Volume (1) No (12), 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Addressing The Idiom Challenge in Machine Translation: A Review Focused on Low-Resource Languages

Manjot Kaur
Department of Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, Punjab
kaur20manjot04@gamil.com

Jasvir Kaur
Department of Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, Punjab
jasvirsaran013@gmail.com

Jasmin Kaur Gahlot
Department of Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, Punjab
gahlotjasmin05@gmail.com

Prof. Palak Sood
Department of Computer Science and Engineering
Guru Nanak Dev Engineering College
Ludhiana, Punjab
palak.sood001@gmail.com

***Abstract*—Machine translation (MT) has made significant progress for high-resource languages, yet idiomatic translation remains a persistent challenge, especially for low-resource languages like Punjabi. Idioms are non-compositional with respect to cultural values, which makes literal translations insufficient. This paper presents a systematic review of idiom translation in pairs of low-resource-to-high-resource languages, focusing on Punjabi-English as a case study. We analyze key challenges—including dataset limitations, figurative-literal ambiguity, structural complexity, and evaluation limitations—and examine existing approaches, including rule-based, statistical, neural, and large language model (LLM)-based methods. We identify gaps in idiom-specific datasets, evaluation frameworks, and multilingual transfer techniques. Finally, we provide some guidance for future research, highlighting hybrid models, community-driven datasets, multimodal translation, and idiom-aware evaluation metrics. This review aims to guide the development of more accurate and culturally aware MT systems for low-resource languages.**

Keywords—Machine Translation, Idiomatic Expressions, Low-Resource Languages, Large Language Models, Survey. Introduction.

I. Introduction

Machine translation has improved a lot in recent years, especially for languages that have a lot of resources. But translating idioms accurately is still a big problem, especially for languages with fewer resources like Punjabi. Idioms are expressions that do not follow the literal meaning of the words but there meaning is depend upon their culture [1]. This is the reason which makes it hard for machine translation tools to translate idioms properly [2][3]. For instance, the Punjabi idiom "ਉੱਚਾ ਬੋਲਣਾ", which literally means ”speak loudly,” actually means ”to speak rudely.” As the direct translation can’t capture the real meaning of idioms, so there is a need for machine translation systems that understand idioms better [4].

This problem is even more noticeable in low resource languages [5][6]. These languages often have fewer translated texts, less cultural understanding, and not enough ways to test how well translations work [7]. The high resource languages with many data and advanced models can have better translation than the low resource language that struggles with lack of resources and good systems [8]. This gap creates limited access of an accurate translation for so many peoples.

This paper provides a detailed look at problems of accuracy in translating idioms for low resource languages as compared to high resource languages.

We focus on major challenges such as limited datasets, confusion between literal and figurative meanings, and the limitation of the evaluation system . We also cover existing methods such as rule-based, neural, and large language model approaches. We also point out where current datasets and testing methods fall short for low resource languages like Punjabi.

The main contributions of this paper are:

- A detailed review of the difficulties in translating idioms for low-resource languages.
- A thorough look at the methods used so far and their problems.
- Suggestions for the future, including combining different methods, creating community-driven datasets, and using more culturally sensitive evaluation methods.

By tackling these issues, this work aims to guide the development of better, more inclusive machine translation systems that can handle idioms in low-resource languages more effectively.

II. Related Work

A. Idiom-Specific Datasets

One of the first steps in making machine translation more aware of idioms has been creating special datasets for idioms [9]. There is a German-English dataset with more than 3,000 idiomatic sentences that is used to assess how well neural machine translation systems handle phrases that don’t mean what their words literally say [10]. Their research showed that standard translation models often give literal translations, which shows the need for better benchmarks focused on idioms. There is not much work done like this in other languages, especially for low resource languages, like Punjabi. There exists an English-Punjabi dataset for idioms and phrases, using automatic alignment and context clues. Even though the dataset is small, it shows how important it is to have language-specific idiom resources for translation research in low-resource languages.

B. Evaluation of Idiom Translation

Evaluating the translation of idioms is not easy, because common MT metrics like BLEU or METEOR do not effectively measure how well idioms are translated [11]. There is special evaluation method created for idioms, that check for literal translation mistakes with automatic tools. Their research found that training models with only one language helps them handle idioms better in machine translation [12]. The other researchers added that evaluations should consider the difference between figurative and literal meanings, since systems often do well with grammar but struggle with meaning.

