International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Integrating Sentiment Analysis and Topic Modeling for Social License to Operate

Syed Arham Akheel[1], Tarun Shrivas[2]

[1]Senior Solutions Architect, Bellevue, WA
[2]Data Scientist, Bellevue, WA
[1]arhamakheel@yahoo.com, [2]tshrivas@gmail.com

Abstract

Social License to Operate (SLO) refers to the informal acceptance or approval that organizations receive from their stakeholders and the broader community. In recent years, sentiment analysis and topic modeling have emerged as powerful tools for understanding the nuances of community perceptions in real time. This literature review examines the technical foundations of sentiment analysis and Latent Dirichlet Allocation (LDA)-based topic modeling, and explores how their hybrid integration can provide richer insights into stakeholder sentiments relevant to SLO. By analyzing scholarly works that investigate machine learning algorithms, natural language processing (NLP) techniques, and case studies across various industries, this paper synthesizes the current methodological landscape and identifies emerging trends and knowledge gaps. The findings underscore the importance of robust data governance, interdisciplinary collaboration, and ethical considerations for effectively deploying hybrid models to maintain and strengthen an organization's Social License to Operate.

Keywords: Social License to Operate, Sentiment Analysis, Topic Modeling, Latent Dirichlet Allocation, Hybrid Approaches, Natural Language Processing

I. INTRODUCTION

Organizations increasingly recognize that beyond fulfilling formal regulatory requirements, they must also secure the tacit approval of local communities and stakeholders, commonly known as the Social License to Operate (SLO) [1], [2].

Failure to maintain positive stakeholder sentiment can manifest in protests, reputational damage, and operational disruptions, affecting both corporate profitability and long-term viability
[3].

Social media platforms, online forums, and other digital channels now serve as rich data sources for understanding community perceptions in real time [4], [5]. Two analytical paradigms stand out in this domain: *sentiment analysis*, which examines the polarity and intensity of public opinion, and *topic modeling*, which identifies key themes or issues under discussion [6], [7]. When combined, these approaches can help organizations pinpoint what specific concerns drive negative or positive sentiment, thereby guiding more informed and proactive engagement strategies [8].

This paper provides a structured literature review of the technical developments in sentiment analysis and LDA-based topic modeling, focusing on how hybrid methods can enhance SLO monitoring. We draw on both foundational studies and recent advancements in natural language processing (NLP) to highlight methodological best practices, identify research gaps, and suggest directions for future work.

A. Role of AI and NLP

The advent of digital communication platforms has transformed how communities express their opinions, with social media and online forums becoming prevalent channels for public discourse [5]. Artificial Intelligence (AI) and Natural Language Processing (NLP) offer powerful tools to analyze vast amounts of unstructured text data generated on these platforms. Sentiment analysis enables the determination of stakeholders' attitudes towards specific topics, while topic modeling identifies prevalent themes within the text [6], [7]. By integrating these techniques, organizations can capture real-time, nuanced insights into community perceptions and potential concerns affecting their SLO.

B. Purpose of the Review

This literature review aims to explore the integration of sentiment analysis with Latent Dirichlet Allocation (LDA)based topic modeling to enhance the assessment of SLO in the mining and manufacturing industries. We analyze scholarly works employing hybrid NLP techniques, examine machine learning algorithms, implementation strategies, and evaluate their effectiveness in real-world applications. The review also identifies emerging trends, such as deep learning models and real-time data analytics, and uncovers knowledge gaps in areas like multilingual sentiment analysis and context-aware modeling.

II. METHODOLOGY OF LITERATURE REVIEW

A. Literature Selection Criteria

The literature review was conducted using a systematic approach to identify relevant studies published between 2000 and 2023. Databases such as IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, and Google Scholar were searched using keywords like *Social License to Operate*, *sentiment analysis*, *topic modeling*, *Latent Dirichlet Allocation*, *hybrid NLP techniques*, *mining industry*, and *real-time monitoring*. Inclusion criteria focused on peer-reviewed articles, conference papers, and industry reports that discuss the integration of sentiment analysis and topic modeling in assessing SLO within the mining and manufacturing sectors.
Studies not directly related to these topics or lacking empirical results were excluded.

B. Analytical Framework

The selected studies were analyzed based on their methodological approaches, algorithms used, implementation strategies, case studies presented, and their effectiveness in real-world applications. Comparative analysis was conducted to evaluate the strengths and limitations of different hybrid models, the tools and platforms employed, and the specific challenges addressed within the industries.

C. Scope and Limitations

While the review aims to be comprehensive, it is limited to English-language publications, which may exclude relevant research in other languages. The focus is primarily on the mining and manufacturing industries, although insights from related sectors are considered where applicable. The rapidly evolving nature of AI and NLP means that some recent developments may not be fully captured.

III. BACKGROUND: SOCIAL LICENSE TO OPERATE

A. Definitions and Dimensions of SLO

SLO is defined as the level of acceptance or approval continually granted to an organization's operations by the local community and stakeholders [1], [2]. It is an unwritten, socially granted permission that goes beyond formal regulatory compliance [3]. The dimensions of SLO include *legitimacy* (adherence to laws and social norms), *credibility* (providing accurate and transparent information), and *trust* (consistency in actions and meeting expectations) [11].

