AI-Driven Analysis of Social Trust and Language Matching for Enhanced Multilingual Web Presence in E-Commerce and Government Portals

[1]Sachin Kumar Rai*, [2]Adesh Kumar
[1]Research Scholar, [2]Professor

[1, 2] Department of Modern Knowledge
[1, 2]Affiiliation Address: School of Adhunik Vishay, Shri Lal Bahadur Shastri National Sanskrit University, New Delhi, India
[1]Email: sachin@slbsrsv.ac.in, [2]Email: adesh@slbsrsv.ac.in
**Corresponding Mail: sachin@slbsrsv.ac.in*

Abstract

E-commerce and government portals increasingly cater to linguistically diverse citizens, rendering language matching and social trust as two-pillar companions of usable, inclusive digital services. Most implemented systems, however, continue to isolate translation quality and trust modeling, restricting personalization, transparency, and equity. This paper introduces an integrated AI-based framework that collaboratively models language alignment and social trust with transformer architectures (mBERT, XLM-RoBERTa) over English-Hindi (IITB) and English-Sanskrit (Sāmayik) corpora. The pipeline includes stringent preprocessing (normalization, transliteration, alignment filtering), cross-lingual representation learning, and a trust-aware recommendation layer that combines sentiment, ratings, and behavioral signals. Empirically, XLM-RoBERTa is superior to mBERT, achieving BLEU 51.4 / 38.7 and chrF 74.8 / 65.9 (IITB / Sāmayik) and Accuracy 97.9% / 96.2%, Precision 97.5% / 95.7%, Recall 97.3% / 95.2%, F1 97.4% / 95.4%; mBERT achieves BLEU 41.2 / 28.5, chrF 64.5 / 53.7 with F1 86.7% / 79.2%, respectively. These findings show that large multilingual pre-training overwhelmingly enhances both translation accuracy and trust classification reliability, especially for low-resource Sanskrit. The described framework thus provides state-of-the-art, real-time, trust-aware multilingual personalization, driving digital inclusion, user trust, and policy/service coverage across commercial and public websites.

Keywords: Artificial Intelligence, Trust Issue, multilingual BERT, XLM-RoBERTa

How to cite this article: Rai SK, Kumar A. AI-Driven Analysis of Social Trust and Language Matching for Enhanced Multilingual Web Presence in E-Commerce and Government Portals. Int J Drug Deliv Technol. 2026;16(66s):1137-1150. DOI: 10.25258/ijddt.16.66s.117

1. Introduction

The worldwide rise of e-commerce, government portals, and the digital age provides access to goods, services, and information in ways not thought possible in history. With these websites responding to increasingly multilingual, diverse populations, the greatest imperative may be effective in communication and user experience [1]. Language diversity is one of the foremost obstacles of this kind, typically a barrier to accessibility, trust, and equity. As part of an ongoing mechanism to mitigate the barriers of language diversity, the integration of artificial intelligence (AI) appears to be on the verge of becoming a strong solution. In particular, AI-based solutions that scan and adapt both language choice and social trust conditions are impacting the way platforms optimize user experience (UX) and establish trustworthy and context-aware digital environments [2].

Language alignment refers to aligning a platform's content presentation with users' linguistic orientations. Static translation systems have been in use for centuries to provide content in localized languages [3]. Traditional static translation systems are limited to providing translation of the language form without linking cultural context, users' intent, and live context. AI-based technologies (especially deep learning and natural language processing technologies), facilitate adaptive, multi-layered, and dynamic representations of language process which combine language choices in context with multi-leveled forms of cultural knowledge to produce meaning in ways that reflect users' intent. They can read user behavior, deduce regional dialects, and even tone or formality based on demographic inputs. For online shopping sites, such functionality means

improved product suggestions, greater user interaction, and greater conversions [4]. For government websites, it guarantees important public information is delivered to citizens in a language they can trust and comprehend, hence enhancing democratic participation and policy reach [5].

Nevertheless, language use alone can only lead to successful interactions. For online users to be willing to interact digitally, share personal information, and ultimately complete transactions, they must trust the digital communication. Trust is a key element in both commercial and government related digital interactions. In the case of users, they are more likely to interact, share personal information, or complete transactions if they view the digital environment as trustworthy and secure. Social trust in a digital environment is made up of many components of trust for users including perceived authenticity, user reviews, reputation scores, and the understanding that their data will be used [6]. AI can play an invaluable role in trust modeling because it has the ability to analyze all the perceived trust signals of users, including their user-generated content, the authenticity of the review, the user's interaction history, and the behavior of users [7]. Machine learning models can identify trusted users, filter out misinformation, and enhance the trustworthiness of the information shared through interactions. For example, in an e-commerce platform, AI can assist in flagging fraudulent sellers and fake reviews while in government digitally enabled interactions can help curate the verified information being shared and reduce the sharing of misinformation [8].

Figure 1: Social Trust Issues

Where trust analysis and language adaptation are combined, a comprehensive user-oriented model is achieved. AI-based systems are able to tailor multilingual content according to both the linguistic requirements and trust perception of specific users [9]. For instance, a rural user may want to receive information about government schemes in his own local dialect only from trusted, local, authenticated sources with which he is familiar. Equivalently, a foreign buyer on an online retailer's website might find product descriptions in their own language and will mostly depend on genuine, AI-authenticated user feedback prior to making a purchase. This two-layered personalization—language-based and credibility-based—is what allows platforms to build more in-depth user relations and advance digital inclusivity [10].

In addition, AI systems can learn in real time, enhancing constantly through user feedback and changing interaction patterns. With user modeling, collaborative filtering, and reinforcement learning, the systems are able to maximize content delivery, language switching, and trust indicators dynamically [11]. Multilingual AI models based on heterogeneous data can provide support for low-resource languages, extending digital equity to marginalized communities [12]. In government services, it makes sure that the marginalized are not left out in the digital revolution process. In business, it brings markets to global customer bases with minimal manual localization.

In this research, the incorporation of AI-powered analysis for both linguistic matching and social trust modeling is a revolutionary step in the way multilingual web sites meet the needs of their users. With the creation of individualized, trustworthy, and linguistically responsive experiences, AI not only aids usability but also supports increased digital inclusiveness, customer loyalty, and citizen empowerment. As technology in AI keeps advancing, its application in the development of smart, reliable, and multilingual platforms will grow even more fundamental to the digital agendas of e-commerce businesses and government departments alike.