C. Neural Architectures and Model Behavior

Recent studies have improves how neural models handle idioms within their processing [13]. Found that Transformer-based machine translation systems tend to show a compositional bias, treating idioms as separate words instead of fixed phrases [14]. This tendency can lead to mistranslations even with enough context to understand the idiom. Additionally, research on multilingual pre- training indicates that large language models are better at understanding idioms, but there are still difficulties when working with languages having low resources [15].

D. Low-Resource Machine Translation and Idioms

Translating idioms is especially hard for languages with few resources [16]. There exists the FLORES dataset to check how well translations work for low resource languages, but this dataset do not include many idioms. Recent studies on machine translation for low-resource languages show that idioms are a big part of language that is often ignored in methods that transfer knowledge between languages [17]. For Punjabi, there aren't many re- sources available, and most research uses small collections of translated texts or manually made word lists. This shows how important it is to develop more resources that include idioms for less commonly spoken languages.

E. Summary and Gap Analysis

The literature reveals consistent limitations for translating idioms of low resource:

- Idiom-specific datasets are available only for high-resource languages.
- Evaluation metrics are improving but remain under-used in Low resource languages.
- Neural models continue to struggle with figurative meaning, especially under data scarcity.
- Punjabi and similar LRLs are severely underrepresented in idiom-focused MT research.

These challenges motivates the present review, which aims to show importance of improving translating system for low resource languages.

III. Challenges And Open Problems

Idiom translation is a big challenge for machine translation systems, especially when dealing with language pairs that have limited resources, such as Punjabi and English. This is because idioms are culturally specific, not follow a direct structure, and can be hard to break down. Here are the main difficulties involved:

A. Data Scarcity and Lack of Idiom-Specific Corpora

Low-resource languages like Punjabi don't have big, annotated datasets that include idioms, which are important for building strong machine translation systems[18]. The parallel corpora that do exist are usually focused on specific areas and don't often include idiomatic phrases in real-life contexts[19]. For instance, the Punjabi idiom "ਸਾਹ ਸੁੱਕਣਾ" (which literally means ”low breadth” but is used figuratively to mean ”being nervous”) needs a deep understanding of context, something that's missing in most training data.

B. Figurative vs. Literal Ambiguity

Idioms can be used in both direct and metaphorical ways, which makes it hard to figure out their ex- act meaning. Machine translation systems, depend on breaking down words individually, provides literal meaning[20]. For example idiom "ਉੱਚਾ ਬੋਲਣਾ " which figurative means disrespecting other but literally means speaking loudly. Although using context-based word representations helps with understanding the right meaning, these systems still have trouble when there's not enough training data, especially in languages that aren't widely used.

C. Structural Complexity of Idioms

Idiomatic expressions does not follows standard alignment rules. For instance, the Punjabi idiom "ਅੱਖਾਂ ਵਿਚ ਧੂੜ ਪਾਉਣਾ" ”to throw dust in the eyes” (figuratively meaning ”to deceive”) has parts that don't sit together in a straightforward way, making it hard for sequence-to- sequence models to identify and match correctly.

D. Fluency and Semantic Coherence

When idioms are translated word for word, the resulting text sound awkward or not make much sense in the target language. A direct translation of a Punjabi idiom into English might be grammatically correct but still seem confusing. To handle this, machine translation systems need to include rules that help with idiomatic expressions

E. Limitations of Evaluation Metrics

Traditional evaluation methods like BLEU, METEOR, and TER don't really measure how well idioms are handled because they focus on matching words or small phrases[21]. Newer tools like BERTScore and COMET are better at catching these issues, but they aren't widely used in situations where there's little data. Without specific tools to test idiomatic accuracy, it's hard to judge how well a translation system is handling idioms.

To solve these issues, we need a combination of better data collection, improved model designs, and more effective ways to test translations. This will help to build more accurate machine translation systems, especially for low resource languages.

