MONTE : Pages:1–12

Original Research

Sequence-to-Sequence Learning for Predicting Procedure Codes from Unstructured Clinical Text

Nirmal Thapa[1]

[1]Karnali Institute of Technology, Department of Computer and Information Systems, Bagmati Highway, Surkhet, Nepal.

Abstract

In many clinical settings, the timely and accurate assignment of procedure codes to free-text notes is an essential task that supports patient record maintenance, billing procedures, and large-scale health analytics. However, clinical text remains inherently unstructured, with potential ambiguities, variations in terminology, and differences in writing styles that hinder traditional rule-based or keyword-driven methods. Recent advances in sequence-to-sequence paradigms offer a robust solution by modeling the entire coding process as a conditional generation task. Through attention mechanisms and latent representations, these approaches capture long-term dependencies, subtle linguistic nuances, and domain-specific terminologies. The focus of this work is on designing and implementing a sequence-to-sequence framework that learns from large corpora of unstructured clinical notes and predicts corresponding procedure codes. By leveraging deep neural architectures, dense embeddings, and structured decoding, the model processes raw text from token-level encodings to final code outputs. This study offers an in-depth exploration of strategies for tokenization, embedding initialization, efficient training via stochastic gradient optimization, and decoding with beam search to handle multiple possible outputs. Rigorous empirical evaluation underscores improved recall and precision compared to more conventional classification-based schemes. Moreover, the system demonstrates notable robustness to common data irregularities and domain-specific jargon. The results highlight the feasibility of automated coding processes in modern healthcare contexts, thereby reducing manual overhead and paving the way for scalable, data-driven clinical documentation solutions.

1. Introduction

The transformation of free-text clinical notes into standardized procedure codes has become increasingly central to the efficient operation of healthcare providers, insurance companies, and biomedical researchers [1]. The robust assignment of codes reduces human error, streamlines administrative workflows, and facilitates reliable knowledge discovery efforts in clinical data mining [2]. Early approaches to automated code assignment often employed linear classification strategies or expert-curated lexicon mappings, which tended to rely on carefully enumerated dictionaries of expressions. While these methods achieved partial success in constrained environments, their limitations emerged when faced with varied clinical note structures, abbreviations, and widely heterogeneous medical vocabularies [3]. Such constraints demanded a more powerful and flexible paradigm that could accommodate the intrinsic variability found in real-world clinical settings.

A sequence-to-sequence strategy arises as a natural way to map the unstructured text input into a structured code output [4]. By conceptualizing the problem as a mapping from one symbol sequence to another, it becomes possible to capture contextual clues across an entire document, even those spanning multiple sentences and paragraphs. The focus rests not on identifying discrete keywords but on learning continuous representations that more effectively generalize across textual variation [5]. In many modern implementations, recurrent neural networks or transformers serve as the core of the encoder-decoder pipeline, providing the means to capture hidden patterns and hierarchical aspects of language [6]. Attention layers often complement these core architectures, permitting the model to selectively emphasize relevant tokens when predicting the subsequent output symbol. Such designs are fully end-to-end differentiable, enabling the direct optimization of target codes without intermediate rule-based filters. [7]

Despite these architectural advantages, training a high-quality sequence-to-sequence model for procedure coding requires more than brute-force application of encoder-decoder modules. The complexities of clinical text, including domain-specific terminology, frequent abbreviations, and ambiguous expressions, necessitate specialized data preparation and pre-processing [8]. Moreover, practical deployments must contend with potentially noisy transcriptions, incomplete sentences, or the presence of extraneous details that do not correspond to any standard code. A robust system must learn which parts of the text to disregard and which to attend to, all while preserving the interpretability demanded by medical professionals [9]. The capacity to pinpoint the textual evidence used in the final code prediction is often as critical as the accuracy of the prediction itself, especially within a clinical setting where accountability and auditing frequently occur.

The present work explores a sequence-to-sequence framework that addresses the multifaceted nature of clinical text [10]. Drawing upon large text corpora and specialized token representations, the system builds a semantic alignment between clinical tokens and codes [11]. The hypothesis is that the latent representation space encapsulates both localized syntactic information and global semantic information, enabling the model to generalize more effectively compared to conventional classification mechanisms. The architecture is trained to decode a sequence of procedure tokens that reflect standardized coding guidelines [12]. We present a rigorous investigation of the data pipeline, encoder-decoder design, training and inference schemes, and the overall performance across real-world clinical notes. The model's capacity for handling ambiguous or incomplete references is examined through controlled experiments, alongside an analysis of efficiency considerations pertinent to large-scale healthcare systems. [13]

The immediate contributions of this study are grounded in bridging deep sequence modeling and practical clinical code assignment. By integrating specialized token embeddings, attention-based decoding, and domain-aligned training curricula, this approach aims to reduce manual efforts that are both time-consuming and susceptible to error [14]. From a methodological perspective, the research situates itself at the intersection of sequence learning, natural language processing, and clinical informatics, proposing new ways of extracting structured insight from otherwise unstructured documents. The subsequent sections explore data representation, model formulation, experimental approaches, and quantitative results, culminating in a discussion of challenges, potential applications, and future research directions that could further refine automated coding systems. [15]

2. Data Representation

The transformation of unstructured text into meaningful representations is crucial for predictive accuracy and interpretability [16]. When dealing with clinical notes, issues such as domain-specific terminology, acronyms, and potential misspellings arise. To address these challenges, a tokenization process is introduced that splits each note into its constituent lexical units [17]. Let the input text be denoted by $x = \{x_1, x_2, \ldots, x_n\}$, where each x_i is a token derived from the original sentence. A vocabulary V is defined, typically consisting of the most frequent words, clinical abbreviations, and special tokens to handle out-of-vocabulary instances. If a token is not in V, it is replaced with a special symbol $\langle \text{unk} \rangle$.

Each token x_i is mapped to a dense embedding vector $\mathbf{e}_{x_i} \in \mathbb{R}^d$, where d is the embedding dimension. Thus, the note is converted to a sequence of embeddings $\{\mathbf{e}_{x_1}, \mathbf{e}_{x_2}, \ldots, \mathbf{e}_{x_n}\}$. Often, these embeddings are initialized with pre-trained vectors derived from large biomedical corpora, thereby offering a specialized semantic grounding [18]. Over the course of training, they are updated to better align with the procedure coding task.

