soloncouncil : Pages:27–40

Original Research

Natural Language Inference Models for Automated Insurance Claim Adjudication

Laura Fernanda Malag´on Navarro[1]

[1]Law graduate, specialist, and researcher in social media and content marketing, San Buenaventura University,Bogota, Colombia.

Abstract
Automated insurance claim adjudication has emerged as a critical challenge in contemporary financial and technological ecosystems, demanding robust methods for assessing the veracity and legitimacy of claims without human intervention. Natural Language Inference (NLI) models, initially deployed for tasks such as textual entailment and question answering, present a compelling opportunity for solving this challenge by systematically interpreting textual information and inferring logical relationships. By capitalizing on the unique capacity of NLI to determine whether a hypothesis is entailed, contradicted, or neutral with respect to a premise, this approach can significantly reduce manual review processes, enhance consistency, and boost overall efficiency. In this work, we examine the theoretical underpinnings and practical implementation of advanced NLI models tailored specifically for insurance claim adjudication. We investigate how model architecture, data encoding, and representation learning can be optimized to address the intricacies of claim documents that often contain specialized, context-dependent terminology and nuanced logical dependencies. To provide a robust foundation, we develop advanced mathematical formulations and inject formal logical reasoning methods to ensure the reliability of automated adjudication decisions. Our experimental findings underscore the feasibility of using NLI-based architectures to automate claim reviews with high accuracy, while also highlighting ongoing challenges. Our ultimate objective is to encourage broader adoption of inference-oriented solutions in the evolving domain of insurance technology.

1. Introduction

The insurance industry, long dependent on intricate procedures for assessing claims, has witnessed significant technological evolution, and one of the most promising developments in this trajectory is the advent of automated solutions [1]. The complexity of insurance claim documents, often spanning medical, automotive, or property-related texts, poses a substantial challenge to machine learning algorithms seeking to emulate or replace human adjudicators [2]. Traditional rule-based systems have provided a baseline for detecting fraudulent statements or identifying potential misinterpretations, yet these systems lack the adaptability and granularity needed in modern claim analysis. In this regard, Natural Language Inference (NLI) models offer a novel approach centered on extracting logical relationships and meaning from text [3]. Instead of relying on a limited set of rules or keywords, such models attempt to reason about the content and draw valid conclusions regarding the compatibility of policy statements with claim statements, as well as the plausibility or legitimacy of information presented by claimants.

The journey toward harnessing NLI in insurance follows a lineage of textual entailment research, where machine learning models endeavor to decide whether one piece of text logically follows from another [4]. In a typical NLI setting, one identifies whether a premise semantically entails a hypothesis, whether it contradicts it, or whether the relationship is neutral. The domain-specific adaptation for insurance demands additional layers of complexity since claim forms often contain unstructured or semi-structured data, specialized jargon, and entangled references to policy clauses [5]. Therefore, effectively applying NLI in this domain requires an amalgamation of robust lexical, syntactic, and contextual understanding, as well as the ability to interpret the semantics of policy documents that may be lengthy and verbose. [6]

The significance of this endeavor lies in its potential to reduce the burden on human auditors and claim handlers, who are typically tasked with reviewing a large volume of claims daily. An effective automated system could curtail processing times, minimize errors or omissions, and ultimately lower operational costs [7]. Yet the hurdles in this space abound: from the need to scale across multiple lines of insurance (health, auto, life, property) to ensuring compliance with various regulations, implementing NLI-based adjudication solutions demands rigorous methodology and thorough validation. Consequently, this paper undertakes an in-depth exploration of how NLI models can be leveraged for automated adjudication, how they can be trained to handle complex textual patterns, and what theoretical frameworks can be applied to enhance reliability and interpretability. [8]

Logic, formal methods, and advanced representations of language form an integral part of this undertaking. The alignment of textual entailment tasks with real-world claim scenarios can be framed through various mathematical and logical constructs [9]. A simplified example might consider a statement p describing a policy clause and a statement q articulating the claim [10]. We then pose the inference question: does p entail q, or do the two statements contradict each other? One may encode these textual statements as vectors in a high-dimensional semantic space, applying transformations such as $q' = T(q)$ under a linear or nonlinear function T, subsequently comparing the representation with p via an inner product or distance metric to assess the level of entailment. Alternatively, a set of logical operators could be used to map textual clauses into formal expressions such that evaluating $p \models q$ becomes feasible with minimal interpretive ambiguity. [11]

The broader goal is not simply to replicate the success of NLI in other contexts but to extend it, bridging the gap between academic research and industrial application. As such, we present a unified perspective that integrates advanced learning algorithms, formal logic, and domain-specific intricacies, offering a cohesive framework for automating claim adjudication [12]. We also delve into how specialized pretraining strategies and curated datasets can bolster system performance in practice. This study aims not only to validate the capabilities of NLI-driven methods but also to illuminate the path ahead for developing comprehensive solutions that are theoretically grounded and practically viable [13, 14]. In doing so, we situate our work at the intersection of language understanding and domain-specific knowledge representation, ultimately advocating for a new frontier in insurance analytics that strongly emphasizes inference-driven processes.

In the ensuing sections, we describe the foundational concepts underpinning NLI, propose architectures suitable for handling the specialized language of insurance, and articulate the mathematical and logical structures that inform our approach [15]. We then proceed to outline experimental protocols, discuss performance metrics that can capture the nuances of automated claim adjudication, and highlight the ethical and practical considerations crucial for large-scale deployment [16]. By the end of this discussion, we aim to establish not only the feasibility of NLI-driven models in this realm but also to illustrate how these solutions can be rigorously tailored to meet real-world demands and regulatory requirements.

2. Foundational Concepts in NLI for Insurance Claims

The field of Natural Language Inference traces its scholarly roots to philosophical explorations of entailment and logic, subsequently evolving into a domain of computational linguistics focused on textual entailment, contradiction, and neutrality [17]. From a practical vantage point, the core principle involves modeling the relationship between two statements and deciding how closely they align in meaning. Within the insurance context, these statements may pertain to a claim's textual description, relevant policy clauses, or supplementary medical records [18]. The task, therefore, transforms into a specialized textual inference problem where each piece of textual evidence is assessed for its logical relation to the policy coverage context.

To understand how these models function, consider the premise text P as the policy clause and the hypothesis text H as the claim statement [19]. We aim to learn a function ϕ such that $\phi(P, H)$ yields a label in {Entailment, Contradiction, Neutral}. In many standard tasks, the relationship is categorical [20]. However, in practical claim adjudication, one might adopt a continuous measure of compatibility, often implemented via a scalar score $s \in \mathbb{R}$, denoting the degree of alignment. For instance, let us formalize an example where p states, "The insurance policy covers damages due to fire," and q states, "The property was flooded." A high degree of contradiction or incompatibility is inferred if the model identifies a mismatch between "fire" and "flood," revealing that the claim does not logically follow from the policy clause.