IV. Existing Approaches

There are several paradigms evolved for idiom containing sentences, beginning with rule-based methods, statistical models, neural machine translation (NMT) and large language models (LLMs). This section reviews the main approaches (Table 1) , highlighting their contributions and limitations in handling idiomatic expressions:

A. Rule-Based and Statistical Machine Translation (SMT)

Early on, people used rule-based systems to translate idioms. These systems had lists of idioms stored manually [22]. This system worked well for fixed phrases but had trouble with paraphrased idioms and non-contiguous idioms.

Later, phrase-based statistical machine translation (SMT) came along [23]. It treated idioms as groups of words [24]. However, it relied too much on matching word positions, which limited how well it could handle new or different expressions. SMT often translated word by word, missing the real meaning of idioms.

B. Neural Machine Translation (NMT)

The introduction of the Transformer model changed things by letting translation consider more context [25]. Models like MarianMT and mBERT help with translating idioms better[26]. But even these systems had problems when an idiom was used in a literal way instead of figuratively, since they didn't have enough examples in their data [27].

New research shows that translating idioms is harder than general translation, especially in languages with little data.

C. Idiom-Aware Enhancements to NMT

To fix this, researchers made special tests and ways to train models more effectively made tools to check how well idioms are translated created automatic ways to measure translation mistakes[28]. Others tried adding idiom information directly into translation models, which helped them understand expressions that don't follow regular word rules.

D. Large Language Models (LLMs)

The emergence of large language models (LLMs), such as GPT-3, ChatGPT, and LLaMA, has introduced new possibilities for idiom translation [29]. LLMs leverage massive pretrained architectures and contextual reasoning capabilities, enabling them to better differentiate between literal and figurative meanings in idiomatic expressions [30][31]. This advancement addresses a key limitation of traditional neural machine translation (NMT) systems, which often struggle with idioms due to their reliance on word-level compositional.

However, the performance of LLMs in idiom translation remains inconsistent, particularly for low-resource languages like Punjabi. This inconsistency arises from two primary challenges:

- **Data Scarcity:** LLMs are trained on vast amounts of text data, but low-resource languages are significantly underrepresented in these datasets [32]. Consequently, LLMs often lack the cultural and contextual understanding required to accurately translate idioms in such languages.
- **Lack of Transparency:** The black-box nature of LLMs complicates error analysis and correction[33]. Unlike rule- based or statistical methods, where errors can be traced to specific rules or alignments, LLM outputs are less interpretable, making it difficult to diagnose and address translation errors.

Despite these challenges, LLMs offer promising directions for future research. For example, few-shot learning and prompt engineering techniques can be explored to improve idiom translation in low-resource settings. By providing LLMs with a small number of idiom examples as context, researchers may guide the model toward more accurate translations [34]. Additionally, fine-tuning LLMs on domain-specific idiom datasets could enhance their ability to capture cultural nuances and figurative meanings. Another potential avenue is the use of multilingual LLMs, such as mT5 or BLOOM, which are pretrained on diverse linguistic data. These models may offer better crosslingual transfer capabilities, enabling them to generalize idiom translations across languages. However, their performance still depends on the availability of high- quality idiom datasets for low-resource languages, which remains a significant limitation.

In summary, while LLMs represent a powerful tool for idiom translation, their effectiveness in low-resource contexts is constrained by data availability and interpret ability challenges. Future research should focus on leveraging LLMs through few-shot learning, prompt engineering, and fine-tuning, while also addressing the need for larger, culturally diverse idiom datasets [16].

TABLE I. Comparison of Existing Approaches For Idiom Translation

Approach	Description	Strengths	Limitations	Low Resource Suitability
Rule-based	Uses handcrafted rules and idiom dictionaries	Accurate for fixed idioms	Poor scalability; fails on paraphrased or non-contiguous idioms	Limited: Requires manual creation
Statistical MT	Phrase-based translation using corpus alignments	Handles multi-word units	Relies on surface-level alignment; often literal	Moderate: Works with small corpora
Neural MT	Context-aware neural networks	Learns better representations and context	Struggles with figurative vs. literal meaning	Moderate: Needs more idiom examples
Large Language Models	Pretrained models such as GPT or LLaMA	Capture figurative meaning; contextual; few-shot capable	Black-box nature; inconsistency for low-resource languages	Limited: Data scarcity remains a challenge

E. Low-Resource and Multilingual Transfer Approaches

For languages such as Punjabi, idiom translation faces unique challenges. Transfer learning from high-resource languages has been applied to mitigate data scarcity. Multilingual pre-trained models (e.g., XLM-R, mBART- 50) provide cross-lingual representations that help preserve idiomatic meaning across resource gaps[35]. Yet, idiom translation remains less reliable when idioms have no direct equivalents in the target language, or when parallel idiom-specific corpora are unavailable.

