedings of the Sixth Italian Conference on Computational Linguistics (Clic-it 2019), volume 2481. CEUR; 2019.
 61. Parisi L, Francia S, and Magnani P. Umberto: an italian language model trained with whole word masking. <https://github.com/musixmatchresearch/umberto>. 2020.
 62. Souza F, Nogueira R, Lotufo R. Bertimbau: Pretrained bert models for Brazilian Portuguese. In Ricardo Cerri and Ronaldo C. Prati, editors, Intelligent Systems. Cham: Springer International Publishing; 2020. p. 403-17
 63. Shibayama N and Shinnou H. Construction and evaluation of Japanese sentence-BERT models. In Proc of the 35th Pacific Asia Conf on Language, Information and Computation, Shanghai, China, 11 2021. Association for Computational Linguistics; 2021. p. 731-8.
 64. Labrak Y, Bazoge A, Dufour R, Rouvier M, Morin E, Daille B, et al. DrBERT: A robust pre-trained model in French for biomedical and clinical domains. In: ACL, editor. Proc of the 61st Ann Meeting of the Ass for Comp Linguistics, Toronto (CA), Canada; 2023.
 65. Kawazoe Y, Shibata D, Shinohara E, Aramaki E, Ohe K. A clinical specific BERT developed using a huge Japanese clinical text corpus. PLoS One 2021 Nov 9;16(11):e0259763. doi: 10.1371/journal.pone.0259763.
 66. Portelli B, Scabro S, Tonino R, Chersoni E, Santus E, Serra G. Monitoring user opinions and side effects on COVID-19 vaccines in the twittersphere: Infodemiology study of tweets. J Med Internet Res 2022 May 13;24(5):e35115. doi: 10.2196/35115.
 67. Xue H, Gong X, Stevens H. COVID-19 vaccine fact-checking posts on facebook: Observational study. J Med Internet Res 2022 Jun 21;24(6):e38423. doi: 10.2196/38423.
 68. Rouillard CJ, Nasser MA, Hu H, Roblin DW. Evaluation of a Natural Language Processing approach to identify social determinants of health in electronic health records in a diverse community cohort. Med Care 2022 Mar 1;60(3):248-55. doi: 10.1097/MLR.0000000000001683.
 69. Rathee S, MacMahon M, Liu A, Katritsis NM, Youssef G, Hwang W, et al. DILI C : An AI-based classifier to search for drug-induced liver injury literature. Front Genet 2022 Jun 29;13:867946. doi: 10.3389/fgene.2022.867946.
 70. Zaman S, Petri C, Vimallesvaran K, Howard J, Bharath A, Francis D, et al. Automatic diagnosis labeling of cardiovascular MRI by using semisupervised natural language processing of text reports. Radiol Artif Intell 2021 Nov 24;4(1):e210085. doi: 10.1148/ryai.210085.
 71. Bhasuran B. BioBERT and similar approaches for relation extraction. Methods Mol Biol 2022;2496:221-35. doi: 10.1007/978-1-0716-2305-3_12.
 72. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems, volume 33. Curran Associates, Inc.; 2020. p. 1877-1901.
 73. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Brief Bioinform 2022 Nov 19;23(6):bbac409. doi: 10.1093/bib/bbac409.
 74. Sirrianni J, Sezgin E, Claman D, Linwood SL. Medical text prediction and suggestion using generative pretrained transformer models with dental medical notes. Methods Inf Med 2022 Dec;61(5-06):195-200. doi: 10.1055/a-1900-7351.
 75. Nath S, Marie A, Ellershaw S, Korot E, Keane PA. New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. Br J Ophthalmol 2022 Jul;106(7):889-92. doi: 10.1136/bjophthalmol-2022-321141.
 76. Habibabadi SK, Haghighi PD, Burstein F, Buttery J. Vaccine adverse event mining of twitter conversations: 2-phase classification study. JMIR Med Inform 2022 Jun 16;10(6):e34305. doi: 10.2196/34305.
 77. Dirkson A, Verberne S, van Oortmerssen G, Gelderblom H, Kraaij W. How do others cope? Extracting coping strategies for adverse drug events from social media. J Biomed Inform 2023 Mar;139:104228. doi: 10.1016/j.jbi.2022.104228.
 78. Yu D, Vydiswaran VGV. An assessment of mentions of adverse drug events on social media with natural language processing: Model development and analysis. JMIR Med Inform 2022 Sep 28;10(9):e38140. doi: 10.2196/38140.
 79. Benítez-Andrades JA, Alija-Pérez JM, Vidal ME, Pastor-Vargas R, García-Ordás MR. Traditional machine learning models and bidirectional encoder representations from transformer (BERT)-based automatic classification of tweets about eating disorders: Algorithm development and validation study. JMIR Med Inform 2022 Feb 24;10(2):e34492. doi: 10.2196/34492.
 80. Metzler H, Baginski H, Niederkrotenthaler T, Garcia D. Detecting potentially harmful and protective suicide-related content on twitter: Machine learning approach. J Med Internet Res 2022 Aug 17;24(8):e34705. doi: 10.2196/34705.
 81. Merhbene G, Nath S, Puttick AR, Kurpicz-Briki M. Burnoutensemble: Augmented intelligence to detect indications for burnout in clinical psychology. Front Big Data 2022 Apr 5;5:863100. doi: 10.3389/fdata.2022.863100.
 82. Golder S, Weissenbacher D, O'Connor K, Hennessey S, Gross R, Gonzalez Hernandez G. Patient-reported reasons for switching or discontinuing statin therapy: A mixed methods study using social media. Drug Saf 2022 Sep;45(9):971-81. doi: 10.1007/s40264-022-01212-0.
 83. Sarker A, Nataraj N, Siu W, Li S, Jones CM, Sumner SA. Concerns among people who use opioids during the COVID-19 pandemic: a natural language processing analysis of social media posts. Subst Abuse Treat Prev Policy 2022 Mar 5;17(1):16. doi: 10.1186/s13011-022-00442-w.
 84. Sarker A, Al-Garadi MA, Ge Y, Nataraj N, Jones CM, Sumner SA. Signals of increasing co-use of stimulants and opioids from online drug forum data. Harm Reduct J 2022 May 25;19(1):51. doi: 10.1186/s12954-022-00628-2.
 85. Segura-Bedmar I, Camino-Perdones D, Guerrero-Aspizua S. Exploring deep learning methods for recognizing rare diseases and their clinical manifestations from texts. BMC Bioinformatics 2022 Jul 6;23(1):263. doi: 10.1186/s12859-022-04810-y.
 86. Chen Z, Zhang H, Yang X, Wu S, He X, Xu J, et al. Assess the documentation of cognitive tests and biomarkers in electronic health records via natural language processing for Alzheimer's disease and related dementias. Int J Med Inform 2023 Feb;170:104973. doi: 10.1016/j.ijmedinf.2022.104973.
 87. Penfold RB, Carrell DS, Cronkite DJ, Pabiniak C, Dodd T, Glass AM, et al. Development of a machine learning model to predict mild cognitive impairment using natural language processing in the absence of screening. BMC Med Inform Decis Mak 2022 May 12;22(1):129. doi: 10.1186/s12911-022-01864-z.
 88. Syed ARP, Anbalagan R, Setlur AS, Karunakaran C, Shetty J, Kumar J, et al. Implementation of ensemble machine learning algorithms on exome datasets for predicting early diagnosis of cancers. BMC Bioinformatics 2022 Nov 18;23(1):496. doi: 10.1186/s12859-022-05050-w.
 89. Song J, Hobensack M, Bowles KH, McDonald MV, Cato K, Rossetti SC, et al. Clinical notes: An untapped opportunity for improving risk prediction for hospitalization and emergency department visit during home health care. Biomed Inform 2022 Apr;128:104039. doi: 10.1016/j.jbi.2022.104039.
 90. Chen PF, Chen L, Lin YK, Li GH, Lai F, Lu CW, et al. Predicting postoperative mortality with deep neural networks and natural language processing: Model development and validation. JMIR Med Inform 2022 May 10;10(5):e38241. doi: 10.2196/38241.
 91. Uronen L, Salanterä S, Hakala K, Hartiala J, Moen H. Combining supervised and unsupervised named entity recognition to detect psychosocial risk factors in occupational health checks. Int J Med Inform 2022 Apr;160:104695. doi: 10.1016/j.ijmedinf.2022.104695.
 92. Wang N, Wang M, Zhou Y, Liu H, Wei L, Fei X, et al. Sequential data-based patient similarity framework for patient outcome prediction: Algorithm development. J Med Internet Res 2022 Jan 6;24(1):e30720. doi: 10.2196/30720.
 93. Fink M, Kades K, ABischoff A, Moll M, Schnell M, Küchler M, et al. Deep learning-based assessment of oncologic outcomes from natural language processing of structured radiology reports. Radiol Artif Intell 2022 Jul 20;4(5):e220055. doi: 10.1148/ryai.220055.
 94. Burnett A, Chen N, Zeritis S, Ware S, McGillivray L, Shand F, et al. Machine learning algorithms to classify self-harm behaviours in New South Wales ambulance electronic medical records: A retrospective study. Int J Med Inform 2022 May;161:104734. doi: 10.1016/j.