The linguistic complexity inherent in insurance documents is addressed by embedding textual statements into vector spaces that capture nuanced semantic features [21]. One popular approach is to adopt a neural architecture that encodes the premises and hypotheses separately, often through advanced language models such as those based on Transformers. Specifically, one might use a function $E(\cdot)$ that transforms each token in the text into an embedding vector, and a self-attention mechanism to generate a combined representation for the entire text [22]. Subsequent layers might incorporate gating functions, residual connections, and specialized attention patterns that highlight relevant passages within the policy text. The final classification or regression layer takes these representations as input and estimates the degree of textual entailment [23]. Numerically, this can be done by computing $z_P = E(P)$ and $z_H = E(H)$, followed by a similarity function, such as $\sigma(z_P, z_H) = z_P^T W z_H$, for some learnable matrix W. The output is then passed through a nonlinear layer to produce a final inference score. [24]

One of the pivotal considerations for insurance claim adjudication is the presence of domain-specific terminology and cross-references to legal or regulatory statutes [25]. Modeling such references may require specialized tokenization strategies and domain-specific embeddings. For instance, the phrase "acute myocardial infarction" might need to be recognized as a single medical concept rather than as three independent words [26]. One approach employs symbol sets $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$, each corresponding to a distinct conceptual domain (e.g., medical, automotive, legal), to create domain-aware embeddings. A logic-based extension might define a set of predicates, such as $\text{Cover}(x, y)$, indicating that the policy x covers the event y, so that textual references to coverage can be mapped to formal statements. In such a scenario, the problem of deciding textual entailment transforms into verifying whether logical statements extracted from the claim's text unify with or contradict the statements extracted from the policy's text. [27]

Moreover, the dynamic interplay between multiple pieces of textual evidence introduces the question of how to combine or aggregate statements. Traditional NLI tasks usually involve a single premise and a single hypothesis [28]. In contrast, insurance claims often come with multiple supporting documents, such as police reports, accident photographs, or hospital discharge summaries [29]. The challenge thus becomes: given premises $\{P_1, P_2, \ldots, P_m\}$ and a single claim statement H, does the collective evidence entail H? Researchers have experimented with hierarchical architectures that first analyze each premise-claim pair, producing intermediate inference scores $\{s_1, s_2, \ldots, s_m\}$. An aggregation function, for example $A(s_1, s_2, \ldots, s_m)$, is then used to produce a final decision. One might define a logic-based aggregator: if any premise strictly contradicts H, the overall decision is contradiction, else if all premises strongly support H, the overall decision is entailment, otherwise we arrive at neutrality or partial coverage [30]. Mathematically, one could formalize this as:

$$D = \begin{cases} \text{Contradiction} & \text{if } \exists i : s_i < \theta_1 \\ \text{Entailment} & \text{if } s_i > \theta_2 \ \forall i \\ \text{Neutral} & \text{otherwise} \end{cases}$$

where θ_1 and θ_2 are domain-specific thresholds determined through empirical validation [31]. This multi-premise approach broadens the scope of NLI research and paves the way for more intricate claim-processing applications.

Another integral aspect is the introduction of interpretability [32]. In high-stakes financial settings, decisions must be explainable to comply with regulations and maintain trust. Hence, neural or statistical

NLI frameworks often include attention or saliency maps that highlight the textual fragments most influential to the inference decision [33]. For instance, if the model concludes that the claim is not covered due to a specific clause in the policy, it is beneficial to display a textual rationale derived from the attention mechanism or from a structured logical derivation [34]. This bridging of purely numeric inference with textual interpretability elevates the user's ability to validate the reasoning steps and ensures accountability in automated adjudication systems.

In summary, the application of NLI to insurance claims is grounded in a solid theoretical and computational framework that must be adapted to domain-specific considerations [35]. This includes specialized embeddings, logic-based formulations, multi-premise aggregation, and interpretability strategies, all of which collectively enable robust inference in the face of varied and complex textual inputs. These foundational concepts lay the groundwork for the more advanced approaches described in the following sections, where we systematically develop methodologies tailored for end-to-end claim processing. [36]

3. Proposed Methodologies for Automated Claim Adjudication

Building upon the foundational concepts, our proposed methodologies emphasize a pipeline that integrates domain-specific language modeling, robust inference mechanisms, and logical validation checks. The design stems from the recognition that an insurance document frequently contains intricate structures, specialized phrases, and references to regulatory norms or medical terminologies that do not always appear in general language corpora [37]. The underlying premise is to capture these complexities, encode them meaningfully, and systematically derive an inference score that represents the consistency between claim statements and policy clauses. [38]

The initial phase of this pipeline involves domain adaptation of language models. Modern large-scale language models are typically trained on wide-ranging corpora that include general web text, news, or books [39]. While these models retain a broad linguistic understanding, they often require fine-tuning or additional training on specialized insurance corpora to accurately process domain-specific vocabulary. Let $\mathcal{D}_{\text{insurance}}$ be a curated corpus comprising claim documents, policies, and relevant regulations. We perform continued pretraining of a baseline Transformer model on $\mathcal{D}_{\text{insurance}}$ using a masked language modeling objective:

$$\max_{\theta} \sum_{x \in \mathcal{D}_{\text{insurance}}} \log P_{\theta}(x_t \mid x_{\setminus t}),$$

where x_t is a masked token and $x_{\setminus t}$ represents the other tokens in the sequence. This stage allows the model to capture the syntactic and semantic regularities of the insurance domain, thus reducing downstream errors in inference tasks. [40]

Once the domain-specific language model is established, the next step involves training an NLI head on top of these representations. Let the premise be denoted P, the claim statement H, and the output be a scalar or vector representing the entailment categories [41]. We feed the token-level representations of P and H into a cross-attention or concatenated attention layer that jointly encodes the semantic interactions between the two sequences. For each pair (P, H), we obtain an output vector $\mathbf{v}_{(P,H)}$. A logistic or softmax function then maps $\mathbf{v}_{(P,H)}$ to the desired classification or continuous score, typically using the function:

$$\hat{y} = \sigma(W\mathbf{v}_{(P,H)} + b),$$

where W and b are learnable parameters [42]. The training objective can be cross-entropy for classification or mean-squared error for regression, depending on whether we adopt a discrete or continuous notion of entailment. [43]

Although neural networks excel at approximating complex relationships, they sometimes falter in scenarios requiring explicit reasoning. To circumvent this, we incorporate a logic-based constraint module that serves as an additional layer of verification [44]. After the neural model produces an initial inference outcome, we translate key components of the premise and claim into symbolic logic. For example, if the premise states "Policy X covers accidents on business premises," we map this to a formal statement Cover(X, Accident, BusinessPremises). If the claim states "The insured reported an accident in a commercial warehouse," we translate this to Event(Accident, Warehouse, Commercial). We then examine whether such statements can be unified under a set of axioms that define coverage conditions [45]. The unified set might hold an axiom stating Cover$(X, Accident, BusinessPremises)$ implies Cover$(X, Accident, Commercial)$ if Commercial $\subset$ BusinessPremises. In linear algebraic parlance, we can think of each logical component as a basis vector in a conceptual subspace. Checking unification corresponds to verifying that a linear combination of these basis vectors remains consistent with the subspace spanned by the policy statements [46]. If the logic-based reasoning conflicts with the neural inference, a conflict resolution mechanism triggers either a fallback to human review or additional explanation [47]. This approach ensures that purely statistical correlations do not override explicit coverage rules established by the policy language.