V. Datasets and Benchmarks

The development of idiom-aware machine translation (MT) systems heavily depends on the availability of high-quality datasets and benchmarks. However, idiom-specific resources remain severely limited, particularly for low-resource languages like Punjabi. While several datasets have been developed for high-resource languages, their coverage of idiomatic expressions is often insufficient for training robust MT models.

Existing idiom-specific datasets include the VNC-Tokens Dataset [36], which contains approximately 2,500 English verb-noun combinations annotated for literal vs. idiomatic usage. Although limited in size, it laid the foundation for idiom detection tasks.

The MAGPIE Corpus is a large-scale English dataset with around 55,000 idiom instances, annotated for figurative and literal meanings [33]. While useful for idiom identification, it is not directly applicable to MT tasks [37]. The PIE Corpus focuses on phrasal idiomatic expressions in English, providing annotations for idiomaticity but offering limited utility for translation purposes.

The PETCI Dataset [38] is a bilingual English-Chinese parallel idiom dataset comprising around 2,000 pairs, specifically designed to evaluate MT systems. It highlights systematic failures of neural MT (NMT) when translating idioms. Additionally, Idiom Test Suites [39] consist of curated idioms embedded in natural sentences, intended for diagnostic evaluation of idiomatic translation rather than training.

Despite these efforts, significant gaps persist for low-resource languages. Most idiom datasets focus on high-resource languages such as English, German, and Chinese, leaving languages like Punjabi, Nepali without large-scale annotated parallel dataset[40]. Existing parallel corpus for low-resource languages are often domain-specific and lack idiomatic expressions in natural contexts.

Furthermore, these datasets frequently suffer from limited contextual diversity, failing to capture regional and cultural variations that are critical for accurate idiom translation. The small size of most idiom datasets (typically <100,000 examples) also renders them insufficient for training modern deep neural or large language model (LLM)-based systems.

To address these gaps, several strategies can be explored. Community-driven annotation platforms, such as Amazon Mechanical Turk can help rapidly expand idiom coverage for low-resource languages. Synthetic data generation using LLMs, followed by human validation, offers another promising approach. Techniques like back-translation and cross-lingual alignment using multilingual embeddings can also help align idioms across languages, even when parallel data is scarce. These methods can facilitate the creation of more comprehensive and culturally diverse idiom datasets for low-resource languages.

VI. Evaluation Methods

In MT research, assessing the quality of idiom translation is a crucial. The non-compositional and context-dependent nature of idioms is frequently overlooked by traditional evaluation tools, which results in inaccurate evaluations of translation quality.

To compare system outputs with reference translations, automatic measures like BLEU, METEOR, and TER use edit distances or n-gram overlaps. Contextual embeddings are used by more sophisticated embedding- based measures, like COMET and BERTScore, to gauge semantic similarity. These metrics still have trouble telling the difference between idiomatic statements' literal and figurative meanings, despite being more adept at identifying paraphrases.

To address these limitations, idiom-specific test suites have been developed [41]. For example, the evaluation framework proposed measures the ability of NMT systems to handle figurative vs. literal usage of idioms, providing fine-grained error analysis. However, such test suites require manual construction and annotation, limiting their scalability. Diagnostic test sets, which include curated idioms embedded in natural sentences, are designed to evaluate idiom adequacy, contextual appropriateness, and preservation of figurative meaning. While these test suites offer valuable insights, their coverage remains limited.

Human evaluation remains the gold standard for assessing idiom translation quality. Native speakers evaluate translations based on fluency (grammatical correctness), adequacy (preservation of meaning), and idiomatically (naturalness in the target language). However, human evaluation is time-consuming, costly, and subjective, making it impractical for large-scale assessments. Error analysis frameworks, categorize idiom translation errors (e.g., literal translations, incorrect figurative meanings) to better understand model failures. Yet, these frameworks are typically manual and non-scalable, highlighting the need for automated idiom-specific error quantification tools.

