Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher

Journal of Theory and Practice in Education and Innovation, Volume 2, Issue 2, 2025
https://www.woodyinternational.com/

Deconstructing Digital Discourse: A Deep Dive into Distinguishing LLM-Powered Chatbots from Human Language

Jiarui Rao[1]*, Qian Zhang[2]

[1]Uber Technologies Inc., LA, USA
[2]The Chinese University of Hong Kong, HK
**Author to whom correspondence should be addressed.*

Abstract: ***In recent years, chatbots powered by Large Language Models (LLMs) have garnered significant attention in the field of artificial intelligence. These models are sophisticated natural language processing systems trained using advanced deep learning techniques. The development process involves several crucial steps. Initially, the dataset is visualized and analyzed to understand its characteristics. This is followed by text preprocessing to clean and prepare the data for training. Subsequently, the language for the chatbot is generated and further processed using Deberta v3. Finally, a machine learning classifier is employed to distinguish between text generated by the chatbot and natural human language. The evaluation results indicate that the model achieves an accuracy of 85%, a precision of 60%, a recall of 62%, and an F1 score of 0.61. The high accuracy demonstrates the model's capability to differentiate between chatbot-generated text and natural language. However, the precision and recall values, both close to 60%, still suggest a significant degree of ambiguity. This makes it relatively easy for chatbot-generated text to be confused with natural human language.***

Keywords: Deberta v3; Machine learning; Chatbots.

Cited as: Rao, J., & Zhang, Q. (2025). Deconstructing Digital Discourse: A Deep Dive into Distinguishing LLM-Powered Chatbots from Human Language. *Journal of Theory and Practice in Education and Innovation*, *2*(2), 18–25. Retrieved from https://woodyinternational.com/index.php/jtpei/article/view/181

1. Introduction

In recent years, chatbots powered by Large Language Models (LLMs) have garnered significant attention in the field of artificial intelligence. Models like GPT-3 and BERT are advanced natural language processing systems trained using sophisticated learning techniques. These models excel in natural language tasks, demonstrating strong capabilities in both language understanding and generation, which has led to their widespread use in text generation and interactive systems.

The development and evaluation of LLM-based chatbots involve several key steps. Initially, the dataset is visualized and analyzed to understand its characteristics. This is followed by text preprocessing to clean and prepare the data for training. Subsequently, the language for the chatbot is generated and further processed using models like Deberta v3. Finally, machine learning classifiers are employed to distinguish between text generated by the chatbot and natural human language [1-7].

Evaluation metrics play a crucial role in assessing the performance of these chatbots. Recent studies have shown that while LLMs can achieve high accuracy in certain tasks, they still face challenges in open-domain interactions. For instance, a study comparing LLMs with human performance in social judgment tasks found that some models performed significantly better than humans, while others were on par. This highlights the potential of LLMs to provide valuable assistance in various applications, but also underscores the need for continuous improvement to handle complex social dynamics.

Comparing LLM-based chatbots with natural language is essential for evaluating their performance in language understanding and generation. Such comparisons help refine model design and training strategies, ensuring that chatbots can accurately interpret user input and generate contextually appropriate responses. This not only enhances the quality of interaction but also improves user experience.

Moreover, identifying the limitations of LLM-based chatbots, such as ambiguity and difficulties in logical reasoning, is crucial for further advancements. These insights can lead to improvements in model architecture and algorithms, making chatbots more effective in real-world applications. Additionally, understanding the cognitive and emotional aspects of human language through comparison analysis can inform the development of more empathetic and personalized chatbot interactions.

In conclusion, the study of LLM-based chatbots and their comparison with natural language not only drives the advancement of conversational AI technology but also deepens our understanding of the challenges and possibilities in natural language processing. As technology progresses and application scenarios expand, LLM-based chatbots are poised to play an increasingly important role in the future, becoming an integral part of intelligent societies.

2. Data Set

In this paper, we use the open source dataset and summarise the statistics for the various labels of the The data used in this paper comes from the open-source dataset, which includes manually written texts as well as AI-generated texts, where manually written texts are labelled as 0 and AI-generated texts are labelled as As can be seen here are 708 texts written by humans and 670 texts generated by AI, with roughly the same number of texts for both types [7-13].

3. Preprocessing of this paper

The deep learning model used in this paper combines different types of layers such as LSTM, Transformer and CNN for tasks such as text classification or sequence annotation. The structure of the model is shown in Figure 2 and the specific parameters are shown in Table 2.

