International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com | A Peer-Reviewed, Refereed and Bimonthly Journal|

|| Volume 5, Issue 6, November – December 2022 ||

DOI: 10.15680/IJCTECE.2022.0506006

Human-Centered AI in Safety-Critical Systems: From Educational Simulations to NLP and Privacy-Aware Redundancy

Amirul Bin Fauzi Mei

Machine Learning Engineer, Johor Bahru, Malaysia

ABSTRACT: Large-scale language understanding increasingly requires systems that operate across modalities, user populations, and deployment environments. This paper presents an integrated approach that combines Natural Language Processing (NLP)–based sentiment analysis of social media with automated sign language interpretation to deliver inclusive, scalable language intelligence. For social media sentiment, we design a pipeline that handles noisy, short-form text (tweets, posts, comments) by combining robust preprocessing (emoji and hashtag normalization, slang lexicons), contextualized language encoders (pretrained transformer models fine-tuned on in-domain data), and multimodal signals (attached images, timestamps, and user metadata) to improve polarity and fine-grained emotion detection. For sign language interpretation, we develop an end-to-end visual–linguistic system that uses spatiotemporal visual encoders (frame-level CNNs + keypoint/pose features), transformer-based sequence models for alignment and translation, and language-model-informed decoders to produce grammatical target-language text. Crucially, we propose a shared engineering and evaluation framework that addresses dataset curation, privacy-aware collection, cross-modal alignment, domain adaptation, and fairness auditing. The integrated system leverages multimodal pretraining, contrastive objectives to align vision and text embeddings, and data-augmentation strategies including synthetic sign generation and back-translation. Experimental protocols describe benchmarking on (1) social media sentiment datasets with manual and weak labels and (2) signer-independent continuous sign translation corpora, with evaluation using accuracy/F1 and BLEU/ROUGE/human intelligibility respectively. We discuss deployment patterns for cloud and edge inference to meet latency and accessibility requirements and describe mitigation strategies for bias amplification and privacy leakage. Results from ablation studies indicate that (a) multimodal fusion improves social-media sentiment F1 on multimodal posts by a measurable margin versus text-only baselines, and (b) pose-informed transformer decoders significantly reduce gloss WER and increase translation fluency for sign interpretation. We conclude with recommended practices for dataset governance, community involvement in sign-language data collection, and technical directions to scale inclusive language understanding systems responsibly.

KEYWORDS: Multimodal learning; social media sentiment analysis; sign language interpretation; transformers; pose estimation; contrastive pretraining; domain adaptation; fairness; privacy; multimodal evaluation.

I. INTRODUCTION

Language understanding at scale must contend with two linked practical challenges: the heterogeneity of user-generated content on social media (short texts, emoji, images, slang) and the accessibility gap for deaf and hard-of-hearing communities who use sign languages that differ linguistically from spoken languages. Addressing both problems in a unified research and engineering agenda yields benefits beyond convenience — it improves inclusivity, broadens user coverage, and fosters multimodal robustness that often generalizes better than unimodal systems.

Social media sentiment analysis faces noisy inputs, domain shifts across platforms, and subtle phenomena such as sarcasm and context-dependent affect. Modern contextual encoders (pretrained transformers) have markedly improved baseline performance, but they still misinterpret multimodal cues (images, emojis) and demographic context. Combining textual models with auxiliary visual and metadata signals—using cross-modal attention or late fusion—reduces ambiguity and increases sensitivity to expression style.

Sign language interpretation (continuous recognition and translation) presents a distinct set of challenges: spatiotemporal complexity, high signer variability, lack of standardized large-scale corpora for many sign languages, and the need to model both manual (hand) and non-manual (facial, body) signals. Recent work with pose estimation, CNN feature extractors, and sequence-to-sequence transformers has advanced the field toward intelligible, large-vocabulary sign-to-text translation, but real-world deployment demands signer-independence, dialectal coverage, and privacy-respecting collection practices.

DOI: 10.15680/IJCTECE.2022.0506006

This paper proposes a combined framework that advances both social media sentiment analysis and automated sign language interpretation under a shared set of methodological principles: multimodal representation learning, temporal alignment, domain adaptation, fairness and privacy safeguards, and deployment-aware engineering (edge/cloud hybrid inference). By sharing lessons across these domains—e.g., contrastive multimodal alignment used for image–text grounding being adapted to align pose trajectories with textual glosses—we aim to accelerate robust, inclusive language understanding that works at scale.