Textual irregularities require additional mechanisms beyond simple token mapping [19]. Common abbreviations such as "PT" for "physical therapy" or "BP" for "blood pressure" may appear alongside unrecognized short forms that vary across institutions. The approach is to treat recognized abbreviations as separate tokens, while unrecognized ones become subwords that retain some morphological structure

[20]. Denote by Π the set of recognized domain abbreviations, then for any substring $s \subseteq x_i$, if $s \in \Pi$, it is considered a single token [21]. Otherwise, subword decomposition occurs, yielding a consistent representation strategy that can handle new or rare substrings. Let the decomposition function be $\delta(x_i)$, mapping a token to its subword units or recognized forms in Π [22]. Then the final embedding vector for token x_i can be seen as a concatenation or summation of subword embeddings, defined as $\mathbf{e}_{x_i} = \gamma(\{\mathbf{e}_s \mid s \in \delta(x_i)\})$, where γ is a pooling function.

Handling synonyms and synonyms-like expansions also requires the data representation to maintain continuity. A symbol x_i that is encountered in multiple forms across different notes (for example, "BP," "blood pressure," or "B.P.") may point to the same concept [23]. Let μ be a mapping from tokens to concept identifiers, such that $\mu(x_i) = c_k$. Then one can augment the embedding vector with concept embeddings $\mathbf{c}_{c_k}$, aiming to unify synonyms under a shared representation. This integrated representation can be expressed as [24]

$$\mathbf{e}'_{x_i} = \mathbf{e}_{x_i} \oplus \mathbf{c}_{\mu(x_i)},$$

where $\oplus$ denotes vector concatenation. By training concept embeddings jointly with token embeddings, it becomes feasible to capture semantic equivalences that otherwise might be obscured by surface-level textual differences. [25]

Beyond token-level features, additional meta-information frequently appears in clinical notes: physician ID, department, or note type [26]. Let $\ell(\cdot)$ be a function assigning each note to its metadata vector $\mathbf{m} \in \mathbb{R}^p$. This metadata can be concatenated to the first or final hidden state in the encoder, injecting context that might improve code prediction. If $\mathbf{h}_1^{(\mathrm{enc})}$ is the first hidden state of the encoder, one might define

$$\tilde{\mathbf{h}}_1^{(\mathrm{enc})} = \mathbf{h}_1^{(\mathrm{enc})} \oplus \mathbf{m}.$$

Such a transformation is particularly helpful when certain departments predominantly handle specific procedures, or when the knowledge of a physician's specialty significantly narrows the possible range of procedure codes. [27]

Throughout the entire data representation pipeline, consistency is maintained by ensuring that each textual and metadata component translates into a numeric vector in $\mathbb{R}^d$. The result is a coherent set of embeddings that feed into the subsequent encoding, alignment, and decoding mechanisms. This strategy accommodates the wide variability of clinical text and leverages domain cues to enhance overall model performance. [28]

3. Model Architecture

The underlying framework is cast as a conditional sequence generation problem, where the model learns a conditional probability distribution $p_\theta(y \mid x)$ over output sequences $y = \{y_1, y_2, \ldots, y_m\}$ given an input text $x = \{x_1, x_2, \ldots, x_n\}$. The sequence y corresponds to the predicted procedure codes, often tokenized in a similar manner to the input. The architecture, composed of an encoder and a decoder, is trained to maximize the log-likelihood of the observed sequences: [29]

$$\mathcal{L}(\theta) = \sum_{(x,y)} \log p_\theta(y \mid x).$$

To capture the complexities of clinical language, a layered encoder transforms the input embeddings into context-rich hidden states [30]. Let $\mathbf{X} = \{\mathbf{e}_{x_1}, \mathbf{e}_{x_2}, \ldots, \mathbf{e}_{x_n}\}$ be the embedded input. The encoder's recurrent or transformer layers produce a sequence of hidden vectors $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\}$. A common approach for recurrent designs uses gated recurrent units or long short-term memory cells to alleviate vanishing or exploding gradient issues. In contrast, a transformer employs self-attention blocks, computing [31]

$$\mathbf{h}_i = \mathrm{TransformerBlock}(\mathbf{h}_{i-1}, \ldots, \mathbf{h}_{i-k}).$$

The result is a hidden representation that preserves local syntactic details as well as global context relationships among tokens.

The decoder is tasked with generating the output tokens one step at a time [32]. For a recurrent decoder, the hidden state at time step t is defined as

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{y}_{t-1}, \mathbf{c}_t),$$

where $\mathbf{y}_{t-1}$ is the embedding of the previously generated token, and $\mathbf{c}_t$ is a context vector derived from the encoder outputs via an attention mechanism. Specifically, the attention mechanism produces weights $\alpha_{t,i}$ that indicate the relevance of encoder state $\mathbf{h}_i$ to the decoder's state $\mathbf{z}_t$. One may define [33]

$$\alpha_{t,i} = \frac{\exp(\mathbf{z}_{t-1}^\top \mathbf{h}_i)}{\sum_{j=1}^{n} \exp(\mathbf{z}_{t-1}^\top \mathbf{h}_j)},$$

and

$$\mathbf{c}_t = \sum_{i=1}^{n} \alpha_{t,i}\, \mathbf{h}_i.$$

The next token distribution is computed through a linear transformation followed by softmax:

$$p_\theta(y_t \mid y_{<t}, x) = \mathrm{softmax}(\mathbf{W}_o\, \mathbf{z}_t + \mathbf{b}_o).$$

For a transformer-based decoder, self-attention and cross-attention layers are stacked, capturing not only relations among the previously generated tokens but also how they align with the encoder states. [34]

From a linear algebra perspective, each attention head in the transformer involves computing a key, query, and value decomposition [35]. Let the encoder output $\mathbf{H}$ be shaped as a matrix of dimension $n \times d$. Then a multi-head attention block calculates

$$\mathbf{Q} = \mathbf{H}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{H}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{H}\mathbf{W}_V,$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d \times d_h}$. The self-attention output is given by [36]

$$\mathrm{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_h}}\right)\mathbf{V}.$$

Multiple attention heads are concatenated, and a final linear mapping $\mathbf{W}_O$ transforms the concatenation back to $\mathbb{R}^d$. This mechanism is replicated in both the encoder and decoder, albeit with a cross-attention step in the decoder that uses the encoder outputs as $\mathbf{K}$ and $\mathbf{V}$, and the decoder hidden states as $\mathbf{Q}$.

