Title
Natural Language Processing for Radiation Oncology: Personalizing Treatment Pathways.

Authors
Lin, Hui
Ni, Lisa
Phuong, Christina
et al.

Publication Date
2024

DOI
10.2147/PGPM.S396971

Pharmacogenomics and Personalized Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Natural Language Processing for Radiation Oncology: Personalizing Treatment Pathways

Hui Lin[1,2], Lisa Ni[1], Christina Phuong[1], Julian C Hong[1,3,4]

[1]Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA; [2]UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, Berkeley and San Francisco, San Francisco, CA, USA; [3]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA; [4]Joint Program in Computational Precision Health, University of California, Berkeley and San Francisco, Berkeley, CA, USA

Correspondence: Julian C Hong, Email julian.hong@ucsf.edu

Abstract: Natural language processing (NLP), a technology that translates human language into machine-readable data, is revolutionizing numerous sectors, including cancer care. This review outlines the evolution of NLP and its potential for crafting personalized treatment pathways for cancer patients. Leveraging NLP's ability to transform unstructured medical data into structured learnable formats, researchers can tap into the potential of big data for clinical and research applications. Significant advancements in NLP have spurred interest in developing tools that automate information extraction from clinical text, potentially transforming medical research and clinical practices in radiation oncology. Applications discussed include symptom and toxicity monitoring, identification of social determinants of health, improving patient-physician communication, patient education, and predictive modeling. However, several challenges impede the full realization of NLP's benefits, such as privacy and security concerns, biases in NLP models, and the interpretability and generalizability of these models. Overcoming these challenges necessitates a collaborative effort between computer scientists and the radiation oncology community. This paper serves as a comprehensive guide to understanding the intricacies of NLP algorithms, their performance assessment, past research contributions, and the future of NLP in radiation oncology research and clinics.

Keywords: artificial intelligence, personalized medicine, radiation therapy, natural language processing

Introduction

Natural Language Processing (NLP), a critical domain in artificial intelligence (AI), has revolutionized a myriad of general language applications, from search engines and recommendation systems to digital personal assistants.[1–3] In the context of healthcare, NLP holds considerable promise, especially in the field of oncology. With the proliferation of Electronic Health Records (EHRs), a wealth of unstructured data is available for exploration. In fact, it is estimated that the US healthcare system has exceeded 2000 exabytes, much of which is unstructured data in clinical notes, demanding sophisticated NLP techniques for utilization.[4] This vast reservoir of EHR data has catalyzed research efforts to unearth meaningful insights for cancer care. For instance, NLP techniques have been employed to identify patients at risk of familial cancers using family history information documented in clinical narratives,[5] to automate the process of extracting cancer staging information from unstructured clinical narratives,[6] leading to more efficient patient stratification and appropriate treatment plans. This horizon is now expanding towards radiation oncology, where personalization and precision play pivotal roles. The remainder of this review presents a comprehensive exploration of the evolution of NLP models and discuss the potential implications of these developments in the context of radiation oncology, more specifically in personalizing treatment pathways. To illustrate the dynamic interplay between data, models, and applications of NLP in radiation oncology, Figure 1 provides a schematic overview that encapsulates the progression from raw clinical data to actionable oncological insights, highlighting the transformative potential of NLP in personalizing radiation therapy pathways.

