DOI: 10.69085/vi20257157

HOW AI COULD TRANSFORM MEDICAL ENGLISH LINGUISTICS: BIOMEDICAL TEXT TOKENIZATION OF THE DISEASE IS ENEMY METAPHOR WITH BioBERT

Slaveyah Goranova
Medical University „Prof. Dr. P. Stoyanov" – Varna

The development of Artificial Intelligence (AI) is rapidly transforming many aspects of our lives. In the present paper, as part of the emerging theories for AI application in healthcare and medical sciences, we suggest that AI can be applied to the study of medical metaphors. In particular, we examine the possibility of biomedical text tokenization using BioBERT of the DISEASE is ENEMY metaphor. The present research is an initial stage of a much larger research in the context of the Conceptual Metaphor Theory and Discourse Analysis as applied to war metaphors in medical discourse.

Keywords: AI, war metaphors, Conceptual Metaphor Theory, Discourse Analysis

The advent of AI

The development of Artificial Intelligence (AI) is transforming our lives at an unprecedented speed. Its cope is also impressive: it is hard to find an area of our lives that has not already felt the effects of AI-involvement.

Being a central sphere of human existence, healthcare cannot be left behind in this AI-enacted transformation. There is a wide array of AI-applications in healthcare: from enhancing diagnostic imaging to elaborating new therapies. The diversity is impressive: AI is already changing clinical trials, maternal care, healthcare robotics, genetics data-driven medicines, equipment such as the stethoscope (Shaheen 2021). While AI-stirred advances in the health sciences are typically associated with disease diagnosis (especially with diagnostic imaging), drug discovery, electronic health records, or telemedicine, a key benefit of AI seems to be often overlooked: it is, namely, the application of AI to the study of medical discourse.

Metaphors in medical discourse

In the emerging field of AI applications in healthcare, we find the **AI-driven analysis of medical discourse** particularly intriguing. Studying medical discourse has been gaining in popularity due to the social significance

of healthcare and the discovery that discourse can actually affect health outcomes. In particular, the way we refer to diseases can influence our ability to deal with them. Our research does not need to rely on non-traditional books such as Jill Edwards' *Conscious Medicine* (Edwards 2010) and Francoise Bourzat's *Consciousness Medicine* (Bourzat 2019): it is enough to pinpoint to the many scientific articles dedicated to the *placebo* and, especially, the *nocebo* effects: „The Nocebo Effect and Its Relevance for Clinical Practice" (Colloca and Miller 2011), „When Words Are Painful: Unveiling the Mechanism of the Nocebo Effect" (Benedetti et al 2007), „A Systematic Review of Factors that Contribute to Nocebo Effects" (Webster et al 2016) to name just a few. In other words, what non-traditional, holistic, and alternative medicine has long been based on has recently found its way into Western medical thought, and scientifically sound evidence has been collected in service of the relationship between the way we think and talk about a medical condition and the healthcare outcome for the patient.

A singular domain of such a relationship can be found in **metaphors**. One of today's most influential critics of the use of medical metaphors, Susan Sontag, claims that it is not currently „morally permissible […] to use cancer as metaphor" (Sontag 1978: 29). Another influential linguist, Elena Semino, endorses the understanding that „the use of cancer as a metaphor is generally best avoided" (Potts and Semino 2019: 94) and reaches the conclusion that only *bad* metaphors must be foregone, i.e. that only appropriate metaphors that are the product of „a well-informed and context-sensitive approach to metaphor selection" (Semino 2021: 2) must be employed. It can be proposed that, currently, there exists a general agreement among researchers that since „metaphors are crucial tools for communication and thinking", they can be especially useful in the field of Medicine and Medicine-related sciences such as Public Health (Semino 2021: 1).

Of all medical metaphors, perhaps the most salient for research are the ones united under the umbrella term ***war metaphors***. In his classical work, *Medicine's Metaphors: Messages and Menaces* (Vaisrub 1977), medical metaphors are classified by various thematic fields such as mythology and war, the latter field incorporating those medical metaphors that reflect the understanding of the disease/condition as the enemy, while the patient is viewed as the warrior. The metaphoric expression *He was a true superhero*/*warrior/champion in his fight with cancer* corresponds to the same conceptual metaphor.

