librasophia : Pages:27–41

Original Research

Adaptive Contextual Embeddings for Detecting Social Determinants of Health in Patient Narratives

Mert Aydoğan[1]

[1]Kırşehir Institute of Technology, Department of Computer Science, Bağbaşı Mah. 112. Sokak No:18, Kırşehir, Turkey.

Abstract

This paper investigates an advanced methodological framework for extracting Social Determinants of Health from patient narratives by leveraging adaptive contextual embeddings. Building upon contemporary natural language processing approaches, it aims to illuminate the mechanisms by which domain-specific context enhances feature representations in neural architectures. The central premise is that embedding spaces can be dynamically aligned with linguistic variability present in clinical text, thereby facilitating robust detection of factors such as socioeconomic status, housing stability, and access to care. Rather than relying on rigid static word vectors, the proposed approach adapts embedding spaces to capture latent relationships within patient descriptions, transcending shallow lexical correlations. The work further explores how auxiliary signals, derived from the semantic composition of clinically relevant terms, can refine the learned representations through iterative alignment techniques. In doing so, it addresses the challenges inherent in modeling subtle language patterns that encode sensitive social characteristics. By integrating advanced linear algebraic formulations and deductive logic statements into the core modeling process, the framework aspires to provide a new layer of interpretability and rigor. This paper will elaborate on the theoretical foundations, architecture, and empirical evaluations that substantiate the effectiveness of the proposed system, offering a blueprint for future innovations in adaptive embeddings for health information extraction and SDOH-driven predictive analytics.

1. Introduction

Social Determinants of Health, often abbreviated as SDOH, play a crucial role in shaping patient outcomes and informing public health strategies [1]. They encompass a broad spectrum of non-medical factors that include socioeconomic context, educational background, housing conditions, cultural influences, and support systems [2]. As healthcare systems shift toward value-based models of care, the ability to capture and measure these determinants gains paramount significance. Patient narratives, whether in the form of clinical notes, physician reports, or patient self-reports, offer rich textual evidence regarding these influences, yet they remain challenging to analyze due to the heterogeneity and complexity of human language [3]. This complexity is especially pronounced in medical contexts where abbreviations, domain-specific terms, and contextual nuances abound.

A central question arises: How can computational models effectively distill critical social information from a voluminous corpus of clinical text without succumbing to superficial lexical patterns? The proposition advanced in this paper is that sophisticated embedding mechanisms—capable of systematically integrating evolving contextual cues—hold great promise [4]. Traditional word embeddings, which rely on static vectors, often fail to capture dynamic relationships spanning wide contextual windows. More advanced models, such as those based on deep attention networks, add some degree of context-awareness, but these may still falter when confronted with specialized healthcare vocabularies that are rife with abbreviations and idiomatic expressions [5]. Consequently, the need for adaptive embeddings emerges, allowing for a constant re-alignment of feature spaces to the shifts in meaning that accompany clinical contexts [6].

Detecting SDOH factors such as economic hardship, employment instability, or lack of social support from free-form text demands an approach that recognizes patterns beyond the purely lexical. Indeed, subtle cues might hinge on interpretative logic, where particular terms or phrases co-occur in ways that require a richer perspective than mere word co-occurrence [7]. For instance, references to chronic stress, economic insecurity, or precarious living conditions may be scattered throughout the patient record, manifesting as implied circumstances rather than explicitly stated conditions. In such scenarios, adaptive embeddings might offer a solution by mapping textual evidence into higher-dimensional manifolds that capture both local and global relationships [8, 9].

In parallel with the methodological challenges, the domain of clinical text imposes stringent requirements concerning model interpretability and rigor. Healthcare providers and other stakeholders must have confidence in the outputs of automated systems, which demands a degree of mathematical transparency and consistency [10]. It becomes necessary to weave in advanced theoretical constructs that can ground the processes of embedding, alignment, and inference in well-defined formal frameworks. By employing algebraic and symbolic tools, it is possible to ensure that transformations in the embedding space are systematically justifiable, particularly when dealing with sensitive patient information that cannot afford misclassification or loss of context [11].

Such concerns highlight the importance of bridging data-driven machine learning with symbolically grounded perspectives [12]. The objective of this work extends beyond a purely empirical demonstration of improved metrics; it aspires to introduce an avenue for explaining results through the lens of structured logical statements. Symbolic reasoning has historically played a pivotal role in areas of artificial intelligence that require high-level inference, but it has often been sidelined in the era of massive neural networks [13]. Here, however, the argument is that large-scale adaptive embeddings can benefit from logical constraints or propositions that refine the underlying manifold. More concretely, a well-designed system should encode statements such as $\forall x \in \mathcal{D}, \exists y \in \mathcal{H} : \text{ContextualAlignment}(x, y)$, expressing that for each textual fragment in the domain $\mathcal{D}$, there exists an appropriate embedding vector in $\mathcal{H}$ that aligns contextual meaning with relevant SDOH factors.

The application of linear algebra also serves as a backbone, particularly in the iterative refinement of representations [14]. For instance, one might define an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ that selectively enhances dimensions of an embedding vector associated with crucial social determinants, while dampening irrelevant aspects. If V is the embedding space of dimension n, we can conceptualize a subspace $U \subseteq V$ devoted to social factors such as housing, nutrition, or psychosocial support, which is iteratively shaped by the transformation T. The rank of T can be strategically constrained so that essential patterns are retained within U, enabling the system to maintain a robust focus on relevant SDOH indicators [15]. Throughout this paper, we integrate such formulations to stress the interpretability and structure of the modeling process.

In recognition of the multifaceted nature of patient narratives, we adopt a perspective that sees language as an evolving entity [16]. In other words, the significance of a particular token or phrase can shift when placed alongside other contextual elements that appear earlier or later in the same text [17]. This perspective motivates the development of embeddings that adapt on the fly, harnessing real-time feedback from context and domain constraints. To validate these ideas, we present empirical results derived from large-scale clinical corpora, illustrating the model's capacity to extract SDOH-related phenomena in diverse textual settings [18]. Beyond standard metrics, we also address how an algebraically sound and logically consistent approach can offer tangible interpretative insights for clinicians and researchers.

The following sections delve deeper into the theoretical underpinnings, the architecture of the adaptive embeddings, the implementation specifics, and a detailed evaluation using controlled experiments [19]. In addition to performance metrics, we emphasize how logic-based alignment procedures and structured algebraic transformations can provide clarity and refinement. Our objective is to illustrate that the detection of social determinants is not merely a matter of pattern recognition, but a scientifically grounded endeavor that merges data-driven learning with formal reasoning in ways that yield robust, transparent, and replicable insights. [20]

2. Theoretical Underpinnings and Structured Representations

The trajectory of modern embedding-based natural language processing often stems from foundational theories that conceptualize linguistic meaning as a point in a high-dimensional vector space [21]. From a linear algebra perspective, let X represent the corpus of all tokens in our system, and let $f : X \longrightarrow \mathbb{R}^n$ be an embedding function that maps each token $x \in X$ to an n-dimensional vector. In classical settings, f might be fixed or slightly fine-tuned, but in this work we propose a dynamic operator $F : X \times C \longrightarrow \mathbb{R}^n$ that not only considers the token x but also the contextual environment $c \in C$. For patient narratives, c might be a collection of preceding and succeeding tokens, domain-specific constraints, or latent variables capturing the semantic intent of the text. The interplay between x and c forms the essence of adaptive embeddings, allowing for the transformation of meaning based on local or global context [22].