B. Conceptual Underpinnings

The concept of Social License to Operate (SLO) has become a pivotal consideration for industries that have significant environmental and social footprints, particularly in the mining and manufacturing sectors. SLO refers to the ongoing acceptance or approval of an organization and its operations by local communities and stakeholders, without which companies may face operational delays, protests, or shutdowns [1], [2]. Unlike formal regulatory licenses, SLO is an informal, dynamic, and non-legally binding agreement that reflects the perceptions, opinions, and trust of the community towards an organization [3]. The term *Social License to Operate* was initially popularized in the mining sector to highlight the community's role in influencing project approvals [1]. Over time, it has expanded into various industries—technology, renewable energy, pharmaceuticals, and others—where public acceptance is pivotal for sustainable operations [9]. Unlike legal permits, an SLO is informal and subject to ongoing renegotiation, making it sensitive to shifts in public opinion.

According to Freeman's Stakeholder Theory [10], organizations must consider the interests and power of diverse groups that can affect or be affected by corporate actions. SLO thus hinges on concepts of *legitimacy*, *credibility*, and *trust* [11], which often revolve around how transparently and ethically organizations manage environmental, social, and governance (ESG) issues [3].

C. Challenges in Monitoring SLO

Monitoring SLO is inherently challenging due to its intangible nature and the dynamic environment in which companies operate. Traditional methods of assessing community perceptions, such as surveys and public meetings, are often time-consuming, resource-intensive, and may not capture real-time sentiments [4]. Furthermore, these methods may suffer from biases and limited reach, failing to represent the broader stakeholder spectrum.

Traditional tools such as surveys and focus groups offer direct insights but can be limited in scale and frequency to measure SLO. By contrast, digital platforms enable largescale, real-time data collection of stakeholder sentiments and emerging issues [5]. The adoption of computational methods—particularly in NLP—has catalyzed a more dynamic approach to tracking public perceptions [4], [6].

IV. TECHNICAL FOUNDATIONS

A. Sentiment Analysis

Sentiment analysis is widely applied in monitoring public opinion on social media platforms, product reviews, and customer feedback, influencing decision-making in marketing, politics, and customer relationship management [7]. In the context of SLO, sentiment analysis helps organizations understand community attitudes toward their operations, policies, and initiatives by analyzing comments, posts, and discussions across various digital platforms [5]. Sentiment analysis involves computational techniques to identify and extract subjective information from textual data, determining the sentiment polarity (positive, negative, neutral) and emotions expressed [6]. Techniques range from rule-based approaches using sentiment lexicons to machine learning methods involving classifiers trained on annotated datasets. Advanced methods utilize deep learning models, such as convolutional neural networks (CNNs) and transformers, to capture complex linguistic patterns [13], [14].

Early sentiment analysis research employed classifiers such as Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression [4], [6]. Common features include bag-of-words, TF-IDF, and part-of-speech tags. While these methods can yield decent performance in structured domains, they often struggle with domain-specific language and sarcasm [12].

Recent advances leverage deep neural networks Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and LSTM architectures—to capture context-dependent semantics [7]. With the advent of BERT [13], GPT, and other transformer-based models, sentiment analysis has become more accurate in handling complex syntactic and semantic relationships [14]. However, issues such as data imbalance and interpretability persist [15].

B. LDA-Based Topic Modeling

Topic modeling is an unsupervised machine learning technique used to identify abstract topics within a collection of documents [16]. Latent Dirichlet Allocation (LDA) is one of the most popular topic modeling algorithms, which assumes that documents are mixtures of topics, and topics are probability distributions over words. LDA helps in summarizing large volumes of text data, revealing hidden thematic structures without prior annotations. LDA is a probabilistic model used to discover latent topics in a corpus [16]. It posits that each document (e.g., a social media post) is a mixture of topics, and each topic is a distribution over words. By inferring these distributions, researchers can identify salient themes—such as environmental concerns or local employment issues—relevant to SLO [8]. Several extensions and variants of LDA address specific data and contextual challenges:

- Labeled LDA: Incorporates labels for a semi-supervised approach [20].
- Hierarchical LDA: Models a hierarchical structure of topics for complex, multi-level corpora [18].
- Dynamic LDA: Tracks topic evolution over time, beneficial for longitudinal SLO monitoring [19].

Metrics such as UMass and UCI coherence scores help evaluate topic quality [16]. Human-in-the-loop assessments, where domain experts validate topic relevance, remain crucial to ensure the model's outputs align with real-world stakeholder concerns [2].

LDA effectively handles large datasets and discovers meaningful topic clusters, aiding in content analysis and information retrieval [16]. However, LDA assumes word independence (bag-of-words model), which may ignore the context and semantics of words. It can struggle with short texts (e.g., tweets), and the number of topics must be predetermined, which may require iterative tuning [18].

V. HYBRID APPROACHES: SENTIMENT + LDA

Integrating sentiment analysis with LDA-based topic modeling provides a comprehensive understanding of both the themes discussed by stakeholders and their associated sentiments. Sentiment analysis identifies the polarity and intensity of opinions expressed in textual data, while LDA-based topic modeling uncovers the hidden thematic structure within documents. Together, these techniques allow organizations to understand not only *what* issues are prevalent but also *how* stakeholders feel about these issues, facilitating targeted responses to specific concerns impacting the SLO [8], [9].