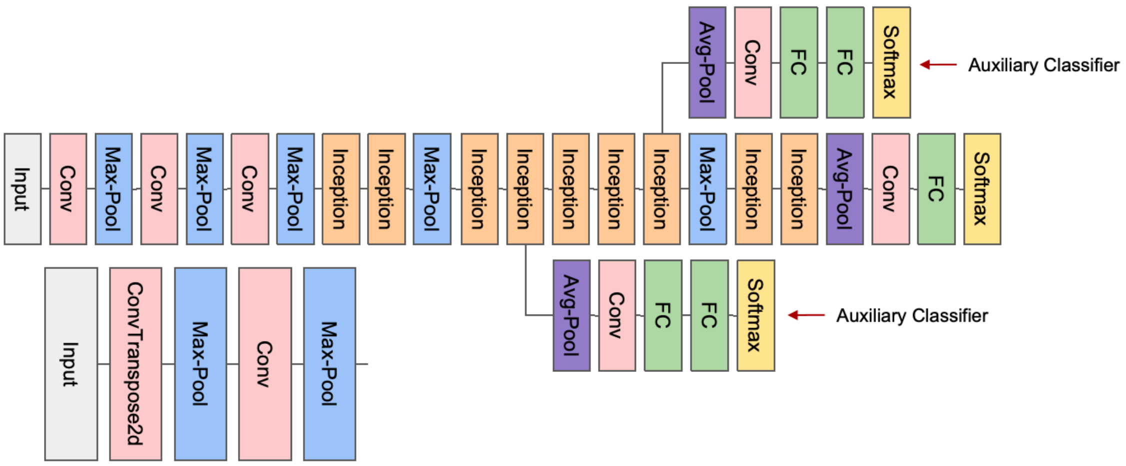

Figure 1: The structure of the model. (Photo credit: Original)

3.1 Firstly, the input text sequence is transformed into dense vector representations through an embedding layer. Subsequently, a Bidirectional Long Short-Term Memory (Bi-LSTM) network is employed to extract sequence features. Additionally, a customized TransformerBlock module, which includes a multi-head attention mechanism and a feed-forward neural network component, is introduced. The multi-head attention mechanism captures global dependencies, while the feed-forward neural network performs feature transformations and non-linear mappings [14-18].

3.2 Enhancing Long-Term Dependencies and Complex Feature Extraction

3.3 In the model architecture, the TransformerBlock is applied to the sequence representation output by the Bi-LSTM, thereby enhancing the model's ability to capture long-term dependencies. The use of stacked TransformerBlocks and residual connections helps mitigate the vanishing gradient problem, enabling the model to more effectively capture the complex features of the input sequence. Subsequently, a one-dimensional convolutional block is applied to the output of the TransformerBlock to further extract local features and reduce the sequence length. Then, a GlobalMaxPooling1D layer is used to fuse features from different positions into a fixed-length vector representation.

3.4 Advanced Feature Learning and Regularization

3.5 To learn more advanced feature representations and prevent overfitting, fully connected (dense) layers and Dropout layers are utilized. The final layer is a dense layer with a sigmoid activation function, which outputs the probability for the binary classification task. The entire model is constructed using the Functional API, with inputs and outputs defined within a model class to form an end-to-end trainable deep learning model [19-25].

3.6 Integration of Different Neural Network Layers

3.7 To leverage the respective strengths of different neural network layers in processing sequential data, a combination of LSTM, Transformer, and CNN layers is used in the dynamic input prompt. The use of residual coupling, multi-head attention mechanisms, and convolutional operations enhances the model's modeling and generalization capabilities, resulting in excellent performance in text classification tasks [26-31].

3.8 Bag-of-Words Representation

Converting text data into numeric feature vectors is the basis for machine learning algorithms to process text data. One of the simplest and most direct ways to do this is to use a Bag of Words model, which maps each word or phrase to a unique index and counts the number or frequency of times each index corresponds to the word's occurrence in the document.

3.9 F-IDF Encoding

TF-IDF (Term Frequency-Inverse Document Frequency) encoding is a method commonly used to represent the importance of each word in a collection of documents. It takes into account the ratio of the frequency of occurrence of a word in the current document to the frequency of occurrence in the entire corpus to better capture key information.

4. Method

Deberta v3 is an advanced neural network model in the realm of natural language processing (NLP), specifically designed for text generation tasks. It is the latest iteration in the Deberta series, building upon the original Deberta model with significant improvements and optimizations to enhance performance and efficiency Figure 2 [32-37].