The model parameters include the embedding matrices, recurrent or transformer weights, and output projections. Let $\Theta = \{\mathbf{E}_x, \mathbf{E}_y, \mathbf{W}_{enc}, \mathbf{W}_{dec}, \mathbf{W}_o, \ldots\}$. Training proceeds by minimizing the negative log-likelihood: [37]

$$\mathcal{L}(\Theta) = -\sum_{(x,y)\in D_{\mathrm{train}}} \sum_{t=1}^{m} \log p_\Theta(y_t \mid y_{<t}, x).$$

A gradient-based optimization algorithm, often Adam or a variant, updates these parameters iteratively. Formally, one can define [38]

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}(\Theta),$$

and the solution typically is approximated through mini-batch stochastic gradient descent. This architecture thus delineates a mechanism that can, in principle, capture the varied linguistic structures in clinical text and produce a final code sequence that aligns with the standardized codes used in healthcare billing and documentation. [39]

4. Experimental Setup

Experiments were conducted on a large dataset of clinical notes, each annotated with one or more procedure codes [40]. The data spanned multiple healthcare institutions, ensuring diversity in writing style and note composition. Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ represent the corpus, where x_i is a tokenized input note, and y_i is the corresponding tokenized code sequence. Before training, a cleaning routine removed personally identifiable information, ensuring compliance with privacy regulations [41]. Additional de-identification processes replaced unique IDs, names, and specific location references with placeholders that do not compromise clinical utility.

A random split was performed to create training, validation, and test subsets in proportions of 70%, 15%, and 15% respectively [42]. The training set was used to update model parameters, while the validation set was used for hyperparameter tuning, early stopping, and learning rate scheduling. The final model evaluation took place on the held-out test set [43]. Hyperparameters were chosen through a grid search over embedding dimensions, learning rates, batch sizes, and model depth. Let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ denote the set of hyperparameters. For each $\lambda_j \in \Lambda$, a candidate model was trained, and the selection was based on minimizing the validation cross-entropy loss [44]. Further refinements included checking the stability of attention distributions and verifying that the model did not overfit to training data. [45]

The data pipeline involved converting all tokens to lowercase, removing extraneous symbols except for medically relevant punctuation, and applying the subword decomposition approach described earlier. Abbreviations within Π were processed to preserve their complete forms, while other unknown tokens were split into subword segments [46]. All embeddings were either randomly initialized or used with domain-specific pre-training from a large collection of clinical text. Let Ω be the pre-trained embedding matrix, where $\Omega \in \mathbb{R}^{|V| \times d}$. If an embedding for a token existed in Ω, it was used as initialization; otherwise, a random vector in $\mathbb{R}^d$ was assigned.

The sequence-to-sequence model was implemented using a deep learning framework that supports automatic differentiation [47]. In the recurrent setup, two-layer bidirectional encoders with gated recurrent units were used, each layer containing d hidden units. The decoder was a unidirectional recurrent layer with an attention mechanism that consumed the final encoder states and produced the procedure code tokens [48]. In the transformer setup, the encoder and decoder each contained multiple self-attention blocks with multi-head attention. The dimensionality of hidden layers and the feed-forward sublayers was set to match that of the embedding dimension to maintain consistency [49]. A dropout probability of 0.3 was applied to reduce overfitting, and gradient clipping was employed to prevent exploding gradients. [50]

Training proceeded for up to 30 epochs, with early stopping triggered if the validation loss did not improve for 3 consecutive epochs. Let r be the initial learning rate for Adam; a warmup schedule increased r linearly for the first few thousand steps, after which it decayed according to an inverse square root schedule [51]. This technique helps mitigate the instability of training large models from scratch. Each epoch iterated over mini-batches of size 32 or 64, depending on the memory constraints of the hardware configuration [52]. The average run time per epoch was approximately 30 minutes on a GPU-equipped server, reflecting the computational intensity of attending to lengthy clinical documents.

During inference, beam search was used to generate the predicted procedure code sequence [53]. If B is the beam size, the model maintains B candidate sequences at each step, scoring them according to their cumulative log probabilities. This increases the chances of recovering the correct sequence of codes, especially when multiple plausible paths exist [54]. Beam sizes ranged from 3 to 10 in different experimental runs, and higher values did not always yield substantial improvements at the cost of increased computational overhead [55]. The final predictions were concatenated if multiple codes were relevant or if the model indicated that multiple procedure symbols should be generated.

To evaluate performance, precision, recall, and F1 metrics were computed at both the token level and the sequence level [56]. The token-level metrics focus on each predicted code token, whereas sequence-level metrics treat the entire predicted sequence as correct only if it perfectly matches the ground truth.

For domain-specific analysis, an exact match on the entire set of codes assigned to a note is considered [57]. Let $\widehat{y}$ denote the model's predicted code sequence. Then exact-match accuracy is

$$\mathrm{ACC} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[\widehat{y}_i = y_i].$$

Additionally, partial credit is given when the predicted set of codes overlaps with the reference set, providing a more nuanced perspective of correctness in multi-code scenarios. [58]

5. Results and Analysis

The model's performance on the held-out test set demonstrated notable improvements over conventional classification-based methods [59]. On average, the sequence-to-sequence approach achieved a token-level F1 score of 0.82, surpassing baseline logistic regression or single-label classifiers that typically reached an F1 score around 0.68 to 0.75. This gap became more pronounced when the clinical notes contained multiple procedures spread across sentences, suggesting that the sequence model's capacity to capture context and dependency was critical [60]. Exact-match accuracy, which demands a perfect reconstruction of all codes, was approximately 0.64 for the best recurrent encoder-decoder model and rose to 0.67 for the transformer-based variant. These results underscore that while the model can achieve robust partial matches, the generation of complete, error-free sequences remains challenging when the domain vocabulary is vast. [61]

Analysis of attention distributions indicated that the model tended to focus on specific key terms, but also on entire spans of text containing relevant clinical actions. For instance, references to "performed," "initiated," or "administered" often prompted the network to look ahead for the procedures themselves [62]. Let us define a logic statement to express the attention coverage. For each encoder token index i, there exists a decoder step t such that $\alpha_{t,i} \geq \delta$ for some threshold δ. Symbolically, $\forall i \in \{1, \ldots, n\}, \exists t \in \{1, \ldots, m\} : \alpha_{t,i} \geq \delta$. This property was observed empirically in many sequences, indicating broad coverage of relevant textual segments, and it also highlighted potential improvements in interpretability.