Logic statements can augment such a representation by imposing constraints on how tokens and contexts interact. For instance, consider a formal rule expressing that certain terms related to employment status cannot coexist with contradictory contextual markers if the discourse is consistent [23]. In a symbolic form, one might encode a proposition: $\forall x \in X, \forall c \in C, \mathrm{Inconsistency}(x, c) \implies \neg \mathrm{ValidAlignment}(F(x, c))$. This indicates that if a token-context pair is flagged as inconsistent by domain logic, the resulting embedding alignment must be nullified or penalized, thus steering the model away from spurious correlations. Symbolic constraints can be integrated into the training objective via regularization terms that promote logically consistent embeddings [24]. Such a structured approach helps ensure that the transformations remain interpretable and preserve clinically valid relationships.

To demonstrate the value of structured representations, one can define a set of relevant attributes for SDOH detection, such as Housing, Employment, SupportNetwork, and HealthcareAccess [25]. Each attribute can be represented as a basis vector in a subspace of $\mathbb{R}^n$, leading to explicit interpretability when evaluating the alignment of a token-context pair (x, c) with these attributes. One potential formulation is to designate a matrix $S \in \mathbb{R}^{n \times k}$, where k is the number of these attributes. The matrix S maps an embedding vector from $\mathbb{R}^n$ to a k-dimensional space that corresponds to the intensities or activations of these attributes. For example, the jth component of $S \cdot F(x, c)$ might indicate the degree to which the token-context pair aligns with attribute j [26].

A robust measure of alignment can be defined using inner products or norms. Consider the expression $\|S \cdot F(x, c)\|_2$ as an indicator of overall alignment with the entire set of SDOH attributes [27]. Alternatively, one can probe the distribution of activation across the k attributes to gauge specific associations. This approach differs fundamentally from unstructured embeddings that only provide a single vector, since it opens the door to transparent mappings between textual evidence and discrete social determinant factors [28]. By specifying a partial order $\preceq$ on the attributes (for example, to represent hierarchical or dependency relationships among them), we can further refine the representation to encode that certain attributes must precede or subsume others.

In the broader landscape of neural architectures, these theoretical and structured perspectives respond to a long-standing challenge: advanced language models often operate as black boxes, capturing contextual cues in ways that defy easy explanation [29]. By integrating formal logic and linear algebraic concepts, we aim to bestow a sense of structure and interpretability without necessarily forfeiting the raw predictive power that massive neural networks can provide. The synergy between symbolic constraints and learned representations becomes especially relevant in a field like healthcare, where justifications for model decisions may be clinically significant [30].

Another dimension of theoretical development arises when considering how to scale these ideas to large corpora [31]. As the number of tokens $|X|$ grows, the complexity of contexts $|C|$ can become exponential, necessitating a more efficient means of representation. Sparse updates, approximate nearest neighbor searches in embedding space, or dimension reduction techniques might be employed to maintain tractability [32]. One might consider approximate embeddings $F'(x, c)$ that skip certain details in contexts deemed non-essential, guided by domain-specific logic rules. For instance, if $\exists r \in R : r \subseteq c \wedge \mathrm{Irrelevant}(r)$ indicates a subset of context r is irrelevant, one could omit it from the embedding calculation to reduce computational load while preserving fidelity to critical SDOH cues.

The intricacies of structured representations underscore the delicate balance between complexity and interpretability [33]. On one hand, deeper and more flexible models hold the promise of capturing nuanced patterns in the data; on the other, explicit constraints and algebraic frameworks can keep the system grounded in domain realism and clinical utility. In the subsequent sections, we examine how these ideas materialize in a full-fledged adaptive embedding architecture tailored to detecting SDOH in patient narratives. [34]

3. Adaptive Embedding Architecture

Building on the theoretical principles outlined previously, this section discusses the specific architecture used to learn adaptive contextual embeddings for the targeted task of SDOH detection in clinical text [35]. The core component is a neural encoder that integrates attention-based mechanisms, transforming token-context pairs into refined embeddings that are subsequently shaped by domain-specific subspaces. Let w_i denote the ith token in a given patient narrative, and let c_i encapsulate the relevant context for w_i [36]. Instead of generating a single embedding for w_i, the architecture outputs $u_i = F(w_i, c_i)$, where F is a learnable function realized through layers of transformations.

One of the critical transformations is the self-attention mechanism, denoted in symbolic form as [37]

$$\alpha_{ij} = \frac{\exp(\mathbf{q}_i^\top \mathbf{k}_j)}{\sum_\ell \exp(\mathbf{q}_i^\top \mathbf{k}_\ell)},$$

where $\mathbf{q}_i = Q\,u_i$ and $\mathbf{k}_j = K\,u_j$ for learned matrices Q and K. The adaptive flavor is introduced by conditioning these query and key projections on domain signals: $Q = Q_0 + Q_{\text{domain}}$ and $K = K_0 + K_{\text{domain}}$, where Q_0, K_0 capture general linguistic patterns, and $Q_{\text{domain}}, K_{\text{domain}}$ incorporate SDOH-related knowledge. The domain-specific components can be initialized from a set of medical ontologies or learned from corpora that are exclusively curated to highlight social factors. By doing so, the self-attention mechanism is no longer purely language-driven but also shaped by knowledge relevant to social contexts [38].

The result of the attention operation, combined with a value transformation V, yields a composite embedding:

$$u_i' = \sum_j \alpha_{ij}(V u_j).$$

Here, $V = V_0 + V_{\text{domain}}$ similarly blends general and domain-specific learned parameters. This updated embedding u_i' is then passed to subsequent layers that integrate logic-based alignment constraints [39]. For instance, if L denotes a logic constraint operator, one might impose a penalty on u_i' that violates domain rules such as $\neg\exists u_i' : \text{Inconsistent}(u_i', \text{HousingIndicator})$. In a practical sense, this penalty is often implemented through an additional term in the loss function that increases whenever the model produces embeddings inconsistent with known domain structures [40].

Additionally, the architecture benefits from iterative refinement. Let $u_i^{(t)}$ be the embedding of token w_i at iteration t. The next iteration's embedding is given by $u_i^{(t+1)} = R\big(u_i^{(t)}, \{u_j^{(t)}\}_{j\neq i}, \Theta\big)$, where R is a recurrent refinement operator and Θ collects the parameters of attention and logic alignment. The process continues for T steps or until convergence, effectively enabling the embedding space to settle into a stable configuration that respects both contextual patterns and domain constraints [41]. Empirically, this iterative process can yield embeddings that are better aligned with subtle SDOH signals, as each pass recalibrates the representation based on the global distribution of domain-relevant cues in the text.

From a geometric point of view, one can interpret each iteration as performing a projection onto the manifold of permissible embeddings in $\mathbb{R}^n$. If M denotes the manifold shaped by domain constraints, the operator R approximates a projection Π_M [42]. One might say $u_i^{(t+1)} \approx \Pi_M(u_i^{(t)})$, with each step ensuring closer adherence to the structural rules. In practice, the manifold can be difficult to characterize explicitly, but domain knowledge and logical constraints act as guiding forces. Whether such constraints are soft or hard influences the shape of the manifold and the feasibility of finding a global optimum [43]. A typical approach is to incorporate the constraints as a differentiable penalty, creating a smooth relaxation of the manifold that can be navigated with standard gradient-based optimizers.