For example, LDA can identify topics such as "environmental concerns" or "employment opportunities" within community discussions, while sentiment analysis determines whether these topics are discussed positively or negatively. This integration provides actionable insights for decisionmakers to address community grievances proactively and strengthen stakeholder relationships [4], [7].

Previous studies have validated the effectiveness of hybrid models. For instance, Lu et al. demonstrated that integrating sentiment analysis with topic modeling improved the detection of consumer preferences and dissatisfaction in online reviews. Similarly, researchers have applied these hybrid methods to social media data to detect emerging trends and public sentiments related to renewable energy projects and mining operations [5], [17]. The integration of sentiment analysis with Latent Dirichlet Allocation (LDA)-based topic modeling provides a dual advantage in analyzing unstructured text data: identifying underlying thematic structures and simultaneously capturing the emotional tone associated with these themes. This combined approach is especially critical in contexts like the Social License to Operate (SLO), where organizations must not only understand the issues being discussed by stakeholders but also assess the emotional undercurrents driving community opinions [8], [9].

The hybridization of these methods bridges the gap between *topic discovery* and *sentiment evaluation*. While LDA provides insights into *what* stakeholders are discussing (e.g., "environmental impacts" or "economic benefits"), sentiment analysis reveals *how* they feel about these issues, providing actionable intelligence to address grievances proactively [4], [7]. Sentiment analysis gains contextual specificity when paired with topic modeling, ensuring that emotional tones are analyzed in relation to the precise themes they address (e.g., negative sentiment about "pollution" versus positive sentiment about "employment opportunities"). The hybrid approach enables organizations to monitor community sentiment and emerging concerns in real time, leveraging the vast amounts of data generated from social media, forums, and news platforms [5]. LDA operates under the assumption that documents are mixtures of topics, and each topic is a distribution over words [16]. This probabilistic modeling approach is unsupervised, meaning it does not require pre-labeled data, making it particularly useful for exploring large-scale text corpora in SLO contexts, where emergent themes may not be predefined.

Sentiment analysis, on the other hand, evaluates the polarity (positive, negative, neutral) and intensity of opinions in textual data. Traditional methods employ lexicon-based approaches or machine learning classifiers, while recent advancements have shifted towards transformer-based models such as BERT for greater semantic understanding [13].

Integrating these two methods ensures that the thematic structure uncovered by LDA is enriched by the sentiment information, thereby contextualizing stakeholder opinions within specific topics.

The combined application of these techniques has demonstrated utility across various domains. By applying LDA to forums and social media posts discussing mining projects, researchers have identified topics like "land use" and "water contamination." Sentiment analysis further revealed that "water contamination" was predominantly discussed with negative sentiment, enabling companies to prioritize corrective actions [8]. In studies analyzing public discourse on wind farms, LDA extracted themes such as "noise pollution" and "aesthetic impacts," while sentiment analysis showed that "noise pollution" was a primary driver of negative community sentiment, correlating with local protests [17]. Lu et al. used sentiment-topic hybrid models to analyze online reviews of CSR initiatives, uncovering both the key areas of consumer concern and their associated emotional tones.

The integration of sentiment analysis and topic modeling addresses several challenges inherent in standalone approaches. Standalone sentiment analysis often misinterprets out-of-context phrases or sarcasm. By tying sentiment analysis to specific topics identified through LDA, the hybrid model reduces noise and improves accuracy [7]. LDA excels at identifying emergent themes that may not yet be part of predefined categories, while sentiment analysis provides real-time sentiment trends associated with these themes [5]. The hybrid approach is scalable to vast datasets generated across multiple platforms, enabling organizations to monitor global stakeholder opinions effectively [6].

For organizations aiming to maintain or enhance their SLO, this hybrid approach provides a strategic advantage. It allows for *risk prioritization* by identifying high-sentiment topics that may lead to stakeholder dissatisfaction if left unaddressed. It enhances *community engagement* by providing insights into the most pressing concerns and their associated emotional drivers. It supports *sustainability and governance efforts* by offering data-driven evidence for decision-making and resource allocation.

The integration of sentiment analysis with LDA-based topic modeling enables organizations to move beyond merely tracking public discourse to understanding the deeper sentimentdriven dynamics that shape stakeholder perceptions. By leveraging this hybrid methodology, decision-makers can anticipate and respond to community concerns with greater precision, ensuring the long-term sustainability of their operations.

A. Sequential Integration

In the sequential approach, topic modeling is first applied to discover thematic clusters, and sentiment analysis is then conducted on the documents associated with each topic. This two-step process ensures that sentiments are analyzed in the context of specific themes, enhancing the relevance of insights.

LDA models a document as a mixture of topics and each topic as a distribution over words. The generative process for LDA is as follows [16]:

1) For each document d:

- Draw a topic distribution $\theta_d \sim$ Dirichlet(α).
- For each word w_n in the document:
  - Draw a topic $z_n \sim$ Categorical(θ_d).
  - Draw a word $w_n \sim$ Categorical(β_{zn}),

where β_{zn} is the distribution over words for topic z_n.

The goal is to infer the posterior distributions of θ (topic distributions for each document) and β (word distributions for each topic), typically achieved using Variational Inference or Gibbs Sampling.