Ablation experiments tested how various design choices impacted the final performance [63]. Removing attention mechanisms significantly degraded accuracy, falling to an F1 score of approximately 0.74 [64]. Eliminating the subword decomposition technique increased the proportion of out-of-vocabulary tokens and led to reduced recall. In contrast, feeding metadata into the encoder's initial state improved results in certain specialized departments, raising F1 scores by 0.01 to 0.02 points in surgical departments but having minimal effect in more general settings [65]. This indicates that while metadata can be beneficial, its impact can be domain-specific. In addition, substituting domain-specific pre-trained embeddings with random initializations consistently lowered performance, emphasizing the importance of capturing biomedical terminology in the representation space. [66]

Several qualitative examples underscored the system's capacity to handle ambiguous expressions. Consider a note describing both a diagnostic imaging procedure and a minor therapeutic intervention [67]. The model's final prediction included two codes, capturing both the imaging code and the procedure code referencing an injection. This underscores the advantage of generating a sequence of codes, rather than merely assigning a single label or predicting from a fixed classification scheme [68]. Errors tended to occur when the note was especially long or included a sequence of unrelated prior procedures that overshadowed the relevant ones [69]. In such cases, the model sometimes generated extraneous codes based on partial matches with older segments of text.

Complex cases involving specialized medical practices illustrated the utility of domain alignments [70]. When the system was trained on a small subset of pediatric oncology notes, it learned to differentiate between chemotherapy administration codes and more routine procedures like intravenous lines. The improvements in recall suggested that the embedding space effectively captured crucial semantic distinctions [71]. Let $\mathbf{u}$ be the embedded representation for "chemotherapy" and $\mathbf{v}$ be the embedded representation for “intravenous,” then the learned dot product $\mathbf{u}^\top \mathbf{v}$ indicated a moderate correlation, reflecting the fact that both terms appeared frequently in overlapping contexts but pertain to distinct coding categories. Where older rule-based methods might conflate these contexts or require explicit dictionary expansions, the learned architecture handles them by adapting during training.

From a computational perspective, training the sequence-to-sequence model required greater resources than simpler classifiers [72]. The extended training time reflects the complexity of capturing multi-step dependencies and the cost of computing attention scores for each pair of encoder-decoder tokens. Nevertheless, once trained, inference can be streamlined through optimized beam search procedures [73]. The experiment showed that a beam size of 5 balanced performance and run-time, taking about 0.5 seconds per note on an average GPU system [74]. Larger beams, while slightly more accurate in certain edge cases, increased latency and might not be practical for real-time coding applications.

These findings demonstrate that the sequence-to-sequence framework not only improves quantitative performance metrics but also captures the nuanced structure of clinical text in ways that simpler classification pipelines cannot [75]. By analyzing attention maps and ablation results, it becomes clear that a careful orchestration of data representation, embedding strategies, and architecture design yields robust predictions. This model moves toward the ultimate objective of enabling more efficient clinical coding, reducing manual overhead, and minimizing errors that could compromise patient records or billing accuracy. [76]

6. Conclusion

This investigation has presented a sequence-to-sequence system for predicting procedure codes from unstructured clinical text. By modeling the task as a conditional sequence generation problem, the approach effectively captures both localized linguistic cues and long-range context, resulting in a notable improvement over baseline classification strategies [77]. The data representation pipeline, encompassing subword decomposition, specialized embeddings, and optional metadata fusion, ensures that domain-specific terminologies are adequately captured, mitigating issues posed by abbreviations, synonyms, and incomplete text. Throughout experiments, attention mechanisms have emerged as a key component, enabling the model to interpret diverse textual inputs and isolate the pertinent details that inform final procedure code predictions. [78]

In many cases, the model exhibits a strong capability to generate sequences of multiple codes, reflecting the inherent complexity of clinical notes that may allude to several procedures [79]. Rigorous evaluations reveal improvements in F1 scores, exact-match accuracy, and robustness against context shifts. The results also highlight some challenges, such as the occasional over-generation of codes and the computational intensity required to train deep encoder-decoder architectures [80]. Nevertheless, through ablation studies, it becomes evident that each architectural and representational choice—ranging from attention-based alignment to domain-specific embedding initialization—plays a vital role in the overall performance.

The outcomes underscore the feasibility of integrating sequence-to-sequence learning into large-scale clinical coding workflows [81]. Such integration can potentially reduce the reliance on manual coders, who often grapple with arduous tasks of navigating lengthy documents. Beyond immediate applications in billing or record-keeping, automated coding pipelines may serve broader goals of healthcare analytics by rapidly rendering unstructured text into standardized representations that are amenable to quantitative analysis [82]. Further explorations may refine or extend the model, introducing advanced regularization schemes, more extensive domain adaptation, or additional interpretability layers to gain deeper insights into the prediction process. As developments in natural language processing and deep learning continue, the potential to streamline clinical documentation processes through automated coding systems will likely expand, making real-time, reliable, and context-aware procedure code assignment a practical reality. [83]

References

[1] A. Sharma and K. Goolsbey, “Identifying useful inference paths in large commonsense knowledge bases by retrograde analysis,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017.

[2] A. Sorbello, S. A. Haque, R. Hasan, R. Jermyn, A. Hussein, A. Vega, K. Zembrzuski, A. Ripple, and M. Ahadpour, “Artificial intelligence-enabled software prototype to inform opioid pharmacovigilance from electronic health records: Development and usability study.,” *JMIR AI*, vol. 2, pp. e45000–e45000, 7 2023.

[3] A. Arora, J. E. Alderman, J. Palmer, S. Ganapathi, E. Laws, M. D. McCradden, L. Oakden-Rayner, S. R. Pfohl, M. Ghassemi, F. McKay, D. Treanor, N. Rostamzadeh, B. Mateen, J. Gath, A. O. Adebajo, S. Kuku, R. Matin, K. Heller, E. Sapey, N. J. Sebire, H. Cole-Lewis, M. Calvert, A. Denniston, and X. Liu, “The value of standards for health datasets in artificial intelligence-based applications.,” *Nature medicine*, vol. 29, pp. 2929–2938, 10 2023.