The way we view war metaphors is harmonious with the Conceptual Metaphor Theory (CMT). The works of George Lakoff and Marc Johnson which form the basis of the CMT while extending the realm of metaphor,

simultaneously narrow it. On one hand, metaphor is all-present: without it, we cannot talk; without it, we cannot embody our thoughts in language. Aristotle first examined metaphor in his *Poetics*: Lakoff and Johnson drastically enlarged its realm in *Metaphors We Live By*. From being a tool to endow something with „a name that belongs to something else" (Aristotle 1995, Poetics XX 1457b 7 – 9), metaphor has become that which we cannot live without: „the way we think, what we experience, and what we do every day is very much a matter of metaphor" (Lakoff and Johnson 1980: 3).

Since we are particularly interested in the way metaphors function in medical discourse, we also rely on Discourse Theory and Discourse Analysis. Here we can't but include Laclau and Mouffe's Discourse Theory (DT), and, in particular, the understanding promoted by DT that is most significant for our research purposes:

> since all social phenomena are mediated through discourse, their meanings can never be permanently fixed […] the aim of discourse analysis, then, is not to discover the 'truth' about reality (for example, to find out which groups exist within society) but to describe how discursive struggle constructs this reality (for example, how people and groups perceive their identity within society) so that it appears natural and neutral (Rear 2013: 5).

Last, but not least, our theoretical framework is also dependent upon emerging theories for AI application in health and health sciences. So far no singular theory has gained popularity: currently, most researchers elaborate mostly descriptive products that raise important issues such as Singhal's „Fairness, Engagement, and Discourse Analysis in AI-Driven Social Media and Healthcare" (Singhal 2023). In fact, it is our intent that the present research makes a contribution, namely, to this young field.

AI for text tokenization of metaphors

The question before us lies in the way AI-driven discourse analysis can partake in metaphor research in order to help with the problem of the metaphor usage in medical discourse. **As part of a much more extensive research project, the present article examines the possibility of performing the first step of this project, namely, text tokenization, by relying on AI**. We demonstrate how AI as represented by BioBERT can be employed in practice.

Our research goal is to demonstrate one such possible application of AI: namely, **tokenization of biomedical texts**, with a current focus on the war metaphor **DISEASE is ENEMY**.

Our motivation in centering on DISEASE is ENEMY metaphor is twofold: on one hand, we felt the need to set limitations to the scope of our

research, hence we picked only one cognitive metaphor; on the other hand, we wanted to tackle a metaphor that is representative of medical discourse metaphors and DISEASE is ENEMY is if not the most common one, then certainly one of the most common ones, its first place being challenged perhaps only by the HEALTH is UP metaphor.

Of course, DISEASE is ENEMY is a subset of the TREATMENT (DISEASE) IS WAR umbrella metaphor. In medical discourse, it is long been admitted that „metaphorical entailments can characterize a coherent system of metaphorical concepts of metaphorical expressions for those concepts" (Lakoff and Johnson 1980: 9). Thus, we are faced with the umbrella metaphor TREATMENT (DISEASE) IS WAR which represents a system of metaphorical concepts that includes DISEASE IS THE ENEMY, PHYSICIAN IS A WARRIOR CAPTAIN, PATIENT IS A BATTLEGROUND (Napolitano 2019: 2). Sometimes, instead of referring metonymically to this system (by using a member of the system, TREATMENT IS WAR, to refer to the whole system of metaphoric concepts), researchers use other umbrella terms such as *Military Metaphors* (Khullar 2014: 2), *Militaristic* or *War Metaphors* (Seixas 2021: 1). Metaphorical expressions of conceptual metaphors of the system include:

> *The doctors at the CDC led the way to victory against the flu epidemic; The cancer patient fought a long, hard, battle, but she eventually succumbed to her disease; A Triumph in the War Against Cancer; With nothing left in their leukemia-fighting arsenal, the doctors were down to Dilaudid* (MetaNet 2013, DISEASE TREATMENT IS WAR).