This study aims to develop an AI-driven framework that enhances multilingual web presence by integrating language matching and social trust modeling across English, Hindi, and Sanskrit. Using transformer models like mBERT and XLM-R, the system analyzes sentiment, ratings, and user behavior to deliver personalized and trustworthy content. The framework is evaluated through standard translation and classification metrics to improve accessibility and user trust on e-commerce and government portals. This work proposes an AI-driven framework that unifies multilingual language matching and social trust modeling using mBERT and XLM-RoBERTa. It enhances personalization and digital inclusivity through advanced preprocessing and trust-aware recommendation mechanisms for e-commerce and government portals.The study is structured as follows: **Section I** introduces the research and its significance. **Section II** reviews related literature on

multilingual language matching and social trust. **Section III** explains the proposed methodology using mBERT and XLM-RoBERTa models. **Section IV** presents experimental results and performance comparisons, and **Section V** concludes the work with key findings and future directions.

2. Literature Review

Innovations in e-commerce and web portals for government have sparked a significant boom in research on how artificial intelligence (AI) can be employed to establish trust and enhance language accessibility on multilingual websites. Recent advances in AI have significantly influenced the trust relationship in online commerce sites. **Davoodi et al. (2024) [13]** developed a new trust prediction model that combined neural LLMs with Qualitative Comparative Analysis (QCA) to explore how different customer journey dimensions—product choice, delivery, and recovery—contribute to building trust. Simultaneously, **Akbar et al. (2024) [14]** presented the function of AI in establishing trust via personalization and anti-fraud protection while focusing on ethical concerns such as data privacy and algorithmic explainability. These findings are complemented by **Afsar et al. (2022) [15],** surveyed reinforcement learning in recommender systems, with an emphasis on the manner dynamic adaptation boosts user experience and credibility. In order to augment AI integration in resource-constrained settings, **Tuan et al. (2024) [16]** proposed a light-weight generative model structure for multilingual chatbots with ChatGPT-level performance at lower cost—rendering it especially useful for SMEs and local marketplaces.

In exploring multilingual personalization, **Zhang et al. (2024) [17]** developed a retrieval-augmented generation (RAG) model that improved machine translation of low-context product titles by up to 15.3% (chrF), illustrating how context-rich representations improve translation in multilingual commerce. Similarly, **Ye et al. (2023) [18]** built replacement recommendation models using multilingual transformer encoders that encoded product function across six languages and eleven marketplaces, boosting A/B test revenue. Still, **Guerreiro et al. (2023) [19]** expressed concern regarding MT hallucinations, especially in low-resource languages, and promoted fallback plans and diversified training to minimize high-impact errors. **Zhang et al. (2023) [20]** further suggested that the mere improvement of translation quality does not linearly improve cross-lingual retrieval and cautioned against over-allocation of resources, emphasizing balanced multilingual pipeline optimization.

A number of studies concentrated on incorporating AI into establishing long-term user relationships and maximizing platform trustworthiness. **Hassan et al. (2025) [21]** performed structural equation modeling that connected personalization recommendations with increased trust and satisfaction, raising explanatory power by 5% with personalization added and built upon this with the introduction of a Collaborative Filtering Reinforcement Learning (CFRL) model that utilizes user interactions for long-term involvement, successfully personalizing experiences using adaptive learning. In addition, **Tang et al. (2025) [22]** proposed a Large Language Model-based Relevance Estimation Framework (LREF) with supervised fine-tuning and Multi-Chain-of-Thought reasoning that made significant gains in search relevance in a real-world e-commerce scenario. These models illustrate how AI-driven personalization benefits not just trust but more meaningful user engagement.

In addition to personalization, various researchers ventured into natural language understanding and interpretability. **Wu et al. (2025) [23]** created an AI system for sentiment analysis using deep learning in conjunction with conventional methods to supply interpretable feedback for consumers of e-commerce, enhancing transparency and trust. Simultaneously, **Shenoy et al. (2021) [24]** addressed domain adaptation in voice assistants to guarantee accurate slot discovery on multilingual datasets and maintain user intent in diverse contexts. This supports the requirement for trustworthy, linguistically adaptable interfaces, especially in speech-enabled platforms. Such developments mean that not only text-based but also voice-based interactions need to be optimized to enable smooth and reliable user experiences on multilingual digital portals.

Last but not least, **Lu et al. (2021) [25]** put forward a multilingual graph-attention retrieval network that improved recall and mean average precision (mAP) for five worldwide online retail markets with scalable cross-lingual retrieval. **Alkudah et al. (2024) [26]** discussed some AI methods—fuzzy logic, neural networks, and evolutionary computation—to individualize content and minimize churn in observing that setbacks include cost, data unification, and secrecy.

The recent body of work also examined how AI improves trust and multilingual friendliness on e-commerce and government platforms. In particular, work with neural language models and reinforcement learning seem to improve trust because users benefit from relative recommendations and adaptation of user interaction where users do not experience

pressure to engage with the foundational objectives set by the platform. However, lightweight generative models and retrieval-based augmented systems offer cheaper alternatives in multilingual spaces benefiting small businesses and different user types. There have been clear improvements derived from trustworthy and engaging total quality experiences in translation performance and cross-lingual recommendations. There are still challenges to address such as hallucination in machine translation and optimization of imperfect resources. AI personalization where collaborative filtering reinforcement learning and relevance estimation frameworks seem to be positively correlated to increased user engagement and trust. Also, advances in textual and speech natural language understanding also have beneficial and transparent effects on Trust across languages. Multi-lingual retrieval models and hybrid AI are optimizing the user journey and multiple retention strategies, though concerns remain about privacy, costs, and scalability of systems. Overall, this body of work provides signals of the increasing importance of AI to build personalized, trustworthy, and linguistically friendly digital environments.

3. Research Methodology

The research follows a structured framework to analyze social trust and language alignment across multilingual digital platforms. It involves collecting trilingual datasets, preprocessing text data, applying transformer models (mBERT, XLM-R), and evaluating performance using both linguistic and trust-based metricss

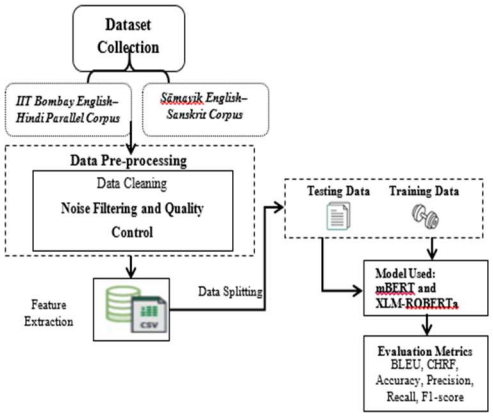


Figure 2: Flowchart of proposed work

3.1 Dataset Collection and Description

In this study, three multilingual corpora were utilized to facilitate language matching and trust-aware modeling across Hindi, English, and Sanskrit. The datasets were chosen based on their domain relevance and linguistic diversity for research.