[4] A. Sharma and K. M. Goolsbey, “Simulation-based approach to efficient commonsense reasoning in very large knowledge bases,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1360–1367, 2019.

[5] J. Mao, G. G. F. Gomez, M. Wang, H. Xu, and T. P. Thyvalikakath, “Prediction of sjögren’s disease diagnosis using matched electronic dental-health record data.,” *BMC medical informatics and decision making*, vol. 24, pp. 43–, 2 2024.

[6] A. M. Simms, A. Kanakia, M. Sipra, B. Dutta, and N. Southall, “A patient safety knowledge graph supporting vaccine product development.,” *BMC medical informatics and decision making*, vol. 24, pp. 10–, 1 2024.

[7] E. C. Seng, S. Mehdipour, S. Simpson, and R. A. Gabriel, “Tracking persistent postoperative opioid use: a proof-of-concept study demonstrating a use case for natural language processing.,” *Regional anesthesia and pain medicine*, vol. 49, pp. 241–247, 4 2024.

[8] I. C. Leist, M. Rivas-Torrubia, M. E. Alarcón-Riquelme, G. Barturen, P. C. Consortium, I. G. Gut, and M. Rueda, “Pheno-ranker: a toolkit for comparison of phenotypic data stored in ga4gh standards and beyond.,” *BMC bioinformatics*, vol. 25, pp. 373–, 12 2024.

[9] A. Gandomi, E. Hasan, J. Chusid, S. Paul, M. Inra, A. Makhnevich, S. Raoof, G. Silvestri, B. C. Bade, and S. L. Cohen, “Evaluating the accuracy of lung-rads score extraction from radiology reports: Manual entry versus natural language processing.,” *International journal of medical informatics*, vol. 191, pp. 105580–105580, 7 2024.

[10] A. Li, O. Jafari, S. Ma, A. Maghsoudi, B. D. Lam, J. Ryu, J. Y. Jiang, M. Ranjan, E. Zhou, R. Bandyo, B. Peng, C. I. Amos, N. R. Fillmore, A. O. Oluyomi, and J. La, “Optimized multi-class <i>vte-bert</i> large language model for prediction of cancer associated thrombosis phenotype,” *Blood*, vol. 144, pp. 161–161, 11 2024.

[11] D. Gangwani and X. Zhu, “Modeling and prediction of business success: a survey,” *Artificial Intelligence Review*, vol. 57, 2 2024.

[12] V. K. Keloth, S. Zhou, L. Lindemann, L. Zheng, G. Elhanan, A. J. Einstein, J. Geller, and Y. Perl, “Mining of ehr for interface terminology concepts for annotating ehrs of covid patients.,” *BMC medical informatics and decision making*, vol. 23, pp. 40–, 2 2023.

[13] C. Zheng, M.-S. Lee, N. Bansal, A. S. Go, C. Chen, T. N. Harrison, D. Fan, A. Allen, E. Garcia, B. Lidgard, D. Singer, and J. An, “Identification of recurrent atrial fibrillation using natural language processing applied to electronic health records.,” *European heart journal. Quality of care & clinical outcomes*, vol. 10, pp. 77–88, 3 2023.

[14] A. Bartal, K. M. Jagodnik, S. J. Chan, and S. Dekel, “Ai and narrative embeddings detect ptsd following childbirth via birth stories.,” *Scientific reports*, vol. 14, pp. 8336–, 4 2024.

[15] J. R. Machireddy, “Machine learning and automation in healthcare claims processing,” *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, vol. 6, no. 1, pp. 686–701, 2024.

[16] G. Zhang, Q. Jin, Y. Zhou, S. Wang, B. Idnay, Y. Luo, E. Park, J. G. Nestor, M. E. Spotnitz, A. Soroush, T. R. Campion, Z. Lu, C. Weng, and Y. Peng, “Closing the gap between open source and commercial large language models for medical evidence summarization.,” *NPJ digital medicine*, vol. 7, pp. 239–, 9 2024.

[17] M. L. Naseem, “Trans-ifft-fgsm: a novel fast gradient sign method for adversarial attacks,” *Multimedia Tools and Applications*, vol. 83, pp. 72279–72299, 2 2024.

[18] W. Rojas-Carabali, R. Agrawal, L. Gutierrez-Sinisterra, S. L. Baxter, C. Cifuentes-González, Y. C. Wei, J. Abisheganaden, P. Kannapiran, S. Wong, B. Lee, A. de-la Torre, and R. Agrawal, "Natural language processing in medicine and ophthalmology: A review for the 21st-century clinician.," *Asia-Pacific journal of ophthalmology (Philadelphia, Pa.)*, vol. 13, pp. 100084–100084, 7 2024.

[19] P. Esmaeilzadeh, "The role of chatgpt in disrupting concepts, changing values, and challenging ethical norms: a qualitative study," *AI and Ethics*, vol. 5, pp. 291–304, 9 2023.

[20] U. Naseem, S. Thapa, Q. Zhang, L. Hu, J. Rashid, and M. Nasim, "Incorporating historical information by disentangling hidden representations for mental health surveillance on social media," *Social Network Analysis and Mining*, vol. 14, 12 2023.

[21] H. L. Amonoo, N. H. Markovitz, P. C. Johnson, A. Kwok, C. Dale, E. C. Deary, E. Daskalakis, J. J. Choe, N. Yamin, M. Gothoskar, K. G. Cronin, C. Fernandez-Robles, W. F. Pirl, Y.-B. Chen, C. Cutler, C. Lindvall, and A. El-Jawahri, "Delirium and healthcare utilization in patients undergoing hematopoietic stem cell transplantation.," *Transplantation and cellular therapy*, vol. 29, pp. 334.e1–334.e7, 2 2023.

[22] K. K. Park, M. Saleem, M. A. Al-Garadi, and A. Ahmed, "Machine learning applications in studying mental health among immigrants and racial and ethnic minorities: an exploratory scoping review.," *BMC medical informatics and decision making*, vol. 24, pp. 298–, 10 2024.

[23] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Big data in cloud computing review and opportunities," *arXiv preprint arXiv:1912.10821*, 2019.

[24] J. R. Vest and O. Mazurenko, "Non-response bias in social risk factor screening among adult emergency department patients.," *Journal of medical systems*, vol. 47, pp. 78–, 7 2023.