The adaptive embedding architecture also necessitates careful initialization [44]. Often, pretrained language models such as those derived from large transformer architectures are used as a starting point [45]. Their parameter weights, derived from generic corpora, provide a baseline embedding function. Domain-specific refinement then proceeds by injecting SDOH-focused data, logic constraints, and supervised signals from labeled narratives [46]. Over time, the parameters deviate from the generic initialization, converging on a specialized model that is more attuned to subtle social factors. Mathematically, this can be seen as shifting the embedding function F along a gradient direction influenced by the domain data [47]. If $\mathcal{L}_{\text{domain}}$ denotes the domain-specific loss, the update rule might be

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\text{domain}}(\theta),$$

where θ represents the trainable parameters across Q, K, V, R, and η is the learning rate. The introduction of logical constraints modifies $\mathcal{L}_{\text{domain}}$ to incorporate penalty terms, effectively shaping the gradient flow to respect domain knowledge.

Crucially, the architecture must handle real-world complexities in clinical text, such as abbreviations, misspellings, and ambiguous phrasing [48]. By maintaining an adaptive approach, the model can shift embeddings to account for unusual or evolving terms [49]. For instance, if w_i is a rarely used abbreviation for a critical social factor, the attention and logic-based constraints can highlight its significance, pushing u_i into a region of the embedding space associated with known synonyms and translations. This approach fosters robustness and prevents the system from ignoring potential SDOH cues simply because they differ from standard medical terminology [50].

Through the lens of advanced mathematics, one can appreciate the architecture as a multi-step mapping from raw text to structured semantic representations, accompanied by constraints that act as a series of transformations in $\mathbb{R}^n$. The synergy of these steps, from attention-based computations to iterative logical alignment, provides a holistic framework. In the following section, we outline the practical implementation details and experiments conducted to assess the performance and interpretability of this approach for SDOH extraction. [51]

4. Implementation and Experimental Analysis

To validate the proposed architecture, we developed an end-to-end pipeline for extracting SDOH information from authentic patient narratives. The datasets used include de-identified clinical notes, physician reports, and patient self-reports, spanning a range of healthcare institutions with varied demographic profiles [52]. The corpus was annotated by domain experts, ensuring reliable gold-standard labels for factors such as housing insecurity, employment issues, and limited access to care. These annotations furnished the supervised signals necessary for domain-specific fine-tuning [53].

In practice, each token w_i and its surrounding context c_i were first encoded through a pretrained language model, yielding an initial representation $v_i = E(w_i, c_i)$ of dimension d [54]. The adaptive embedding function F extended this initial representation to incorporate domain constraints, effectively transforming v_i into u_i. We employed iterative refinement with T iterations, during which the attention and logic-based modules continuously refined u_i based on feedback from the entire sequence [55]. This iterative mechanism required careful hyperparameter tuning for stability, including choices for T, the learning rate η, and the weighting of constraint penalties in the loss function.

The logic constraints used in this study spanned both high-level and granular rules [56]. A high-level rule might state that certain references to housing instability must be supported by mentions of either eviction risk or homelessness. Symbolically, this could be formalized as [57]

$$\forall i : \text{HousingInstability}(u_i) \implies \exists j : \text{EvictionOrHomelessness}(u_j),$$

suggesting that a token labeled as indicative of housing instability should be contextually aligned with at least one token referencing eviction or homelessness. Violation of this rule introduced a penalty term proportional to the mismatch between the model's predictions and the rule's requirement [58]. A more granular example pertains to contradictory contexts: if a patient narrative references stable employment in one sentence, it should not simultaneously map to an "unemployed" embedding subspace unless annotated as inconsistent or outdated information [59]. Handling such temporal contradictions necessitated specialized logic constraints:

$$\forall i, j : \text{TemporalContradiction}(u_i, u_j) \implies \neg\text{CoOccur}(u_i, u_j).$$

These constraints enforced consistency within each narrative, balancing the model's capacity to detect fine-grained cues [60].

Empirically, we evaluated the system on held-out sets of clinical narratives, measuring precision, recall, and F1-score in identifying SDOH factors. Additionally, we included interpretability metrics designed to assess the alignment between the model's internal representations and domain rules [61]. For instance, we quantified the frequency of rule violations and the alignment between certain tokens and their designated subspace vectors in S. The results consistently demonstrated that the adaptive embedding approach achieved higher recall on subtle SDOH indicators than baseline transformer models that lacked logic constraints [62].

One illustrative experiment involved diagnosing model performance in cases where textual cues were notably indirect [63]. For example, a narrative might mention that a patient was "worried about missing rent," a phrase that strongly implies housing insecurity but does not explicitly confirm homelessness. Traditional embeddings might overlook such indirect references, but the adaptive approach, equipped with domain logic, pushed the representation of this phrase closer to the housing instability subspace, thereby increasing the recall of relevant factors [64]. A deeper error analysis revealed that certain domain constraints were pivotal in disambiguating references that might otherwise yield false positives or false negatives.

We also explored the degree to which these enhancements carry over to new contexts or institutions [65]. Using data from a distinct hospital system, the model retained a significant portion of its performance advantage, although a slight decline was observed, attributable to domain shift in linguistic patterns. Notably, the iterative refinement mechanism appeared less sensitive to shifts, suggesting that repeated alignment with partial context can accommodate new usage patterns, provided the underlying logic constraints remain pertinent [66]. This result underscores the importance of systematically crafted constraints that can generalize to different writing styles or demographic factors.

In a parallel set of experiments, we probed the effect of ignoring logical constraints by setting the penalty term in the loss function to zero, effectively reverting to a standard attention-based transformer [67]. As anticipated, performance dropped on subtle SDOH cues, though for heavily signposted cues, the difference was minimal [68]. This dichotomy highlights the importance of domain constraints primarily in capturing nuanced references that do not occur frequently enough to be learned purely from data. Furthermore, the logic-driven approach yielded embeddings that were more stable across iterative refinements, converging faster to a consistent configuration [69].

Beyond raw performance, interpretability was a major area of interest for clinicians involved in the study. By examining activation patterns in the $S \in \mathbb{R}^{n \times k}$ matrix, one could easily trace why certain tokens were mapped to specific subspace components, linking them to domain knowledge. For example, words indicating tenuous employment were heavily mapped onto the Employment subspace, a phenomenon that was readily explainable by referencing the relevant logic constraints [70]. This transparency is vital when making decisions that may impact patient care, lending legitimacy to machine-generated inferences in the eyes of healthcare professionals.

Such empirical findings lend credence to the notion that domain-centered logic statements and advanced linear algebraic transformations, when woven into neural architectures, can materially improve the detection of subtle social factors [71]. These gains extend beyond performance metrics to include better interpretability and a robust defense against domain shifts. The remainder of this paper discusses how these insights might be generalized, the remaining limitations, and prospective directions for future research. [72]

5. Discussion of Extended Implications

The adaptive contextual embedding framework presented here transcends a mere incremental improvement over standard transformer-based approaches [73]. At its core, it underscores a paradigm shift that intertwines computational learning with formal structures such as logic rules and algebraic operators. This shift is especially salient in the healthcare domain, where interpretability, reliability, and domain compatibility are non-negotiable [74]. By adopting a vantage point that sees embedding spaces as malleable, shaped by both data-driven signals and structured constraints, the methodology opens avenues for deeper synergy between symbolic and sub-symbolic paradigms of artificial intelligence.