To locate metaphorical expressions of the conceptual metaphor DISEASE is ENEMY, we investigated the current possibilities for AI text tokenization and identified **BioBERT as the current most suitable candidate for carrying out the task of biomedical text tokenization**.

BioBERT for biomedical text tokenization

BioBERT, *Biomedical Bidirectional Encoder Representations from Transformers*, is particularly suited for text mining of texts in the field of the health sciences. Text mining can be defined as „a truly interdisciplinary method drawing on information retrieval, machine learning, statistics, computational linguistics and especially data mining" (Hotho et al 2005: 19). Such mining of biomedical texts is particularly useful when examining medial text corpora, performing sentiment analysis of patient utterances, etc.

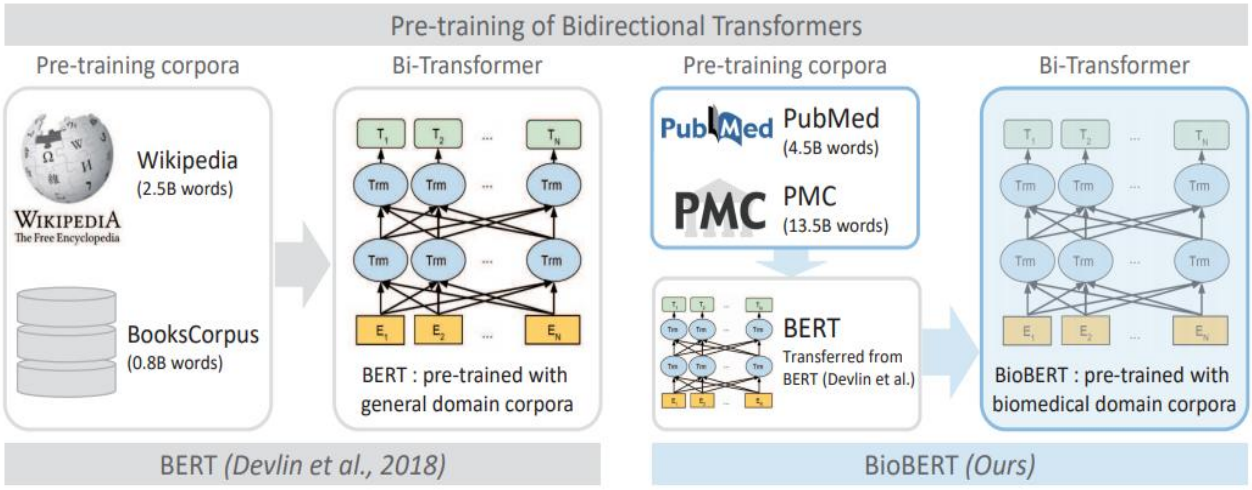
BioBERT can be viewed as the medical scion of the BERT family.

BERT is an acronym of *Bidirectional Encoder Representations from Transformers*, which „unlike recent language representation models […] is

designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers“ (Devlin et al 2019: 1). A breakthrough in Natural Language Processing, BERT is soon followed by RoBERTa, *A Robustly Optimized BERT Pretraining Approach*, whose „improved training recipe“ is to rectify the severe undertraining of BERT (Liu et al 2019: 1).

Shortly after, there appears BioBERT, which is aimed specifically at addressing biomedical issues. BioBERT is a specialized version of BERT, which allows that „the pre-trained language model BERT […] be adapted for biomedical corpora“ (Lee et al 2019: 1).

Image 1. *A comparison between BERT and BioBERT in terms of pre-training of bidirectional transformers*



Source: https://images.app.goo.gl/a4yRmHFRQQUZVgNz7

BioBERT has been employed in pursuit of pharmaceutical interests. For instance, when BioBERT, „an output-modified bidirectional transformer“ is combined with „a bidirectional gated recurrent unit layer (BiGRU) to obtain the vector representation of sentences“, it can offer precise data on drug-drug interaction entries in biomedical literature (Zhu et al 2020: 1,7).