Sentiment analysis is applied to the documents grouped by topics to determine the polarity (positive, negative, neutral) of

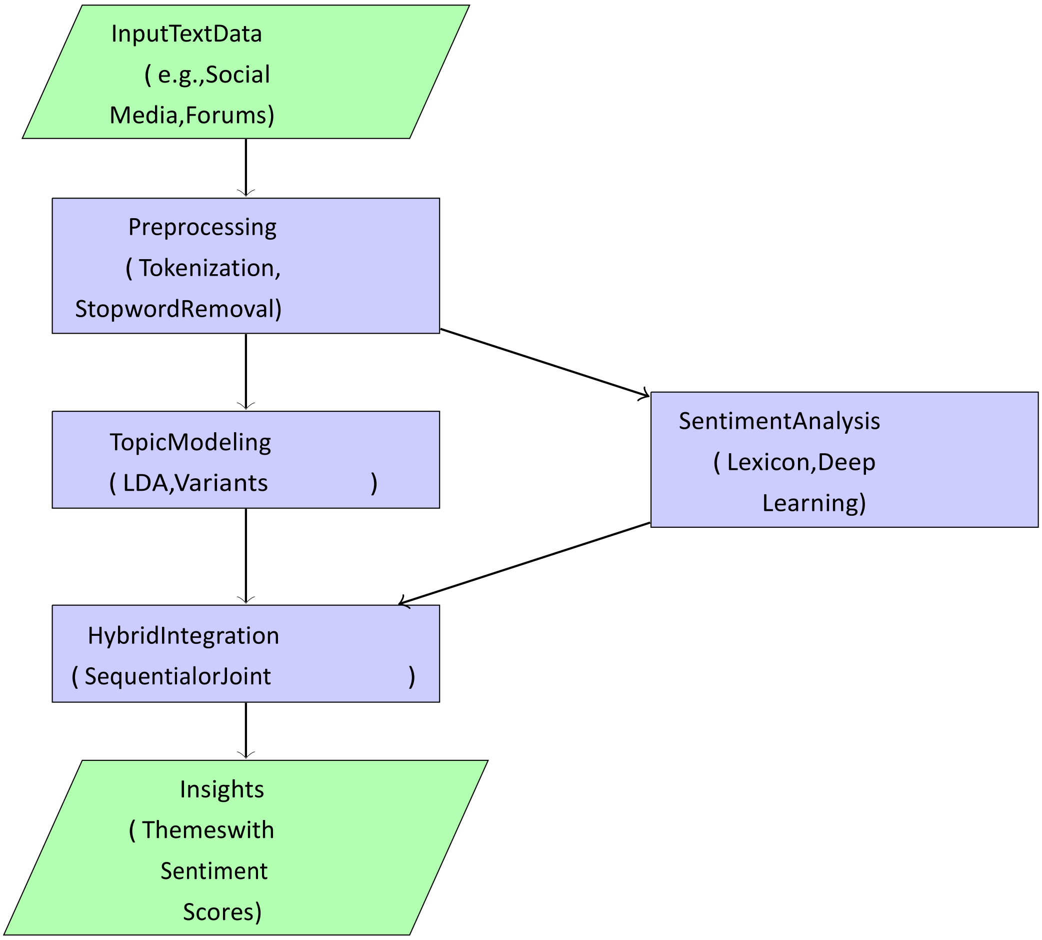

Fig.1.HybridApproach

sentiments associated with each theme. The sentiment score S for a document d is calculated as:

$$S_d = \frac{\sum_{w \in d} s(w) \cdot p(w|\text{topic})}{|d|}$$

where $s(w)$ is the sentiment score of word w, and $p(w|\text{topic})$ represents the probability of w in the associated topic distribution. This ensures that sentiment analysis focuses on the most relevant words within each topic [6].

B. Joint Models for Simultaneous Learning

In the joint approach, models such as Sentiment-LDA and Joint Sentiment/Topic (JST) integrate sentiment analysis directly into the topic modeling process. These models extend LDA by incorporating sentiment priors, enabling the simultaneous learning of topics and their associated sentiments. The JST model modifies the LDA framework by introducing an additional sentiment layer. The generative process for JST is as follows [24]:

1) For each document d:

- Draw a sentiment distribution $\pi_d \sim$ Dirichlet(γ).
- For each sentiment s, draw a topic distribution $\theta_{d,s} \sim$ Dirichlet(α).

- For each word w_n in the document:
  - Draw a sentiment $s_n \sim$ Categorical(π_d).
  - Draw a topic $z_n \sim$ Categorical($\theta_{d,sn}$).
  - Draw a word $w_n \sim$ Categorical(β_{zn}).

Inference in JST is typically performed using Gibbs Sampling or Variational Inference. The process iteratively estimates the posterior distributions of latent variables (sentiment s_n, topic z_n, and distributions π_d, $\theta_{d,s}$, β_{zn}) until convergence. The joint modeling of sentiment and topics increases the computational cost, especially for large datasets. The performance of JST is sensitive to the choice of hyperparameters (α,β,γ), requiring careful tuning. JST relies on a bag-of-words assumption, which may limit its ability to capture nuanced linguistic patterns such as sarcasm or implicit sentiments.

Despite their potential, hybrid approaches face several challenges: Large datasets require efficient algorithms for real-time processing [6]. Current models struggle with context-specific language, including industry jargon and regional dialects [7]. Ensuring data privacy and transparency in model decision-making is critical for building stakeholder trust [23].