In parallel, we introduce a multi-document aggregator [48]. Often, insurance claims rely on supplementary pieces of textual evidence that may confirm or dispute the claim statement. Instead of concatenating these documents into one large text, we process each (P_i, H) pair separately, thereby generating a set of intermediate results $\{r_1, r_2, \ldots, r_m\}$. We then define an aggregator function $A(r_1, \ldots, r_m)$ that accounts for potential synergy or conflict among the premises [49]. A straightforward implementation might sum the confidence scores, while a more sophisticated approach might weigh them based on the reliability of each premise, as estimated through calibration data. Symbolically, if each premise leads to an entailment probability α_i, then the final score could be: [50]

$$\alpha_{\text{final}} = \frac{\sum_{i=1}^{m} w_i \alpha_i}{\sum_{i=1}^{m} w_i},$$

where w_i is a weight reflecting the trustworthiness of premise P_i, possibly derived from the length of the claim history or the perceived credibility of its source.

Evaluation of this pipeline requires carefully curated datasets that simulate real-world claim scenarios [51]. We can define ground-truth labels for each claim-premise pair by consulting domain experts [52]. The final system's performance is gauged through an array of metrics, including accuracy for discrete entailment decisions, F1 scores for multi-class classification, and aggregated measures that capture the proportion of correctly adjudicated claims. Additionally, we examine the interpretability by assessing how effectively the system can highlight the textual portions and logical rules that underlie each decision [53]. To facilitate such evaluations, advanced annotation schemes are designed, requiring human annotators to label textual segments signifying coverage clauses and linking them to the relevant claim statements.

In essence, this proposed methodology combines the strengths of neural representation learning, domain adaptation, logic-based validation, and multi-document reasoning to construct a holistic pipeline for automated insurance claim adjudication [54]. The synergy between statistical and symbolic techniques serves to mitigate common pitfalls in purely data-driven or purely rule-based systems. By systematically integrating each component, we enable a higher degree of reliability, interpretability, and adaptability to the diverse set of challenges encountered in real-world claim processing. [55]

4. Experimental Framework and Mathematical Underpinnings

To rigorously assess the capabilities and limitations of our proposed system, we design an experimental framework grounded in principles of reproducibility, formal verification, and statistical generalization [56]. Experiments revolve around curated datasets reflecting realistic insurance claims, policy documents, and associated external reports, ensuring that models are tested against authentic complexity rather than contrived examples. In tandem, we develop a set of mathematical tools and metrics aimed at scrutinizing not only predictive performance but also logical consistency and interpretability. [57]

We begin by assembling a corpus of claim documents sampled from multiple insurance lines, such as health, auto, and property, each containing textual data in structured (claim forms) and unstructured (medical reports, email correspondences) formats. This corpus, denoted as C, is carefully partitioned into training C_{train}, validation C_{val}, and test sets C_{test}. For each claim record $d \in C$, domain experts annotate the relevant policy clauses and label whether the claim is covered or not. Additionally, intermediate entailment labels are provided for claim-coverage pairs [58]. This annotation process yields a fine-grained view of the logical relationships within each claim scenario.

Next, the domain-adapted language model and the NLI classifier described in the previous section are trained on C_{train}. The training objective, typically cross-entropy, is augmented with a logic consistency regularizer Ω [59]. Let $\mathcal{L}_{\text{data}}$ be the empirical loss function computed over the training set, for example,

$$\mathcal{L}_{\text{data}} = -\sum_{(P,H,y)\in C_{\text{train}}} \log P_\theta(y \mid P, H),$$

where y denotes the entailment label [60]. We define Ω as the expected penalty for logical inconsistency:

$$\Omega(\theta) = \mathbb{E}_{(P,H)\sim C_{\text{train}}}\left[\text{LC}(\phi_\theta(P,H), \mathcal{R}(P,H))\right],$$

where $\phi_\theta(P,H)$ is the neural model output, and $\mathcal{R}(P,H)$ is the symbolic representation of the premise and claim. The function $\text{LC}(\cdot)$ computes a score quantifying the mismatch between the neural inference and the result of the logic unification step. Minimizing $\Omega(\theta)$ encourages the network to align with rules extracted from the textual data [61]. The overall loss is then:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{data}}(\theta) + \lambda\Omega(\theta),$$

where λ is a hyperparameter balancing data fidelity and logical consistency. [62]

The performance evaluation proceeds in two stages. In the first stage, we measure standard metrics such as accuracy, precision, recall, and F1 score on C_{test}. We also track confusion matrices that show how often claims that should be covered are incorrectly classified as uncovered, and vice versa [63]. These metrics shed light on the system's predictive correctness but do not fully capture its reasoning capabilities. In the second stage, we conduct a logic-consistency evaluation by examining how often the model's predictions contradict formal statements in $\mathcal{R}(P,H)$. Define the logical consistency rate as: [64]

$$\text{LCR} = 1 - \frac{\sum_{(P,H)\in C_{\text{test}}} \mathbf{1}[\phi_\theta(P,H) \nLeftrightarrow \mathcal{R}(P,H)]}{|C_{\text{test}}|},$$

where $\mathbf{1}[\phi_\theta(P,H) \nLeftrightarrow \mathcal{R}(P,H)]$ indicates a mismatch between the model output and the logical unification result, and $\nLeftrightarrow$ denotes a logical inequivalence. Values closer to 1 indicate high alignment with domain-specific logic. [65]

To probe interpretability, we conduct a post-hoc analysis involving localized explanations. For each test instance, we extract attention-weight matrices or gradient-based saliency maps that highlight the tokens most crucial for the model's inference [66]. By comparing these highlights against the manually annotated segments in C_{test}, we obtain a measure of alignment between automated and human-identified rationales. A high overlap suggests that the model's internal inference pathways align with human interpretable evidence, thus reinforcing trust in automated decisions.

The mathematics of vector space transformations also plays a role in analyzing how well the system separates the manifold of covered claims from that of uncovered claims [67]. One can map the final embedding $\mathbf{v}_{(P,H)}$ into a 2D or 3D space via dimensionality reduction techniques and examine cluster separations. If the embeddings corresponding to distinct coverage decisions form well-separated clusters, it implies that the model has learned consistent semantic boundaries. Alternatively, if we find significant overlap, that signals potential confusion in boundary cases [68]. Furthermore, we can explore the eigenvalues of the covariance matrix of $\mathbf{v}_{(P,H)}$ for each category of claims, investigating whether certain principal components align with domain-relevant features such as coverage type or claim type.

Finally, we address generalization to new policy documents or claim scenarios by evaluating domain-shift resilience [69]. This test involves introducing C_{novel}, a set of claims derived from lines of insurance or textual styles not seen during training. We measure how abruptly the performance metrics degrade, focusing particularly on the logical consistency dimension. A graceful degradation suggests that the system's underlying representations are robust and can extrapolate to new textual domains with minimal retraining. [70, 71]

Taken together, this experimental framework and mathematical analysis form a multi-faceted lens for assessing the system's real-world viability. By balancing empirical performance, logical consistency, and interpretability, we aim to demonstrate both the feasibility of NLI-based models for automated insurance claim adjudication and the rigorous underpinnings that guide system design and evaluation. [72]

5. Discussion of Results and Ongoing Challenges

The outcomes of the experiments conducted thus far underscore the promise and complexity of applying NLI models to the domain of automated insurance claim adjudication. In scenarios where policies and claims exhibit straightforward textual connections, the system achieves high accuracy and a high logical consistency rate, illustrating that domain-adapted language models can successfully learn and enforce coverage rules [73]. Despite these positive results, several ongoing challenges merit deeper investigation and refinement to ensure that the technology matures to a level suitable for large-scale deployment.