Received: 23 August 2023
Accepted: 29 January 2024
Published: 13 February 2024

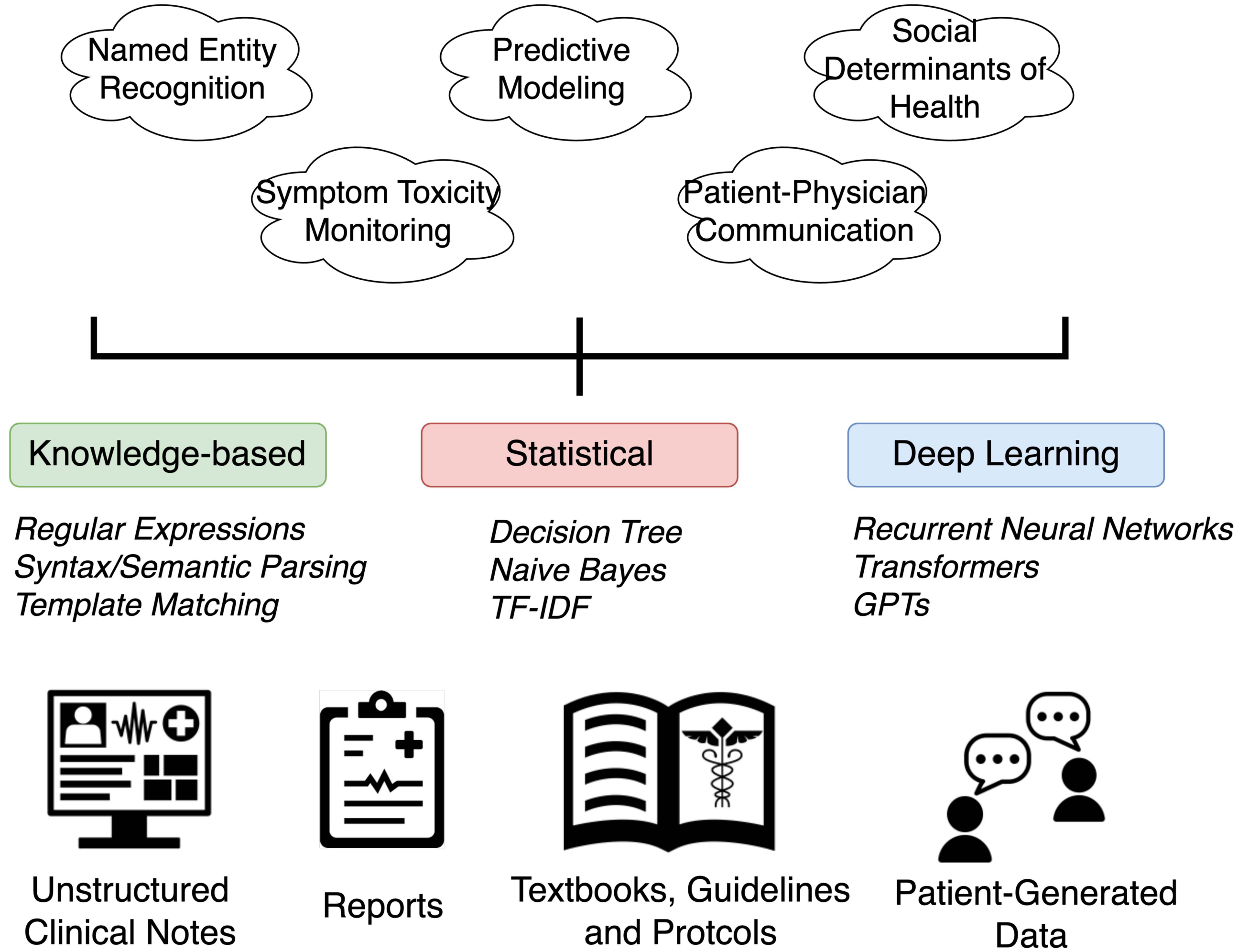

Figure 1 A schematic overview of the flow from foundational data to diverse applications in the radiation oncology domain empowered by NLP methods. The bottom layer represents various foundational data sources used in radiation oncology. The middle layer categorizes the predominant NLP methodologies into three classes: knowledge-based, statistical, and deep learning, where knowledge-based methods rely on domain-specific rules, statistical methods employ algorithms to infer patterns from data, and deep learning utilizes complex neural network architectures for more nuanced language understanding. The top layer displays the key applications of these NLP methods in radiation oncology.

Roadmap of Natural Language Processing Models and Recent Advances in Large Language Models

The rudimentary phase of NLP, dating back to the 1950s, was dominated by rule-based systems.[7] These models were built around predefined rules and grammatical structures, often resulting in inflexibility and limited contextual understanding. The 1980s and 1990s marked a shift towards statistical methods, such as Hidden Markov Models and Naive Bayes, that made use of real-world data for model training.[8] However, these models were often limited by their inability to capture semantic relations and contextual subtleties in language. The introduction of machine learning ushered in a new era for NLP, particularly with the advent of neural networks. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) models provided a means to understand context over long texts.[9–11] Despite the advancement, these models were still faced with significant limitations, such as difficulty in handling long sequences and the vanishing gradient problem.[12]

The advent of the Transformer model in 2017 marked a profound shift in the NLP landscape.[13] Distinguished by its innovative self-attention mechanism, this model enabled the efficient capture of long-range dependencies in text. This mechanism, by weighting different words in a sequence based on their relevance to a given context, allowed the Transformer model to move beyond the linear sequential processing of RNNs, instead processing all words simultaneously and gaining a more holistic understanding of the textual context. This feature was particularly beneficial in managing the inherently complex and context-dependent nature of human language. This groundbreaking development paved the way for the emergence of larger, pre-trained models like Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT).[14,15] BERT introduced the concept of bidirectional training to transformers, allowing each word to be conditioned on its preceding and following words, thus capturing left and right context simultaneously. This bidirectional understanding led to significant improvements in language tasks such as question answering,[16] relation extraction[17] and named entity recognition.[18] On the other hand, GPT, starting with GPT-1 and leading up to the latest GPT-4,[19] has demonstrated remarkable language generation capabilities. Unlike BERT, GPT utilizes a unidirectional approach, where words are conditioned only on their preceding words, making it more suitable for text generation tasks. GPT-4, in particular, exhibits a compelling ability to generate coherent and contextually accurate responses, heralding a new era for NLP applications. However, it is crucial to approach the use of GPT models, including ChatGPT, with a degree of caution. As generative AI models, they are built on large datasets compiled from diverse and undisclosed sources. This lack of transparency about the input data and the proprietary nature of the algorithms can lead to a heightened risk of bias in the generated content. Moreover, the data used to train these models may not always be representative or free from inherent biases present in the highly personalized cancer care data. Therefore, while GPT models offer great potentials in clinical notes analysis and generation, their applications in radiation oncology should be accompanied by a thorough understanding of these limitations.

Over the recent years, the scale and performance of these language models, collectively known as Large Language Models (LLMs), have grown dramatically. The pre-training aspect of these models, where they are trained on a large corpus of text data before being fine-tuned for specific tasks, has been a cornerstone of their success. The process enables these models to learn a rich understanding of language semantics and syntax, allowing them to generalize better to unseen data. LLMs have shown significant potential in the tasks such as machine translation[20,21] and sentiment analysis,[22] a technique used to determine the emotional tone behind words. The BERT-based models, for instance, have shown promising results on the SemEval-2017 Task 4,[23] a popular benchmark for sentiment analysis tasks.

In the realm of cancer care, these LLMs have spurred more sophisticated applications. For example, they have been used to generate medical text for educational purposes, simulate patient-doctor conversations for training, and even support real-time clinical decision-making.[24–26] Their ability to accurately interpret and generate language has led to improved diagnostic accuracy, enhanced patient engagement, and more efficient healthcare operations. This marks a significant progression from traditional models, which were often hampered by their inability to effectively handle the complexity and nuances of clinical language.