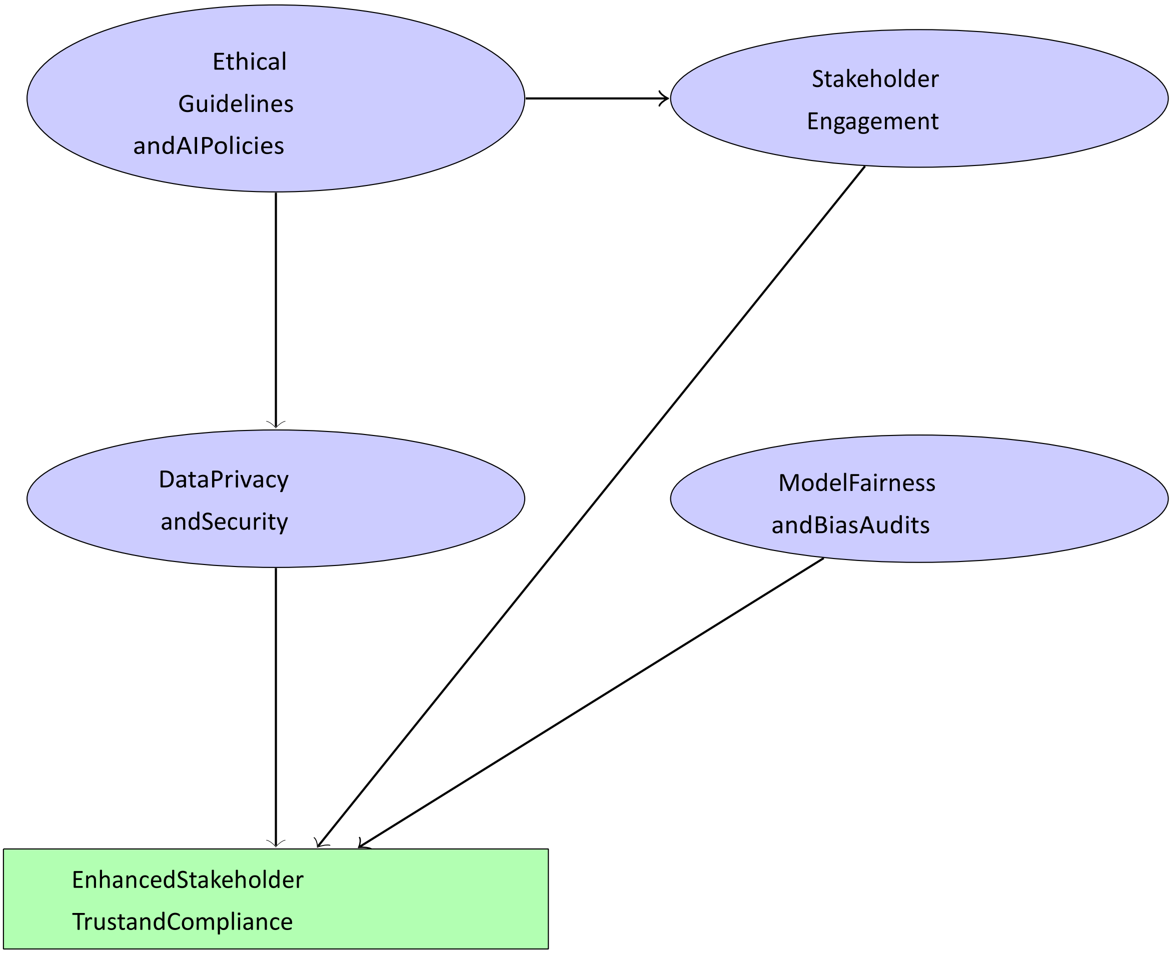

Fig. 2. Ethical AI Governance for Hybrid Models.

VI. CASE STUDIES AND APPLICATIONS

Hybrid approaches combining sentiment analysis and topic modeling have been extensively applied across various industries to address specific challenges, uncover actionable insights, and facilitate stakeholder engagement. This section discusses the practical applications of these techniques in the mining and renewable energy sectors, highlighting their impact on organizational strategies and decision-making.

A. Mining Sector

The mining industry has historically faced significant challenges related to environmental and social impacts, making the Social License to Operate (SLO) a critical factor for sustainable operations. Communities often express concerns about land use changes, pollution, and economic disruptions caused by mining activities. Hybrid sentiment analysis and topic modeling methods have proven instrumental in addressing these challenges by enabling companies to monitor and respond to stakeholder concerns effectively.

1) Case Study: Community Forums and Environmental Concerns: Thomson and Boutilier [1] demonstrated the use of hybrid techniques to analyze online community forums where local residents discussed mining projects. By applying Latent Dirichlet Allocation (LDA), the researchers identified recurring topics such as "environmental degradation," "water contamination," and "employment opportunities." Sentiment analysis revealed that discussions surrounding environmental degradation and water contamination were predominantly negative, whereas employment opportunities elicited mixed sentiments.

These insights prompted the mining company to implement targeted environmental protection measures and enhance community outreach programs. The company introduced regular updates on water quality monitoring and organized town hall meetings to address community concerns. As a result, trust levels among stakeholders increased, leading to improved community relations and reduced project delays.