[25] A. K. Saxena, "Evaluating the regulatory and policy recommendations for promoting information diversity in the digital age," *International Journal of Responsible Artificial Intelligence*, vol. 11, no. 8, pp. 33–42, 2021.

[26] W. Z. Kariampuzha, G. Alyea, S. Qu, J. Sanjak, E. Mathé, E. Sid, H. Chatelaine, A. Yadaw, Y. Xu, and Q. Zhu, "Precision information extraction for rare disease epidemiology at scale.," *Journal of translational medicine*, vol. 21, pp. 157–, 2 2023.

[27] M. Y. Mustafa, A. Tlili, G. Lampropoulos, R. Huang, P. Jandrić, J. Zhao, S. Salha, L. Xu, S. Panda, null Kinshuk, S. López-Pernas, and M. Saqr, "A systematic review of literature reviews on artificial intelligence in education (aied): a roadmap to a future research agenda," *Smart Learning Environments*, vol. 11, 12 2024.

[28] A. E. Krumm, E. Ötleş, K. M. Marcotte, B. L. Spencer, S. Izadi, B. C. George, and B. Zendejas, "Strategies for evaluating predictive models: examples and implications based on a natural language processing model used to assess operative performance feedback," *Global Surgical Education - Journal of the Association for Surgical Education*, vol. 3, 12 2023.

[29] E. Lin, R. Zwolinski, J. T.-Y. Wu, J. La, S. Goryachev, L. Huhmann, C. Yildrim, D. P. Tuck, D. C. Elbers, M. T. Brophy, N. V. Do, and N. R. Fillmore, "Machine learning-based natural language processing to extract pd-l1 expression levels from clinical notes.," *Health informatics journal*, vol. 29, pp. 14604582231198021–, 8 2023.

[30] S. Yang, X. Yang, T. Lyu, J. L. Huang, A. Chen, X. He, D. Braithwaite, H. J. Mehta, Y. Wu, Y. Guo, and J. Bian, "Extracting pulmonary nodules and nodule characteristics from radiology reports of lung cancer screening patients using transformer models.," *Journal of healthcare informatics research*, vol. 8, pp. 463–477, 5 2024.

[31] S. Sivarajkumar, H. A. Mohammad, D. Oniani, K. Roberts, W. Hersh, H. Liu, D. He, S. Visweswaran, and Y. Wang, "Clinical information retrieval: A literature review.," *Journal of healthcare informatics research*, vol. 8, pp. 313–352, 1 2024.

[32] V. Chittajallu, E. Mansoor, J. Perez, Y. A. Omar, S. A. Firkins, H. Yoo, B. Baggott, and R. Simons-Linares, "Association of bariatric surgery with risk of incident obesity-associated malignancies: a multi-center population-based study.," *Obesity surgery*, vol. 33, pp. 4065–4069, 11 2023.

[33] M. Kia, S. Sadeghi, H. Safarpour, M. Kamsari, S. J. Ghoushchi, and R. Ranjbarzadeh, "Innovative fusion of vgg16, mobilenet, efficientnet, alexnet, and resnet50 for mri-based brain tumor identification," *Iran Journal of Computer Science*, vol. 8, pp. 185–215, 12 2024.

[34] C. Pais, J. Liu, R. Voigt, V. Gupta, E. Wade, and M. Bayati, "Large language models for preventing medication direction errors in online pharmacies.," *Nature medicine*, vol. 30, pp. 1574–1582, 4 2024.

[35] S. Nerella, S. Bandyopadhyay, J. Zhang, M. Contreras, S. Siegel, A. Bumin, B. Silva, J. Sena, B. Shickel, A. Bihorac, K. Khezeli, and P. Rashidi, “Transformers and large language models in healthcare: A review.,” *Artificial intelligence in medicine*, vol. 154, pp. 102900–102900, 6 2024.

[36] D. MacMath, M. Chen, and P. Khoury, “Artificial intelligence: Exploring the future of innovation in allergy immunology.,” *Current allergy and asthma reports*, vol. 23, pp. 351–362, 5 2023.

[37] S. Borna, C. A. Gomez-Cabello, S. M. Pressman, S. A. Haider, A. Sehgal, B. C. Leibovich, D. Cole, and A. J. Forte, “Comparative analysis of artificial intelligence virtual assistant and large language models in post-operative care.,” *European journal of investigation in health, psychology and education*, vol. 14, pp. 1413–1424, 5 2024.

[38] V. Joshi, P. Mitra, and S. Bose, “Multi-modal multi-head self-attention for medical vqa,” *Multimedia Tools and Applications*, vol. 83, pp. 42585–42608, 10 2023.

[39] S. Bhanvadia, B. R. Saseendrakumar, J. Guo, M. Spadafore, M. Daniel, L. Lander, and S. L. Baxter, “Evaluation of bias and gender/racial concordance based on sentiment analysis of narrative evaluations of clinical clerkships using natural language processing.,” *BMC medical education*, vol. 24, pp. 295–, 3 2024.

[40] K. Araki, N. Matsumoto, K. Togo, N. Yonemoto, E. Ohki, L. Xu, Y. Hasegawa, H. Inoue, S. Yamashita, and T. Miyazaki, “Real-world treatment response in japanese patients with cancer using unstructured data from electronic health records,” *Health and Technology*, vol. 13, pp. 253–262, 2 2023.

[41] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Federated query processing for big data in data science,” in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 6145–6147, IEEE, 2019.

[42] S. Boligarla, E. K. E. Laison, J. Li, R. Mahadevan, A. Ng, Y. Lin, M. Y. Thioub, B. Huang, M. H. Ibrahim, and B. Nasri, “Leveraging machine learning approaches for predicting potential lyme disease cases and incidence rates in the united states using twitter.,” *BMC medical informatics and decision making*, vol. 23, pp. 217–, 10 2023.

[43] H. Nakai, G. Suman, D. A. Adamo, P. J. Navin, C. A. Bookwalter, J. D. LeGout, F. K. Chen, C. V. Wellnitz, A. C. Silva, J. V. Thomas, A. Kawashima, J. W. Fan, A. T. Froemming, D. J. Lomas, M. R. Humphreys, C. Dora, P. Korfiatis, and N. Takahashi, “Natural language processing pipeline to extract prostate cancer-related information from clinical notes.,” *European radiology*, vol. 34, pp. 7878–7891, 6 2024.