In addition, BioBERT can be of use in medical diagnosis. BioBERT is proven to „improve performance of named entity recognition (NER)“ for Electronic Medical Records (EMR), related to „automatic annotation of clinical problems, treatments and tests in EMR“, which in turn can influence the medical diagnosis (Yu et al 2019: 1).

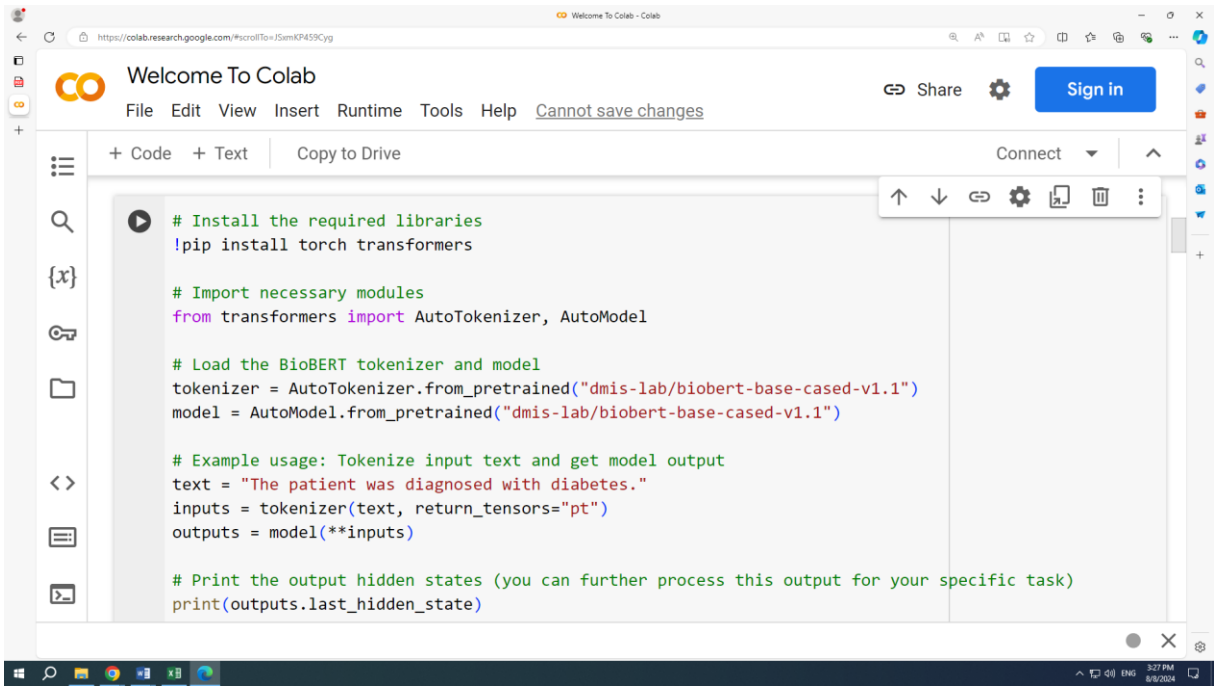
BioBERT can be successfully employed in the study of War metaphors due to its particular advantage of **the efficient medical corpus**

tokenization, which is key for a precise analysis of the DISEASE IS ENEMY metaphor.

BioBERT in practice

In order to use BioBERT, one must rely on the programming language *Python*. An example of how to code in *Python* for the installation of BioBERT is provided below:

Image 2. *Python Script for Installing and Using BioBERT*

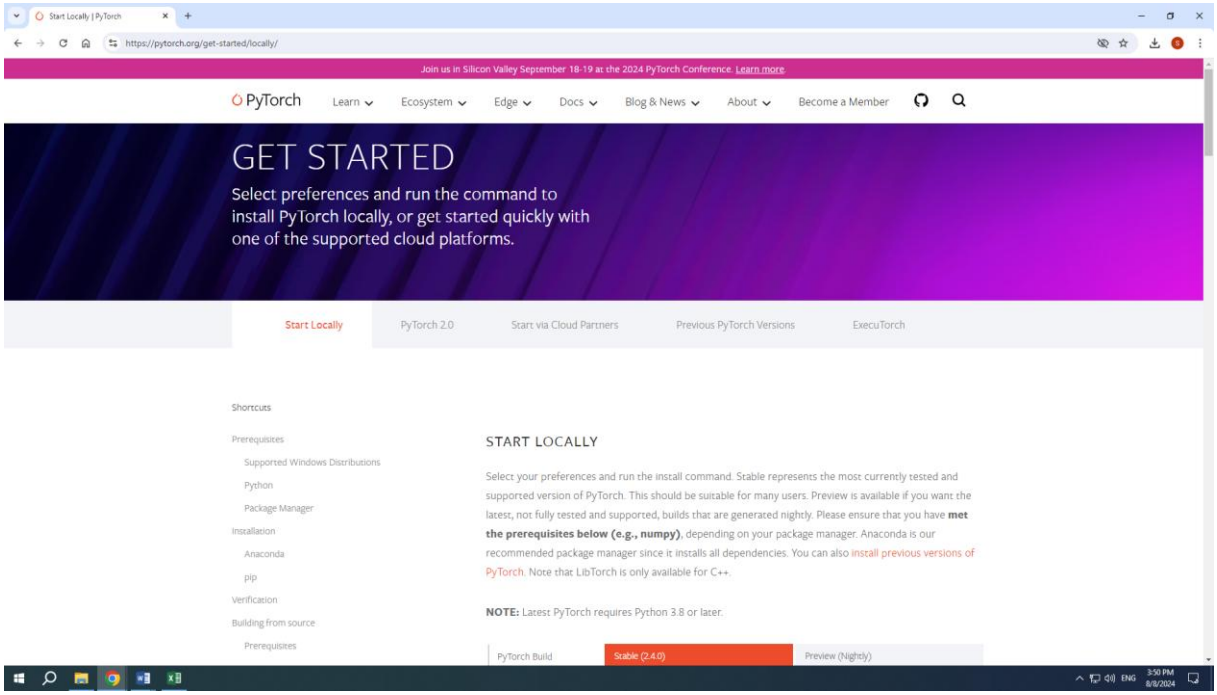


Source: https://colab.research.google.com/

We used the conversational AI model ChatGPT to create the Python code, which we then run in *Colab*, *Google Colaboratory*, to ensure the needed *Jupyter* notebook environment.

Then libraries such as *PyTorch*, *TensorFlow*, *NumPY* must be installed. The image below gives an example of the installation of such a library:

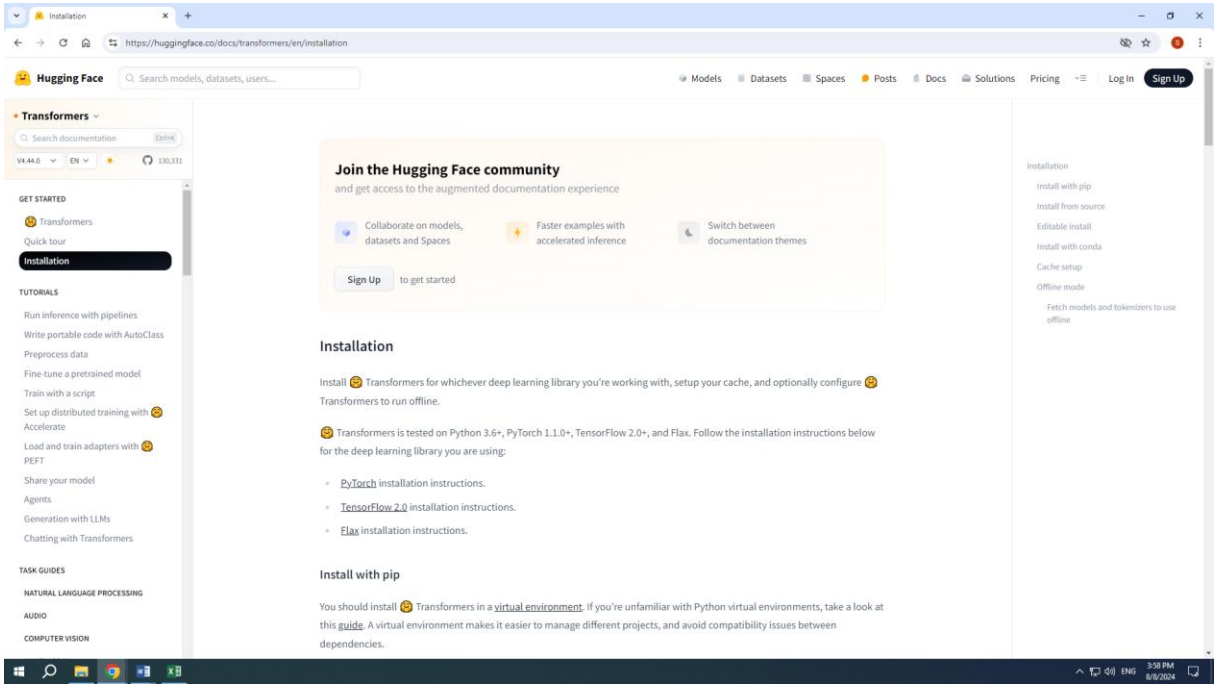
Image 3. *Installing PyTorch*



Source: https://pytorch.org/get-started/locally/

Additionally, *Hugging Face Transformers Library* (or, simply, *Transformers*) must also be installed.

Image 4. *Installing Transformers*



Source: https://huggingface.co/docs/transformers/en/installation

The architecture for the pretrained BioBERT model includes *AutoTokenizer* and *AutoModel*, which can be obtained from the *Autoclasses* of *Transformers*. **BioBERT can now be used to tokenize the input medical text**.

We can presently offer a sample tokenization of a war metaphor. We can tokenize a sentence such as *The patient lost the battle with cancer* in the following manner: ['[CLS]', 'the', 'patient', 'lost', 'the', 'battle', 'with', 'cancer', '[SEP]'], the token [CLS] marking the beginning of the sentence, while the token [SEP] denoting its end point (Prakash 2021: 1).

Once the corpus is tokenized, BioBERT allows us to execute our War metaphor search thanks to the contextual embeddings of the processed tokens.

Conclusion

BioBERT is a great new tool for tokenization of medical corpora and so it offers significant advantages to the initial stage of the examination of metaphors such as War metaphors. In this way, BioBERT can be of great aid to the study of medical discourse, and further research which explores the advantages that AI can offer medical discourse analysis seems called for.

BIBLIOGRAPHY

Aristotle 1995: Aristotle. Poetics. // *The Complete Works of Aristotle*, Barnes, J., ed., Princeton: Princeton University Press, vol.2, 1995.

Benedetti et al 2007: Benedetti, F., Lanotte, M., Lopiano, L., and Colloca, L. When words are painful: Unraveling the mechanisms of the nocebo effect. <https://www.sciencedirect.com/science/article/pii/S0306452207001819#:~:text=Recent%20experimental%20evidence%20indicates%20that%20negative%20verbal%20suggestions,cholecystokinin%20%28CCK%29%20which%2C%20in%20turn%2C%20facilitates%20pain%20transmission>, 10 August, 2024.

Bourzat 2019: Bourzat, F. (with K. Hunter). *Consciousness Medicine: Indigenous Wisdom, Entheogens, and Expanded States of Consciousness for Healing and Growth*. Berkeley: North Atlantic Books, 2019.

Devlin et al 2019: Devlin, J., Chang, M., Lee, K., and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://arxiv.org/pdf/1810.04805>, 7 August, 2024.

Colloca and Miller 2011: Colloca, L. and Miller, F. The nocebo effect and its relevance for clinical practice. <https://pubmed.ncbi.nlm.nih.gov/21862825/>, 10 August, 2024.

Edwards 2010: Edwards, J. *Conscious Medicine: Creating Health and Well-Being in a Conscious universe. London: Little Brown Book Group*, 2010.

Hotho et al 2005: Hotho, A, Nurnberger, A., and Paab, G. *A brief Survey of Text Mining.* <https://jlcl.org/article/download/68/66>, 7 August, 2024.