- ijmedinf.2022.104734.
95. Gillespie A, Reader TW. Online patient feedback as a safety valve: An automated language analysis of unnoticed and unresolved safety incidents. *Risk Anal* 2022 Aug 9. doi: 10.1111/risa.14002.
96. Barnett J, Bjarnadóttir MV, Anderson D, Chen C. Understanding gender biases and differences in web-based reviews of sanctioned physicians through a machine learning approach: Mixed methods study. *JMIR Form Res* 2022 Sep 8;6(9):e34902. doi: 10.2196/34902.
97. Seltzer EK, Guntuku SC, Lanza AL, Tufts C, Srinivas SK, Klinger EV, et al. Patient experience and satisfaction in online reviews of obstetric care: Observational study. *JMIR Form Res* 2022 Mar 31;6(3):e28379. doi: 10.2196/28379.
98. Himmelstein G, Bates D, Zhou L. Examination of stigmatizing language in the electronic health record. *JAMA Netw Open* 2022 Jan 4;5(1):e2144967. doi: 10.1001/jamanetworkopen.2021.44967.
99. Ferrari A, Pirrotta L, Bonciani M, Venturi G, Vainieri M. Higher readability of institutional websites drives the correct fruition of the abortion pathway: A cross-sectional study. *PLoS One* 2022 Nov 4;17(11):e0277342. doi: 10.1371/journal.pone.0277342.
100. Saeed N, Naveed H. Medical terminology-based computing system: a lightweight post-processing solution for out-of-vocabulary multi-word terms. *Front Mol Biosci* 2022 Aug 12;9:928530. doi: 10.3389/fmolb.2022.928530.
101. Hendawi R, Alian S, Li J. A smart mobile app to simplify medical documents and improve health literacy: System design and feasibility validation. *JMIR Form Res* 2022 Apr 1;6(4):e35069. doi: 10.2196/35069.
102. Wang Z, Fan Y, Lv H, Deng S, Xie H, Zhang L, et al. The gap between self-rated health information literacy and internet health information-seeking ability for patients with chronic diseases in rural communities: Cross-sectional study. *J Med Internet Res* 2022 Jan 31;24(1):e26308. doi: 10.2196/26308.