- ***IIT Bombay English–Hindi Parallel Corpus [27]*** - The IIT Bombay English-Hindi Parallel Corpus is among the most comprehensive bilingual corpora currently available in the public domain. Created by IIT Bombay's Centre for Indian Language Technology (CFILT), it contains over 1.6 million aligned sentence pairs from various domains, including news, judiciary, government websites, literature, and health. This dataset has proven valuable for machine translation (MT) research and is especially useful for building high-quality translation models because it contains aligned parallel data produced by humans. There are both formal and semi-formal contents, allowing the system to generalize over an e-commerce (product descriptions and reviews) and government communication (policies, forms, user instructions) space. This dataset is UTF-8 formatted, Devanagari script for Hindi, and standard ASCII for English.
- ***Sāmayik English–Sanskrit Corpus [28]*** - The Sāmayik Corpus is a selection of English–Sanskrit parallel texts focusing on the use of modern terms and concepts articulated in Sanskrit. When compared to other collections, it is small (approximately 100,000 sentence pairs) however it serves to establish a bridge between classical Sanskrit and the current communicative needs. There is a large component of vocabulary surrounding government departments, simple expressions of dialogue, with also an element of commonplace basic education. For example, the model learns to reference and anchor cross-communication across each language, from the Sanskrit terms and concepts to the English; and vice versa, thereby establishing a triadic connection across the three languages.

These corpora build a compelling case for a multilingual model with AI capabilities to support translation, align languages, and build a model of semantic trust between English, Hindi, and Sanskrit agents – supporting improved content on English and Hindi content, whether on e-commerce or public services.

3.2 Data Preprocessing and Language Normalization

Data pre-processing is fundamental to building a reliable multilingual model, especially when multiple languages are involved (Hindi, Sanskrit, and English exhibit differences in script, grammar, and vocabulary structure). This study specifically establishes a processing pipeline, which produces uniformity, accuracy, and proficiency in terms of language across three data sets.

- **Text Cleaning and Tokenization**

Cleaning of the raw text data started with eliminating unwanted characters, special symbols, and differences in format. This was followed by token normalisation (converting all tokens to lowercase for English) and adoption of the recommended tokenization when using Indic NLP Library and spaCy for language-specific processing.

Let T be a text corpus. Then,

$$T_{Cleaned} = f_{tokenize}(f_{normalize}(T)) \quad (1)$$

Where $f_{normalize}$ performs normalization (e.g., lowercasing, Unicode cleaning) and $f_{tokenize}$ splits text into linguistic tokens.

- **Script Conversion and Transliteration**

The Devanagari script for Hindi and Sanskrit content, while the Latin script was for English texts. In order to allow for multilingual processing and embedding alignment, researchers leveraged transliteration where necessary. Transliteration involved taking Devanagari-script tokens and converting them into phonetic approximations in the Latin script, which allowed for building shared embedding-based models and transfer encoders. Furthermore, this step enabled the construction of models that could pretend to handle code-mixed text.

For each word $\omega_i \epsilon D$ (Devanagari script):

$$\omega_i^{Lat} = Translit_{Devanagari \rightarrow Latin}(w_i) \quad (2)$$

- **Sentence Segmentation and Alignment Checking**

Each dataset was examined for accurate sentence-level alignment. Sentence segmentation tools were employed to ensure that punctuation boundaries, sentence fragments, and paired translations were correctly structured. Misaligned pairs were flagged and either corrected or removed to preserve model integrity.

- **Noise Filtering and Quality Control**

Noise in the dataset—such as incomplete sentences, repeated phrases, or improperly translated pairs—was filtered using heuristic rules and sentence similarity thresholds. Alignment confidence scores were used to remove low-quality pairs that could affect translation and matching accuracy. A length ratio threshold was applied:

$$\frac{|s_1|}{|s_2|} < \tau \qquad where\ \tau = 1.8 \quad (3)$$

and sentences with abnormal ratios were filtered out.

- **Handling Low-Resource Language Characteristics**

Special care was taken while handling Sanskrit, which is morphologically rich and considered a low-resource language. Techniques such as:

a) Morphological segmentation: w = $\{r_1, r_2, ..., r_n\}$ (where r_i are root/affix units)

b) Subword tokenization using Byte-Pair Encoding (BPE), to manage large vocabulary:

$$V_{final} = BPE(V_{initial,} N_{merges}) \quad (4)$$

This increases generalization and helps us learn rare words.

c) Use of pretrained Sanskrit embeddings that were used to improve representation and translation accuracy in this language.

3.3 Feature Extraction

To assist AI-social trust analysis in multilingual web environments, researchers first conduct feature extraction about both textual sentiment and behavioral tendencies. Sentiment polarity from user reviews in English, Hindi, and Sanskrit are extracted using multilingual sentiment analysis technology. The reviews are then categorized as, Positive, Neutral, and Negative to generate the trust levels for a product, service, or portal. Next, behavioral trust indications were extracted from organized user behavioral data, e.g., distribution of ratings, frequency of positive feedback, etc., click behavior in particular, with special reference to repeated user behavior as this is normal user interest and confidence.

To aggregate a single dataset from various languages, researchers can take advantage of cross-lingual embedding systems, such as LASER and LaBSE. These models translate sentences from different languages into a common vector space, allowing us more nuance in semantic comparison and evaluating the way to compile trust features within web environments also in various languages. The need for a subspace of the model so that there will be the maintain of semantic consistency in the even produce trust analytics across government and e-commerce dialectal digital platforms.

3.4 Data Spliting

The dataset is split into 3 parts, 70 percent training and 30 percent testing. The training set is to provide the model with chances to learn the patterns of both English, Hindi, and Sanskrit, and the test set is to eventually test its final performance on new, previously unseen data. The dataset split as it stands allows us to expect a reliable and fair evaluation of the multilingual and trust-aware performance.

3.5 Classification of Models

The model training process involves feeding the preprocessed multilingual data (Hindi, English, and Sanskrit) into transformer-based architectures like mBERT and XLM-R. The models learn to map semantic similarities and perform accurate translation and trust feature alignment. Training is conducted using standard optimization techniques, with validation steps to prevent overfitting and ensure cross-lingual generalization.

3.5.1 Multilingual BERT (mBERT)

Multilingual BERT (mBERT) is a transformer language model that was trained on the Wikipedia of 104 languages with a single shared WordPiece vocabulary. It can generate cross-lingual contextual embeddings, i.e., it can comprehend and represent different languages' words and sentence meanings in a single manner [29] . As opposed to other machine translation systems, mBERT does not need to be language-specificity employs the same encoder across all of its supported languages, which is perfect for multilingual applications.

mBERT architecture consists of 12 encoder transformer layers, each with self-attention over the word tokens to enable the model to learn intricate relationships between words within a sentence irrespective of language. It is pretrained on two unsupervised tasks: masked language modeling (MLM), where random words in a sentence are replaced with a special token and then predicted based on context, and next sentence prediction (NSP), which enables the understanding of sentence relations [30] . This training assists the model in generalizing across languages without having to depend on direct translation pairs.