Khullar 2014: Khullar, D. *The Trouble With Medicine's Metaphors*. <https://www.theatlantic.com/health/archive/2014/08/the-trouble-with-medicines-metaphors/374982/>, 7 August, 2024.

Lakoff and Johnson 1980: Lakoff, G., Johnson, M. *Metaphors We Live By*. Chicago: The University of Chicago Press, 1980.

Lee et al 2019: Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., Ho, C., and Kang, J. *BioBERT: a Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining.* <https://arxiv.org/pdf/1901.08746>, 8 August, 2024.

Liu et al 2019: Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. <https://openreview.net/forum?id=SyxS0T4tvS>, 7 August, 2024.

MetaNet 2013: MetaNet. *DISEASE TREATMENT IS WAR.* <https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:DISEASE_TREATMENT_IS_WAR>, 7 August, 2024.

Napolitano 2019: Napolitano, S. *Metaphors in Healthcare: to Better Comprehend and to Improve Awareness.* <https://www.medicinanarrativa.eu/metaphors-in-healthcare-to-better-comprehend-and-to-improve-awareness>, 7 August, 2024.

Potts and Semino 2019: Potts, A., Semino, E, Cancer as a Metaphor // *Metaphor and Symbol*, vol. 34., num. 2, 2019.

Prakash 2021: Prakash, P. *An Explanatory Guide to BERT Tokenizer*. <https://www.analyticsvidhya.com/blog/2021/09/an-explanatory-guide-to-bert-tokenizer/>, 7 August, 2024.

Rear 2013: Rear, D. *Laclau and Mouffe's Discourse Theory and Fairclough's Critical Discourse Analysis: An Introduction and Comparison*. <https://www.academia.edu/2912341/Laclau_and_Mouffe_s_Discourse_Theory_and_Faircloughs_Critical_Discourse_Analysis_An_Introduction_and_Comparison>, 7 August, 2024.

Seixas 2021: Seixas, E. *War Metaphors in Political Communication on COVID-19*. <https://pubmed.ncbi.nlm.nih.gov/33869509/>, 7 August, 2024.

Semino 2021: Semino, E. *COVID-19: A Forest Fire Rather than a Wave?*. <https://eprints.lancs.ac.uk/id/eprint/151450/1/Semino_METODE_piece_Rev_final.pdf>, 10 August, 2024.

Shaheen 2021: Shaheen, M. *Applications of Artificial Intelligence (AI) in Healthcare: A Review*. <https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-.PPVRY8K.v1>, 8 August 2024.

Singhal 2023: Singhal, A. *Fairness, Engagement, and Discourse Analysis in AI-Driven Social Media and Healthcare*. <https://knowledgecommons.lakeheadu.ca/bitstream/handle/2453/5204/SinghalA2023m-1a.pdf?sequence=1&isAllowed=y>, 10 August 2024.

Sontag 1978: Sontag S. *Illness as Metaphor*. <https://buddhistuniversity.net/exclusive_01/Illness%20As%20Metaphor%20-%20Susan%20Sontag.pdf>, 6 August 2024.

Vaisrub 1977: Vaisrub, S. *Medicine's Metaphors: Messages and Menaces*. New Jersey: Medical Economics Company, 1997.

Webster et al 2016: Webster, R., Weinman, J., and Rubin, G. A systematic review of factors that contribute to nocebo effects <https://pubmed.ncbi.nlm.nih.gov/27657801/>, 10 August, 2024.

Yu et al 2019: Yu, X., Hu, W., Lu, S., Sun, X., and Yuan, Z. *BioBERT Based Named Entity Recognition in Electronic Medical Record*. <https://ieeexplore.ieee.org/abstract/document/8965108>, 10 August, 2024.

Zhu et al 2020: Zhu, Y., Li, L., Lu, H., Zhou, A., and Qin, X. *Extracting Drug-Drug Interactions from Texts with BioBERT and Multiple Entity-Aware Attentions*. <https://www.sciencedirect.com/science/article/pii/S1532046420300794>, 8 August, 2024.