One possible extension involves coupling the current method with knowledge graph embeddings, wherein established relationships between social factors and health outcomes are stored in graph form [75]. A node representing "HousingInstability" could be connected to a node "HealthRisks" with an edge denoting the correlation or causal link, while additional nodes reflect living arrangements or financial stress. Embeddings derived from this knowledge graph can be treated as a prior, guiding the adaptation of textual embeddings so that they remain consistent with well-documented domain knowledge [76]. Symbolically, a statement such as $\forall x(\text{HousingInstability}(x) \rightarrow \text{ElevatedRisk}(x))$ might impose constraints on how these nodes align in embedding space.

Linear algebraic extensions can also be envisioned [77]. For instance, when decomposing the embedding space into domain-specific subspaces, one might employ a more nuanced decomposition approach, such as singular value decomposition or spectral decomposition, to isolate factors that consistently correlate with certain social determinants. If $M \in \mathbb{R}^{N \times n}$ is the matrix of embeddings for an entire corpus of N tokens, then computing a rank-k approximation M_k can highlight the dominant axes of variation that correspond to the k primary SDOH attributes. The residual $M - M_k$ can be interpreted as noise or less relevant variation, thereby sharpening the focus on the subspace that drives SDOH detection [78]. This approach can be especially valuable when dealing with large, heterogeneous datasets.

Further theoretical work could refine the logical constraints to accommodate probabilistic truths [79]. In many patient narratives, it is not always clear whether a particular reference unambiguously signals a social factor. A proposition such as $\text{HousingInstability}(w_i)$ might carry a probability p that depends on contextual cues. In such circumstances, one might unify symbolic logic with probabilistic inference, leading to a framework akin to Markov Logic Networks [80]. There, each logical rule would be associated with a weight, and the system would strive to maximize the likelihood of the data under the constraints imposed by those weighted rules. Mathematically, this introduces an energy function that merges the negative log-likelihood of the neural network outputs with penalty terms for constraint violations, each scaled by a learned or predefined weight [81]. The iterative refinement procedure could then be reinterpreted as a gradient-based search on this energy function in the space of embeddings [82, 83].

Another far-reaching implication relates to how these adaptive embeddings can inform downstream tasks beyond detection. In a clinical setting, once social factors are identified, they might be used to predict patient outcomes, resource allocation, or intervention effectiveness [84]. If the same embeddings that detect SDOH also serve as input to a prognostic model, the constraints that guide the embeddings can yield improved outcome predictions. Suppose one has a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ that predicts a clinical risk score from the embedding. If g operates on logically constrained embeddings, the interpretability and consistency of g's outputs may also improve [85]. This synergy is reflected by the statement $\forall x \in \mathcal{X}, \exists y \in \mathcal{Y} : (\text{DomainConstraints}(x) \wedge \text{PredictiveConsistency}(g(y)))$. In simpler terms, it means that if the embedding is domain-compliant, the predictions are more likely to be accurate and reliable.

Despite these prospective benefits, challenges remain [86]. The articulation of logic rules demands expertise from clinicians, social workers, and domain specialists, and the process of converting intuitive knowledge about social determinants into formal logical statements can be time-consuming. Furthermore, over-constraining the system might lead to conflicts among rules or might prevent the model from discovering novel correlations [87]. The solution is often to employ a carefully balanced approach, possibly incorporating preference orderings among rules, or adopting a multi-tier constraint system where some rules are enforced strictly and others are enforced as soft constraints [88].

Another area of ongoing research pertains to data bias. Because SDOH reflect societal inequalities, the narratives themselves may contain biased language or incomplete portrayals of certain populations [89]. If the model inadvertently amplifies these biases, it might lead to skewed representations that disadvantage marginalized communities. Paradoxically, the same logic constraints that enhance interpretability could also perpetuate or mask biases if the constraints are implicitly biased [90]. Tools from fairness in machine learning might help here, imposing additional constraints that ensure parity or limit disparate impact. Algebraically, one might define constraints on the norm or distribution of embeddings across demographic groups, ensuring that no group is systematically misrepresented in the SDOH subspace [91].

Finally, scaling the system to extremely large collections of clinical narratives requires attention to computational efficiency [92]. Iterative refinement with multiple logic constraints can be expensive. Optimizations such as parallelization, approximate search in embedding space, and gradient caching become critical [93]. Nevertheless, the conceptual framework of adaptive embeddings remains intact even under large-scale scenarios, provided suitable engineering optimizations are in place.

Overall, the successful detection of subtle social factors in patient narratives has critical real-world ramifications [94]. Identifying patients at risk due to social determinants can inform targeted interventions, guide public health policy, and uncover societal trends that might otherwise remain hidden in unstructured text. By marrying advanced NLP techniques with formal logic and algebraic insights, this methodology provides a step toward robust, explainable, and domain-aligned computational solutions in healthcare [95]. The concluding section will distill key takeaways and prospective pathways for applying these insights in broader contexts.

6. Conclusion

This paper presented a rigorous and comprehensive exploration of adaptive contextual embeddings for the detection of Social Determinants of Health within clinical narratives [96]. Throughout the discussion, it became evident that traditional static embeddings or even partially context-aware models encounter limitations when grappling with the multifaceted nature of patient text, which often encodes subtle references to socioeconomic and psychosocial conditions [97]. By integrating domain-specific logic constraints, linear algebraic formulations, and iterative refinement, the proposed framework achieves a more nuanced representation of language, one that remains faithful to the realities of healthcare.

The architecture described herein leverages attention-based transformations that are systematically adapted to social determinants, thereby avoiding the pitfalls of one-size-fits-all token encodings [98]. Through logical propositions, the model learns to align its internal representations with clinically validated statements, reinforcing the interpretability of its predictions. These symbolic constraints were integrated as differentiable penalty terms, guiding the model's convergence toward embeddings that simultaneously respect domain knowledge and capture empirical language patterns [99]. Empirical evaluations demonstrated improvements in the detection of subtle SDOH cues, including housing instability, employment challenges, and insufficient healthcare access, reinforcing the value of a structured approach.

Moreover, the paper highlighted the broader implications for patient care, public health research, and policy-making [100]. Automated systems capable of reliably extracting social determinants can streamline clinical workflows, direct resource allocation, and illuminate systemic disparities that underlie health outcomes. The interpretability facilitated by structured embeddings fosters trust among clinicians who require transparent explanations for automated decisions [101]. In addition, the discussed algebraic constructs, such as subspace projections, spectral decompositions, and manifold approximations, offer avenues to maintain or enhance computational efficiency and refine domain alignment [102].

Despite these encouraging findings, the outlined framework is but a step toward fully realizing the potential of context-adaptive NLP in healthcare. Future research may delve deeper into probabilistic logic statements, multi-tier constraint systems, and fairness constraints that protect against biases embedded in clinical text [103]. Knowledge graph embeddings, domain adaptation across disparate institutions, and prospective integration with predictive modeling pipelines also stand out as promising next directions. Scaling the approach to massive, real-time healthcare databases remains an engineering challenge requiring careful orchestration of high-performance computing, distributed processing, and memory-efficient data structures [104].