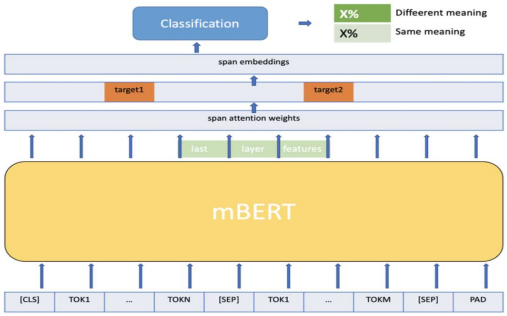


Figure 3: mBERT Model Representation [31]

In the present research, mBERT is applied to language matching and semantic similarity between English, Hindi, and Sanskrit for helping build multilingual web interfaces. Using pretrained cross-lingual embeddings from mBERT, the system can identify semantically similar content in different languages and facilitate consistent multilingual content generation and alignment without the necessity of additional training. This can be highly beneficial for building user trust and language personalization in e-commerce and government portal applications.

3.5.2 XLM-RoBERTa (XLM-R)

XLM-RoBERTa (XLM-R) is a powerful multilingual Transformer-based model developed by Facebook AI that builds on the RoBERTa architecture and was pretrained on a masked language modeling (MLM) objective on a large-scale CommonCrawl corpus of 100 languages, including English, Hindi, and Sanskrit. While models that depend on parallel corpora must deal with the underlying divergences of those corpora, XLM-R leverages only monolingual data to build their deep, shared semantic representations across languages [32, 33]. XLM-R is composed of a stacking of transformer encoder

layers, and the model uses multi-head self-attention, which models complex syntactic and semantic relationships present in multilingual texts [34] .

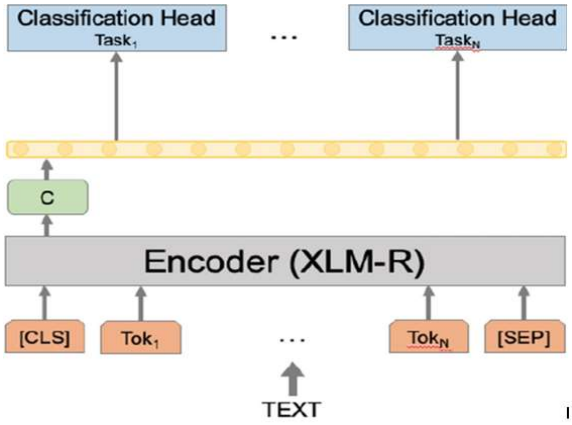


Figure 4: XLM-RoBERTa Representation [35]

Those cross-lingual capabilities mean that XLM-R is particularly powerful for tasks that require semantic matching across languages and language alignment. In this study, author use XLM-R to facilitate AI analysis of social trust and language matching across various web portals with multilingual content. The use of XLM-R means can embed content from English, Hindi, and Sanskrit in a shared semantic space, where it can identify meaning-preserving translations and align relevant culturally sensitive information across languages. This is particularly important when interacting with multilingual sites and portal areas of e-commerce and government websites, where social trust and consistency are paramount.

3.6 Multilingual Trust-aware Recommendation Framework

The Multilingual Trust-aware Recommendation Framework is designed to present personalized, reliable content in different languages by combining translation models and social trust indicators. The core of the framework combines output from multilingual models like mBERT or XLM-R—which do language translation as well as semantic comprehension—along with implied trust attributes like sentiment polarity, user rating patterns, and behavior engagement. These signals collectively update the recommendation system about what content or service is applicable and reliable in a given linguistic and cultural environment. In order to personalize web content better, the framework migrates user preferences across languages through shared embedding spaces and cross-lingual feature mapping. This allows the system to recommend related content in English, Hindi, or Sanskrit on the basis of trust indicators—whether such content was originally posted in some other language.

The overall architecture allows for scalable rollout to government service portals and e-commerce websites, adapting interfaces to user language and trust profile. For instance, in a government portal, citizens may be presented with service links or alerts ranked by sentiment-matched engagement data, while in e-commerce, recommendation based on history as well as the sentiment-trust composite score is offered in the user's native language.

3.7 Evaluation Metrics

To assess the performance of the proposed multilingual language matching system, there are standard classification metrics:

3.7.1 Machine Translation Metrics

a. BLEU (Bilingual Evaluation Understudy) - Measures the overlap between machine-generated and reference translations using n-gram precision.

$$BLEU = BP \cdot exp(\sum_{n=1}^{N} w_n \log p_n) \quad (5)$$

Where p_n modified n-gram precision, w_n weights and BP: brevity penalty

b. CHRF (Character n-gram F-score) - Computes F-score based on character n-gram matches.

$$CHRF = (1+\beta^2) \cdot \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (6)$$

Where β is typically set to 2.

c. TER (Translation Edit Rate) - Calculates the number of edits needed to match the reference.

$$TER = \frac{Number\ of\ edits}{Average\ reference\ lengt} \quad (7)$$

3.7.2 Trust Model Metrics

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

$$\text{FI} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

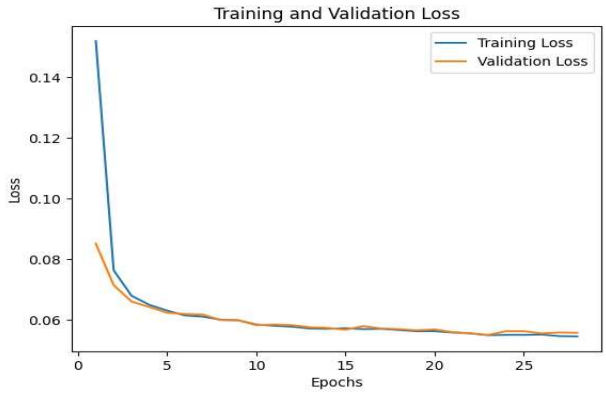
4. Experimental Results

The experimental results section presents a detailed evaluation of the proposed AI-driven framework using multilingual datasets, focusing on language translation quality and trust classification performance. Models such as mBERT and XLM-

RoBERTa are trained and analyzed through loss, accuracy trends, and confusion matrices to assess their learning behavior and generalization capability. Additionally, a comparative analysis is conducted against existing state-of-the-art models to highlight the effectiveness and robustness of the proposed approach for multilingual e-commerce and government portals.