First, consider the intricacies introduced by domain-specific terminology [74]. In experiments focusing on health insurance claims, the model often encountered medical terms with ambiguous or context-dependent significance [75]. While domain adaptation through additional pretraining alleviated some issues, the system occasionally failed to capture the subtlety that certain terms might imply partial coverage or coverage contingent upon other conditions. For instance, a phrase like "elective procedure" must be interpreted within the legal definitions outlined in policy documents, which might specify conditions under which an "elective" intervention becomes "medically necessary." Incorporating advanced ontological resources and deeper symbolic frameworks can provide more robust semantic grounding for such specialized terms. [76]

Another considerable challenge is the variability of claim submissions, especially in lines like property or auto insurance. Claimants often describe events in narratives that are imprecise, subjective, or laden with extraneous detail [77]. Although the multi-premise aggregator helps to synthesize evidence from various documents, it remains susceptible to conflicting or incomplete information. In certain cases, contradictory evidence from different documents, such as a discrepancy between a policyholder's statement and an official incident report, confuses the aggregation mechanism [78]. A more refined approach might leverage weighting strategies that dynamically adjust premise reliability in real-time [79]. One could develop a partial order or ranking model where each piece of evidence has an associated credibility score, derived from the claim history or known veracity of sources.

Additionally, interpretability has emerged as a focal point, both from a legal compliance standpoint and for end-user trust [80]. While attention-based explanations and logic constraints offer some transparency, end-users and regulators often require explicit justifications. Achieving a high level of fidelity in these justifications, akin to a step-by-step derivation showing why a claim qualifies or disqualifies under a policy, remains an area of active research [81]. Future systems may integrate formal theorem provers with neural networks to produce structured proofs, ensuring that each inference step is transparently documented. The logic-based constraint module is an embryonic form of such an integrated system, highlighting the synergy between data-driven and rule-based reasoning. [82]

Scalability also poses a logistical challenge. Insurance carriers often process thousands of claims daily, each potentially accompanied by lengthy policy documents [83]. A naive approach that pairs each claim statement with every policy clause in an exhaustive manner quickly becomes computationally prohibitive [84]. Techniques such as caching vector embeddings for frequently cited clauses, or implementing approximate nearest-neighbor searches for semantically similar policy statements, can mitigate computational costs. However, ensuring that these optimizations do not degrade inference quality demands careful balancing [85]. Preliminary benchmarks suggest that advanced indexing schemes, combined with concurrency in large-scale cloud computing environments, can handle these volumes, but practical deployment requires robust engineering pipelines that integrate seamlessly with existing claim management systems.

An additional factor concerns the ethical and regulatory dimensions [86]. Automated systems must ensure fairness, particularly when dealing with health or life insurance claims, where decisions can profoundly impact individuals. Models inadvertently inheriting biases from training data could disproportionately harm certain demographics [87]. Although fairness metrics have been proposed in other machine learning contexts, applying them to the specialized language and structure of insurance claims necessitates further adaptation [88]. One could consider constraints of the form: for any pair of individuals i and j that meet identical coverage prerequisites under the policy text, the system's predicted outcome for their claims should remain consistent. Mathematically, this can be expressed as $\phi_\theta(P, H_i) = \phi_\theta(P, H_j)$ if $\mathcal{R}(H_i) \equiv \mathcal{R}(H_j)$, meaning the symbolic representations of the claims are equivalent. Operationalizing this principle at scale is a formidable but essential undertaking. [89]

Finally, we observe that real-world claims often contain intangible or context-dependent factors that resist purely textual analysis. Adjudicators may consider the claimant's prior history, local regulations, or even photographic evidence that is not readily translated into textual form [90]. Extending the scope of NLI to multimodal inputs, such as image and text, represents an exciting trajectory for future work. Some initial prototypes employ convolutional or Vision Transformer-based modules for images and combine their outputs with text embeddings for an integrated coverage decision [91]. However, bridging the semantic gap between textual statements and visual evidence remains non-trivial, particularly when trying to maintain the level of interpretability demanded in insurance processes.

In summary, while the results of our experiments validate the feasibility of NLI-based automated claim adjudication, the domain presents unique challenges related to specialized terminology, variable data quality, interpretability, scalability, fairness, and multimodal evidence [92]. Addressing these challenges requires a confluence of advanced language modeling, symbolic reasoning, logical formalisms, robust engineering solutions, and a deep appreciation for ethical and regulatory constraints [93]. We envision that the continuing evolution of hybrid AI frameworks, combining neural networks with logical inference engines, will progressively narrow the gap between prototype systems and the complex demands of real-world insurance adjudication.

6. Conclusion

The pursuit of automated insurance claim adjudication via Natural Language Inference models stands at the nexus of advanced language processing, domain-specific knowledge representation, and formal logic [94]. Our investigation into domain adaptation, multi-document inference, and symbolic reasoning reveals that state-of-the-art NLI architectures hold considerable promise in interpreting intricate textual data to produce coherent and justifiable coverage decisions. By unifying the representational power of neural networks with rule-based logic checks, we inch closer to a system that can rapidly and reliably handle large volumes of claims while retaining the capacity for transparency and interpretability demanded by regulators and end-users alike. [95]

The development process highlighted critical issues that must guide future research agendas. The specialized language of insurance introduces complexities that surpass typical NLI tasks, necessitating sophisticated embeddings, deeper symbolic frameworks, and high-fidelity logic modules [96]. Scaling these approaches to accommodate multiple lines of insurance, integrating evidence from non-textual sources, and mitigating biases pose formidable challenges [97]. Experimental results suggest that while high accuracy and logical consistency are within reach, further refinements in model architectures, algorithmic optimization, and fairness auditing are essential for consistent real-world performance. Additionally, interpretability remains a non-negotiable cornerstone, as decisions in insurance claim adjudication directly affect individuals and organizations, requiring clear rationales and compliance with legal standards. [98]

In reflecting on these findings, we argue for a stronger synergy between computational linguistics and formal logic, extended by robust engineering practices that address the domain's scalability and reliability requirements. Hybrid solutions that combine Transformer-based neural encoders with symbolic inference will likely continue to evolve, integrating external knowledge bases, ontologies, or theorem provers to more effectively handle the myriad contingencies embedded in policy documents [99]. We further anticipate that fairness and ethical considerations will rise in prominence, spurring new methods for ensuring equitable treatment across diverse populations of claimants.