A hybrid approach that combines automatic metrics with human evaluation can balance scalability and accuracy. For instance, idiom-specific test suites can be used to pre-screen system outputs, followed by human evaluation for fine-grained analysis. Such hybrid frameworks are increasingly adopted in NMT evaluations, as they provide a practical compromise between efficiency and linguistic quality.

TABLE II. Comparison of Different Evaluation Methods

Method	Description	Strengths	Limitations
Automatic Metrics (BLEU, METEOR, TER)	N-gram or edit-distance–based metrics comparing system output with references	Fast, widely adopted, reproducible	Fail to capture idiomatic meaning; penalize paraphrased but correct translations
Embedding-Based Metrics (BERTScore, COMET)	Use contextual embeddings to measure semantic similarity	Better at recognizing paraphrases; more sensitive to idiomatic equivalence	Still not idiom-specific; may misinterpret figurative contexts
Idiom-Specific Test Suites	Benchmarks designed for idiom evaluation (e.g., literal vs. figurative usage)	Fine-grained, idiom-focused evaluation; good for error analysis	Limited coverage; require manual construction and annotation
Human Evaluation	Native speakers judge fluency, adequacy, and idiomaticity	Gold standard; captures cultural and contextual nuances	Expensive, time-consuming, subjective
Error Analysis Frameworks	Categorize idiom translation errors systematically	Provides insight into model weaknesses; helps refine systems	Few standardized tools; lack automation for large-scale evaluation

Hybrid Approaches	Combination of automatic and human evaluation	Balances scalability with qualitative depth	Requires coordination; still partly resource-intensive

VII. Discussion And Open Issues

While machine translation (MT) has achieved remarkable progress in recent years, idiom translation between low-resource and high-resource language pairs remains an unsolved challenge. This paper has reviewed the key challenges, approaches, datasets, and evaluation methods for idiom translation, with a focus on low-resource languages like Punjabi. Our analysis reveals several open issues that require further attention from the research community.

One of the most significant gaps is distance between evaluation metrics and real-world performance. Traditional automatic metrics such as BLEU and METEOR fail to capture idiomatic meaning, often penalizing correct but paraphrased translations. While idiom-specific test suites provide a more nuanced evaluation, they cover only a limited range of idiomatic expressions. For example, a Punjabi idiom like "ਉੱਚਾ ਬੋਲਣਾ " (disrespecting someone) may be correctly translated, but standard metrics like BLEU will underestimate its performance because the translation is not literal. This evaluation bottleneck underscores the need for idiom-aware metrics that can assess cultural appropriateness.

Another critical issue is the lack of multi modal and conversational context in current MT systems. Idioms often carry cultural, visual, or situational meanings that are lost in text-only translation. For instance, the Punjabi idiom "ਅੱਖਾਂ ਵਿਚ ਧੂੜ ਪਾਉਣਾ" (to deceive) may require visual or contextual cues to fully convey its meaning. While recent advances in multi modal MT and conversational AI offer potential solutions, research in this area remains limited, particularly for low-resource languages.

The cultural and contextual nuances of idioms present additional challenges. Idioms are not merely linguistic expressions but also reflect cultural values, humor, and social norms. Preserving these nuances in translation is difficult, especially when cultural equivalents do not exist in the target language. This issue is amplified in low-resource languages, where idioms often lack widely accepted equivalents in high-resource languages. For ex- ample, translating a Punjabi idiom like "ਸਾਹ ਸੁੱਕਣਾ" (being nervous) into English requires deep cultural understanding, which current MT systems lack.

The scalability of idiom translation to low-resource languages is another open issue. While parallel idiomatic corpora exist for high-resource languages, low- resource languages like Punjabi, Nepali, and Sinhala lack such resources. This scarcity hinders the development of idiom-aware MT systems for these languages. Addition- ally, human evaluation, though accurate, is costly and non-scalable, making it impractical for large-scale assessments.