Personalized Treatment Pathways for Cancer Patients

The treatment of cancer has become more individualized over time as it is a complex disease process, necessitating increasing oversight of patients' unique treatment pathways through clinical presentation, workup, diagnosis, treatment, outcomes and toxicities. Even within specific cancer subtypes, there is significant variability in the clinical presentation across patients, and the use of standard treatment paradigms can lead to vastly different outcomes per patient. Over the past few decades, there has been a tremendous increase in interest for personalizing health care for the individual. In the modern era, advancements in molecular analysis and the development of "omic" approaches (eg genomic, proteomic, metabolomic) have allowed for personalized treatment recommendations for patients with cancer.[27,28] Similarly, as treatment pathways diverge to become unique for each patient, managing toxicity grows increasingly complex, requiring individualized approaches as well.[29] Due to the heterogeneity in patients' disease presentation, treatments, toxicities, and outcomes, the use of artificial intelligence will be vital in order to scale the ability of clinicians to take all of these factors into account in order to provide the best comprehensive patient care.

Advancing upon the strengths of NLP technologies and their impressive growth in scope and performance, their application in the field of radiation oncology can redefine the landscape of personalized cancer care. Given the capability of digesting vast training text data and the proficiency in understanding language semantics and syntax, NLP models are uniquely positioned to learn and process the intricate, multi-dimensional data associated with individual cancer patient profiles. In the context of personalized treatment pathways, these advanced NLP technologies can transform raw, high-dimensional data into actionable insights, and can harness the power of "omic" approaches by effectively integrating and interpreting complex genomic, proteomic, and metabolomic data to guide the selection of precise, patient-tailored treatment strategies.[30,31] Moreover, in an environment where treatment pathways are diverging to become unique for each patient, managing toxicity grows increasingly complex. NLP can play a significant role in capturing and understanding patient-specific treatment responses, enabling early detection and management of treatment-related toxicities. By continuously monitoring and interpreting diverse data streams, these technologies can predict potential adverse events and provide critical guidance on treatment modifications. In essence, with the potential to provide a more comprehensive, accurate, and personalized approach in radiation oncology, NLP stands at the frontier of a new era of medical practice. Its potential in numerous applications promise to redefine the patient experience and cancer care outcomes. In the following sections, we delve deeper into these intriguing applications and explore how they are shaping the future of personalized cancer care in radiation oncology.

The literature search was primarily conducted using PubMed and Google Scholar. The search was centered around key phrases such as "radiation oncology" AND "Natural Language Processing", "cancer" AND "Natural Language Processing", allowing us to capture a broad spectrum of relevant studies. We employed a combination of these keywords, using standard Boolean operators to refine our search. We focused on studies published in the prior decade to concentrate more on modern NLP techniques leveraging recent advances in machine learning and deep learning. The search results were systematically screened by title and abstract before a full-text review to identify studies discussing unique applications at the intersection of NLP and radiation therapy. We supplemented search results with manual examination of reference lists from recent reviews and selected articles to identify additional relevant citations.

Applications of Natural Language Processing to Personalize Radiation Oncology

Dynamic Personalized Care: Symptom and Toxicity Monitoring

Advancements in NLP have highlighted the immense potential for integrating detailed radiation therapy data, typically hidden in free-text sections of EHR, into databases and clinical summaries. This refined process of data extraction can potentially revolutionize decision-making protocols, elevate the efficacy of patient follow-up, and enhance overall patient management.[32] A recent study evaluated Apache Clinical Text Analysis Knowledge Extraction System (cTAKES),[33] an open-source NLP pipeline designed for clinical text, for its capacity to extract toxicity data from unstructured cancer therapy treatment visit notes.[34] Notably, cTAKES demonstrated impressive efficacy in identifying present symptoms, even outperforming human inter-rater variability. However, the system's performance in recognizing negated symptoms was less satisfactory. These findings draw attention to the potential and limitations of NLP tools in the context of data extraction from clinical notes. The superior performance of cTAKES in identifying certain symptoms exemplifies the substantial promise of NLP in enhancing healthcare data management. However, the limitations observed in recognizing negated symptoms also underscore the need for further optimization of these tools.[34]

The progressive strides as deep transformer models are increasingly showing potential for enhancing the recognition of radiation therapy entities and facilitating relation extraction, thereby underlining the potential of advanced NLP methods in augmenting data extraction and standardization. A recent study demonstrated the use of a deep learning model to extract symptoms from unstructured clinical notes in EHR.[35] The researchers trained the model to identify a wide range of symptoms relevant to cancer care, showing its feasibility for large-scale health system applications. Another research study developed NLP models to identify the presence and severity of esophagitis from the notes of patients treated with thoracic radiation therapy.[36] This study provided proof-of-concept for automated, detailed toxicity monitoring in expanded domains using NLP. Building on these advancements, LLMs such as the GatorTron,[37] with its extensive training on clinical notes, could be harnessed to detect and monitor symptoms and toxicities in real-time. Its deep understanding of medical language might facilitate the extraction of nuanced information about patient symptoms and treatment-related toxicities, offering a more comprehensive view of patient health. Another exemplary study utilized Optum's proprietary NLP tool to mine unstructured data from EHRs for a nationally representative study. The authors leveraged the NLP tool to facilitate the extraction of biomarker status and pneumonitis incidence from physician notes, allowing for an in-depth analysis of treatment-related adverse events in non-small cell lung cancer patients.[38] The study successfully identified several predictors of pneumonitis, highlighting NLP's potential role in predicting adverse events and facilitating the development of mitigation strategies. Yet, the application of NLP in radiation therapy is still in its early stages, thus necessitating continued collaboration between NLP researchers and the radiation oncology community. An NLP tool capable of detecting potential risk factors for radiation therapy-induced toxicities, like connective-tissue disorders, pregnancy, pacemakers, and prior radiation therapy records, could serve as a robust quality assurance mechanism. Moreover, NLP models capable of extracting toxicity and outcome data could be potentially used for real-time EHR monitoring.