Overall, this exploration underscores the transformative potential of inference-driven models in automating insurance claim workflows while carefully managing the complexities and responsibilities inherent in such high-impact applications [100]. We envision a not-too-distant future where insurers routinely rely on robust, transparent, and logically consistent NLI systems for adjudicating claims, ultimately streamlining processes, reducing costs, and enhancing the customer experience. The trajectory of ongoing and future developments suggests that, with the right blend of technical rigor and domain-awareness, this vision can be realized, ushering in a new era of accuracy and trustworthiness in automated insurance claim adjudication. [101]

References

[1] A. Vaid, I. Landi, G. Nadkarni, and I. Nabeel, "Using fine-tuned large language models to parse clinical notes in musculoskeletal pain disorders.," *The Lancet. Digital health*, vol. 5, pp. e855–e858, 10 2023.

[2] M. Z. Suchsland, L. Kowalski, H. A. Burkhardt, M. G. Prado, L. G. Kessler, M. Yetisgen, M. A. Au, K. A. Stephens, F. Farjah, A. M. Schleyer, F. M. Walter, R. D. Neal, K. Lybarger, C. A. Thompson, M. A. Achkar, E. A. Sarma, G. Turner, and M. Thompson, "How timely is diagnosis of lung cancer? cohort study of individuals with lung cancer presenting in ambulatory care in the united states.," *Cancers*, vol. 14, pp. 5756–5756, 11 2022.

[3] S. Raza, B. Schwartz, S. Lakamana, Y. Ge, and A. Sarker, "A framework for multi-faceted content analysis of social media chatter regarding non-medical use of prescription medications.," *BMC digital health*, vol. 1, 8 2023.

[4] V. Roche, J.-P. Robert, and H. Salam, "A holistic ai-based approach for pharmacovigilance optimization from patients behavior on social media.," *Artificial intelligence in medicine*, vol. 144, pp. 102638–102638, 8 2023.

[5] R. Avula, "Applications of bayesian statistics in healthcare for improving predictive modeling, decision-making, and adaptive personalized medicine," *International Journal of Applied Health Care Analytics*, vol. 7, no. 11, pp. 29–43, 2022.

[6] F. Shah-Mohammadi, W. Cui, and J. Finkelstein, "Comparison of acm and clamp for entity extraction in clinical notes.," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2021, pp. 1989–1992, 11 2021.

[7] B. Sun, R. Zhang, C.-L. Chi, and R. Straka, "318 optimizing the identification and prediction of statin intolerance to improve statin adherence using natural language processing and machine learning," *Journal of Clinical and Translational Science*, vol. 7, pp. 95–95, 4 2023.

[8] R. Winnenburg, A. Sorbello, and O. Bodenreider, "Exploring adverse drug events at the class level," *Journal of biomedical semantics*, vol. 6, pp. 18–18, 5 2015.

[9] A. B. Chapman, K. Cordasco, S. Chassman, T. Panadero, D. Agans, N. Jackson, K. Clair, R. Nelson, A. E. Montgomery, J. Tsai, E. Finley, and S. Gabrielian, "Assessing longitudinal housing status using electronic health record data: a comparison of natural language processing, structured data, and patient-reported history.," *Frontiers in artificial intelligence*, vol. 6, pp. 1187501–, 5 2023.

[10] 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.

[11] A. Sharma, K. M. Goolsbey, and D. Schneider, "Disambiguation for semi-supervised extraction of complex relations in large commonsense knowledge bases," in *7th Annual Conference on Advances in Cognitive Systems*, 2019.

[12] D. A. Lightfoot, A. Srour, J. Afzal, and N. Saini, "The multigeneic rhg1 locus: A model for the effects on root development, nematode resistance and recombination suppression.," *Nature Precedings*, pp. 1–1, 12 2008.

[13] Z. Liu and W. W. Chu, "Knowledge-based query expansion to support scenario-specific retrieval of medical free text," *Information Retrieval*, vol. 10, pp. 173–202, 1 2007.

[14] 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.

[15] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, pp. 13521–13617, 4 2023.

[16] Z. Li, H. Lu, D. Liu, A. N. C. Yu, and M. Gendron, "Emotional event perception is related to lexical complexity and emotion knowledge.," *Communications psychology*, vol. 1, pp. 45–, 12 2023.

[17] R. Sajdeya, M. T. Mardini, P. J. Tighe, R. L. Ison, C. Bai, S. Jugl, G. Hanzhi, K. Zandbiglari, F. I. Adiba, A. G. Winterstein, T. A. Pearson, R. L. Cook, and M. Rouhizadeh, "Developing and validating a natural language processing algorithm to extract preoperative cannabis use status documentation from unstructured narrative clinical notes.," *Journal of the American Medical Informatics Association : JAMIA*, vol. 30, pp. 1418–1428, 4 2023.

[18] J. S. Barrett, R. K. Goyal, J. Gobburu, S. Baran, and J. Varshney, "An ai approach to generating midd assets across the drug development continuum.," *The AAPS journal*, vol. 25, pp. 70–, 7 2023.

[19] P. B. Pruitt, A. M. Naidech, J. V. Ornam, P. Borczuk, and W. K. Thompson, "A natural language processing algorithm to extract characteristics of subdural hematoma from head ct reports.," *Emergency radiology*, vol. 26, pp. 301–306, 1 2019.

[20] H. Chen, E. Cohen, D. Wilson, and M. Alfted, "Improving patient safety event report classification with machine learning and contextual text representation.," *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting. Human Factors and Ergonomics Society. Annual meeting*, vol. 67, pp. 1063–1069, 10 2023.

[21] A. Abhishek and A. Basu, "A framework for disambiguation in ambiguous iconic environments," in *AI 2004: Advances in Artificial Intelligence: 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004. Proceedings 17*, pp. 1135–1140, Springer, 2005.

[22] S. Jung, C. C. Stahl, A. A. Rosser, A. S. Kraut, B. H. Schnapp, M. Westergaard, A. G. Hamedani, R. M. Minter, and J. A. Greenberg, "Multi-disciplinary assessment of the entrustable professional activities of surgery residents.," *Global surgical education : journal of the Association for Surgical Education*, vol. 1, pp. 28–, 7 2022.

[23] A. Pinar, "An integrated sentiment analysis and q-rung orthopair fuzzy mcdm model for supplier selection in e-commerce: a comprehensive approach," *Electronic Commerce Research*, 10 2023.

[24] F. Liu and D. Panagiotakos, "Real-world data: a brief review of the methods, applications, challenges and opportunities.," *BMC medical research methodology*, vol. 22, pp. 287–, 11 2022.

[25] P. A. Wortman and J. A. Chandy, "A framework for evaluating security risk in system design," *Discover Internet of Things*, vol. 2, 12 2022.

[26] S. Wang, L. Tang, A. Majety, J. F. Rousseau, G. Shih, Y. Ding, and Y. Peng, "Trustworthy assertion classification through prompting.," *Journal of biomedical informatics*, vol. 132, pp. 104139–104139, 7 2022.

[27] G. T. Gobbel, M. E. Matheny, R. R. Reeves, J. M. Akeroyd, A. Turchin, C. M. Ballantyne, L. A. Petersen, and S. S. Virani, "Leveraging structured and unstructured electronic health record data to detect reasons for suboptimal statin therapy use in patients with atherosclerotic cardiovascular disease.," *American journal of preventive cardiology*, vol. 9, pp. 100300–100300, 12 2021.