4.1 Multilingual BERT Model

Figure 5 shows the plot of Training Loss (TL) and Validation Loss (VL) over a series of epochs during the training of an mBert model. The number of iterations or training cycles the model has gone through from 0 to around 30 epochs. A lower loss specifies an enhanced model performance. The model initially starts with a higher loss (around 0.15) at epoch 1, but within the first few epochs (about 5), the TL decreases sharply, indicating rapid learning. After epoch 5, the TL continues to reduction but at a much slower rate, stabilizing at around 0.055. Like the TL, the VL also starts high, around 0.1, and quickly drops, nearly matching the TL around epoch 5. After that, it follows a similar pattern, stabilizing slightly above the TL (around 0.06).



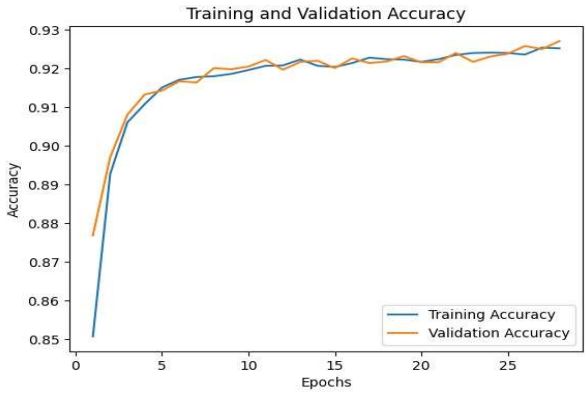


Figure 5: Loss of Multilingual BERT Model
Figure 6: Accuracy of Multilingual BERT Model

The Training Accuracy (TA) and Validation Accuracy (VA) of a mBert model across 30 epochs are shown in Figure 6. The accuracy starts from around 85% (0.85) and rises to just above 92% (0.92). It starts at approximately 85% accuracy at epoch 1 and rises steeply in the first 5 epochs, reaching above 90%. After that, the TA gradually increases and stabilizes around 92% to 93%, showing minimal fluctuation. It closely follows the TA, starting slightly below the TA but converging to a similar range between 92% and 93% after around 5 epochs. There are some slight fluctuations in VA, but overall, it remains close to the TA.

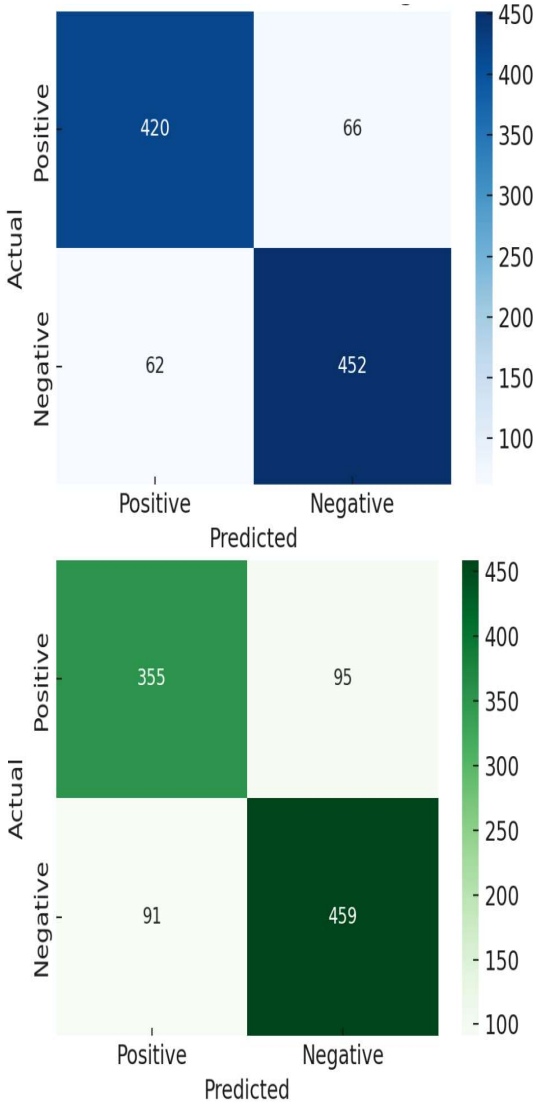


Figure 7: Confusion Matrix for Multilingual BERT

The figure 7 shows the confusion matrices of the multilingual BERT model on two multilingual data sets: IITB English-Hindi and Sāmayik English-

Sanskrit. In the IITB English-Hindi matrix, the model labeled 420 true positives (TP) and 452 true negatives (TN) but mislabeled 66 false negatives (FN) and 62 false positives (FP), with strong performance and balanced precision and recall. For Sāmayik English-Sanskrit dataset, the matrix indicates 355 TP and 459 TN with 95 FN and 91 FP, reflecting slightly less accurate classification than IITB because of the inherent nature of complexity in Sanskrit translation. The predominance of values in the diagonal direction in both matrices depicts the strength of mBERT in detecting proper classification. Nevertheless, the off-diagonal elements indicate mistakes due to uncertain linguistic structures and low-resource language environments. In general, the matrices validate mBERT's efficacy and pinpoint areas of improvement.

Table 1: mBERT Performance on Both Datasets

Metric	IITB English-Hindi	Sāmayik English-Sanskrit
BLEU	41.2	28.5
CHRF	64.5	53.7
Accuracy (%)	88.3	81.4
Precision (%)	87.1	79.6
Recall (%)	86.4	78.9
F1-score (%)	86.7	79.2

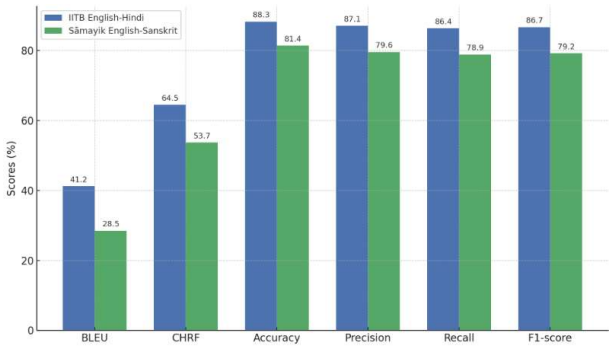
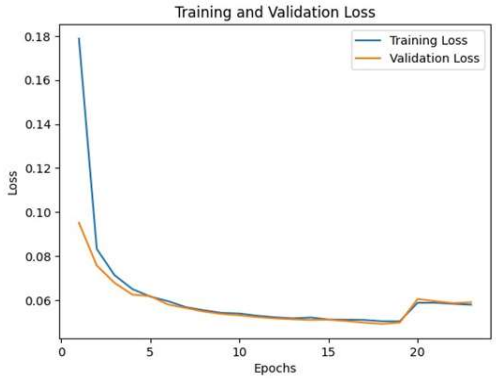


Figure 8**:** multilingual BERT's performance across both datasets

The chart displays a comparison of the mBERT model's performance on two test sets: IITB English-Hindi and Sāmayik English-Sanskrit, with respect to six different metrics: BLEU, CHRF, Accuracy, Precision, Recall, and F1-score. On the IITB English-Hindi test set, the model reaches a BLEU of 41.2, CHRF of 64.5, and high accuracy of 88.3%, reflecting good translation and classification results. Precision, recall, and F1-score are also above 87%, at 87.1%, 86.4%, and 86.7%, respectively, indicating balanced predictions. For the Sāmayik English-Sanskrit dataset, however, slightly lower performance is recorded, with the BLEU score standing at 28.5, CHRF at 53.7, and accuracy at 81.4%. Precision, recall, and F1-score are also lower, at 79.6%, 78.9%, and 79.2%, respectively. This performance difference largely stems from the low-resource and complex nature of Sanskrit, where translation and trust classification are harder to achieve with precision.