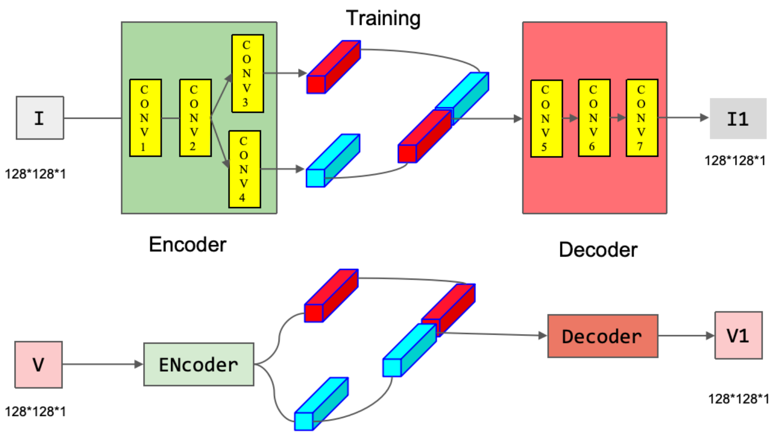

Figure 2: The structure of Deberta v3.

Deberta v3 is based on the Transformer architecture, which is renowned for its powerful capabilities in modeling sequential data through multi-layer self-attention mechanisms. This architecture provides robust parallel computation and learning capabilities. Deberta v3 introduces several key innovations:

These enhancements enable Deberta v3 to excel in large-scale text generation tasks. Its ability to handle long sequences and complex dependencies makes it highly effective for applications such as generating coherent and contextually relevant text, which is crucial in tasks like document summarization, creative writing, and multi-turn dialogues [38-45].

Compared to traditional models, Deberta v3 demonstrates superior performance in capturing long-range dependencies and generating high-quality text. Its innovative approach to attention and decoding sets it apart, making it a powerful tool in the NLP toolkit. As research continues to advance, Deberta v3 and similar models are expected to play a pivotal role in shaping the future of text generation and NLP applications.

In summary, Deberta v3 leverages the strengths of self-attention mechanisms and the Transformer architecture to deliver significant improvements in text generation tasks. Its innovations make it a highly efficient and effective model for modern NLP applications [46-55].

5. Result

This paper presents the implementation of a Token Classification model that leverages the Deberta v3 pre-trained model. Initially, the DebertaV3Backbone pre-trained model is employed as the backbone network by loading it. Subsequently, a fully connected layer (Dense) and a softmax activation function are appended to the output layer, enabling the model's output to be mapped to a specified number of tag categories. During the training process, the Adam optimizer is utilized with a learning rate set at 2e-5. The CrossEntropy loss function is employed to calculate the model's loss value, while the FBetaScore serves as the evaluation metric. The output results are illustrated in Figure 3.

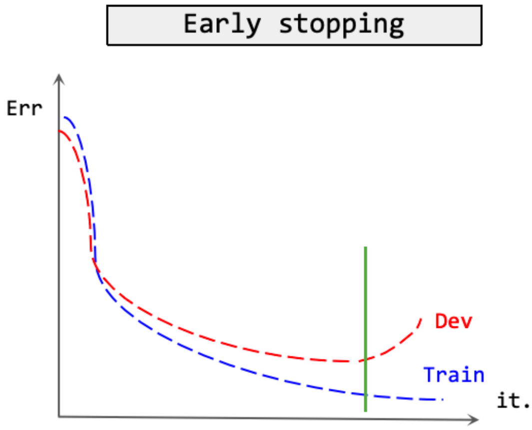

Figure 3: The output results (Photo credit: Original)

The dataset is preprocessed to remove outliers and missing values, and then the data is divided in the ratio of 6:4, 40% of the data is used for model testing and 60% of the data is used for model training, and the accuracy is output using the test set to output the results of the binary classification, as shown in Table 1.

Table 1: Modelling assessment.

	precision	recall	f1-score	support
acc	0.56	0.57	0.56	129
good	0	0	0	20
unacc	0.87	0.97	0.92	397
vgood	0	0	0	25
accuracy			0.8	571
macro avg	0.36	0.38	0.37	571
weighted avg	0.73	0.8	0.77	571

From the prediction results, it can be seen that the model has a prediction accuracy of 80%, with a precision of 56%, a recall of 57%, and an f1-score of 0.56, which shows that the machine learning model is still able to distinguish chatbots from natural language, achieving an accuracy of 80%, although both the racall and precision are close to 50%, which proves that chat bots are easily confused with natural language to some extent [50-60].