[28] J. van Stekelenborg, J. Ellenius, S. Maskell, T. Bergvall, O. Caster, N. Dasgupta, J. Dietrich, S. Gama, D. J. Lewis, V. Newbould, S. Brosch, C. E. Pierce, G. Powell, A. Ptaszynska-Neophytou, A. F. Z. Wiśniewski, P. Tregunno, G. N. Norén, and M. Pirmohamed, "Recommendations for the use of social media in pharmacovigilance: Lessons from imi web-radr," *Drug safety*, vol. 42, pp. 1393–1407, 8 2019.

[29] J. S. Chen and S. L. Baxter, "Applications of natural language processing in ophthalmology: present and future.," *Frontiers in medicine*, vol. 9, pp. 906554–, 8 2022.

[30] Q. Li, L. Deléger, T. Lingren, H. Zhai, M. Kaiser, L. Stoutenborough, A. G. Jegga, K. B. Cohen, and I. Solti, "Mining fda drug labels for medical conditions.," *BMC medical informatics and decision making*, vol. 13, pp. 53–53, 4 2013.

[31] C. Chiu, F. Villena, K. Martin, F. Núñez, C. Besa, and J. Dunstan, "Training and intrinsic evaluation of lightweight word embeddings for the clinical domain in spanish.," *Frontiers in artificial intelligence*, vol. 5, pp. 970517–, 9 2022.

[32] K. D. Forbus, C. Riesbeck, L. Birnbaum, K. Livingston, A. B. Sharma, and L. C. Ureel, "A prototype system that learns by reading simplified texts.," in *AAAI Spring Symposium: Machine Reading*, pp. 49–54, 2007.

[33] C. G. Arnold, B. Sonn, F. J. Meyers, A. Vest, R. Puls, E. Zirkler, M. Edelmann, I. M. Brooks, and A. A. Monte, "Accessing and utilizing clinical and genomic data from an electronic health record data warehouse.," *Translational medicine communications*, vol. 8, 3 2023.

[34] E. Hsu, I. Malagaris, Y.-F. Kuo, R. Sultana, and K. Roberts, "Deep learning-based nlp data pipeline for ehr-scanned document information extraction.," *JAMIA open*, vol. 5, pp. ooac045–, 4 2022.

[35] L. L. Guo, E. Steinberg, S. L. Fleming, J. Posada, J. Lemmon, S. R. Pfohl, N. Shah, J. Fries, and L. Sung, "Ehr foundation models improve robustness in the presence of temporal distribution shift.," *Scientific reports*, vol. 13, pp. 3767–, 3 2023.

[36] P. B. Ryan, M. J. Schuemie, E. Welebob, J. Duke, S. Valentine, and A. G. Hartzema, "Defining a reference set to support methodological research in drug safety," *Drug safety*, vol. 36, pp. 33–47, 10 2013.

[37] I. Banerjee, M. A. Davis, B. L. Vey, S. Mazaheri, F. Khan, V. Zavaletta, R. Gerard, J. W. Gichoya, and B. Patel, "Natural language processing model for identifying critical findings-a multi-institutional study.," *Journal of digital imaging*, vol. 36, pp. 105–113, 11 2022.

[38] K. B. Ramadi, R. Mehta, D. He, S. Chao, Z. Chu, R. Atun, and F. T. Nguyen, "Grass-roots entrepreneurship complements traditional top-down innovation in lung and breast cancer.," *NPJ digital medicine*, vol. 5, pp. 10–, 1 2022.

[39] M. Matero, S. Giorgi, B. Curtis, L. H. Ungar, and H. A. Schwartz, "Opioid death projections with ai-based forecasts using social media language.," *NPJ digital medicine*, vol. 6, pp. 35–, 3 2023.

[40] A. C. Pickett and D. Valdez, "Mining online discourse related to transgender exclusive policies in interscholastic sport: an exploratory natural language processing study," *Sexuality Research and Social Policy*, vol. 20, pp. 936–949, 10 2022.

[41] J. Reps, R. D. Williams, S. C. You, T. Falconer, E. P. Minty, A. Callahan, P. B. Ryan, R. W. Park, H.-S. Lim, and P. R. Rijnbeek, "Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation," *BMC medical research methodology*, vol. 20, pp. 1–10, 5 2020.

[42] S. Terzioğlu, K. N. Çoğalmış, and A. Bulut, "Ad creative generation using reinforced generative adversarial network," *Electronic Commerce Research*, vol. 24, pp. 1491–1507, 5 2022.

[43] J. A. Pacheco, L. V. Rasmussen, K. Wiley, T. N. Person, D. J. Cronkite, S. Sohn, S. Murphy, J. H. Gundelach, V. Gainer, V. M. Castro, C. Liu, F. Mentch, T. Lingren, A. S. Sundaresan, G. Eickelberg, V. Willis, A. Furmanchuk, R. Patel, D. S. Carrell, Y. Deng, N. Walton, B. A. Satterfield, I. J. Kullo, O. Dikilitas, J. C. Smith, J. F. Peterson, N. Shang, K. Kiryluk, Y. Ni, Y. Li, G. N. Nadkarni, E. A. Rosenthal, T. L. Walunas, M. S. Williams, E. W. Karlson, J. E. Linder, Y. Luo, C. Weng, and W. Wei, "Evaluation of the portability of computable phenotypes with natural language processing in the emerge network.," *Scientific reports*, vol. 13, pp. 1971–, 2 2023.

[44] S. Purohit, P. Calyam, M. L. Alarcon, N. R. Bhamidipati, A. S. M. Mosa, and K. Salah, "Honestchain: Consortium blockchain for protected data sharing in health information systems," *Peer-to-peer networking and applications*, vol. 14, pp. 1–17, 5 2021.

[45] H. Y. Liu, M. Alessandri-Bonetti, J. A. Arellano, and F. M. Egro, "Can chatgpt be the plastic surgeon's new digital assistant? a bibliometric analysis and scoping review of chatgpt in plastic surgery literature.," *Aesthetic plastic surgery*, vol. 48, pp. 1644–1652, 10 2023.

[46] N. Tavabi, M. Raza, M. Singh, S. Golchin, H. Singh, G. D. Hogue, and A. M. Kiapour, "Disparities in cannabis use and documentation in electronic health records among children and young adults.," *NPJ digital medicine*, vol. 6, pp. 138–, 8 2023.

[47] L. Chen, Z. Fan, J. Chang, R. Yang, H. Hou, H. Guo, Y. Zhang, T. Yang, C. Zhou, Q. Sui, Z. Chen, C. Zheng, X. Hao, K. Zhang, R. Cui, Z. Zhang, H. Ma, Y. Ding, N. Zhang, X. Lu, X. Luo, H. Jiang, S. Zhang, and M. Zheng, "Sequence-based drug design as a concept in computational drug design.," *Nature communications*, vol. 14, pp. 4217–, 7 2023.

[48] Y. Hua, L. Wang, V. Nguyen, M. Rieu-Werden, A. McDowell, D. W. Bates, D. Foer, and L. Zhou, "A deep learning approach for transgender and gender diverse patient identification in electronic health records.," *Journal of biomedical informatics*, vol. 147, pp. 104507–104507, 9 2023.