Finally, the limited integration of large language models (LLMs) in idiom research presents an opportunity for future work. While LLMs like GPT-3 and LLaMA have shown promise in handling figurative language, their ap- plication to low-resource idiom translation remains underexplored. Systematic research on few-shot learning, prompt engineering, and fine-tuning for idioms could yield significant improvements, particularly in zero-shot or few- shot settings .

These open issues highlight the need for a multidisciplinary research agenda that integrates linguistic theory, computational modeling, and cultural studies. Addressing these challenges will not only advance idiom translation but also promote exclusivity for underrepresented languages in global NLP research.

VIII. Future Direction

Advancing the translation of idioms between languages with few resources and those with plenty requires several key research areas. First, building big collections of idioms is really important. People can help by contributing, using online crowd sourcing, and linking idioms across languages. This helps gather more idioms and reflects different regions and cultures, making the data more varied and useful.

Second, large language models like GPT, LLaMA, and mT5 can be great for translating idioms. Training these models with a few examples or none at all, and using prompts to guide them, can help them better understand meanings that are not literal or culturally specific. Combining rules from language with these models can also help, as it covers rare expressions while keeping the system effective and accurate.

Third, looking into methods that understand context and use different forms of information, like speech and images, can improve idiom translation. Idioms often depend on the situation or what's around them, so using these different types of data can help models tell the difference between literal and figurative language, making translations more accurate in spoken or visual contexts.

Fourth, creating special ways to measure how well idioms are translated is necessary. Tools that check for idiomatic accuracy, along with human checks, can help assess trans- lation quality in a way that's both wide and precise. Working together between language experts, teachers, and machine learning researchers can ensure translations are culturally meaningful and make sense to people. Sharing data and setting standards can also speed up progress for languages that aren't widely used.

Finally, using idiom-aware translation systems in education, language preservation, and digital tools is important. These systems can help people learn languages, keep their languages alive, and communicate better across cultures. This can also help reduce the gap between people who have access to technology and those who don't, while sup- porting the languages of communities that are not widely spoken. Together, these efforts can lead to better, more inclusive, and more culturally aware machine translation systems.

IX. Conclusion

This paper offers a detailed look at how idioms are translated between low resource languages and high re- source languages, focusing especially on Punjabi and English. Idioms are tricky because they don't follow the usual rules of language and are tied to specific cultures. This makes them hard for translation systems to handle properly. Our review points out several key problems, such as not having enough data, confusion between literal and figurative meanings, complex structures, and a lack of evaluation metrics. We looked at different methods used so far, from old rule-based systems to newer neural networks and large language models, and found what works and what doesn't when it comes to translating idioms.

Our findings show there are still big gaps in the research. For example, there's not enough large and varied collections of idioms for languages with little data, and the tools used to evaluate translations aren't always aware of idioms. Also, there's not much work on using multiple forms of information like images or context to improve translations. By bringing all these points together, we offer clear suggestions for future work, such as mixing different translation methods, training large models to better understand idioms, using techniques that trans- fer knowledge between languages, and creating better, culture-aware ways to measure translation quality.

In short, this survey not only summarizes the latest developments in translating idioms but also helps researchers and translators build more effective and respectful systems. By tackling these issues, future translation tools can bet- ter connect people across languages, help keep cultural expressions alive, and make technology more accessible for speakers of less commonly used languages.

REFERENCES

[1] C. Baziotis, P. Mathur, and E. Hasler, "Automatic Evaluation and Analysis of Idioms in Neural Machine Translation," *EACL 2023 - 17th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Conf.*, pp. 3664–3682, 2023, doi: 10.18653/v1/2023.eacl-main.267.

[2] S. B. Shah, "Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.

[3] R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730334.

[4] W. Zhang, X. Li, Y. Yang, and R. Dong, "Pre-Training on Mixed Data for Low-Resource Neural Machine Translation," *Information*, vol. 12, no. 3, 2021, doi: 10.3390/info12030133.

[5] S. Garg, "AI-Driven Innovations in Storage Quality Assurance and Manufacturing Optimization," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 1, no. 1, pp. 143–147, 2020, doi: 10.54660/.IJMRGE.2020.1.1.143-147.