II. LITERATURE REVIEW

Multimodal approaches to language tasks have gained traction as researchers realize the limits of text-only models on socially generated content. Surveys of multimodal machine learning categorize fusion strategies (early, late, hybrid), alignment challenges, and tasks such as sentiment recognition, visual question answering, and multimodal translation. For social media sentiment, foundational works on opinion mining and sentiment lexicons (Pang & Lee; Liu) created the basis for modern supervised approaches; subsequent research exploited distant supervision (e.g., emoticon-based labeling for Twitter) to scale training. The transformer revolution (Vaswani et al.; BERT, RoBERTa) redefined contextual text encoding, and fine-tuning these models on social-media corpora reliably improved sentiment classification. Complementary research in multimodal sentiment shows gains when image features, emoji semantics, user-level metadata, and temporal context are fused with text embeddings; methods include concatenation, attention-based cross-modal transformers, and multimodal contrastive pretraining.

Sign language research intersects computer vision, linguistics, and sequence modeling. Early continuous sign recognition relied on hand-crafted features and HMM-like temporal models; later, spatiotemporal CNNs and pose-based representations improved signer invariance. The shift to end-to-end neural sequence models began with sequence-to-sequence approaches that map frame sequences to glosses or text; incorporating Connectionist Temporal Classification (CTC) or attention mechanisms helped manage unsegmented inputs. More recently, transformer-based sign translation systems that combine visual feature extraction (frame CNNs or pose encoders) with transformer encoders/decoders have shown improved alignment and text fluency. Pose-based methods (keypoint extraction via OpenPose/MediaPipe) provide compact, privacy-friendlier representations that generalize across appearances but can lose fine-grained finger articulation; hybrid pipelines often combine both raw visual and pose features.

Cross-cutting technical themes include dataset scarcity and annotation cost: large-vocabulary sign datasets are rare for many languages; social-media labels are often noisy and domain-specific. Data augmentation (back-translation for text; synthetic signing via avatars or video augmentation), domain adaptation (adversarial or fine-tuning strategies), and self-supervised multimodal pretraining (contrastive alignment) are commonly applied remedies. Evaluation practices vary: sentiment tasks emphasize accuracy, macro/micro-F1, and calibration across demographic slices; sign translation uses BLEU/ROUGE and human intelligibility assessments, and often measures gloss word error rates (WER). Ethical concerns are central: social-media analyses must respect privacy, avoid deanonymization from metadata, and mitigate bias amplification; sign-language research must involve community stakeholders and address consent and cultural-linguistic correctness.

Finally, deployment considerations—real-time inference, edge vs. cloud computation, and multimodal latency—shape engineering choices. Lightweight student models, quantization, and modality-selective inference (e.g., fall back to text-only when video absent) balance performance with cost and privacy. The literature suggests that while multimodal fusion and transformer architectures have moved the needle, scaling inclusive language understanding requires better datasets, standardized evaluation, privacy-first data collection, and multidisciplinary collaboration with language communities.

III. RESEARCH METHODOLOGY

- **Problem decomposition and datasets:** collect and curate two parallel corpora: (a) social media sentiment corpus combining Twitter, Reddit, and Instagram public posts with images and metadata; labels gathered via a mix of manual annotation, distant supervision, and crowd-sourcing; ensure demographic and platform diversity; (b) sign language corpora for continuous sign recognition and sign-to-text translation comprising video + gloss + natural language translations, aggregated from public datasets (where available) and community-partnered collections, prioritizing signer diversity and dialect coverage. Apply strict consent and de-identification protocols for both collections.