Ultimately, this work substantiates the thesis that bridging data-driven neural models and formal symbolic mechanisms can yield more robust, interpretable, and clinically aligned solutions for SDOH detection. By grounding language representations in mathematically coherent and logically consistent frameworks, it becomes feasible to capture the richness of patient narratives without forfeiting rigor or adaptability [105]. The insights gleaned hold relevance not only for extracting social determinants but also for a broad range of information extraction tasks in domains where contextual nuance and interpretability are paramount. The continued refinement and expansion of these adaptive embedding techniques promise to amplify the role of computational tools in understanding and addressing the social dimensions that shape individual and population health. [106]

References

[1] F. Li, W. Liu, and H. Yu, "Extraction of information related to adverse drug events from electronic health record notes: Design of an end-to-end model based on deep learning.," *JMIR medical informatics*, vol. 6, pp. e12159–, 11 2018.

[2] S. J. Fodeh, M. A. Al-Garadi, O. Elsankary, J. Perrone, W. C. Becker, and A. Sarker, "Utilizing a multi-class classification approach to detect therapeutic and recreational misuse of opioids on twitter.," *Computers in biology and medicine*, vol. 129, pp. 104132–, 11 2020.

[3] J. H. Moore and J. H. Holmes, "The golden era of biomedical informatics has begun.," *BioData mining*, vol. 9, pp. 15–15, 4 2016.

[4] K. Ferar, T. O. Hall, D. C. Crawford, R. Rowley, B. A. Satterfield, R. Li, L. Gragert, E. W. Karlson, M. de Andrade, I. J. Kullo, C. A. McCarty, A. Kho, M. G. Hayes, M. D. Ritchie, P. K. Crane, D. B. Mirel, C. Carlson, J. J. Connolly, H. Hakonarson, A. T. Crenshaw, D. Carrell, Y. Luo, O. Dikilitas, J. C. Denny, G. P. Jarvik, and D. R. Crosslin, "Genetic variation in the human leukocyte antigen region confers susceptibility to clostridioides difficile infection.," *Scientific reports*, vol. 13, pp. 18532–, 10 2023.

[5] D. Singer, B. Chastek, A. Sargent, J. C. Johnson, S. Shetty, C. Conoscenti, and E. J. Bernstein, "Impact of chronic fibrosing interstitial lung disease on healthcare use: association between fvc decline and inpatient hospitalization.," *BMC pulmonary medicine*, vol. 23, pp. 337–, 9 2023.

[6] D. Brandon, "Data mining twitter for covid-19 sentiments concerning college online education," *Future Business Journal*, vol. 9, 12 2023.

[7] R. Avula, "Addressing barriers in data collection, transmission, and security to optimize data availability in healthcare systems for improved clinical decision-making and analytics," *Applied Research in Artificial Intelligence and Cloud Computing*, vol. 4, no. 1, pp. 78–93, 2021.

[8] F. R. Tewes, "Artificial intelligence in the american healthcare industry: Looking forward to 2030," *Journal of Medical Research and Surgery*, vol. 3, pp. 107–108, 10 2022.

[9] J. R. Machireddy, "Harnessing ai and data analytics for smarter healthcare solutions," *International Journal of Science and Research Archive*, vol. 08, no. 02, pp. 785–798, 2023.

[10] S. Hassanpour and C. P. Langlotz, "Unsupervised topic modeling in a large free text radiology report repository.," *Journal of digital imaging*, vol. 29, pp. 59–62, 9 2015.

[11] 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.

[12] A. M. Griffith, P. Ockerse, A. M. Shaaban, and C. Kelly, "Effect of the covid-19 pandemic on ct scans ordered from the emergency department for abdominal complaints.," *Emergency radiology*, vol. 28, pp. 485–495, 1 2021.

[13] M. S. YILDIZ and A. ALPER, "Potential functions of artificial intelligence chatbot chatgpt in health management: Scoping review," *Turkish Bulletin of Hygiene and Experimental Biology*, vol. 80, no. 4, pp. 523–530, 2023.

[14] M. Moh, T.-S. Moh, Y. Peng, and L. Wu, "On adverse drug event extractions using twitter sentiment analysis," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 6, pp. 18–, 9 2017.

[15] Y. Wu, J. C. Denny, S. T. Rosenbloom, R. A. Miller, D. A. Giuse, M. Song, and H. Xu, "A preliminary study of clinical abbreviation disambiguation in real time," *Applied clinical informatics*, vol. 6, pp. 364–374, 6 2015.

[16] Y. Ahuja, J. Wen, C. Hong, Z. Xia, S. Huang, and T. Cai, "A semi-supervised adaptive markov gaussian embedding process (samgep) for prediction of phenotype event times using the electronic health record.," *Scientific reports*, vol. 12, pp. 17737–, 10 2022.

[17] M. Dahiya, E. A. Eboreime, A. Hyde, S. Rahman, M. Sebastianski, M. Carbonneau, E. B. Tapper, and P. Tandon, "International classification of diseases codes are useful in identifying cirrhosis in administrative databases," *Digestive diseases and sciences*, vol. 67, pp. 1–16, 6 2021.

[18] C. Deng, A. Reddy, B. K. Kumar, M. Babu, and B. A. Babu, "Forecasting length of stay: Will it be clear or cloudy today?," *Intelligence-Based Medicine*, vol. 6, pp. 100078–100078, 2022.

[19] A. Sharma and K. D. Forbus, "Graph-based reasoning and reinforcement learning for improving q/a performance in large knowledge-based systems," in *2010 AAAI Fall Symposium Series*, 2010.

[20] J. W. Bear, T. Long, D. E. Skinner, and J. H. McKerrow, "Predictions of novel schistosoma mansoni - human protein interactions consistent with experimental data," *Scientific reports*, vol. 8, pp. 13092–13092, 8 2018.

[21] D. A. Hanauer, Q. Mei, V. G. V. Vydiswaran, K. Singh, Z. Landis-Lewis, and C. Weng, "Complexities, variations, and errors of numbering within clinical notes: the potential impact on information extraction and cohort-identification," *BMC medical informatics and decision making*, vol. 19, pp. 59–69, 4 2019.

[22] S. Hassanpour, G. H. Bay, and C. P. Langlotz, "Characterization of change and significance for clinical findings in radiology reports through natural language processing.," *Journal of digital imaging*, vol. 30, pp. 314–322, 1 2017.

[23] X. Luo, P. Gandhi, Z. Zhang, W. Shao, Z. Han, V. Chandrasekaran, V. Turzhitsky, V. Bali, A. Roberts, M. Metzger, J. Baker, C. L. Rosa, J. Weaver, P. R. Dexter, and K. Huang, "Applying interpretable deep learning models to identify chronic cough patients using ehr data.," *Computer methods and programs in biomedicine*, vol. 210, pp. 106395–, 9 2021.

[24] K. E. Affolter, K. M. Gligorich, N. J. Samadder, W. S. Samowitz, and K. Curtin, "Feasibility of large-scale identification of sessile serrated polyp patients using electronic records: A utah study," *Digestive diseases and sciences*, vol. 62, pp. 1455–1463, 3 2017.

[25] C. K. Patel, E. Carreras, R. H. Henderson, S. C. Wong, and S. Berg, "Evolving outcomes of surgery for retinal detachment in retinopathy of prematurity: the need for a national service in the united kingdom : An audit of surgery for acute tractional retinal detachment complicating rop in the uk.," *Eye (London, England)*, vol. 36, pp. 1–7, 7 2021.