Personalizing Care Around Social Determinants of Health

Social determinants of health are known to significantly impact patient risk factors, treatment, and clinical outcomes. A few considerations for those undergoing radiation therapy include access to specific radiotherapy techniques, the need for often daily outpatient visits over a period of time, and toxicity management. Increasing use of the EHR for access to medical information and communication with the medical team has exposed disparities within the health system. Older age, black or Hispanic race, English as a second language, and lower socioeconomic status are but a few factors that have been shown to be negatively associated with use of internet based technologies such as patient portals that are increasingly being used in oncologic care.[39–41]

There remains limited understanding of underlying causes of cancer disparities resulting in limited ability to address these disparities. Underlying causes of cancer disparities are not easily captured within the EHR in a structured format. Rather, this information is often documented as clinical narratives. NLP systems are being developed to more accurately extract this data to construct clinical profiles for each patient.[42,43] For example, researchers have expanded upon the existing schema to develop an NLP model to improve extraction of clinical information and pertinent non-medical named entities in lung cancer patients.[42] Additionally, an NLP algorithm utilizing a generated list of lexicon for social isolation was developed to improve evaluation for social isolation in prostate cancer patients, which was tested against domain expert manual review with high level of precision and recall.[43]

More recently, NLP models to identify factors underlying social determinants of health for those undergoing radiation therapy are being explored.[44] Linguistic differences have been identified in clinical notes among patients of different social contexts.[45] These tools can help identify disparities in care and drive appropriate intervention, including appropriate allocation of resources and support. Further, they have the potential to improve our understanding of how these underlying factors affect cancer management and outcomes.

Personalizing Patient Resources: Patient-Physician Communication and Patient Education

The plethora of medical information that can be access through the internet can be overwhelming for patients. Methods to provide patients with relevant and accurate information represent an ongoing area of focus, especially in radiation oncology where many patients have limited baseline insight. Various strategies to compute recommended patient education material exist including information retrieval approach and the recommendation algorithm, which are based on user queries and assessment of a patient's potential needs. NLP techniques have been explored to improve the relevance of patient education material based on their health data, though are still in early development.[46] Although this study was focused on education material for chronic disease, there is potential to expand to other areas including oncology and radiation oncology. As physicians are often limited in time, an NLP application to facilitate distribution of patient resources can be of great assistance.

Just as patient education helps facilitate care, the healthcare team's understanding of patient values and goals of care are critical components of cancer management. Effective discussions around prognosis, values, advanced care planning, and code status help ensure treatments are aligned with patient goals. NLP offers a scalable method of identifying serious illness communication (SIC) in EHR using semi-structured data, which can serve as a baseline.[47] Further advances can potentially extract more detailed SIC information in unstructured data as well. There is potential to identify temporal changes in patient goals as diseases progress to improve care. As goals of care discussions are dynamic ongoing conversations, domain knowledge of how often SIC occur and in what context can inform intervention to improve goals of care discussions with patients and families.

Predictive Modeling to Personalize Treatment Strategies

The application of NLP for predictive modeling in the context of cancer has emerged as a significant research focus. The potential to discern critical insights from the wealth of information embedded in unstructured EHR data has been proven instrumental in refining predictive modeling efforts.[48,49] In the realm of traditional NLP techniques, significant contributions to predictive modeling in cancer care have been made. One compelling study highlighted the role of NLP in determining the smoking status of patients from information found in their discharge records, an important predictor for lung cancer and its treatment outcomes. As a part of the i2b2 (Informatics for Integrating Biology to the Bedside) project,[50] the authors organized a challenge wherein a variety of machine learning and rule-based algorithms were employed by participating teams. Despite the diversity in approaches, many of the systems submitted to the challenge delivered promising results. Analysis of the results underscored that discharge summaries express smoking status using a limited number of textual features, from which many of the successful smoking status identifiers benefited.[51] This study underscores the potential of even simple, rule-based NLP systems in predictive modeling in healthcare.

Following the success of these traditional methods, recent advancements have been made utilizing transformer-based NLP methods for predictive modeling in cancer care. For instance, a deep learning model, PPES-Met, was proposed to estimate the short-term life expectancy of metastatic cancer patients, employing free-text clinical notes while maintaining the temporal visit sequence.[52] The model achieved impressive accuracy, which holds promise for its utilization as a decision support tool in personalizing metastatic cancer treatment. MetBERT, another transformer-based model, was specifically fine-tuned to predict cancer metastasis from clinical notes. It yielded high performance on an in-house validation dataset, demonstrating its potential for deployment as a tool for early metastasis diagnosis. MetBERT was also found useful in identifying the date of cancer metastasis from clinical notes, thus underscoring the potential of NLP methods in early diagnosis.[53] OncoBERT, a deep transformer transfer learning framework, was built for cancer outcome prediction using unstructured clinical notes from various cancer sites, including Glioma, Prostate, and Breast.[54] Compared to conventional NLP methods and stage-based outcome models, OncoBERT outperformed them in predicting cancer outcomes across three distinct cancer types. The interpretability pipeline introduced alongside OncoBERT improves transparency and comprehensibility, empowering clinicians and patients to make more informed decisions.[54] Using a similar BERT-based infrastructure, a study showcased the potential of a large language model NYUTron,[55] which was trained using unstructured clinical notes from EHR and was fine-tuned for various clinical and operational predictive tasks. The study emphasized the importance of clinical predictive models in assisting physicians and administrators in making time-constrained decisions. NYUTron demonstrated impressive performance, especially in tasks like 30-day all-cause readmission prediction, in-hospital mortality prediction, and more, highlighting the growing potential of transformer-based models in healthcare. These traditional and advanced NLP techniques together underscore the utility of NLP in predictive modeling. While transformer-based models like BERT variants can handle complex tasks, traditional rule-based and machine learning techniques remain integral for their simplicity, interpretability, and proven effectiveness. Therefore, continuous advancements and integrations of both methods can provide more comprehensive and reliable predictive modeling in cancer care.