[49] S. Monteith, T. Glenn, J. Geddes, P. C. Whybrow, and M. Bauer, "Commercial use of emotion artificial intelligence (ai): Implications for psychiatry.," *Current psychiatry reports*, vol. 24, pp. 203–211, 2 2022.

[50] V. A. Shetty, S. Durbin, M. S. Weyrich, A. D. Martínez, J. Qian, and D. L. Chin, "A scoping review of empathy recognition in text using natural language processing.," *Journal of the American Medical Informatics Association : JAMIA*, vol. 31, pp. 762–775, 12 2023.

[51] H. Huang, O. Zheng, D. Wang, J. Yin, Z. Wang, S. Ding, H. Yin, C. Xu, R. Yang, Q. Zheng, and B. Shi, "Chatgpt for shaping the future of dentistry: the potential of multi-modal large language model.," *International journal of oral science*, vol. 15, pp. 29–, 7 2023.

[52] M. Srikumar, R. Finlay, G. Abuhamad, C. Ashurst, R. Campbell, E. Campbell-Ratcliffe, H. Hongo, S. R. Jordan, J. Lindley, A. Ovadya, and J. Pineau, "Advancing ethics review practices in ai research," *Nature Machine Intelligence*, vol. 4, pp. 1061–1064, 12 2022.

[53] Y. Wang, Y. Xu, Y. Sun, H. Wang, J. Qi, B. Wan, W. Ye, Y. Lin, Y. Shao, S. Dong, B. M. Tyler, and Y. Wang, "Leucine-rich repeat receptor-like gene screen reveals that nicotiana rxeg1 regulates glycoside hydrolase 12 mamp detection," *Nature communications*, vol. 9, pp. 594–594, 2 2018.

[54] S. Kang, L. Patil, A. Rangarajan, A. Moitra, T. Jia, D. Robinson, and D. Dutta, "Automated feedback generation for formal manufacturing rule extraction," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 33, pp. 289–301, 3 2019.

[55] E. Cano-Marin, D. Ribeiro-Soriano, A. Mardani, and C. B. Gonzalez-Tejero, "Exploring the challenges of the covid-19 vaccine supply chain using social media analytics: A global perspective," *Sustainable Technology and Entrepreneurship*, vol. 2, no. 3, pp. 100047–100047, 2023.

[56] M. A. Weinzierl, S. Hopfer, and S. M. Harabagiu, "Scaling up the discovery of hesitancy profiles by identifying the framing of beliefs towards vaccine confidence in twitter discourse.," *Journal of behavioral medicine*, vol. 46, pp. 253–275, 5 2022.

[57] R. Avula, "Assessing the impact of data quality on predictive analytics in healthcare: Strategies, tools, and techniques for ensuring accuracy, completeness, and timeliness in electronic health records," *Sage Science Review of Applied Machine Learning*, vol. 4, no. 2, pp. 31–47, 2021.

[58] D. F. Pethani, M. A. Chapman, D. M. Conway, D. X. Dai, M. D. Bishay, M. V. Choh, M. A. He, M. S.-E. Lim, M. H. Y. Ng, M. T. Mahony, A. P. A. Yaacoub, D. S. Karimi, P. H. Spallek, and P. A. Dunn, "Extracting social determinants of health from dental records: feasibility study," *International Dental Journal*, vol. 73, pp. S27–S27, 2023.

[59] A. Yan, J. McAuley, X. Lu, J. Du, E. Y. Chang, A. Gentili, and C.-N. Hsu, "Radbert: Adapting transformer-based language models to radiology.," *Radiology. Artificial intelligence*, vol. 4, pp. e210258–, 6 2022.

[60] A. Chattopadhyay, S. Slocum, B. D. Haeffele, R. Vidal, and D. Geman, "Interpretable by design: Learning predictors by composing interpretable queries.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, pp. 7430–7443, 5 2023.

[61] M. R. Banaji, S. T. Fiske, and D. S. Massey, "Systemic racism: individuals and interactions, institutions and society.," *Cognitive research: principles and implications*, vol. 6, pp. 82–, 12 2021.

[62] A. G. Russo, A. Ciarlo, S. Ponticorvo, F. D. Salle, G. Tedeschi, and F. Esposito, "Explaining neural activity in human listeners with deep learning via natural language processing of narrative text.," *Scientific reports*, vol. 12, pp. 17838–, 10 2022.

[63] 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.

[64] I. Spasic, M. Greenwood, A. Preece, N. A. Francis, and G. Elwyn, "Flexiterm: a flexible term recognition method," *Journal of biomedical semantics*, vol. 4, pp. 27–27, 10 2013.

[65] 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.

[66] T. Ekin, O. Kocadağlı, N. D. Bastian, L. V. Fulton, and P. M. Griffin, "Fuzzy decision making in health systems: a resource allocation model," *EURO Journal on Decision Processes*, vol. 4, no. 3, pp. 245–267, 2016.

[67] J. Jha, M. Almagro, and H. Tissot, "Designing nlp applications to support icd coding: an impact analysis and guidelines to enhance baseline performance when processing patient discharge notes," *Journal of Digital Health*, pp. 63–81, 10 2023.

[68] H. Zhou, G. Silverman, Z. Niu, J. Silverman, R. Evans, R. Austin, and R. Zhang, "Annotating music therapy, chiropractic and aquatic exercise using electronic health record.," *Proceedings. IEEE International Conference on Healthcare Informatics*, vol. 2022, pp. 610–611, 9 2022.

[69] Y. Wang, N. Afzal, S. Fu, L. Wang, F. Shen, M. Rastegar-Mojarad, and H. Liu, "Medsts: A resource for clinical semantic textual similarity," *Language Resources and Evaluation*, vol. 54, pp. 57–72, 10 2018.

[70] A. Syrtash, D. Nugent, O. S. Demirel, E. Akbas, F. Kohlhepp, and J. Lamb, "P-508why i left my fertility clinic for another art provider – it's not what you think: A realtalk patient survey," *Human Reproduction*, vol. 37, 6 2022.

[71] 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.

[72] L. Pei, M. Ak, N. H. M. Tahon, S. Zenkin, S. Alkarawi, A. Kamal, M. Yilmaz, L. Chen, M. Er, N. Ak, and R. Colen, "A general skull stripping of multiparametric brain mris using 3d convolutional neural network.," *Scientific reports*, vol. 12, pp. 10826–, 6 2022.

[73] 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.

[74] 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.

[75] R. C. Kessler, M. S. Bauer, T. M. Bishop, R. M. Bossarte, V. M. Castro, O. V. Demler, S. M. Gildea, J. L. Goulet, A. J. King, C. J. Kennedy, S. J. Landes, H. Liu, A. Luedtke, P. Mair, B. P. Marx, M. K. Nock, M. V. Petukhova, W. R. Pigeon, N. A. Sampson, J. W. Smoller, A. Miller, G. Haas, J. Benware, J. Bradley, R. R. Owen, S. House, S. Urosevic, and L. M. Weinstock, "Evaluation of a model to target high-risk psychiatric inpatients for an intensive postdischarge suicide prevention intervention.," *JAMA psychiatry*, vol. 80, pp. 230–230, 3 2023.