[26] Y. Wang, X. Fan, L. Chen, E. Chang, S. Ananiadou, J. Tsujii, and Y. Xu, "Mapping anatomical related entities to human body parts based on wikipedia in discharge summaries.," *BMC bioinformatics*, vol. 20, pp. 430–430, 8 2019.

[27] L. V. Rasmussen, "The electronic health record for translational research," *Journal of cardiovascular translational research*, vol. 7, pp. 607–614, 7 2014.

[28] E. Mahmoudi, W. Wu, C. Najarian, J. Aikens, J. Bynum, and V. G. V. Vydiswaran, "Identifying caregiver availability using medical notes with rule-based natural language processing: Retrospective cohort study.," *JMIR aging*, vol. 5, pp. e40241–e40241, 9 2022.

[29] C. Caliskan, "How does "a bit of everything american" state feel about covid-19? a quantitative twitter analysis of the pandemic in ohio.," *Journal of computational social science*, vol. 5, pp. 1–27, 4 2021.

[30] A. Tariq, O. Kallas, P. Balthazar, S. J. Lee, T. Desser, D. Rubin, J. W. Gichoya, and I. Banerjee, "Transfer language space with similar domain adaptation: a case study with hepatocellular carcinoma.," *Journal of biomedical semantics*, vol. 13, pp. 8–, 2 2022.

[31] Y. Xiang, J. Xu, Y. Si, Z. Li, L. Rasmy, Y. Zhou, F. Tiryaki, F. Li, Y. Zhang, Y. Wu, X. Jiang, W. J. Zheng, D. Zhi, C. Tao, and H. Xu, "Time-sensitive clinical concept embeddings learned from large electronic health records.," *BMC medical informatics and decision making*, vol. 19, pp. 58–58, 4 2019.

[32] A. Can, V. M. Castro, Y. H. Ozdemir, S. Dagen, D. Dligach, S. Finan, S. Yu, V. S. Gainer, N. A. Shadick, G. Savova, S. N. Murphy, T. Cai, S. T. Weiss, and R. Du, "Alcohol consumption and aneurysmal subarachnoid hemorrhage," *Translational stroke research*, vol. 9, pp. 13–19, 7 2017.

[33] I. Bordino, A. Ferretti, F. Gullo, and S. Pascolutti, "Garnlp: A natural language processing pipeline for garnishment documents," *Information Systems Frontiers*, vol. 23, pp. 101–114, 3 2020.

[34] J. Liu, M. T. Islam, S. Sang, L. Qiu, and L. Xing, "Biology-aware mutation-based deep learning for outcome prediction of cancer immunotherapy with immune checkpoint inhibitors.," *NPJ precision oncology*, vol. 7, pp. 117–, 11 2023.

[35] F. A. Batarseh and L. J. Freeman, "A survey on ai assurance," *Journal of Big Data*, vol. 8, pp. 1–30, 4 2021.

[36] B. I. Reiner, "Innovation opportunities in critical results communication: practical solutions.," *Journal of digital imaging*, vol. 26, pp. 830–837, 8 2013.

[37] G. M. Gray, A. Zirikly, L. M. Ahumada, M. Rouhizadeh, T. Richards, C. Kitchen, I. Foroughmand, and E. Hatef, "Application of natural language processing to identify social needs from patient medical notes: development and assessment of a scalable, performant, and rule-based model in an integrated healthcare delivery system.," *JAMIA open*, vol. 6, pp. ooad085–, 10 2023.

[38] T. H. McCoy, V. M. Castro, L. A. Snapper, K. L. Hart, and R. H. Perlis, "Efficient genome-wide association in biobanks using topic modeling identifies multiple novel disease loci.," *Molecular medicine (Cambridge, Mass.)*, vol. 23, pp. 285–294, 8 2017.

[39] Y. Jiang, P. C.-I. Pang, D. Wong, and H. Y. Kan, "Natural language processing adoption in governments and future research directions: A systematic review," *Applied Sciences*, vol. 13, pp. 12346–12346, 11 2023.

[40] N. Afzal, S. Sohn, S. Abram, C. G. Scott, R. Chaudhry, H. Liu, I. J. Kullo, and A. M. Arruda-Olson, "Mining peripheral arterial disease cases from narrative clinical notes using natural language processing," *Journal of vascular surgery*, vol. 65, pp. 1753–1761, 2 2017.

[41] N. Dewaswala, D. Chen, H. Bhopalwala, V. C. Kaggal, S. P. Murphy, J. M. Bos, J. B. Geske, B. J. Gersh, S. R. Ommen, P. A. Araoz, M. J. Ackerman, and A. M. Arruda-Olson, "Natural language processing for identification of hypertrophic cardiomyopathy patients from cardiac magnetic resonance reports.," *BMC medical informatics and decision making*, vol. 22, pp. 272–, 10 2022.

[42] X. Mei, Z. Liu, A. Singh, M. Lange, P. Boddu, J. Q. X. Gong, J. Lee, C. DeMarco, C. Cao, S. Platt, G. Sivakumar, B. Gross, M. Huang, J. Masseaux, S. Dua, A. Bernheim, M. Chung, T. Deyer, A. Jacobi, M. Padilla, Z. A. Fayad, and Y. Yang, "Interstitial lung disease diagnosis and prognosis using an ai system integrating longitudinal data.," *Nature communications*, vol. 14, pp. 2272–, 4 2023.

[43] A. Alzu'bi, H. Najadat, W. Doulat, O. M. Al-Shari, and L. Zhou, "Predicting the recurrence of breast cancer using machine learning algorithms," *Multimedia Tools and Applications*, vol. 80, pp. 13787–13800, 1 2021.

[44] C. A. Austin-Tse, V. Jobanputra, D. L. Perry, D. Bick, R. J. Taft, E. Venner, R. A. Gibbs, T. Young, S. Barnett, J. W. Belmont, N. Boczek, S. Chowdhury, K. A. Ellsworth, S. Guha, S. Kulkarni, C. Marcou, L. Meng, D. R. Murdock, A. U. Rehman, E. Spiteri, A. Thomas-Wilson, H. M. Kearney, H. L. Rehm, and null null, "Best practices for the interpretation and reporting of clinical whole genome sequencing.," *NPJ genomic medicine*, vol. 7, pp. 27–, 4 2022.

[45] W. Jankowski, H. D. Lagassé, W. C. Chang, J. R. McGill, K. I. Jankowska, A. D. Gelinas, N. Janjic, and Z. E. Sauna, "Modified aptamers as reagents to characterize recombinant human erythropoietin products.," *Scientific reports*, vol. 10, pp. 18593–18593, 10 2020.

[46] L. Jamian, L. Wheless, L. J. Crofford, and A. Barnado, "Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record," *Arthritis research & therapy*, vol. 21, pp. 305–305, 12 2019.

[47] E. Sirohi and P. L. Peissig, "Pacific symposium on biocomputing - study of effect of drug lexicons on medication extraction from electronic medical records.," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 308–318, 2004.

[48] J. M. Havrilla, A. Singaravelu, D. M. Driscoll, L. Minkovsky, I. Helbig, L. Medne, K. Wang, I. Krantz, and B. R. Desai, "Phenominal: an ehr-integrated web application for structured deep phenotyping at the point of care.," *BMC medical informatics and decision making*, vol. 22, pp. 198–, 7 2022.

[49] D. Riskin, R. Cady, A. Shroff, N. A. Hindiyeh, T. Smith, and S. Kymes, "Using artificial intelligence to identify patients with migraine and associated symptoms and conditions within electronic health records.," *BMC medical informatics and decision making*, vol. 23, pp. 121–, 7 2023.