2) Case Study: Social Media Sentiment Tracking: Another notable application of hybrid methods involved tracking social media sentiment regarding a proposed mining project in Australia. Using LDA, researchers identified key themes such as "economic benefits," "environmental risks," and "community displacement." Sentiment analysis revealed a stark divide: positive sentiments were associated with economic benefits, while environmental risks and displacement generated negative sentiments [9].

The findings informed the mining company's communication strategy, enabling it to highlight its commitment to minimizing environmental risks and supporting displaced communities. For example, the company allocated funds for reforestation projects and established retraining programs for displaced workers. These actions not only mitigated community resistance but also enhanced the company's reputation for corporate responsibility.

B. Renewable Energy

The renewable energy sector, particularly wind and solar energy projects, often encounters resistance from local communities due to perceived environmental and aesthetic impacts. Hybrid sentiment analysis and topic modeling techniques have been pivotal in identifying the root causes of stakeholder opposition and facilitating proactive engagement strategies.

1) Case Study: Wind Farm Proposals and Noise Pollution: Fast et al. [17] applied hybrid methods to analyze public discourse surrounding wind farm proposals in Ontario, Canada. LDA uncovered dominant themes such as "noise pollution," "property value impacts," and "green energy benefits." Sentiment analysis revealed that noise-related discussions were overwhelmingly negative, correlating with local protests and community resistance.

By addressing these concerns, wind energy developers implemented several noise mitigation measures, including improved turbine designs and strategic placement of wind farms away from residential areas. The developers also launched informational campaigns to educate communities about the environmental benefits of renewable energy, which helped shift public sentiment toward a more positive outlook.

2) Case Study: Aesthetic Impact on Landscapes: In another study, researchers analyzed social media posts and survey data to explore community opposition to solar farms in rural regions. LDA identified topics such as "landscape aesthetics," "agricultural land use," and "clean energy benefits." Sentiment analysis revealed a nuanced perspective: while the majority supported clean energy, there was significant negative sentiment associated with the aesthetic and agricultural impacts of large-scale solar installations [3].

These findings encouraged solar energy companies to engage directly with local communities, offering solutions such as incorporating natural vegetation around solar farms to reduce visual impacts and compensating farmers for lost agricultural income. The proactive measures led to a reduction in opposition and expedited project approvals.

C. Impact of Hybrid Methods on Industry Practices

Hybrid sentiment and topic modeling techniques have transformed how industries engage with stakeholders. By providing a detailed understanding of community concerns and sentiments, organizations can design more effective engagement strategies, build trust, and strengthen their Social License to Operate.

The ability to monitor public discourse in real time allows companies to identify emerging issues before they escalate into major conflicts. For example, by detecting early signs of community dissatisfaction with environmental practices, organizations can implement corrective measures and prevent protests or regulatory interventions.

The integration of these analytical techniques into organizational workflows has fostered a data-driven approach to decision-making. By leveraging insights from sentiment and topic analysis, companies can prioritize initiatives, allocate resources effectively, and communicate transparently with stakeholders.

VII. EVALUATION

The evaluation of hybrid sentiment and topic modeling approaches is a crucial aspect of ensuring their effectiveness and reliability in capturing nuanced insights into stakeholder sentiments and thematic concerns. This section discusses performance evaluation metrics, ethical considerations, and governance frameworks to maintain stakeholder trust and achieve meaningful outcomes in Social License to Operate (SLO) monitoring.

A. Performance Evaluation

To evaluate hybrid sentiment and topic modeling frameworks, it is essential to use metrics that capture both the sentiment classification performance and the quality of topic discovery. This dual evaluation ensures that the hybrid model effectively integrates sentiment and topic information for actionable insights.

1) Sentiment Analysis Metrics: The performance of sentiment analysis is typically evaluated using classification metrics derived from a confusion matrix. The common metrics include:

- Accuracy: Measures the overall correctness of the model. Defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively [4].

- Precision: Evaluates the proportion of correctly predicted positive sentiments out of all predicted positives:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- Recall: Measures the proportion of correctly predicted positive sentiments out of all actual positives:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- F1 Score: The harmonic mean of precision and recall, offering a balance between the two:

$$F1 = 2 \text{ x } \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics are critical for assessing the accuracy and reliability of sentiment classification models, particularly in detecting stakeholder attitudes toward specific topics.

2) Topic Modeling Metrics: The quality of topics generated by LDA or its variants is often evaluated using coherence metrics, which measure the semantic interpretability of topics. The most common coherence scores include:

- UMass Coherence: Measures the co-occurrence of words within a topic based on document-level word counts. Defined as:

$$\text{UMass} = \frac{1}{|W|(|W|-1)} \sum_{w_i, w_j \in W} \log \frac{P(w_i, w_j) + \epsilon}{P(w_j)}$$

where $P(w_i,w_j)$ represents the co-occurrence probability of words w_i and w_j, $P(w_j)$ is the marginal probability of w_j, and ϵ is a smoothing term [16].

- C_V Coherence: Based on a sliding window and cosine similarity of word embeddings, C_V coherence measures the semantic similarity of topic words in a vector space:

$$C_V = \frac{\sum_{i<j} \cos(\vec{w}_i, \vec{w}_j)}{\binom{|W|}{2}}$$

where w_i and w_j are word embeddings for words w_i and w_j, respectively, and $|W|$ is the size of the topic word set.