[76] E. S. Bett, T. C. Frommeyer, T. Reddy, and J. T. Johnson, "Assessment of patient perceptions of technology and the use of machine-based learning in a clinical encounter," *Intelligence-Based Medicine*, vol. 7, pp. 100096–100096, 2023.

[77] X. Li, J. Chen, B. J. Aguilar, E. Shishova, P. J. Morin, D. Berlowitz, D. R. Miller, M. K. O'Connor, A. H. Nguyen, R. Zhang, A. A. T. Monfared, Q. Zhang, and W. Xia, "Building a neuropsychiatric testing database for veterans with mild cognitive impairment and alzheimer's disease with unstructured electronic health records," *Alzheimer's & Dementia*, vol. 19, 12 2023.

[78] Z. Javed, H. Kundi, R. Chang, A. Titus, and H. Arshad, "Polysocial risk scores: Implications for cardiovascular disease risk assessment and management.," *Current atherosclerosis reports*, vol. 25, pp. 1059–1068, 12 2023.

[79] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Automatic visual recommendation for data science and analytics," in *Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2*, pp. 125–132, Springer, 2020.

[80] W. Semere, S. A. Crossley, A. J. Karter, C. R. Lyles, W. Brown, M. E. Reed, D. S. McNamara, J. Y. Liu, and D. Schillinger, "Secure messaging with physicians by proxies for patients with diabetes: Findings from the eclippse study.," *Journal of general internal medicine*, vol. 34, pp. 2490–2496, 8 2019.

[81] C. A. Morioka, F. Meng, R. K. Taira, J. Sayre, P. Zimmerman, D. Ishimitsu, J. Huang, L. Shen, and S. El-Saden, "Automatic classification of ultrasound screening examinations of the abdominal aorta.," *Journal of digital imaging*, vol. 29, pp. 742–748, 7 2016.

[82] Q.-Y. Zhong, E. W. Karlson, B. Gelaye, S. Finan, P. Avillach, J. W. Smoller, T. Cai, and M. A. Williams, "Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing," *BMC medical informatics and decision making*, vol. 18, pp. 30–30, 5 2018.

[83] B. Kösoglu-Kind, R. Loredo, M. Grossi, C. Bernecker, J. M. Burks, and R. Buchkremer, "A biological sequence comparison algorithm using quantum computers.," *Scientific reports*, vol. 13, pp. 14552–, 9 2023.

[84] M. Masalkhi, J. Ong, E. Waisberg, N. Zaman, P. Sarker, A. G. Lee, and A. Tavakkoli, "Chatgpt to document ocular infectious diseases.," *Eye (London, England)*, vol. 38, pp. 826–828, 11 2023.

[85] D. Demner-Fushman and N. Elhadad, "Aspiring to unintended consequences of natural language processing: A review of recent developments in clinical and consumer-generated text processing.," *Yearbook of medical informatics*, vol. 25, pp. 224–233, 11 2016.

[86] P. Rahman, C. Ye, K. F. Mittendorf, M. Lenoue-Newton, C. Micheel, J. Wolber, T. Osterman, and D. Fabbri, "Accelerated curation of checkpoint inhibitor-induced colitis cases from electronic health records.," *JAMIA open*, vol. 6, pp. ooad017–, 1 2023.

[87] D. Valdez, L. Mena-Meléndez, B. L. Crawford, A. Arvind, and K. N. Jozkowski, "Online social media reactions to the overturn of roe v. wade: Public health implications and policy insights," *Sexuality Research and Social Policy*, vol. 21, pp. 616–631, 10 2023.

[88] J. Li, W.-H. Chen, Q. Xu, N. Shah, J. C. Kohler, and T. K. Mackey, "Detection of self-reported experiences with corruption on twitter using unsupervised machine learning," *Social Sciences & Humanities Open*, vol. 2, no. 1, pp. 100060–, 2020.

[89] S.-F. Chen, S. Loguercio, K.-Y. Chen, S. E. Lee, J.-B. Park, S. Liu, H. J. Sadaei, and A. Torkamani, "Artificial intelligence for risk assessment on primary prevention of coronary artery disease," *Current Cardiovascular Risk Reports*, vol. 17, pp. 215–231, 11 2023.

[90] A. Sharma, M. Witbrock, and K. Goolsbey, "Controlling search in very large commonsense knowledge bases: a machine learning approach," *arXiv preprint arXiv:1603.04402*, 2016.

[91] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines.," *NPJ digital medicine*, vol. 6, pp. 74–, 4 2023.

[92] M. A. Alvi, L. M. Elkaim, J. J. Levett, A. Pando, S. Roy, N. Samuel, N. M. Alotaibi, and G. Zadeh, "Current landscape of social media use pertaining to glioblastoma by various stakeholders.," *Neuro-oncology advances*, vol. 5, pp. vdad039–, 1 2023.

[93] S. B. Coopey, E. Mazzola, J. M. Buckley, J. Sharko, A. K. Belli, E. M. H. Kim, F. Polubriaginof, G. Parmigiani, J. Garber, B. L. Smith, M. A. Gadd, M. C. Specht, A. J. Guidi, C. A. Roche, and K. S. Hughes, "The role of chemoprevention in modifying the risk of breast cancer in women with atypical breast lesions," *Breast cancer research and treatment*, vol. 136, pp. 627–633, 11 2012.

[94] E. Vashishtha and H. Kapoor, "Enhancing patient experience by automating and transforming free text into actionable consumer insights: A natural language processing (nlp) approach," *International Journal of Health Sciences and Research*, vol. 13, pp. 275–288, 10 2023.

[95] J. Valinejad, Z. Guo, J.-H. Cho, and I.-R. Chen, "Social media-based social–psychological community resilience analysis of five countries on covid-19," *Journal of Computational Social Science*, vol. 6, pp. 1001–1032, 8 2023.

[96] J. R. Vest, R. J. Hinrichs, and H. Hosler, "How legal problems are conceptualized and measured in healthcare settings: a systematic review.," *Health & justice*, vol. 11, pp. 48–, 11 2023.

[97] D. Redd, T. E. Workman, Y. Shao, Y. Cheng, S. Tekle, J. H. Garvin, C. A. Brandt, and Q. Zeng-Treitler, "Patient dietary supplements use: Do results from natural language processing of clinical notes agree with survey data?," *Medical sciences (Basel, Switzerland)*, vol. 11, pp. 37–37, 5 2023.

[98] 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.

[99] K. Mermin-Bunnell, Y. Zhu, A. Hornback, G. Damhorst, T. Walker, C. Robichaux, L. Mathew, N. Jaquemet, K. Peters, T. M. Johnson, M. D. Wang, and B. Anderson, "Use of natural language processing of patient-initiated electronic health record messages to identify patients with covid-19 infection.," *JAMA network open*, vol. 6, pp. e2322299–e2322299, 7 2023.

[100] 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.

[101] S. Yang, X. Yang, T. Lyu, X. He, D. Braithwaite, H. J. Mehta, Y. Guo, Y. Wu, and J. Bian, "A preliminary study of extracting pulmonary nodules and nodule characteristics from radiology reports using natural language processing.," *Proceedings. IEEE International Conference on Healthcare Informatics*, vol. 2022, pp. 618–619, 9 2022.