[44] R. Lotfi, A. M. Nasrabadi, S. S. Ali, N. Mardani, S. M. R. Davoodi, and S. Aghakhani, “A viable relief supply chain network design by considering risk and robustness for disaster and crisis management,” *Central European Journal of Operations Research*, 7 2024.

[45] A. Sharma and K. Forbus, “Automatic extraction of efficient axiom sets from large knowledge bases,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 27, pp. 1248–1254, 2013.

[46] J. R. Machireddy, “Automation in healthcare claims processing: Enhancing efficiency and accuracy,” *International Journal of Science and Research Archive*, vol. 09, no. 01, pp. 825–834, 2023.

[47] H.-J. Lee, L. H. Schwamm, L. H. Sansing, H. Kamel, A. de Havenon, A. C. Turner, K. N. Sheth, S. Krishnaswamy, C. Brandt, H. Zhao, H. Krumholz, and R. Sharma, “Strokeclassifier: ischemic stroke etiology classification by ensemble consensus modeling using electronic health records.,” *NPJ digital medicine*, vol. 7, pp. 130–, 5 2024.

[48] S. A. Haider, S. Borna, C. A. Gomez-Cabello, S. M. Pressman, C. R. Haider, and A. J. Forte, “The algorithmic divide: A systematic review on ai-driven racial disparities in healthcare.,” *Journal of racial and ethnic health disparities*, 12 2024.

[49] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Context-aware query performance optimization for big data analytics in healthcare,” in *2019 IEEE High Performance Extreme Computing Conference (HPEC-2019)*, pp. 1–7, 2019.

[50] S. Steyaert, M. Pizurica, D. Nagaraj, P. Khandelwal, T. Hernandez-Boussard, A. J. Gentles, and O. Gevaert, “Multimodal data fusion for cancer biomarker discovery with deep learning.,” *Nature machine intelligence*, vol. 5, pp. 351–362, 4 2023.

[51] B. K. Beaulieu-Jones, F. Frau, S. Bozzi, K. J. Chandross, M. J. Peterschmitt, C. Cohen, C. Coulovrat, D. Kumar, M. J. Kruger, S. L. Lipnick, L. Fitzsimmons, I. S. Kohane, and C. R. Scherzer, “Disease progression strikingly differs in research and real-world parkinson’s populations.,” *NPJ Parkinson’s disease*, vol. 10, pp. 58–, 3 2024.

[52] P. Ma, Y. Cheng, J. L. Goulet, F. Sandbrink, C. Brandt, C. Spevak, J. T. Kean, W. Becker, A. Libin, N. Shara, H. M. Sheriff, J. S. Houston, J. Butler, E. T. Workman, R. M. Agrawal, J. Kupersmith, and Q. Zeng-Treitler, “Guideline concordant opioid therapy in veterans receiving va and community care.,” *BMC health services research*, vol. 24, pp. 1284–, 10 2024.

[53] B. Zaidat, J. Tang, V. Arvind, E. A. Geng, B. Cho, A. H. Duey, C. Dominy, K. D. Riew, S. K. Cho, and J. S. Kim, "Can a novel natural language processing model and artificial intelligence automatically generate billing codes from spine surgical operative notes?," *Global spine journal*, vol. 14, pp. 2022–2030, 3 2023.

[54] T. A. Miller, A. J. McMurry, J. Jones, D. Gottlieb, and K. D. Mandl, "The smart text2fhir pipeline.," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2023, pp. 514–, 3 2023.

[55] Y. Wang, J. A. Bannon, N. Roszkowska, E. L. Wilson, S. Bonett, E. Lazarus, N. Dowshen, and R. Stevens, "From virus to viral: content analysis of hiv-related twitter messages among young men in the u.s.," *BMC Digital Health*, vol. 2, 7 2024.

[56] N. N. El-Den, M. Elsharkawy, I. Saleh, M. Ghazal, A. Khalil, M. Z. Haq, A. Sewelam, H. Mahdi, and A. El-Baz, "Ai-based methods for detecting and classifying age-related macular degeneration: a comprehensive review," *Artificial Intelligence Review*, vol. 57, 8 2024.

[57] D. Bertsimas and B. Öztürk, "Global optimization via optimal decision trees," *Journal of Global Optimization*, 7 2023.

[58] S. Somani, S. Balla, A. W. Peng, R. Dudum, S. Jain, K. Nasir, D. J. Maron, T. Hernandez-Boussard, and F. Rodriguez, "Contemporary attitudes and beliefs on coronary artery calcium from social media using artificial intelligence.," *NPJ digital medicine*, vol. 7, pp. 83–, 3 2024.

[59] I. H. Cerda, E. Zhang, M. Dominguez, M. Ahmed, M. Lang, S. Ashina, M. E. Schatman, R. J. Yong, and A. C. G. Fonseca, "Artificial intelligence and virtual reality in headache disorder diagnosis, classification, and management.," *Current pain and headache reports*, vol. 28, pp. 869–880, 6 2024.

[60] M. Wu, Y. Zhang, M. Markley, C. Cassidy, N. Newman, and A. Porter, "Covid-19 knowledge deconstruction and retrieval: an intelligent bibliometric solution.," *Scientometrics*, vol. 129, pp. 1–7259, 5 2023.

[61] C. Xiang and M. Abouelyazid, "The impact of generational cohorts and visit environment on telemedicine satisfaction: A novel investigation," 2020.

[62] E. Hermann, G. Y. Williams, and S. Puntoni, "Deploying artificial intelligence in services to aid vulnerable consumers," *Journal of the Academy of Marketing Science*, vol. 52, pp. 1431–1451, 11 2023.

[63] A. Bowles, C. Perez, A. Vachani, J. Steltz, B. Rose, A. K. Bryant, H. Eyre, S. L. DuVall, J. A. Lynch, and P. R. Alba, "An nlp framework for the extraction of concept measurements from radiology and pathology notes.," *Studies in health technology and informatics*, vol. 310, pp. 1446–, 1 2024.

[64] K. Slater, P. N. Schofield, J. Wright, P. Clift, A. Irani, W. Bradlow, F. Aziz, and G. V. Gkoutos, "Talking about diseases; developing a model of patient and public-prioritised disease phenotypes.," *NPJ digital medicine*, vol. 7, pp. 263–, 9 2024.