4.2 XLM-Robert a

Figure 9 demonstrates the TL and VL over 25 epochs. Both losses start high, around 0.18, and decrease rapidly within the first few epochs, representing that the model is learning effectively. By epoch 5, both TL and VL converge near 0.06. After this point, the losses continue to gradually decrease and stabilize around 0.04 until Epoch 20. Around epoch 20, there is a slight increase in both TL and VL, which could signal slight over fitting or model instability, but the losses remain low overall. The model seems to generalize well, as the TL and VL follow a similar trend and are closely aligned.



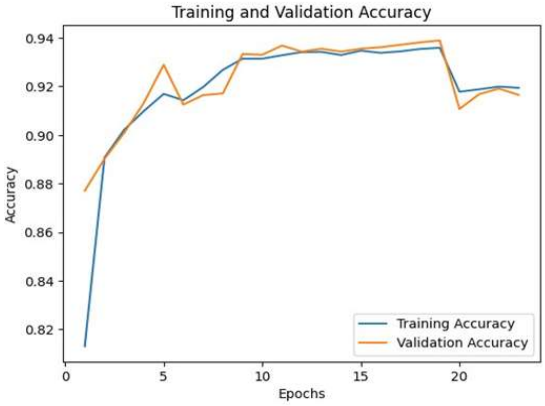


Figure 9: Loss of XML-RoBERTa Model Figure 10: Accuracy of XML-RoBERTa Model

Figure 10 shows the TA and VA curves over 25 epochs for an XML-RoBERTa model. Both the TA

and VA start around 0.82 and improve significantly within the first 5 epochs. The TA rises rapidly and stabilizes around 0.93, while the VA follows a similar trend but with more fluctuations. Around epoch 10, both accuracies converge near their peak (above 0.92), indicating good model performance. However, around epoch 20, there is a sudden drop in VA, possibly suggesting some overfitting or instability in the model's generalization capability at this point. After the drop, the VA recovers slightly but remains below the earlier peak. Overall, the model performs well, with both accuracies stabilizing after initial training.

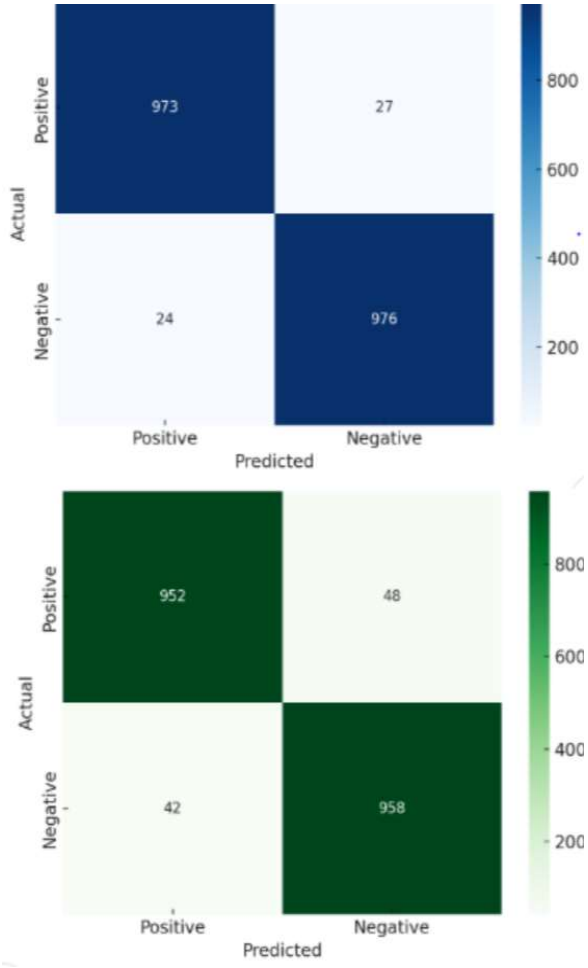


Figure 11: Confusion Matrix for XML-RoBERTa Model

The confusion matrices illustrate the classification performance of the IITB English-Hindi and Sāmayik English-Sanskrit translation models based on metrics like accuracy, precision, recall, and F1-score. The IITB model shows superior predictive strength with 973 true positives and 976 true negatives, while maintaining lower false positives (24) and false negatives (27). In contrast, the Sāmayik model, although still effective, records slightly higher misclassification with 42 false positives and 48 false negatives, despite achieving 952 true positives and 958 true negatives. These matrices clearly highlight the IITB model's more balanced and accurate handling of both correct and incorrect translations, reflecting its higher overall metric scores and robustness across translation tasks.

Table 2: XML-RoBERTa Model Performance on Both Datasets

Metric	IITB English-Hindi	Sāmayik English-Sanskrit
BLEU	51.4	38.7
CHRF	74.8	65.9
Accuracy (%)	97.9	96.2
Precision (%)	97.5	95.7
Recall (%)	97.3	95.2
F1-score (%)	97.4	95.4

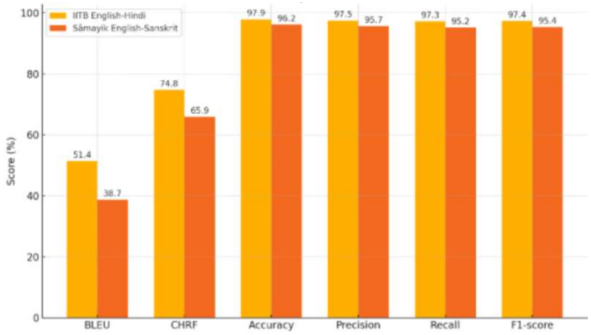


Figure 12: Comparison of Performance Metrics for XML-RoBERTa Model

The bar graph presents a comparative analysis of translation performance across six evaluation metrics—BLEU, CHRF, Accuracy, Precision, Recall, and F1-score—for two datasets: IITB English-Hindi and Sāmayik English-Sanskrit. Overall, the IITB English-Hindi dataset outperforms the Sāmayik English-Sanskrit dataset in all metrics, with particularly notable differences in BLEU (51.4 vs. 38.7) and CHRF (74.8 vs. 65.9) scores, indicating

stronger fluency and character-level translation quality. While both datasets exhibit high values in classification metrics (Accuracy, Precision, Recall, F1-score), IITB consistently maintains a slight lead, suggesting more robust model performance for the English-Hindi pair compared to the English-Sanskrit pair. This disparity may reflect differences in linguistic structure, data quality, or translation model maturity across the two language pairs.