[50] M. D. Murray, "Use of data from electronic health records for pharmacoepidemiology," *Current Epidemiology Reports*, vol. 1, pp. 186–193, 9 2014.

[51] F. Xie, C. Mercado, S. S. Kim, W. Yu, C. Zheng, D. Getahun, S. J. Jacobsen, and W. Chen, "Identifying spontaneous abortion from clinical notes within a large integrated healthcare system," *SN Computer Science*, vol. 3, 5 2022.

[52] D. Oniani, J. Hilsman, Y. Peng, R. K. Poropatich, J. C. Pamplin, G. L. Legault, and Y. Wang, "Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare.," *NPJ digital medicine*, vol. 6, pp. 225–, 12 2023.

[53] V. Herrington and O. T. Adedeji, "At the intersection of medical robotic surgery and drug discovery with quantum computing," *Journal of Electrical Electronics Engineering*, vol. 2, 8 2023.

[54] V. K. Yeruva, S. Junaid, and Y. Lee, "Contextual word embeddings and topic modeling in healthy dieting and obesity," *Journal of healthcare informatics research*, vol. 3, pp. 159–183, 6 2019.

[55] R. Avula, "Overcoming data silos in healthcare with strategies for enhancing integration and interoperability to improve clinical and operational efficiency," *Journal of Advanced Analytics in Healthcare Management*, vol. 4, no. 10, pp. 26–44, 2020.

[56] E. A. C. Hubal, J. J. Frank, R. M. Nachman, M. M. Angrish, N. C. Deziel, M. M. Fry, R. Tornero-Velez, A. D. Kraft, and E. Lavoie, "Advancing systematic-review methodology in exposure science for environmental health decision making.," *Journal of exposure science & environmental epidemiology*, vol. 30, pp. 906–916, 5 2020.

[57] 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.

[58] Z. Shen, D. Schutte, Y. Yi, A. Bompelli, F. Yu, Y. Wang, and R. Zhang, "Classifying the lifestyle status for alzheimer's disease from clinical notes using deep learning with weak supervision.," *BMC medical informatics and decision making*, vol. 22, pp. 88–, 7 2022.

[59] S. Zhou, H. Kang, B. Yao, and Y. Gong, "An automated pipeline for analyzing medication event reports in clinical settings," *BMC medical informatics and decision making*, vol. 18, pp. 1–11, 12 2018.

[60] T. Cai, F. Cai, K. Dahal, G. Cremone, E. Lam, C. Golnik, T. Seyok, C. Hong, T. Cai, and K. P. Liao, "Improving the efficiency of clinical trial recruitment using an ensemble machine learning to assist with eligibility screening.," *ACR open rheumatology*, vol. 3, pp. 593–600, 7 2021.

[61] S. Wu, V. C. Kaggal, D. Dligach, J. J. Masanz, P. Chen, L. Becker, W. W. Chapman, G. Savova, H. Liu, and C. G. Chute, "A common type system for clinical natural language processing," *Journal of biomedical semantics*, vol. 4, pp. 1–12, 1 2013.

[62] B. Slostad, A. Karnik, V. Appadurai, and A. Narang, "Applications of artificial intelligence in echocardiography," *Current Cardiovascular Risk Reports*, vol. 17, pp. 123–132, 5 2023.

[63] D. Danks and D. Trusilo, "The challenge of ethical interoperability," *Digital Society*, vol. 1, 8 2022.

[64] A. E. Kristoffersen, S. A. Quandt, and T. Stub, "Use of complementary and alternative medicine in norway: a cross-sectional survey with a modified norwegian version of the international questionnaire to measure use of complementary and alternative medicine (i-cam-qn).," *BMC complementary medicine and therapies*, vol. 21, pp. 93–93, 3 2021.

[65] M. Abouelyazid and C. Xiang, "Machine learning-assisted approach for fetal health status prediction using cardiotocogram data," *International Journal of Applied Health Care Analytics*, vol. 6, no. 4, pp. 1–22, 2021.

[66] A. P. Nunes, A. M. Loughlin, Q. Qiao, S. M. Ezzy, L. Yochum, C. R. Clifford, R. Gately, D. D. Dore, and J. D. Seeger, "Tolerability and effectiveness of exenatide once weekly relative to basal insulin among type 2 diabetes patients of different races in routine care," *Diabetes therapy : research, treatment and education of diabetes and related disorders*, vol. 8, pp. 1349–1364, 10 2017.

[67] B. C. Coleman, D. Finch, R. Wang, S. L. Luther, A. Heapy, C. Brandt, and A. J. Lisi, "Extracting pain care quality indicators from u.s. veterans health administration chiropractic care using natural language processing.," *Applied clinical informatics*, vol. 14, pp. 600–608, 5 2023.

[68] R. Avula, "Optimizing data quality in electronic medical records: Addressing fragmentation, inconsistencies, and data integrity issues in healthcare," *Journal of Big-Data Analytics and Cloud Computing*, vol. 4, no. 5, pp. 1–25, 2019.

[69] H. Kilicoglu, A. B. Abacha, Y. Mrabet, S. E. Shooshan, L. Rodriguez, K. Masterton, and D. Demner-Fushman, "Semantic annotation of consumer health questions," *BMC bioinformatics*, vol. 19, pp. 34–34, 2 2018.

[70] D. E. Malden, S. Y. Tartof, B. K. Ackerson, V. Hong, J. Skarbinski, V. Yau, L. Qian, H. Fischer, S. Shaw, S. Caparosa, and F. Xie, "Natural language processing for improved covid-19 characterization: Evidence from more than 350,000 patients in a large integrated health care system," *SSRN Electronic Journal*, 1 2022.

[71] 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.

[72] J. Kalyanam and T. K. Mackey, "A review of digital surveillance methods and approaches to combat prescription drug abuse," *Current Addiction Reports*, vol. 4, pp. 397–409, 9 2017.

[73] D. Berlowitz, B. J. Aguilar, X. Li, E. Shishova, G. Jasuja, P. J. Morin, D. R. Miller, M. K. O'Connor, A. H. Nguyen, R. Zhang, A. A. T. Monfared, Q. Zhang, and W. Xia, "Comparative identification of veterans with mild cognitive impairment or alzheimer's disease extracted from clinical notes and administrative data," *Alzheimer's & Dementia*, vol. 19, 12 2023.

[74] G. Jia, X. Zhong, H. K. Im, N. Schoettler, M. Pividori, D. K. Hogarth, A. I. Sperling, S. R. White, E. T. Naureckas, C. S. Lyttle, C. Terao, Y. Kamatani, M. Akiyama, K. Matsuda, M. Kubo, N. J. Cox, C. Ober, A. Rzhetsky, and J. Solway, "Discerning asthma endotypes through comorbidity mapping.," *Nature communications*, vol. 13, pp. 6712–, 11 2022.

[75] A. B. Shams, E. H. Apu, A. Rahman, M. S. Raihan, N. Siddika, R. B. Preo, M. R. Hussein, S. Mostari, and R. Kabir, "Web search engine misinformation notifier extension (seminext): A machine learning based approach during covid-19 pandemic.," *Healthcare (Basel, Switzerland)*, vol. 9, pp. 156–, 2 2021.