[6] K. Murugandi and R. Seetharaman, "Sustainable Manufacturing Practices in the Apparel Industry : Balancing Innovation and Environmental Responsibility," *TIJER*, vol. 11, no. 12, pp. 908–914, 2024.

[7] K. Garg, A. Kumar, and V. Goyal, "Development of Punjabi-English (PunEng) Parallel Corpus for Machine Translation System," *Int. J. Eng. Technol.*, vol. 7, p. 690, May 2018, doi: 10.14419/ijet.v7i2.10762.

[8] S. Singh, "Performance Evaluation of Machine Learning Regression Models for 5G Network Resource Allocation Optimization," in *2025 International Conference on Multimedia Computing, Networking and Applications (MCNA)*, 2025, pp. 47–52. doi: 10.1109/MCNA65829.2025.11124374.

[9] M. Fadaee, A. Bisazza, and C. Monz, "Examining the tip of the iceberg: A data set for idiom translation," *Lr. 2018 - 11th Int. Conf. Lang. Resour. Eval.*, no. 2015, pp. 925–929, 2018.

[10] D. Patel, "AI-Enhanced Natural Language Processing for Improving Web Page Classification Accuracy," vol. 4, no. 1, pp. 133–140, 2024, doi: 10.56472/25832646/JETA-V4I1P119.

[11] V. Pal and I. Researcher, "Bias Detection and Mitigation in Foundation AI Models : A Human-Centric Approach," *TIJER*, vol. 8, no. 2, pp. 4–10, 2021.

[12] J. Briskilal and C. N. Subalalitha, "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa," *Inf. Process. Manag.*, vol. 59, no. 1, p. 102756, 2022, doi: https://doi.org/10.1016/j.ipm.2021.102756.

[13] V. Dankers, E. Bruni, and D. Hupkes, "The Paradox of the Compositionality of Natural Language: A Neural Machine Translation Case Study," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 4154–4175. doi: 10.18653/v1/2022.acl-long.286.

[14] Hemish Prakashchandra Kapadia, "Generative AI for Real Time Conversational Agents," *Int. J. Curr. Sci.*, vol. 13, no. 3, pp. 201–208, 2023.

[15] A. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 8440–8451, 2020, doi: 10.18653/v1/2020.acl-main.747.

[16] N. Goyal *et al.*, "The FLORES-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation," *Trans. Assoc. Comput. Linguist.*, vol. 10, pp. 522–538, 2022, doi: 10.1162/tacl_a_00474.

[17] K. S. B. Lohit Sai Andra1, Sandeep Kumar Srivastava2, Shruthi Krishna3, Phaneendra Siddana1, "Assessing User Confidence and Acceptance of AI Enhanced Through Multi-LLM Consensus Mechanisms," *Int. Trans. Electr. Eng. Comput. Sci.*, vol. 4, no. 4, 2025.

[18] R. Karne, P. K. Pativada, and A. Dudhipala, "DFIR-Chain - Integrating Memory Forensics, YARA Scanning, and LLM Summarization for Automated Triage," in *2025 9th International Conference on Inventive Systems and Control (ICISC)*, 2025, pp. 1263–1268. doi: 10.1109/ICISC65841.2025.11187513.

[19] R. Karne, A. Dudhipala, and P. K. Pativada, "LoRA-Fine Tuned LLaMA-3 for Low-Resource Multilingual Machine Translation," in *2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2025, pp. 39–45. doi: 10.1109/ICIMIA67127.2025.11200663.

[20] V. Verma, "Security Compliance and Risk Management in AI-Driven Financial Transactions," *Int. J. Eng. Sci. Math.*, vol. 12, no. 7, pp. 1–15, 2023.

[21] V. Prajapati, "Enhancing Supply Chain Resilience through Machine Learning- Based Predictive Analytics for Demand Forecasting," pp. 345–354, 2025.

[22] K. Mohiuddin *et al.*, "Retention Is All You Need," *Int. Conf. Inf. Knowl. Manag. Proc.*, no. Nips, pp. 4752–4758, 2023, doi: 10.1145/3583780.3615497.

[23] S. P. Kalava, "Innovative API Design with GraphQL: Ensuring Scalability and Flexibility," *IJCEM*, vol. 7, no. 9, p. 5, 2025.