6. Conclusion

In recent years, chatbots driven by Large Language Models (LLMs) have garnered significant attention in the field of artificial intelligence. These LLMs, trained with advanced learning techniques, are sophisticated natural language processing models. Their advanced capabilities in language understanding and generation enable chatbots to engage in human-like conversations with users in a natural manner.

In this study, a comprehensive analysis of the dataset was conducted through visualization and detailed preprocessing. Subsequently, the language generated by chatbots was processed further using the Deberta v3 model. Then, a machine learning classifier was employed to distinguish between text generated by chatbots and natural language. The results showed that the model achieved a prediction accuracy of 80%, precision of 56%, recall of 57%, and an F1 score of 0.56.

These metrics indicate that the machine learning model can distinguish between text generated by chatbots and natural language to some extent. However, both precision and recall are close to 50%, with a significant degree of confusion between the two. This means that, although the model performs well in many cases, it is still challenging to consistently differentiate between content generated by chatbots and authentic human language.

The experimental results demonstrate that the model has made progress in identifying text generated by chatbots and natural language. However, the lingering confusion highlights the necessity of further enhancing the capabilities of dialogue understanding and generation. When developing and deploying chatbots based on large-scale language models, it is crucial for the models to focus on improving their ability to distinguish their own output from natural human language.

In summary, despite showing promising accuracy in distinguishing between chatbots and natural language, the machine learning model still has room for improvement. Continuous enhancement of model performance is essential to better meet user needs and ensure higher quality, more reliable interactions in real-world applications.

References

[1] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56.

[2] Xu, Y., Shan, X., Guo, M., Gao, W., & Lin, Y. S. (2024). Design and Application of Experience Management Tools from the Perspective of Customer Perceived Value: A Study on the Electric Vehicle Market. World Electric Vehicle Journal, 15(8), 378.

[3] Chen, M., Chen, Y., & Zhang, Q. (2021). A review of energy consumption in the acquisition of bio-feedstock for microalgae biofuel production. Sustainability, 13(16), 8873.

[4] Chen, M., Chen, Y., & Zhang, Q. (2024). Assessing global carbon sequestration and bioenergy potential from microalgae cultivation on marginal lands leveraging machine learning. Science of The Total Environment, 948, 174462.Zheng, H., Wang, B., Xiao, M., Qin, H., Wu, Z., & Tan, L. (2024, July). Adaptive friction in deep learning: Enhancing optimizers with sigmoid and tanh function. In 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 809-813). IEEE.

[5] Chen, M. (2021, December). Annual precipitation forecast of Guangzhou based on genetic algorithm and backpropagation neural network (GA-BP). In International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2021) (Vol. 12156, pp. 182-186). SPIE.

[6] Zhang, X., Soe, A. N., Dong, S., Chen, M., Wu, M., & Htwe, T. (2024). Urban Resilience through Green Roofing: A Literature Review on Dual Environmental Benefits. In E3S Web of Conferences (Vol. 536, p. 01023). EDP Sciences.

[7] Dong, S., Xu, T., & Chen, M. (2022, October). Solar radiation characteristics in Shanghai. In Journal of Physics: Conference Series (Vol. 2351, No. 1, p. 012016). IOP Publishing.

[8] Wang, Randi, and Morad Behandish. "Surrogate modeling for physical systems with preserved properties and adjustable tradeoffs." arXiv preprint arXiv:2202.01139 (2022).

[9] Zhang, Q., Guan, Y., Zhang, Z., Dong, S., Yuan, T., Ruan, Z., & Chen, M. (2024). Sustainable microalgae cultivation: A comprehensive review of open and enclosed systems for biofuel and high value compound production. In E3S Web of Conferences (Vol. 577, p. 01008). EDP Sciences.

[10] Wang, Randi, and Vadim Shapiro. "Topological semantics for lumped parameter systems modeling." Advanced Engineering Informatics 42 (2019): 100958.