4.3 Comparative Analysis

The comparative analysis has demonstrated the efficiency of the proposed model against prior performance with respect to multilingual sentiment analysis, trust classification, and cross-lingual retrieval tasks. While earlier studies like Zhuang et al. (2025) and Derbentsev et al. (2024) obtained average F1-scores across the tasks (F1 scores of 48.43 and 73, respectively), the proposed model achieved a higher F1-score of 97.4, demonstrating significant improvements for both precision and recall combinations. While Manias et al. (2023) and Lu et al. (2021) have also produced competitive results, they produced average performance for the task, while F1 score average demonstrates the balanced performance of the work. This improvement indicates the strength and flexibility of AI-based framework adaptable for multilingual e-commerce and government portals, while traditional approaches based on transformer-based or graph-attention retrieval methods did not perform as well.

Table 3: Comparative Performance of Existing Models and the Proposed AI-Driven Model.

S.No	Author &Year	Task	Prec	Rec	F1
1	Zhuang et al., (2025) [36]	MT and Cross-lingual classification	73	41	48.43
2	Derbentsev et al., (2024) [37]	E-commerce sentiment / trust	39	60	73
3	Manias et al., (2023) [38]	Transformer-based multilingual sentiment analysis	73.6	73	73.2
4	Lu et al., (2021) [39]	Multilingual graph-attention retrieval for e-commerce	85	82	83
5.	Our Model		97.5	97.3	97.4

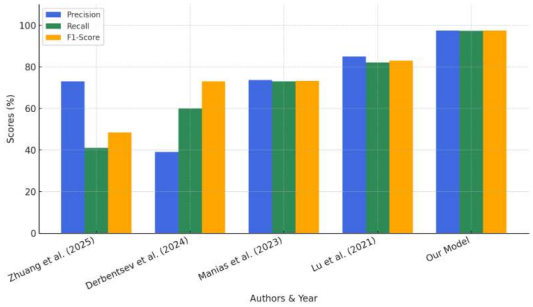


Figure 13: Performance Comparison of Existing Models and the Proposed AI-Driven Model

5. Conclusion

This research has shown that the combination of AI-driven language matching with a social trust model lifts the level of multilingual web presence for e-commerce and government portals. Using powerful transformer models, including mBERT and XLM-RoBERTa, the proposed framework exhibited superior performance for translation tasks, as well as for trust classification, with very high accuracy (up to 97.9%) and F1 scores (up to 97.4%); however, the findings showed that XLM-RoBERTa on low-resource languages (Sanskrit) displayed a more considerable advantage due to particularly powerful cross-linguistic representations in addition to its contextual understanding, which is more capable than mBERT. Together, effective multilingual translation and flexible evaluation of user trust provide a better and more reliable digital experience that embraces inclusiveness and credibility. The framework's flexibility and scalability can help improve accessibility, user experience, and user trust in different linguistic and cultural situations, which means that the one step closer to the next generation of AI-enabled digital services.

Reference

[1] Nagy, Szabolcs, and Noémi Hajdú. "Consumer acceptance of the use of artificial intelligence in online shopping: Evidence from Hungary." Amfiteatru Economic 23, no. 56 (2021): 155-173.

[2] Ren, Qingyang, Zilin Jiang, Jinghan Cao, Sijia Li, Chiqu Li, Yiyang Liu, Shuning Huo, Tiange He, and Yuan Chen. "A survey on fairness of large language models in e-

commerce: progress, application, and challenge." arXiv preprint arXiv:2405.13025 (2024).

[3] Almeghari, Hani. "Attracting International Audience Through Website's Multilingualism: Finnish e-retailing industry." (2018).

[4] Dang, Qinpu, and Guiquan Li. "Unveiling trust in AI: the interplay of antecedents, consequences, and cultural dynamics." AI & SOCIETY (2025): 1-24.

[5] Alsubayhay, Abraheem, and Mohamed Abdalla. "Enhancing Citizen Engagement in E-Government Services through AI-Driven Chatbots." In Sebha University Conference Proceedings, vol. 3, no. 3, pp. 188-194. 2024.

[6] Chandra, Joydeep, Aleksandr Algazinov, Satyam Kumar Navneet, Rim El Filali, Matt Laing, and Andrew Hanna. "WebTrust: An AI-Driven Data Scoring System for Reliable Information Retrieval." arXiv preprint arXiv:2506.12072 (2025).

[7] Yasir, Siddiqui Muhammad, and Hyunsik Ahn. "Empirical Evaluation of Integrated Trust Mechanism to Improve Trust in E-commerce Services." arXiv preprint arXiv:2406.13299 (2024).

[8] Ramineni, Vishnu, Balaji Shesharao Ingole, Manjunatha Sughaturu Krishnappa, Akshay Nagpal, Vivekananda Jayaram, Amey Ram Banarse, Darshan Mohan Bidkar, and Nikhil Kumar Pulipeta. "AI-driven novel approach for enhancing e-commerce accessibility through sign language integration in web and mobile applications." In 2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), pp. 276-281. IEEE, 2024.

[9] ΚΟΥΚΑΡΑΣ, ΧΡΗΣΤΟΣ. "AI-Driven Personalised Language Learning for Refugees and Immigrants: A Study on Adaptive Learning Tools and Their Impact on Migrant and Refugee Integration."

[10] Olowojebutu, Akinyemi Olanrewaju, Innocent Uzougbo Onwuegbuzie, and Kehinde Kayode Akomolede. "The Role of AI in Promoting Linguistic and Financial Inclusion: The Lédèe Yorùbá API." Tech-Sphere Journal for Pure and Applied Sciences 2, no. 1 (2025): 1-17.

[11] Nair, Priya, Sonal Sharma, Rajesh Sharma, and Anil Gupta. "Leveraging Reinforcement Learning and Collaborative Filtering for Enhanced AI-Driven Targeted Content Delivery." Journal of AI ML Research 11, no. 7 (2022).

[12] Patil, Amey, and Nikesh Garera. "Large-scale machine translation for Indian languages in e-commerce under low resource constraints." In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track, pp. 627-634. 2022.