[76] J. T. Wu, F. Dernoncourt, S. Gehrmann, P. D. Tyler, E. T. Moseley, E. T. Carlson, D. W. Grant, Y. Li, J. Welt, and L. A. Celi, "Behind the scenes: A medical natural language processing project.," *International journal of medical informatics*, vol. 112, pp. 68–73, 12 2017.

[77] Y.-C. Lo, S. Varghese, S. Blackley, D. L. Seger, K. G. Blumenthal, F. R. Goss, and L. Zhou, "Reconciling allergy information in the electronic health record after a drug challenge using natural language processing.," *Frontiers in allergy*, vol. 3, pp. 904923–, 5 2022.

[78] H. A. Elmarakeby, P. S. Trukhanov, V. M. Arroyo, I. B. Riaz, D. Schrag, E. M. V. Allen, and K. L. Kehl, "Empirical evaluation of language modeling to ascertain cancer outcomes from clinical text reports.," *BMC bioinformatics*, vol. 24, pp. 328–, 9 2023.

[79] 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.

[80] Y. Gao, D. Dligach, T. Miller, M. M. Churpek, and M. Afshar, "Overview of the problem list summarization (probsum) 2023 shared task on summarizing patients' active diagnoses and problems from electronic health record progress notes.," *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2023, pp. 461–467, 2023.

[81] L. Shen, X. Shi, Z. Zhao, and K. Wang, "Informatics and machine learning methods for health applications.," *BMC medical informatics and decision making*, vol. 20, pp. 342–, 12 2020.

[82] E. V. Goldstein, E. V. Bailey, and F. A. Wilson, "Poverty and suicidal ideation among hispanic mental health care patients leading up to the covid-19 pandemic.," *Hispanic health care international : the official journal of the National Association of Hispanic Nurses*, vol. 22, pp. 6–10, 6 2023.

[83] 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.

[84] 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.

[85] R. R. Bond, M. D. Mulvenna, C. Potts, S. O'Neill, E. Ennis, and J. Torous, "Digital transformation of mental health services.," *Npj mental health research*, vol. 2, pp. 13–, 8 2023.

[86] M. N. O. Sadiku, Y. Zhou, and S. M. Musa, "Natural language processing in healthcare," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, pp. 39–42, 6 2018.

[87] V. P. Gurupur, M. Shelleh, C. Leone, D. Schupp-Omid, R. Azevedo, and S. Dubey, "Thnn - a neural network model for telehealth data incompleteness prediction.," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2023, pp. 1–5, 7 2023.

[88] B. Meskó and E. J. Topol, "The imperative for regulatory oversight of large language models (or generative ai) in healthcare.," *NPJ digital medicine*, vol. 6, pp. 120–, 7 2023.

[89] J. Cunningham, P. Singh, C. Reeder, B. Claggett, P.-M. M. Castellote, E. Lau, S. Khurshid, P. Batra, S. Lubitz, M. Maddah, A. A. Philippakis, P. Finn, A. S. Desai, P. T. Ellinor, O. Vardeny, S. Solomon, and J. E. Ho, "Abstract 13880: Natural language processing for adjudication of heart failure hospitalizations in a multi-center clinical trial," *Circulation*, vol. 148, 11 2023.

[90] H. Kan, K. Dyagilev, P. Schulam, S. Saria, H. Kharrazi, D. P. Bodycombe, C. T. Molta, and J. R. Curtis, "Factors associated with physicians' prescriptions for rheumatoid arthritis drugs not filled by patients," *Arthritis research & therapy*, vol. 20, pp. 79–79, 5 2018.

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

[92] A. Ramanathan, L. L. Pullum, T. C. Hobson, C. A. Steed, S. Quinn, C. Chennubhotla, and S. Valkova, "Orbit: Oak ridge biosurveillance toolkit for public health dynamics," *BMC bioinformatics*, vol. 16, pp. 1–12, 12 2015.

[93] F. Soboczenski, T. A. Trikalinos, J. Kuiper, R. G. Bias, B. C. Wallace, and I. J. Marshall, "Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study.," *BMC medical informatics and decision making*, vol. 19, pp. 96–96, 5 2019.

[94] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Data virtualization for analytics and business intelligence in big data," in *CS & IT Conference Proceedings*, vol. 9, CS & IT Conference Proceedings, 2019.

[95] 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.

[96] A. Sharma and K. Forbus, "Modeling the evolution of knowledge in learning systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, pp. 669–675, 2012.

[97] A. Jagannatha and H. Yu, "Acl - calibrating structured output predictors for natural language processing.," *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2020, pp. 2078–2092, 2020.

[98] T. Hartman, M. D. Howell, J. Dean, S. Hoory, R. Slyper, I. Laish, O. Gilon, D. Vainstein, G. S. Corrado, K. Chou, M. J. Po, J. Williams, S. Ellis, G. E. Bee, A. Hassidim, R. Amira, G. Beryozkin, I. Szpektor, and Y. Matias, "Customization scenarios for de-identification of clinical notes," *BMC medical informatics and decision making*, vol. 20, pp. 14–14, 1 2020.

[99] E. Helderop, J. Huff, F. Morstatter, A. Grubesic, and D. Wallace, "Hidden in plain sight: A machine learning approach for detecting prostitution activity in phoenix, arizona," *Applied Spatial Analysis and Policy*, vol. 12, pp. 941–963, 11 2018.

[100] T. Champagne-Langabeer, M. W. Swank, S. Manas, Y. Si, and K. Roberts, "Dramatic increases in telehealth-related tweets during the early covid-19 pandemic: A sentiment analysis," *Healthcare (Basel, Switzerland)*, vol. 9, pp. 634–, 5 2021.

[101] B. S. Greenberg and L. S. Lasdon, "Using global optimization to estimate population class sizes," *Journal of Global Optimization*, vol. 36, pp. 319–338, 6 2006.

[102] S. Razdan, A. R. Siegal, Y. Brewer, M. Sljivich, and R. J. Valenzuela, "Assessing chatgpt's ability to answer questions pertaining to erectile dysfunction: can our patients trust it?," *International journal of impotence research*, vol. 36, pp. 734–740, 11 2023.

[103] S. C. Davoud and V. P. Kovacheva, "On the horizon: Specific applications of automation and artificial intelligence in anesthesiology.," *Current anesthesiology reports*, vol. 13, pp. 31–40, 4 2023.

[104] E. Alsentzer, M. J. Rasmussen, R. Fontoura, A. L. Cull, B. Beaulieu-Jones, K. J. Gray, D. W. Bates, and V. P. Kovacheva, "Zero-shot interpretable phenotyping of postpartum hemorrhage using large language models.," *NPJ digital medicine*, vol. 6, pp. 212–, 11 2023.

[105] R. C. Cury, I. Megyeri, R. Macedo, J. Batlle, S. Kim, B. Baker, R. Harris, and R. H. Clark, "Natural language processing and machine learning for detection of respiratory illness by chest ct imaging and tracking of covid-19 pandemic in the us," *SSRN Electronic Journal*, 1 2020.

[106] R. F. Dougherty, P. Clarke, M. Atli, J. Kuc, D. Schlosser, B. W. Dunlop, D. J. Hellerstein, S. T. Aaronson, S. Zisook, A. H. Young, R. Carhart-Harris, G. M. Goodwin, and G. A. Ryslik, "Psilocybin therapy for treatment resistant depression: prediction of clinical outcome by natural language processing.," *Psychopharmacology*, 8 2023.