Challenges and Future Directions of NLP in Radiation Oncology

Privacy and Security of Clinical Data

The power and potential of NLP and particularly LLMs in radiation oncology are underpinned by the extensive use of healthcare data, most prominently in the form of EHR unstructured data. However, this revolutionary trend is not without its challenges. Prominent among them are the concerns over data privacy and security, which have taken center stage in discussions about the ethical use of NLP in healthcare.[56,57] The essence of these concerns lies in the inherently sensitive nature of health data. Personal health information (PHI) is a type of personal data that warrants particularly high levels of protection due to the risk of stigma, discrimination, or harm that could result if it is mishandled or disclosed inappropriately.[58] Despite best efforts to de-identify EHR data, the possibility of re-identification cannot be entirely eliminated, particularly in instances where unique or uncommon patient histories are involved. This risk is amplified when the models are trained on large-scale, diverse, and detailed patient data.[59]

Data security is a closely related issue. EHR systems house vast quantities of sensitive patient data, making them a prime target for malicious attacks. Ensuring robust data protection measures are in place at every stage of the data handling process – from data collection and preprocessing to model training, validation, and deployment – is of paramount importance.[60] Adding to these challenges is the growing interest in building multi-institution or even international databases to improve the generalizability and effectiveness of NLP tools. However, the uneven landscape of data privacy laws and guidelines across countries and institutions adds complexity to this task.[61] Furthermore, challenges related to language discrepancies, variations in healthcare delivery systems, and inconsistent data formatting across institutions can hamper this ambitious goal.

Despite these obstacles, the pursuit of secure and privacy-preserving NLP tools for radiation oncology continues with the exploration of innovative solutions. Federated learning, a decentralized machine learning approach, has emerged as a potential way forward. It enables model training across multiple decentralized edge devices or servers holding local data samples, without the need to exchange raw data.[62,63] This not only helps overcome data sharing restrictions, but also significantly enhances the privacy protection of patient data. Another common mitigation strategy against the privacy concerns is data de-identification or anonymization. This involves removing, encrypting, or replacing personally identifiable information within the health records to prevent the identification of individuals, while still retaining the valuable medical data for research.[64] While this process can significantly reduce the risk of privacy breaches, it is not infallible. For instance, sophisticated methods of re-identification can potentially link de-identified data back to the individual.[65] Therefore, ongoing research into more robust de-identification and anonymization methods is crucial. The pathway to effective and ethical implementation of NLP in radiation oncology is far from straightforward, but the journey is crucial. Addressing data privacy and security concerns will not only satisfy legal and ethical requirements, but it will also build patient trust – a factor that will ultimately determine the success of these advanced technologies in healthcare. Future research should continue to prioritize these aspects, exploring innovative solutions and frameworks that can bring the promise of NLP to fruition without compromising patient privacy and data security.

Promise and Pitfalls of Large-Scale Diverse Clinical Datasets

Having established the ethical and legal concerns over patient confidentiality and the security of PHI, it becomes evident that these challenges are inextricably tied to the realities of dealing with large-scale and diverse clinical datasets. Not only must we secure and preserve privacy within these vast data reserves, but also manage inherent complexities in their size, diversity, and quality. The challenges of data redundancy, heterogeneity, and the need for standardization underscore the obstacles that researchers and clinicians face in optimizing these datasets for NLP. Moreover, there are other limitations specific to EHR corpora. One example is redundancy in longitudinal records, as physicians tend to copy and paste text directly from prior notes. The diversity in large-scale EHR text corpora can thus be severely limited due to the redundancy from patients with several notes in their longitudinal records and may risk introducing unwanted bias. Methods for mitigating this redundancy include using metadata when available, or using fingerprinting algorithms in de-identified corpora.[66]

Another major limitation to the use of EHR data is the heterogeneity of the data, across institutions and even across providers within the same institution, including the data models, naming conventions, and degree of detail included when representing similar data.[67] This makes access to and aggregation of data very difficult. Recognizing this challenge, datasets like MIMIC-III[68] and MIMIC-IV[69] have been developed, offering large-scale clinical notes for research. However, while these datasets are invaluable, they are primarily single-institutional and may be restricted to specific clinical specialties, not being predominantly oncology-focused. Thus, efforts have been made to standardize EHR data from across providers and healthcare organizations, in order to generate structured datasets for interoperable secondary use.[70,71] Careful collection and annotation of large-scale EHR datasets will be fundamental for reliable training of language models to be used in clinical decision-making.[72]

Bias in NLP Language Models

One major limitation of language models is the risk for bias. The quality and diversity of data used for training language models can affect their performance drastically. Since most text corpora are based on real world data, many exhibit problematic stereotypical biases, which reflect negative generalizations based on social constructs such as race, gender, and religion.[73–75] These biases can then be amplified and propagated in language models trained on such data. For example, a language model that exhibits gender bias may assign a higher likelihood to the final token of "he worked as a [doctor]" than "she worked as a [doctor]".[76] In medicine, incorporation of such biases could lead to exacerbation of disparities in healthcare. In oncology, there may not be enough data about rare cancers or underrepresented populations, leading to blind spots in language models. It is critical to develop methods to identify and mitigate these biases as these models are increasingly used for decision-making in the real world.