[24] Y. Liu *et al.*, "Multilingual denoising pre-training for neural machine translation," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 726–742, 2020, doi: 10.1162/tacl_a_00343.

[25] X. Deng and Z. Yu, "A Systematic Review of Machine-Translation-Assisted Language Learning for Sustainable Education," *Sustainability*, vol. 14, no. 13, 2022, doi: 10.3390/su14137598.

[26] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, pp. 16–22, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[27] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating Text Generation With Bert," *8th Int. Conf. Learn. Represent. ICLR 2020*, pp. 1–43, 2020.

[28] Y. Fu, Z. Huang, L. Yang, Y. Lu, and Z. Dai, "CHENGYU-BENCH: Benchmarking Large Language Models for Chinese Idiom Understanding and Use," *arXiv*, pp. 2355–2366, 2025, doi: 10.18653/v1/2025.emnlp-main.119.

[29] A. Dudhipala, R. Karne, and P. K. Pativada, "Prompt2Graph: Leveraging LLMs to Construct Knowledge Graphs from Technical Manuals," in *2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2025, pp. 912–919. doi: 10.1109/ICIMIA67127.2025.11200177.

[30] R. Appicharla, B. Gain, S. Pal, and A. Ekbal, "Beyond the Sentence: A Survey on Context-Aware Machine Translation with Large Language Models," *arXiv*, vol. 2, 2025.

[31] U. Dodda, H. Volikatla, and J. R. Vummadi, "Exploring the Role of AI-Enhanced Chatbots in Automating Recruitment Processes in Human Capital Management Systems," *Int. J. Emerg. Trends Comput. Sci. Inf. Technol.*, vol. 6, no. 3, pp. 28–36, 2025.

[32] F. Guzmán *et al.*, "The {FLORES} Evaluation Datasets for Low-

Resource Machine Translation: Nepali-- English and Sinhala--English," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 6098–6111. doi: 10.18653/v1/D19-1632.

[33] G. D. Salton, R. J. Ross, and J. D. Kelleher, "Idiom token classification using sentential distributed semantics," *54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap.*, vol. 1, pp. 194–204, 2016, doi: 10.18653/v1/p16-1019.

[34] V. Rajendran, D. Besiahgari, S. C. Patil, M. Chandrashekaraiah, and V. Challagulla, "A Multi-Agent LLM Environment for Software Design and Refactoring: A Conceptual Framework," in *SoutheastCon 2025*, 2025, pp. 488–493. doi: 10.1109/SoutheastCon56624.2025.10971563.

[35] R. Patel, "A Survey on PLC Programming Languages and Their Applications in Modern Industrial Automation Industry," *Int. J. Eng. Sci. Math.*, vol. 10, 2021.

[36] P. Cook, A. Fazly, and S. Stevenson, "The VNC-Tokens Dataset," *Proc. Lr. Work. Towar. a Shar. Task Multiword Expressions*, 2008.

[37] H. Haagsma, J. Bos, and M. Nissim, "{MAGPIE}: A Large Corpus of Potentially Idiomatic Expressions," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., Marseille, France: European Language Resources Association, May 2020, pp. 279–287.

[38] K. Tang, "PETCI: A Parallel English Translation Dataset of Chinese Idioms," 2022.

[39] T. Adewumi, R. Vadoodi, A. Tripathy, K. Nikolaidou, F. Liwicki, and M. Liwicki, "Potential Idiomatic Expression (PIE)-English: Corpus for Classes of Idioms," *2022 Lang. Resour. Eval. Conf. Lr. 2022*, pp. 689–696, 2022.

[40] D. Patel and R. Tandon, "Recent Advances in Distributed Systems: Addressing Latency, Consistency, and Scalability in Modern Applications," *Int. J. Res. Anal. Rev.*, vol. 8, no. 2, pp. 1–7, 2021.

[41] M. Popović, "chr{F}: character n-gram {F}-score for automatic {MT} evaluation," in *Proceedings of the Tenth Workshop on Statistical Machine Translation*, O. Bojar, R. Chatterjee, C. Federmann, B. Haddow, C. Hokamp, M. Huck, V. Logacheva, and P. Pecina, Eds., Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 392–395. doi: 10.18653/v1/W15-3049.