- Perplexity: Measures how well the model predicts unseen data. A lower perplexity indicates better generalization:

$$\text{Perplexity} = \exp\left(-\frac{\sum_{d=1}^{D} \log P(w_d)}{\sum_{d=1}^{D} N_d}\right)$$

where $P(w_d)$ is the likelihood of document d under the model, and N_d is the number of words in d.

3) Hybrid Model Evaluation: For hybrid models, researchers often use a combination of the metrics above to evaluate both the sentiment and topic modeling components. Additional metrics for hybrid evaluation include:

- Sentiment-Topic Coherence: A custom metric that evaluates the alignment between sentiment scores and topic relevance. For each topic, the sentiment distribution is compared to the document-level sentiment to measure consistency.
- Real-Time Performance: Measures the latency and throughput of the model in processing large-scale, realtime data streams. Metrics such as time-per-inference and throughput (documents per second) are used to assess scalability.

B. Ethical and Governance Considerations

1) Data Privacy and Compliance: As organizations increasingly rely on digital footprints to monitor stakeholder perceptions, data privacy has become a critical concern. Regulations such as the General Data Protection Regulation (GDPR) mandate the responsible collection, storage, and processing of personal data [21]. Organizations must:

- Implement robust data anonymization techniques to remove personally identifiable information (PII).
- Ensure data transparency by clearly communicating data usage policies to stakeholders.
- Enable mechanisms for individuals to request data deletion or opt out of data collection.

2) Bias and Fairness in Models: AI models are prone to biases originating from training data or algorithmic design. For instance, sentiment lexicons may not adequately represent cultural or regional variations in language, leading to skewed sentiment scores [6]. Strategies to mitigate bias include:

- Using diverse and representative training datasets to minimize demographic or linguistic biases.
- Employing fairness-aware machine learning techniques to detect and correct biased outcomes.
- Regularly auditing model outputs to ensure equitable treatment of all stakeholder groups.

3) Explainability and Transparency: Ensuring that hybrid models are explainable is essential for building stakeholder trust and compliance with ethical AI guidelines. Techniques such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) can help visualize and explain model predictions [23]. For example, in sentiment analysis, SHAP values can indicate which words contributed most to the sentiment classification.

4) Governance Frameworks: Organizations should establish governance frameworks that align with ethical principles and regulatory requirements. These frameworks should include:

- Ethical Review Boards: Regularly review AI applications to ensure compliance with ethical guidelines.

- Stakeholder Engagement: Involve community representatives in model development and deployment processes to incorporate diverse perspectives.
- Continuous Monitoring: Implement systems to monitor model performance and ethical compliance over time.

VIII. CONCLUSION AND FUTURE DIRECTIONS

This paper reviewed the integration of sentiment analysis and Latent Dirichlet Allocation (LDA)-based topic modeling as a hybrid framework to enhance the assessment of the Social License to Operate (SLO) across industries. By leveraging the strengths of both methodologies, organizations can gain deeper insights into the underlying themes driving stakeholder perceptions and the associated sentiments, enabling more effective decision-making and stakeholder engagement.

The integration of sentiment analysis and LDA provides several key benefits, while sentiment analysis captures the emotional tone of stakeholders, LDA uncovers thematic structures, enabling organizations to understand both *what* is being discussed and *how* it is perceived. Hybrid approaches allow organizations to prioritize stakeholder concerns by identifying high-sentiment topics, such as “environmental degradation” or “employment opportunities”. Applications in the mining and renewable energy sectors have demonstrated the utility of hybrid models in addressing environmental concerns, social resistance, and operational risks [1], [17]. The ability to analyze large-scale data from social media, forums, and surveys ensures real-time monitoring of stakeholder sentiments and emerging trends [4], [7].

Despite these advantages, the hybrid approach faces several challenges, including computational complexity, data privacy concerns, and the need for contextual language understanding. Addressing these limitations will be critical for the continued evolution of sentiment-topic modeling techniques. The integration of sentiment analysis and topic modeling has the potential to transform stakeholder engagement strategies across multiple domains. Beyond the mining and renewable energy sectors, hybrid models can be applied to monitor public perceptions of emerging technologies, urban development projects, healthcare initiatives, and more. By providing real-time, granular insights into stakeholder concerns, these techniques empower organizations to foster trust, address grievances proactively, and align their operations with community expectations.

As organizations face increasing scrutiny from stakeholders and communities, the ability to understand and address public concerns has become paramount. Hybrid sentiment and topic modeling methods offer a robust framework for navigating this complex landscape, combining the strengths of NLP and machine learning to generate actionable insights. However, realizing the full potential of these techniques will require continued innovation, interdisciplinary collaboration, and a commitment to ethical AI practices. By addressing these challenges and embracing emerging opportunities, organizations can leverage hybrid models to build stronger relationships with stakeholders, ensuring sustainable growth and long-term success.

Building on the findings of this paper, future research should focus on addressing current limitations and exploring novel applications of hybrid sentiment-topic models. The following areas represent promising avenues for further investigation:

A. Advancing Multilingual and Multicultural Models

Current sentiment and topic modeling frameworks often fail to account for linguistic and cultural diversity, limiting their applicability in global contexts. Future work should emphasize:

- Developing cross-lingual models that incorporate multilingual sentiment dictionaries and contextual embeddings to analyze stakeholder sentiments across diverse languages [5].
- Investigating cultural nuances in sentiment expression and adapting models to account for region-specific language usage and idiomatic expressions.