There is now a significant body of work that has been done to quantify bias in language models, including in the medical context.[77,78] For example, datasets have been created to test language models for bias across domains such as race, gender, profession, and religion.[79,80] There are many proposed metrics for quantifying bias in text corpora, and it is likely that different metrics can reveal different types of bias.[81,82] In addition to choosing the appropriate measures to identify bias within datasets used for training, involving clinicians will be helpful to identify potential sources for bias and guide the selection of high quality data sources.[77,83]

Once biases are identified, developing methods for mitigating bias is critical, and this is still an active area of research.[82,84] Periodic evaluation and monitoring of model performance across various subgroups will be essential in identifying and reducing bias. Model documentation frameworks will allow for transparent reporting of choices and assumptions made when curating and evaluating datasets.[85] Moving forward, it will also be important to establish best practices for identifying and debiasing language models, in order to improve the fairness of language models and reducing harm to traditionally marginalized groups of patients.[86]

Interpretability and Generalizability of NLP Models

As NLP models become increasingly complex, understanding their decision-making processes—known as model interpretability—becomes more challenging yet increasingly important.[87,88] Large-scale transformer models, like BERT and its variants, have demonstrated impressive predictive capabilities in diverse tasks, but their "black-box" nature makes the interpretability elusive.[89] While these models can generate precise outputs, understanding why they arrive at these outputs is challenging, which limits their trustworthiness and potential for deployment in clinical settings, including radiation oncology.[90] Emerging research has aimed to address this issue by developing methods for interpreting NLP models. These methods range from visualization techniques for attention mechanisms[91] to methods like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) which provide instance-level explanations.[92,93] Despite the advancements, the interpretability of NLP models, particularly transformer-based models, remains an area requiring further research.

Generalizability is another significant challenge for NLP applications in radiation oncology. Many machine learning models are developed and validated using data from a single institution or country. These models, therefore, may not perform as well when applied to a different patient population or healthcare setting, which can vary greatly in terms of demographics, clinical practices, and documentation styles. The lack of multi-institutional data also inhibits the development of models that can truly represent diverse patient populations and healthcare contexts. This underscores the need for establishing multi-institutional international databases that can yield more generalizable NLP models. However, the creation of such databases faces numerous logistical and data privacy-related hurdles as discussed in the previous section, which require careful considerations and innovative solutions.[62,64,94] To tackle these issues, multi-task learning provides an innovative solution. Multi-task learning trains a single model to perform multiple related tasks concurrently, enabling it to learn a shared representation that can be beneficial across tasks. This methodology can be particularly effective in a healthcare context where each institution may have unique patient demographics and disease prevalence. By training a model to perform well across various tasks, the probability of the learned representation generalizing well to new institutions is increased.[95,96] Nonetheless, research is still ongoing to optimize and refine this approach, underscoring the dynamic nature of the field of NLP in radiation oncology.

Conclusion

The use of Natural Language Processing presents a transformative opportunity in radiation oncology. It has the potential to decipher and leverage the wealth of information embedded within the unstructured notes of EHR, thereby facilitating big data analytics and advancing precision medicine. However, as we seize the benefits, it is critical to navigate the multifaceted challenges that emerge, including those pertinent to privacy, security and model bias. Robust safeguards and algorithms should be in place to protect patient data and ensure the technology serves all populations equitably. As NLP technology matures and its integration into clinical settings becomes more prevalent, it is imperative for radiation oncologists to actively collaborate with computer scientists to introduce the advanced technology to the clinic. Such multidisciplinary partnerships can focus on high-priority tasks, contribute to the assembly of clinical corpora, establish guidelines, and define acceptable performance metrics for both research and clinical applications. This coordinated approach will ensure the true potential of NLP is realized, improving both research outcomes and patient care in the field of radiation oncology.

Disclosure

JH: Research Funding from American Cancer Society, American Society for Clinical Oncology/Conquer Cancer Foundation, American Society for Radiation Oncology, National Institutes of Health/National Cancer Institute, Prostate Cancer Foundation, Radiation Oncology Institute, Roche. HL: Research Funding from DOD, Uncle Kory Foundation. The authors report no other conflicts of interest in this work.

References

1. Socher R, Chen D, Manning CD, Ng A. Reasoning with neural tensor networks for knowledge base completion. *Advan Neur Informat Process Syst*. 2013;2013:26.
2. Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. Paper presented at: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); 2014.
3. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. *IEEE Comput Intel Mag*. 2018;13 (3):55–75. doi:10.1109/MCI.2018.2840738
4. Coughlin S, Roberts D, O'Neill K, Brooks P. Looking to tomorrow's healthcare today: a participatory health perspective. *Internal Med J*. 2018;48 (1):92–96. doi:10.1111/imj.13661
5. Warner JL, Jain SK, Levy MA. Integrating cancer genomic data into electronic health records. *Genome Med*. 2016;8:1–13. doi:10.1186/s13073-016-0371-3
6. Savova GK, Danciu I, Alamudun F, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. *Cancer Res*. 2019;79(21):5463–5470. doi:10.1158/0008-5472.CAN-19-0579
7. Hutchins WJ. *Machine Translation: Past, Present, Future*. Ellis Horwood Chichester; 1986.
8. Christopher DM, Hinrich S. Foundations of statistical natural language processing; 1999.
9. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–1780. doi:10.1162/neco.1997.9.8.1735
10. Rumelhart DE, Hinton GE, Williams RJ. *Learning Internal Representations by Error Propagation*. Institute for Cognitive Science, University of California, San Diego La; 1985.
11. Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv*. 2014;2014:1.
12. Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks. Paper presented at: Neural Networks, 1993, IEEE International Conference on1993; 1993.
13. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advan Neur Informat Process Syst*. 2017;2017:30.
14. Kenton JDM, Toutanova LK. BERT: pre-training of deep bidirectional transformers for language understanding. Paper presented at: Proceedings of NAACL-HLT; 2019.
15. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training; 2018.
16. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. Xlnet: generalized autoregressive pretraining for language understanding. *Advan Neur Informat Process Syst*. 2019;32.
17. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234–1240. doi:10.1093/bioinformatics/btz682
18. Akbik A, Blythe D, Vollgraf R. Contextual string embeddings for sequence labeling. Paper presented at: Proceedings of the 27th international conference on computational linguistics; 2018.
19. Koubaa A. GPT-4 vs. GPT-3.5: a concise showdown; 2023.
20. Fan A, Bhosale S, Schwenk H, et al. Beyond English-centric multilingual machine translation. *J Mach Learn Res*. 2021;22(1):4839–4886.
21. Aharoni R, Johnson M, Firat O. Massively Multilingual Neural Machine Translation. Paper presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2019.
22. Naseem U, Razzak I, Musial K, Imran M. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Gener Comput Syst*. 2020;113:58–69. doi:10.1016/j.future.2020.06.050
23. Rosenthal S, Farra N, Nakov P. SemEval-2017 task 4: sentiment analysis in Twitter. *arXiv preprint arXiv*. 2019;2019:1.
24. Zech J, Pain M, Titano J, et al. Natural language–based machine learning models for the annotation of clinical radiology reports. *Radiology*. 2018;287(2):570–580. doi:10.1148/radiol.2018171093
25. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *Npj Digital Med*. 2018;1(1):18. doi:10.1038/s41746-018-0029-1
26. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inf Assoc*. 2018;25(9):1248–1258. doi:10.1093/jamia/ocy072