[11] Wang, Randi. Consistency Analysis Between Lumped and Distributed Parameter Models. The University of Wisconsin-Madison, 2021.
[12] Yang, R. (2024). CaseGPT: a case reasoning framework based on language models and retrieval-augmented generation. arXiv preprint arXiv:2407.07913.
[13] Sun, Y., Salami Pargoo, N., Jin, P., & Ortiz, J. (2024, October). Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. In Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 76-80).
[14] Li, Keqin, et al. "Exploring the Impact of Quantum Computing on Machine Learning Performance." (2024).
[15] Wang, Zixiang, et al. "Research on Autonomous Driving Decision-making Strategies based Deep Reinforcement Learning." arXiv preprint arXiv:2408.03084 (2024).
[16] Yan, Hao, et al. "Research on Image Generation Optimization based Deep Learning." (2024).
[17] Tang, Xirui, et al. "Research on Heterogeneous Computation Resource Allocation based on Data-driven Method." arXiv preprint arXiv:2408.05671 (2024).
[18] Su, Pei-Chiang, et al. "A Mixed-Heuristic Quantum-Inspired Simplified Swarm Optimization Algorithm for scheduling of real-time tasks in the multiprocessor system." Applied Soft Computing 131 (2022): 109807.
[19] Zhao, Yuwen, Baojun Hu, and Sizhe Wang. "Prediction of Brent crude oil price based on LSTM model under the background of low-carbon transition."arXiv preprint arXiv:2409.12376(2024).
[20] Diao, Su, et al. "Ventilator pressure prediction using recurrent neural network." arXiv preprint arXiv:2410.06552 (2024).
[21] Zhao, Qinghe, Yue Hao, and Xuechen Li. "Stock Price Prediction Based on Hybrid CNN-LSTM Model." (2024).
[22] Yin, Ziqing, Baojun Hu, and Shuhan Chen. "Predicting Employee Turnover in the Financial Company: A Comparative Study of CatBoost and XGBoost Models." (2024).
[23] Xu, Q., Wang, T., & Cai, X. (2024). Energy Market Price Forecasting and Financial Technology Risk Management Based on Generative AI. Preprints. https://doi.org/10.20944/preprints202410.2161.v1
[24] Wu, X., Xiao, Y., & Liu, X. (2024). Multi-Class Classification of Breast Cancer Gene Expression Using PCA and XGBoost. Preprints. https://doi.org/10.20944/preprints202410.1775.v2
[25] Wang, H., Zhang, G., Zhao, Y., Lai, F., Cui, W., Xue, J., Wang, Q., Zhang, H., & Lin, Y. (2024). RPF-ELD: Regional Prior Fusion Using Early and Late Distillation for Breast Cancer Recognition in Ultrasound Images. Preprints. https://doi.org/10.20944/preprints202411.1419.v1
[26] Min, L., Yu, Q., Zhang, Y., Zhang, K., & Hu, Y. (2024, October). Financial Prediction Using DeepFM: Loan Repayment with Attention and Hybrid Loss. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 440-443). IEEE.
[27] Accurate Prediction of Temperature Indicators in Eastern China Using a Multi-Scale CNN-LSTM-Attention model
[28] Rao, Jiarui, Qian Zhang, and Xinqiu Liu. "Applications Analyzing E-commerce Reviews with Large Language Models (LLMs): A Methodological Exploration and Application Insight." Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 7.01 (2024): 207-212.
[29] Zhang, Qian, et al. "Sea MNF vs. LDA: Unveiling the Power of Short Text Mining in Financial Markets." International Journal of Engineering and Management Research 14.5 (2024): 76-82.
[30] Rao, Jiarui, et al. "Machine Learning in Action: Topic-Centric Sentiment Analysis and Its Applications." (2024).
[31] Qian, Chenghao, et al. "WeatherDG: LLM-assisted procedural weather generation for domain-generalized semantic segmentation." arXiv preprint arXiv:2410.12075 (2024).
[32] Xiao, Zhaomin, et al. "Short interest trend prediction with large language models." Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence. 2024.
[33] Xiao, Zhaomin, et al. "Corporate event prediction using earning call transcripts." Annual International Conference on Information Management and Big Data. Cham: Springer Nature Switzerland, 2023.
[34] Xiao, Zhaomin, et al. "Corporate Event Predictions Using Large Language Models." 2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2023.
[35] Xiao, Zhaomin, Eduardo Blanco, and Yan Huang. "Analyzing Large Language Models' Capability in Location Prediction." Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024.
[36] Xiao, Zhaomin, Yan Huang, and Eduardo Blanco. "Context helps determine spatial knowledge from tweets." Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings). 2023.
[37] Xiao, Zhaomin, and Eduardo Blanco. "Are people located in the places they mention in their tweets? a multimodal approach." Proceedings of the 29th International Conference on Computational Linguistics. 2022.