B. Incorporating Temporal Dynamics

Static models fail to capture the temporal evolution of topics and sentiments, which is essential for tracking changes in stakeholder perceptions over time. Enhancements such as Dynamic Topic Models (DTM) and time-series sentiment analysis can provide insights into:

- How community concerns and sentiments evolve in response to significant events or corporate actions [19].
- Predictive modeling to forecast shifts in public opinion and pre-emptively address potential conflicts.

C. Integration with Explainable AI (XAI) Techniques

As hybrid models become more complex, ensuring transparency and interpretability will be crucial for building stakeholder trust. Future work should focus on:

- Incorporating explainability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to provide actionable and transparent insights into model decisions [23].
- Developing intuitive visualizations that enable decisionmakers to explore sentiment-topic relationships interactively.

D. Real-Time and Scalable Solutions

The demand for real-time insights necessitates the development of scalable hybrid models that can process vast amounts of data efficiently. Key research directions include:

- Leveraging distributed computing frameworks, such as Apache Spark, to enhance processing capabilities.
- Exploring online learning algorithms to enable incremental model updates as new data streams are ingested [6].

E. Addressing Ethical Challenges

The increasing reliance on hybrid models raises ethical concerns related to data privacy, fairness, and accountability. To ensure responsible deployment, future research should:

- Develop privacy-preserving techniques, such as differential privacy, to safeguard stakeholder data [21].
- Conduct fairness audits to mitigate biases in sentiment classification and topic discovery.
- Establish ethical guidelines and governance frameworks to regulate the use of sentiment-topic analytics in sensitive contexts.

REFERENCES

[1] I. Thomson and R. G. Boutilier, “The Social License to Operate,” in *SME Mining Engineering Handbook*, P. Darling, Ed., pp. 1779–1796, Society for Mining, Metallurgy, and Exploration, 2011.

[2] S. Bice, “What Gives You a Social Licence? An Exploration of the Social Licence to Operate in the Australian Mining Industry,” *Resources*, vol. 3, no. 1, pp. 62–80, 2014.

[3] N. Gunningham, R. A. Kagan, and D. Thornton, “Social License and Environmental Protection: Why Businesses Go Beyond Compliance,” *Law & Social Inquiry*, vol. 29, no. 2, pp. 307–341, 2004.

[4] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[5] G. C. Kane, M. Alavi, G. Labianca, and S. P. Borgatti, “What’s Different About Social Media Networks? A Framework and Research Agenda,” *MIS Quarterly*, vol. 38, no. 1, pp. 275–304, 2014.

[6] B. Liu, *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2012.

[7] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New Avenues in Opinion Mining and Sentiment Analysis,” *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[8] K. Moffat and A. Zhang, “The Paths to Social Licence to Operate: An Integrative Model Explaining Community Acceptance of Mining,” *Resources Policy*, vol. 39, pp. 61–70, 2014.

[9] R. G. Boutilier, “Frequently Asked Questions About the Social Licence to Operate,” *Impact Assessment and Project Appraisal*, vol. 32, no. 4, pp. 263–272, 2014.

[10] R. E. Freeman, *Strategic Management: A Stakeholder Approach*. Boston: Pitman, 1984.

[11] M. C. Suchman, “Managing Legitimacy: Strategic and Institutional Approaches,” *Academy of Management Review*, vol. 20, no. 3, pp. 571– 610, 1995.

[12] M. Ott, C. Cardie, and J. T. Hancock, “Estimating the Prevalence of Deception in Online Review Communities,” in *Proceedings of the 21st International Conference on World Wide Web*, pp. 201–210, 2012.

[13] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of NAACL*, pp. 4171–4186, 2019.

[14] C. Sun, X. Qiu, Y. Xu, and X. Huang, “How to Fine-Tune BERT for Text Classification?,” in *Chinese Computational Linguistics*, vol. 11856 of LNCS, pp. 194–206, Springer, 2019.

[15] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The Ethics of Algorithms: Mapping the Debate,” *Big Data & Society*, vol. 3, no. 2, pp. 1–21, 2016.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[17] S. Fast, W. E. Mabee, and J. Blair, “The Role of Community Acceptance in Planning Outcomes for Renewable Energy in Ontario, Canada,” *Energy Research & Social Science*, vol. 10, pp. 135–143, 2015.

[18] D. M. Blei, “Nested Hierarchical Dirichlet Processes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 228–245, 2010.

[19] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 113– 120, 2006.

[20] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, “Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 248–256, 2009.

[21] M. J. Culnan and P. K. Armstrong, "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization Science*, vol. 10, no. 1, pp. 104–115, 1999.

[22] R. Heeks and J. Renken, "Data Justice for Development: What Would It Mean?," *Information Development*, vol. 34, no. 1, pp. 90–102, 2018.

[23] L. Floridi, "Soft Ethics and the Governance of the Digital," *Philosophy & Technology*, vol. 31, no. 1, pp. 1–8, 2018.

[24] C. Lin, Y. He, and R. Everson, "Joint Sentiment/Topic Model for Sentiment Analysis," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, pp. 375–384, 2009.