27. Stricker T, Catenacci DV, Seiwert TY. Molecular profiling of cancer--the future of personalized cancer medicine: a primer on cancer biology and the tools necessary to bring molecular testing to the clinic. *Semin Oncol*. 2011;38(2):173–185. doi:10.1053/j.seminoncol.2011.01.013
28. Gambardella V, Tarazona N, Cejalvo JM, et al. Personalized medicine: recent progress in cancer therapy. *Cancers*. 2020;12(4):1009. doi:10.3390/cancers12041009
29. Esfahani K, Elkrief A, Calabrese C, et al. Moving towards personalized treatments of immune-related adverse events. *Nat Rev Clin Oncol*. 2020;17(8):504–515. doi:10.1038/s41571-020-0352-8
30. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nature Med*. 2019;25(1):24–29. doi:10.1038/s41591-018-0316-z
31. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharmaceut*. 2016;13(7):2524–2530. doi:10.1021/acs.molpharmaceut.6b00248
32. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inf Assoc*. 1994;1(2):161–174. doi:10.1136/jamia.1994.95236146
33. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. *J Am Med Inf Assoc*. 2010;17(5):507–513. doi:10.1136/jamia.2009.001560
34. Hong JC, Fairchild AT, Tanksley JP, Palta M, Tenenbaum JD. Natural language processing for abstraction of cancer treatment toxicities: accuracy versus human experts. *JAMIA Open*. 2020;3(4):513–517. doi:10.1093/jamiaopen/ooaa064
35. Lindvall C, Deng C-Y, Agaronnik ND, et al. Deep learning for cancer symptoms monitoring on the basis of electronic health record unstructured clinical notes. *JCO Clin Cancer Inform*. 2022;6:e2100136.
36. Chen S, Guevara M, Ramirez N, et al. Natural language processing to automatically extract the presence and severity of esophagitis in notes of patients undergoing radiotherapy. *arXiv preprint arXiv*. 2023;2023:1.
37. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *Npj Digital Med*. 2022;5(1):194. doi:10.1038/s41746-022-00742-2
38. Tyczynski JE, Potluri R, Kilpatrick R, Mazumder D, Ghosh A, Liede A. Incidence and risk factors of pneumonitis in patients with non-small cell lung cancer: an observational analysis of real-world data. *Oncol Therapy*. 2021;9(2):471–488. doi:10.1007/s40487-021-00150-8
39. Sinha S, Garriga M, Naik N, et al. Disparities in electronic health record patient portal enrollment among oncology patients. *JAMA Oncol*. 2021;7(6):935–937. doi:10.1001/jamaoncol.2021.0540
40. Gerber DE, Laccetti AL, Chen B, et al. Predictors and intensity of online access to electronic medical records among patients with cancer. *J Oncol Pract*. 2014;10(5):e307–e312. doi:10.1200/JOP.2013.001347
41. Girault A, Ferrua M, Lalloué B, et al. Internet-based technologies to improve cancer care coordination: current use and attitudes among cancer patients. *Eur J Cancer*. 2015;51(4):551–557. doi:10.1016/j.ejca.2014.12.001
42. Hong J, Davoudi A, Yu S, Mowery DL. Annotation and extraction of age and temporally-related events from clinical histories. *BMC Med Inf Decis Making*. 2020;20(11):1–15. doi:10.1186/s12911-020-01333-5
43. Zhu VJ, Lenert LA, Bunnell BE, Obeid JS, Jefferson M, Halbert CH. Automatically identifying social isolation from clinical narratives for patients with prostate Cancer. *BMC Med Inf Decis Making*. 2019;19(1):1–9. doi:10.1186/s12911-018-0723-6
44. Derton A, Murray A, Liu D, et al. Exploring methods to understand cancer disparities using natural language processing of clinical notes. *Int J Radiat Oncol Biol Phys*. 2022;114(3):S21. doi:10.1016/j.ijrobp.2022.07.369
45. Derton A, Guevara M, Chen S, et al. Natural language processing methods to empirically explore social contexts and needs in cancer patient notes. *JCO Clin Cancer Inform*. 2023;7:e2200196. doi:10.1200/CCI.22.00196
46. Wang Z, Huang H, Cui L, et al. Using natural language processing techniques to provide personalized educational materials for chronic disease patients in China: development and assessment of a knowledge-based health recommender system. *JMIR Med Inform*. 2020;8(4):e17642. doi:10.2196/17642
47. Davoudi A, Tissot H, Doucette A, et al. Using natural language processing to classify serious illness communication with oncology patients. Paper presented at: AMIA Annual Symposium Proceedings; 2022.
48. Bitterman DS, Miller TA, Mak RH, Savova GK. Clinical natural language processing for radiation oncology: a review and practical primer. *Int J Radiat Oncol Biol Phys*. 2021;110(3):641–655. doi:10.1016/j.ijrobp.2021.01.044
49. Ni L, Phuong C, Hong J. Natural Language Processing for Radiation Oncology. In: *Artificial Intelligence in Radiation Oncology*. World Scientific; 2023:157–183.
50. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inf Assoc*. 2010;17(2):124–130. doi:10.1136/jamia.2009.000893
51. Uzuner Ö, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Am Med Inf Assoc*. 2008;15(1):14–24. doi:10.1197/jamia.M2408
52. Banerjee I, Gensheimer MF, Wood DJ, et al. Probabilistic prognostic estimates of survival in metastatic cancer patients (PPES-Met) utilizing free-text clinical narratives. *Sci Rep*. 2018;8(1):10037.
53. Liu K, Kulkarni O, Witteveen-Lane M, Chen B, Chesla D. MetBERT: a generalizable and pre-trained deep learning model for the prediction of metastatic cancer from clinical notes. Paper presented at: AMIA Annual Symposium Proceedings; 2022.
54. Lin H, Ginart JB, Chen W, et al. OncoBERT: building an interpretable transfer learning bidirectional encoder representations from transformers framework for longitudinal survival prediction of cancer patients; 2023.
55. Jiang LY, Liu XC, Nejatian NP, et al. Health system-scale language models are all-purpose prediction engines. *Nature*. 2023;2023:1–6.
56. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J Royal Soc Interface*. 2018;15(141):20170387. doi:10.1098/rsif.2017.0387
57. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–1358. doi:10.1056/NEJMra1814259
58. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318. doi:10.1001/jama.2017.18391
59. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2017;22(5):1589–1604. doi:10.1109/JBHI.2017.2767063
60. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *New Engl J Med*. 2018;378(11):981. doi:10.1056/NEJMp1714229
61. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574. doi:10.1038/psp.2013.52

62. Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: system design. *Proceed Mach Learn Syst*. 2019;1:374–388.
63. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. Paper presented at: Artificial intelligence and statistics; 2017.
64. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. *BMJ*. 2015;2015:350.
65. Rocher L, Hendrickx JM, De Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(1):1–9. doi:10.1038/s41467-018-07882-8
66. Cohen R, Elhadad M, Elhadad N. Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. *BMC Bioinf*. 2013;14(1):1–15. doi:10.1186/1471-2105-14-10
67. Sujansky W. Heterogeneous database integration in biomedicine. *J Biomed Informat*. 2001;34(4):285–298. doi:10.1006/jbin.2001.1024
68. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Scientific Data*. 2016;3(1):1–9. doi:10.1038/sdata.2016.35
69. Johnson AE, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*. 2023;10(1):1. doi:10.1038/s41597-022-01899-x
70. Rea S, Pathak J, Savova G, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. *J Biomed Informat*. 2012;45(4):763–771. doi:10.1016/j.jbi.2012.01.009
71. Fernández-Breis JT, Maldonado JA, Marcos M, et al. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *J Am Med Inf Assoc*. 2013;20(e2):e288–e296. doi:10.1136/amiajnl-2013-001923
72. Albright D, Lanfranchi A, Fredriksen A, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative. *J Am Med Inf Assoc*. 2013;20(5):922–930. doi:10.1136/amiajnl-2012-001317
73. Bolukbasi T, Chang K-W, Zou JY, Saligrama V, Kalai AT. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advan Neur Informat Process Syst*. 2016;29:1.
74. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017;356 (6334):183–186. doi:10.1126/science.aal4230
75. Lauscher A, Glavaš G. Are we consistently biased? Multidimensional analysis of biases in distributional word vectors. Paper presented at: Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (* SEM 2019); 2019.
76. Liang PP, Wu C, Morency L-P, Salakhutdinov R. Towards understanding and mitigating social biases in language models. Paper presented at: International Conference on Machine Learning; 2021.
77. Zhang H, Lu AX, Abdalla M, McDermott M, Ghassemi M. Hurtful words: quantifying biases in clinical contextual word embeddings. Paper presented at: proceedings of the ACM Conference on Health, Inference, and Learning; 2020.
78. Prabhakaran V, Hutchinson B, Mitchell M. Perturbation Sensitivity Analysis to Detect Unintended Model Biases. Paper presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019.
79. Nadeem M, Bethke A, Reddy S. StereoSet: measuring stereotypical bias in pretrained language models. Paper presented at: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021.
80. Robinson R. Assessing gender bias in medical and scientific masked language models with StereoSet. *arXiv preprint arXiv*. 2021.
81. Recasens M, Danescu-Niculescu-Mizil C, Jurafsky D. Linguistic models for analyzing and detecting biased language. Paper presented at: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2013.
82. Bordia S, Bowman S. Identifying and reducing gender bias in word-level language models. Paper presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop; 2019.
83. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;2023:1–9.
84. Lauscher A, Lueken T, Glavaš G. Sustainable modular debiasing of language models. *arXiv preprint arXiv*. 2021;2021:1.
85. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. Paper presented at: Proceedings of the conference on fairness, accountability, and transparency; 2019.
86. Blodgett SL, Barocas S, Daumé III H, Wallach H. Language (Technology) is Power: a Critical Survey of "Bias" NLP. Paper presented at: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020.
87. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv*. 2018;51(5):1–42. doi:10.1145/3236009
88. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv*. 2017;2017:1.
89. Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: what we know about how BERT works. *Transact Assoc Comput Linguist*. 2021;8:842–866.
90. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *New Engl J Med*. 2017;376(26):2507. doi:10.1056/NEJMp1702071
91. Vig J. A multiscale visualization of attention in the transformer model. *arXiv preprint arXiv*. 2019;2019:1.
92. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. Paper presented at: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016.
93. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advan Neur Informat Process Syst*. 2017;2017:30.
94. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Trans Med*. 2010;2(57):57cm29–57cm29. doi:10.1126/scitranslmed.3001456
95. Ruder S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv*. 2017;2017:1.
96. Zhang K, Yu J, Yan Z, et al. BiomedGPT: a unified and generalist biomedical generative pre-trained transformer for vision, language, and multimodal tasks. *arXiv preprint arXiv*. 2023;2023:1.