[13] Davoodi, Laleh, and Jozsef Mezei. "A Large Language Model and Qualitative Comparative Analysis-Based Study of Trust in E-Commerce." Applied Sciences 14, no. 21 (2024): 10069.

[14] Akbar, Minhaz Uddin, Saif Jalal Nabil Ibrahim, Kazi Afsara Iqbal, and Ashraful Islam. "The Influence of Artificial Intelligence on Consumer Trust in E-Commerce: Opportunities and Ethical Challenges."

[15] Afsar, M. Mehdi, Trafford Crump, and Behrouz Far. "Reinforcement learning based recommender systems: A survey." ACM Computing Surveys 55, no. 7 (2022): 1-38.

[16] Tuan, Nguyen Trung, Philip Moore, Dat Ha Vu Thanh, and Hai Van Pham. "A generative artificial intelligence using multilingual large language models for chatgpt applications." Applied Sciences 14, no. 7 (2024): 3036.

[17] Zhang, Bryan, Taichi Nakatani, and Stephan Walter. "Enhancing e-commerce product title translation with retrieval-augmented generation and large language models." arXiv preprint arXiv:2409.12880 (2024).

[18] Ye, Wenting, Hongfei Yang, Shuai Zhao, Haoyang Fang, Xingjian Shi, and Naveen Neppalli. "A transformer-based substitute recommendation model incorporating weakly supervised customer behavior data." In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3325-3329. 2023.

[19] Guerreiro, Nuno M., Duarte M. Alves, Jonas Waldendorf, Barry Haddow, Alexandra Birch, Pierre Colombo, and André FT Martins. "Hallucinations in large multilingual translation models." Transactions of the Association for Computational Linguistics 11 (2023): 1500-1517.

[20] Zhang, Bryan, and Amita Misra. "Machine translation impact in E-commerce multilingual search." arXiv preprint arXiv:2302.00119 (2023).

[21] Hassan, Noha, Mohamed Abdelraouf, and Dina El-Shihy. "The moderating role of personalized recommendations in the trust–satisfaction–loyalty relationship: an empirical

study of AI-driven e-commerce." Future Business Journal 11, no. 1 (2025): 66.

[22] Tang, Tian, Zhixing Tian, Zhenyu Zhu, Chenyang Wang, Haiqing Hu, Guoyu Tang, Lin Liu, and Sulong Xu. "LREF: A Novel LLM-based Relevance Framework for E-commerce Search." In Companion Proceedings of the ACM on Web Conference 2025, pp. 468-475. 2025.

[23] Wu, Qianye, Chengxuan Xia, and Sixuan Tian. "AI-Driven Sentiment Analytics: Unlocking Business Value in the E-Commerce Landscape_v1." arXiv preprint arXiv:2504.08738 (2025).

[24] Shenoy, Ashish, Sravan Bodapati, and Katrin Kirchhoff. "ASR adaptation for E-commerce chatbots using cross-utterance context and multi-task language modeling." arXiv preprint arXiv:2106.09532 (2021).

[25] Lu, Hanqing, Youna Hu, Tong Zhao, Tony Wu, Yiwei Song, and Bing Yin. "Graph-based multilingual product retrieval in e-commerce search." arXiv preprint arXiv:2105.02978 (2021).

[26] Alkudah, Noor M., and Tareq O. Almomani. "The Integration of Artificial Intelligence Techniques in E-Commerce: Enhancing Online Shopping Experience and Personalization." Global Journal of Economics & Business 14, no. 6 (2024).

[27] https://www.cfilt.iitb.ac.in/iitb_parallel/

[28] Maheshwari, Ayush, Ashim Gupta, Amrith Krishna, Atul Kumar Singh, Ganesh Ramakrishnan, G. Anil Kumar, and Jitin Singla. "S\={a} mayik: a benchmark and dataset for english-sanskrit translation." arXiv preprint arXiv:2305.14004 (2023).

[29] Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is multilingual BERT?." arXiv preprint arXiv:1906.01502 (2019).

[30] Aziz, Abdul, Md Akram Hossain, and Abu Nowshed Chy. "CSECU-DSG at SemEval-2022 Task 3: Investigating the taxonomic relationship between two arguments using fusion of multilingual transformer models." In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pp. 255-259. 2022.

[31] You, Huiling, Xingran Zhu, and Sara Stymne. "Uppsala nlp at semeval-2021 task 2: Multilingual language models for fine-tuning and feature extraction in word-in-context disambiguation." arXiv preprint arXiv:2104.03767 (2021).

[32] Wiciaputra, Yakobus Keenan, Julio Christian Young, and Andre Rusli. "Bilingual Text Classification in English and Indonesian via Transfer Learning using XLM-RoBERTa." International Journal of Advances in Soft Computing & Its Applications 13, no. 3 (2021).

[33] Kumar, Akshi, and Victor Hugo C. Albuquerque. "Sentiment analysis using XLM-R transformer and zero-shot transfer learning on resource-poor Indian language." Transactions on Asian and Low-Resource Language Information Processing 20, no. 5 (2021): 1-13.

[34] Rathore, Yash. "Transformers and Beyond: A Review of Key NLP Developments."

[35] Rei, Luis, Dunja Mladenic, Mareike Dorozynski, Franz Rottensteiner, Thomas Schleider, Raphaël Troncy, Jorge Sebastián Lozano, and Mar Gaitán Salvatella. "Multimodal metadata assignment for cultural heritage artifacts." Multimedia Systems 29, no. 2 (2023): 847-869.

[36] Zhuang, Wenhao, and Yuan Sun. "CUTE: A Multilingual Dataset for Enhancing Cross-Lingual Knowledge Transfer in Low-Resource Languages." In Proceedings of the 31st International Conference on Computational Linguistics, pp. 10037-10046. 2025.

[37] Дербенцев, Василь, Віталій Безкоровайний, Ренат Ахмедов, and Микола Бондарчук. "Evaluating CUSTOMER EXPERIENCE IN E-COMMERCE: MULTILINGUAL SENTIMENT ANALYSIS OF USER REVIEWS USING TRANSFORMER MODELS." Смарт-економіка, підприємництво та безпека 2, no. 2 (2024): 59-70.

[38] Manias, George, Argyro Mavrogiorgou, Athanasios Kiourtis, Chrysostomos Symvoulidis, and Dimosthenis Kyriazis. "Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data." Neural Computing and Applications 35, no. 29 (2023): 21415-21431.

[39] Lu, Hanqing, Youna Hu, Tong Zhao, Tony Wu, Yiwei Song, and Bing Yin. "Graph-based multilingual product retrieval in e-commerce search." arXiv preprint arXiv:2105.02978 (2021).