[65] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, "Mapa bilstm-bert: multi-aspects position aware attention for aspect level sentiment analysis," *The Journal of Supercomputing*, vol. 79, pp. 11452–11477, 2 2023.

[66] A. K. AlShami, R. Rabinowitz, K. Lam, Y. Shleibik, M. Mersha, T. Boult, and J. Kalita, "Smart-vision: survey of modern action recognition techniques in vision," *Multimedia Tools and Applications*, 12 2024.

[67] A. K. Saxena, R. R. Dixit, and A. Aman-Ullah, "An lstm neural network approach to resource allocation in hospital management systems," *International Journal of Applied Health Care Analytics*, vol. 7, no. 2, pp. 1–12, 2022.

[68] L. Y. Jiang, X. C. Liu, N. P. Nejatian, M. Nasir-Moin, D. Wang, A. Abidin, K. Eaton, H. A. Riina, I. Laufer, P. Punjabi, M. Miceli, N. C. Kim, C. Orillac, Z. Schnurman, C. Livia, H. Weiss, D. Kurland, S. Neifert, Y. Dastagirzada, D. Kondziolka, A. T. M. Cheung, G. Yang, M. Cao, M. Flores, A. B. Costa, Y. Aphinyanaphongs, K. Cho, and E. K. Oermann, "Health system-scale language models are all-purpose prediction engines.," *Nature*, vol. 619, pp. 357–362, 6 2023.

[69] P. G. Anderson, H. Tarder-Stoll, M. Alpaslan, N. Keathley, D. L. Levin, S. Venkatesh, E. Bartel, S. Sicular, S. Howell, R. V. Lindsey, and R. M. Jones, "Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest x-rays.," *Scientific reports*, vol. 14, pp. 25151–, 10 2024.

[70] J. Wang, J. S. de Vale, S. Gupta, P. Upadhyaya, F. A. Lisboa, S. A. Schobel, E. A. Elster, C. J. Dente, T. G. Buchman, and R. Kamaleswaran, "Clotcatcher: a novel natural language model to accurately adjudicate venous thromboembolism from radiology reports.," *BMC medical informatics and decision making*, vol. 23, pp. 262–, 11 2023.

[71] E. Hatef, C. Kitchen, G. M. Gray, A. Zirikly, T. Richards, L. M. Ahumada, and J. P. Weiner, "Enhancement of a social risk score in the electronic health record to identify social needs among medically underserved patients: using structured data and free-text provider notes.," *JAMIA open*, vol. 7, pp. ooae117–, 10 2024.

[72] S.-Y. J. Lee, M. Alzeen, and A. Ahmed, “Estimation of racial and language disparities in pediatric emergency department triage using statistical modeling and natural language processing.,” *Journal of the American Medical Informatics Association : JAMIA*, vol. 31, pp. 958–967, 2 2024.

[73] Q. Song, C. Ni, J. L. Warner, Q. Chen, L. Song, S. T. Rosenbloom, B. A. Malin, and Z. Yin, “Optimizing word embeddings for small dataset: a case study on patient portal messages from breast cancer patients.,” *Scientific reports*, vol. 14, pp. 16117–, 7 2024.

[74] V. LeBaron, T. Flickinger, D. Ling, H. Lee, J. Edwards, A. Tewari, Z. Wang, and L. E. Barnes, “Feasibility and acceptability testing of commsense: A novel communication technology to enhance health equity in clinician-patient interactions.,” *Digital health*, vol. 9, pp. 20552076231184991–, 7 2023.

[75] X. Zhang, G. Zhao, Y. Ren, W. Wang, W. Cai, Y. Zhao, X. Zhang, and J. Liu, “Data augmented large language models for medical record generation,” *Applied Intelligence*, vol. 55, 12 2024.

[76] M. Shahin, F. F. Chen, A. Hosseinzadeh, H. K. Koodiani, H. Bouzary, and A. Shahin, “Enhanced safety implementation in 5s+1 via object detection algorithms,” *The International Journal of Advanced Manufacturing Technology*, vol. 125, pp. 3701–3721, 2 2023.

[77] H. Y. Liu, M. A. Bonetti, F. D. Lorenzi, M. L. Gimbel, V. T. Nguyen, and F. M. Egro, “Consulting the digital doctor: Google versus chatgpt as sources of information on breast implant-associated anaplastic large cell lymphoma and breast implant illness.,” *Aesthetic plastic surgery*, vol. 48, pp. 590–607, 10 2023.

[78] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Approximate query processing for big data in heterogeneous databases,” in *2020 IEEE international conference on big data (big data)*, pp. 5765–5767, IEEE, 2020.

[79] A. Moglia, K. Georgiou, P. Cerveri, L. Mainardi, R. M. Satava, and A. Cuschieri, “Large language models in healthcare: from a systematic review on medical examinations to a comparative analysis on fundamentals of robotic surgery online test,” *Artificial Intelligence Review*, vol. 57, 8 2024.

[80] S. J. Burnett, R. Stemerman, J. C. Innes, M. C. Kaisler, R. P. Crowe, and B. M. Clemency, “Social determinants of health in ems records: A mixed-methods analysis using natural language processing and qualitative content analysis.,” *The western journal of emergency medicine*, vol. 24, pp. 878–, 9 2023.

[81] D. E. Malden, I.-L. A. Liu, L. Qian, L. S. Sy, B. J. Lewin, D. T. Asamura, D. S. Ryan, C. Bezi, J. T. B. Williams, R. Kaiser, M. F. Daley, J. C. Nelson, D. L. McClure, O. Zerbo, M. L. Henninger, C. C. Fuller, E. S. Weintraub, S. Saydah, and S. Y. Tartof, “Post-covid conditions following covid-19 vaccination: a retrospective matched cohort study of patients with sars-cov-2 infection.,” *Nature communications*, vol. 15, pp. 4101–, 5 2024.

[82] K. D. Forbus, K. Lockwood, A. B. Sharma, and E. Tomai, “Steps towards a second generation learning by reading system.,” in *AAAI Spring Symposium: Learning by Reading and Learning to Read*, pp. 36–43, 2009.

[83] A. Sharma and K. D. Forbus, “Modeling the evolution of knowledge and reasoning in learning systems,” in *2010 AAAI Fall Symposium Series*, 2010.