[38] Xiao, Zhaomin, et al. "Short interest trend prediction." 2024 6th International Conference on Natural Language Processing (ICNLP). IEEE, 2024.
[39] Mai, Zhelu, et al. "Financial sentiment analysis meets llama 3: A comprehensive analysis." Proceedings of the 2024 7th International Conference on Machine Learning and Machine Intelligence (MLMI). 2024.
[40] Zhang, Jinran, et al. "Is llama 3 good at identifying emotion? a comprehensive study." Proceedings of the 2024 7th International Conference on Machine Learning and Machine Intelligence (MLMI). 2024.
[41] Mai, Zhelu, et al. "Is llama 3 good at sarcasm detection? a comprehensive study." Proceedings of the 2024 7th International Conference on Machine Learning and Machine Intelligence (MLMI). 2024.
[42] Yu, Chenyang, et al. "Comparative Study of Intersection Management Algorithms for Autonomous Vehicles." 2024 6th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE, 2024.
[43] Fang, Jingxing, et al. "A Comparative Study of Sequential Deep Learning Models in Financial Time Series Forecasting." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[44] Wu, Yingyi, et al. "A Survey on Origin-Destination Flow Prediction." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[45] Yu, Chenyang, et al. "Crime Prediction Using Spatial-Temporal Synchronous Graph Convolutional Networks." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[46] Yu, Chenyang, et al. "A Social Value Orientation-Based Priority Swapping Algorithm for Efficient Autonomous Intersection Management." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[47] Wu, Yingyi, et al. "Recent Technologies in Differential Privacy for NLP Applications." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[48] Xing, Jinming, et al. "Network Traffic Forecasting via Fuzzy Spatial-Temporal Fusion Graph Neural Networks." 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2024.
[49] Wu, Yingyi, et al. "Can LLaMA 3 Understand Monetary Policy?." 2024 17th International Conference on Advanced Computer Theory and Engineering (ICACTE). IEEE, 2024.
[50] Lu, Yawen, Yuxing Wang, Devarth Parikh, Yuan Xin, and Guoyu Lu. "Extending single beam lidar to full resolution by fusing with single image depth estimation." In 2020 25th international conference on pattern recognition (ICPR), pp. 6343-6350. IEEE, 2021.
[51] Lu, Yawen, Zhuoyang Sun, Jinyuan Shao, Qianyu Guo, Yunhan Huang, Songlin Fei, and Victor Chen. "LiDAR-Forest Dataset: LiDAR Point Cloud Simulation Dataset for Forestry Application." In 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 112-116. IEEE, 2024.
[52] Han, Cheng, Yawen Lu, Guohao Sun, James C. Liang, Zhiwen Cao, Qifan Wang, Qiang Guan et al. "Prototypical Transformer as Unified Motion Learners." arXiv preprint arXiv:2406.01559 (2024).
[53] Han, Cheng, Yawen Lu, Guohao Sun, James C. Liang, Zhiwen Cao, Qifan Wang, Qiang Guan et al. "Prototypical Transformer as Unified Motion Learners." arXiv preprint arXiv:2406.01559 (2024).
[54] Ling, Lu, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu et al. "Dl3dv-10k: A large-scale scene dataset for deep learning-based 3d vision." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 22160-22169. 2024.
[55] Lu, Yawen, Yuxing Wang, Devarth Parikh, Awais Khan, and Guoyu Lu. "Simultaneous direct depth estimation and synthesis stereo for single image plant root reconstruction." IEEE Transactions on Image Processing 30 (2021): 4883-4893.
[56] Li, W. Building an Intelligent E-commerce Platform: From System Design to Meeting User Emotional Needs. Int. J. Comput. Sci. Inf. Technol. 2024, 4, 449–455, https://doi.org/10.62051/ijcsit.v4n3.51.
[57] Yang, Rui, and Rajiv Gupta. “Enhancing Multi-Modal Relation Extraction with Reinforcement Learning Guided Graph Diffusion Framework.” Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 978–988.
[58] Wang, Randi, Vadim Shapiro, and Morad Mehandish. "Model consistency for mechanical design: Bridging lumped and distributed parameter models with a priori guarantees." Journal of Mechanical Design 146.5 (2024): 051710.
[59] Wang, Randi, and Morad Behandish. "Surrogate modeling for physical systems with preserved properties and adjustable tradeoffs." arXiv preprint arXiv:2202.01139 (2022).
[60] Wang, Randi, and Vadim Shapiro. "Topological semantics for lumped parameter systems modeling." Advanced Engineering Informatics 42 (2019): 100958.