Correspondence to:

Cyril Grouin
 Université Paris-Saclay
 CNRS, LISN
 Campus universitaire
 91405 Orsay, France
 E-mail: cyril.grouin@lisn.upsaclay.fr

Appendix: Content Summaries of Best Papers for the Natural Language Processing Section of the 2023 IMIA Yearbook

Ahne A, Khetan V, Tannier X, Rizvi MdIH, Czernichow T, Orchard F, Bour C, Aano A, Fagherazzi G

Extraction of explicit and implicit cause-effect relationships in patient-reported diabetes-related tweets from 2017 to 2021: Deep learning approach

JMIR Med Inform 2022;10(7):e37201. doi:10.2196/37201

In this paper, the authors aim at providing a deep learning-based method to extract implicit and explicit relations of cause-effect for diabetes from tweets, and a methodology to understand opinions/feelings reported by patients from a causality perspective. They fine-tuned a BERTweet model on 562,000 tweets annotated with emotion information in order to detect causal sentences. Then, they designed a CRF model using BERTweet features to identify

possible cause-effects associations from 265,000 causal sentences. This method allows the authors to obtain several clusters for cause-effect (diabetes, death, insulin), including emotions (anger, fear, sadness) reported by diabetes patients.

Li Y, Wehbe RM, Ahmad FS, Wang H, Luo Y

A comparative study of pretrained language models for long clinical text

J Am Med Inform Assoc 2023 Jan 18;30(2):340-7. doi:10.1093/jamia/ocac225

This paper proposes to enrich Transformers models with clinical knowledge, which allows the authors to achieve state-of-the-art results on biomedical NLP tasks. Nevertheless, the authors highlight that the self-attention mechanism uses a lot of memory and does not allow to process long texts (limitation of 512 sub-units: e.g., discharge summaries from MIMIC have 2,984 tokens on average). They produced two domain-enriched language models based on Longformer (Clinical-Longformer) and BigBird (Clinical-BigBird) to process up to 4,096 sub-units. Those models outperformed existing models (BERT, RoBERTa, BioBERT, and ClinicalBERT) on three

tasks (NLI @ medNLI ; QA @ emrQA-relations ; NER @ i2b2 2014). We notice that the source code is available.

Phatak A, Savage DW, Ohle R, Smith J, Mago V

Medical Text Simplification Using Reinforcement Learning (TESLEA): Deep Learning-Based Text Simplification Approach

JMIR Med Inform 2022 Nov 18;10(11):e38095. doi: 10.2196/38095

The authors of this paper highlight that abstracts of scientific papers are publicly available, but they are hard to understand due to the use of medical vocabulary. They develop a text simplification method based on deep-learning trained on 3,568 complex-simple paragraphs (training) and evaluated on 480 paragraphs. Several scores are used to evaluate all aspects: FKGL (Flesch-Kincaid Grade Level), ROUGE, SARI (Simplified Automatic Readability Index), Likert scale. In addition, several examples of generated medical paragraphs are given in the paper, including texts generated by other systems (BART fine-tuned, BART-UL, MUSS, Keep-It-Simple, PEGASUS), which allows to compare all produced outputs.