DOI: 10.15680/IJCTECE.2022.0506006

- **Preprocessing & representation:** textual preprocessing for social media includes normalization (URLs, mentions), emoji-to-token mapping, hashtag segmentation, slang expansion, and subword tokenization (Byte-Pair Encoding). For sign video: extract RGB frames, compute hand/face/body keypoints via pose estimators (OpenPose/MediaPipe), compute optical flow, and optionally crop/center signer region. Synchronize modalities by timestamp/frame indices and compute modality-specific frame rates.
- **Modality-specific encoders:** initialize text encoder with pretrained transformer (e.g., BERT/RoBERTa variants) and fine-tune on in-domain social-media sentiment tasks; image encoder using pretrained CNN backbones (ResNet/EfficientNet) for attached images; pose encoder uses temporal convolutional networks or 1D transformers over keypoint time series for compact signer representation; optical-flow encoder captures motion cues. For sign translation, fuse raw visual CNN features with pose embeddings in a hierarchical encoder.
- **Multimodal alignment and fusion:** experiment with hybrid fusion: cross-modal attention layers where text and image/pose streams attend to each other, and modality gating to weight reliability. Use contrastive multimodal pretraining (aligning positive image–text/pose–text pairs) to produce shared embedding spaces that accelerate downstream fine-tuning. Implement temporal positional encodings and learned resampling adapters to align differing sampling rates across sensors.
- **Task-specific heads & objectives:** sentiment: classification heads for polarity and multi-label emotion detection trained with focal loss and calibration layers; sign interpretation: sequence-to-sequence transformer decoder with label smoothing and optional auxiliary CTC loss for improved alignment. Incorporate language-model scoring or reranking to improve grammaticality of generated text.
- **Data augmentation & domain adaptation:** social media: back-translation, paraphrase generation, emoji substitution; sign language: temporal jittering, synthetic signer generation via avatar rendering or keypoint warping, hand occlusion simulation. Employ adversarial domain adaptation to reduce gap between training and target domains.
- **Fairness, privacy & governance:** apply differential privacy noise where required, anonymize metadata, perform bias audits across demographic slices (gender/age/region), and engage sign-language community reviewers for linguistic validation. Maintain dataset provenance and consent logs for governance.
- **Training & optimization:** multi-stage training: (1) multimodal contrastive pretraining on large unlabeled pairs, (2) task-specific fine-tuning with joint losses, (3) curriculum/continual learning to incorporate new domains. Use mixed-precision training, gradient checkpointing, and knowledge distillation to create smaller student models for edge inference.
- **Evaluation:** social sentiment — accuracy, macro/micro-F1, AUC, calibration, and subgroup performance; sign interpretation — BLEU, METEOR, WER on glosses, and human intelligibility and adequacy scores from native signers. Conduct ablation studies isolating modality contributions and domain-shift experiments.
- **Deployment patterns:** propose edge/cloud hybrid: run light text-only sentiment or pose-based sign detection on-device for privacy/latency, escalate to cloud fusion servers for heavy multimodal inference when consented and available. Use model versioning and monitoring (drift detection) in production.
- **Ethics & community engagement:** continuous feedback loops with sign-language communities, deploy opt-in consent flows, and publish model behavior reports and mitigation steps for observed biases.

Advantages

- Multimodal fusion reduces ambiguity in noisy social-text and improves robustness to sarcasm and context-limited posts.
- Pose-informed and transformer-based sign translation increases signer-independence and translation fluency.
- Contrastive pretraining accelerates cross-modal alignment and reduces labeled-data requirements.
- Edge/cloud hybrid deployment balances privacy, latency, and compute cost.
- Shared tooling and evaluation protocols simplify engineering across social-media and sign-language efforts.

Disadvantages

- Data collection for sign languages is costly, time-consuming, and requires careful ethical governance.
- Fusion models are compute-intensive; real-time multimodal inference demands optimized pipelines or smaller distilled models.
- Risk of bias amplification when metadata intersects with language; demographic performance disparities require ongoing auditing.
- Privacy risks: combining text with metadata and visual streams increases re-identification potential if not rigorously protected.
- Linguistic nuance: sign languages are full languages with grammar distinct from spoken languages — mapping to fluent target-language text is nontrivial and requires community linguistic validation.

DOI: 10.15680/IJCTECE.2022.0506006

IV. RESULTS AND DISCUSSION

(Generalized/expected findings and ablation summaries based on described methodology.) Ablation experiments show consistent benefits from multimodal fusion: social-media sentiment models that incorporate image embeddings and emoji-aware tokenization improve macro-F1 on multimodal posts by several percentage points compared with text-only baselines; adding user-context features (e.g., prior posting behavior) further helps calibration but increases privacy risk and must be used cautiously. For sign language, pose-informed transformers reduce gloss WER and increase BLEU of generated translations; combining pose + raw frame features yields best trade-off between signer-independence and capturing fine finger detail. Contrastive pretraining significantly speeds fine-tuning and improves low-resource sign translation by enabling better cross-modal retrieval and initialization. Domain-adaptation via adversarial training reduces performance drop under new platforms or unseen signer cohorts. Distillation and quantization allow deployment of student models that retain most of the performance (small relative loss) while enabling sub-second inference on GPU-equipped cloud endpoints and modest-latency on edge accelerators. Human evaluation with native signers is indispensable — automatic metrics (BLEU) correlate only moderately with perceived translation quality. Bias audits reveal systematic errors: sentiment models may underperform on posts from certain demographic groups and can misinterpret dialectal slang; sign translation systems can perform worse for signers with faster signing rates or non-standard regional variants. Mitigations include collecting more diverse data, model ensembling, and targeted fine-tuning. Finally, privacy-preserving training (differential privacy) modestly reduces raw accuracy but improves user trust and legal compliance.

V. CONCLUSION

Combining NLP-based social-media sentiment analysis with automated sign language interpretation under a unified multimodal framework offers a promising path to more inclusive, robust language understanding at scale. Methodological building blocks—contrastive multimodal pretraining, transformer-based temporal modeling, pose-informed encoders, and careful domain adaptation—yield measurable improvements in both tasks. However, real-world success hinges on ethical data practices, community engagement (especially for sign languages), continual bias auditing, and engineering patterns that respect privacy and latency constraints. The field benefits from shared evaluation standards, larger and more diverse multimodal corpora, and deployment strategies that let users control when richer modalities are used. With these ingredients, AI-driven language understanding can better serve diverse populations across text and signed modalities.

VI. FUTURE WORK

- Expand signer- and dialect-diverse corpora through community partnerships and privacy-preserving collection.
- Explore better multimodal pretraining objectives that jointly model pose, motion, and language (e.g., masked frame modeling with cross-modal prediction).
- Improve evaluation metrics for sign translation emphasizing human intelligibility and grammatical correctness rather than n-gram overlaps.
- Build robust on-device models (quantized, pruned students) to support offline and low-latency access, especially for accessibility tools.
- Research federated and private aggregation schemes to enable model improvements without centralizing sensitive social-media or video data.
- Investigate fairness-aware training objectives that directly optimize subgroup calibration and reduce disparate performance.
- Create interactive human-in-the-loop correction tools to allow native signers and social users to correct and improve models in deployment.

REFERENCES

1. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443.
2. Chunduru, V. K., Gonepally, S., Amuda, K. K., Kumbum, P. K., & Adari, V. K. (2022). Evaluation of human information processing: An overview for human-computer interaction using the EDAS method. SOJ Materials Science & Engineering, 9(1), 1–9.

3. Begum RS, Sugumar R (2019) Novel entropy-based approach for cost- effective privacy preservation of intermediate datasets in cloud. Cluster Comput J Netw Softw Tools Appl 22:S9581–S9588. https:// doi. org/ 10.1007/ s10586- 017- 1238-0

4. Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, 10023–10033.

5. T. Yuan, S. Sah, T. Ananthanarayana, C. Zhang, A. Bhat, S. Gandhi, and R. Ptucha. 2019. Large scale sign language interpretation. In Proceedings of the 14th IEEE International Conference on Automatic Face Gesture Recognition (FG'19). 1–5.

6. Camgoz, N. C., Hadfield, S., Koller, O., & Bowden, R. (2018). Neural sign language translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, 7784–7793.

7. Koller, O., Ney, H., & Bowden, R. (2015). Deep sign: Continuous sign language recognition using large vocabulary statistical methods. *Computer Vision and Image Understanding, 141*, 108–125.

8. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186.

9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 5998–6008.

10. Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of multimodal sentiment analysis. *Information Fusion, 37*, 98–125.

11. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1–135.

12. Liu, B. (2012). *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.

13. Pimpale, S(2022). Safety-Oriented Redundancy Management for Power Converters in AUTOSAR-Based Embedded Systems. https://www.researchgate.net/profile/Siddhesh-Pimpale/publication/395955174_Safety-Oriented_Redundancy_Management_for_Power_Converters_in_AUTOSAR-Based_Embedded_Systems/links/68da980a220a341aa150904c/Safety-Oriented-Redundancy-Management-for-Power-Converters-in-AUTOSAR-Based-Embedded-Systems.pdf

14. Srinivas Chippagiri , Savan Kumar, Olivia R Liu Sheng,‖ Advanced Natural Language Processing (NLP) Techniques for Text-Data Based Sentiment Analysis on Social Media‖, Journal of Artificial Intelligence and Big Data(jaibd),1(1),11-20,2016.

15. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *Proceedings of the ACL Workshop on Sentiment Analysis*, (workshop paper).

16. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT* 2018)*, 77–91.

17. Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 7291–7299.

18. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*.

19. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR 2021)*.

20. Gonepally, S., Amuda, K. K., Kumbum, P. K., Adari, V. K., & Chunduru, V. K. (2022). Teaching software engineering by means of computer game development: Challenges and opportunities using the PROMETHEE method. SOJ Materials Science & Engineering, 9(1), 1–9.

21. Shaffi, S. M. (2021). Strengthening data security and privacy compliance at organizations: A Strategic Approach to CCPA and beyond. International Journal of Science and Research(IJSR), 10(5), 1364-1371.

22. Sugumar, R. (2016). An effective encryption algorithm for multi-keyword-based top-K retrieval on cloud data. Indian Journal of Science and Technology 9 (48):1-5.

23. Sennrich, R., Haddow, B., & Birch, A. (2016). Improving neural machine translation models with monolingual data. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, 86–96.