WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Scope of ChatGPT using Deep Learning: BERT Model

Nadamuni. Hari Chandana[1], Dr. K. Kavitha[2]

[1]UG Student, [2]Sr. Assistant Prof, GMRIT

ABSTRACT:

ChatGPT is one of the most trending topic and widely used platform by most of the people working in various sectors. It can be built by using these approaches based on part of deep learning-specifically a type of Neural network called transformer, LLM models, usage of AI tool in ESP, BERT (Bidirectional Encoders Representations from Transformers) model, etc. One of the main reason to use deep learning in ChatGPT is its intelligence in understanding the relationships and patterns from the user input text or data and generate a predicted output which is similar to the input text given by the user. ChatGPT is mainly used for replicating the human behaviour or the way of responding to the user. They made any kind of descriptive tasks easy to complete within short time. Now a days ChatGPT is widely used in social media platforms also, to get an accurate explanation or reply regarding any kind of situation. This BERT model is used in social media platforms. By using this approach ChatGPT is widely improved and is helpful in any aspect regarding social media platform. The aim of this paper is to bring the improvement in replying the tweets with correct phrases using ChatGPT according to the situation of having the accuracy of 96.49%.

Keywords: *ChatGPT, Deep learning, BERT* (Bidirectional Encoders Representations from Transformers), *ESP* (Example-based Sensor Predictions), *LLM (Large Language Models), AI (Artificial Intelligence), Machine Learning*

Introduction

HUGE impacting innovations in the NLP domain include the AI language model ChatGPT, developed by OpenAI. While BERT revolutionized the face of NLP with contextual understanding via bidirectional training, it is only matched by chatGPT in terms of transformer architecture in the development of secure conversation agents. The next section explores the relation of ChatGPT to deep learning models, largely regarding the scale of impact caused by BERT. By comparing architecture, use cases, and strengths of ChatGPT and BERT, this talk flags where the design principles of BERT have shaped and extended the capabilities of conversational AI systems to enable more accurate, coherent, and contextually aware interactions across different applications. Recent years have seen a surge in Transformer applications beyond traditional NLP tasks. That is why BERT's use of the bidirectional approach makes it able to perform so significantly outstandingly in this task that requires deep sentence structure understanding, such as question answering and language inference. Its full-word context consideration in both forward and backward directions sets new standards for text comprehension in the NLP field. In contrast, ChatGPT adopts somewhat similar transformations in the transformer architecture, with a focus on using a unidirectional or autoregressive mode toward enhancing generative capabilities, supporting long, dynamic conversations where context has to be preserved over several turns.

The influence that BERT has on other models, such as ChatGPT, shows just how revolutionary the transformer architectures are to NLP. In essence, the underlying philosophies behind BERT-the application of attention mechanisms and transfer learning-have made it possible to realize much more sophisticated responsive AI. These models are therefore extremely widely applied today in customer services, health and education systems, and content creation, where accuracy, reliability, and awareness of context within interactions are critical. Looking at the design and strengths of both BERT and ChatGPT reveals the role each of them has in the increasing applications of conversational AI, as proposed here, advocating depth and versatility to transformer-based approaches in state-of-the-art NLP applications.

Literature review

1.“Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT”, Qihuang Zhong, LiangDing, JuhuaLiu, BoDu , DachengTaoarXiv:2302.10198v2 [cs.CL] 2 Mar 2023. The paper discusses the evolution of Transformer-based pretrained language models (PLMs) and categorizes them into three groups: encoder-only (e.g., BERT), decoder-only (e.g., GPT-3), and encoder-decoder. It highlights the success of BERT-style models in natural language understanding (NLU) tasks due to their bidirectional masked language modeling objective, which allows them to encode context effectively.

2.“Fighting Fire with Fire: Can ChatGPT Detect AI-generated Text?”, Amrita Bhattacharjee, Huan Liu, arXiv:2308.01284v2 [cs.CL] 17 Aug 2023. The paper investigates the performance of ChatGPT in detecting AI-generated text, building on existing methods that utilize language models for text

classification. Previous works have explored feature-based classifiers and fine-tuned models for distinguishing between human and AI-generated text. The study uses the TuringBench dataset, which includes various AI-generated texts from multiple generators, to evaluate ChatGPT's detection capabilities.

3."Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling" Linyao Yang, Hongyang Chen, Senior Member, IEEE, Zhao Li, Xiao Ding, Xindong Wu, Fellow, IEEE**,** arXiv:2306.11489v2 [cs.CL] 30 Jan 2024. The paper reviews existing studies on enhancing pre-trained language models (PLMs) with knowledge graphs (KGs), categorizing them into implicit and explicit incorporation methods. However, previous reviews have only covered a limited subset of knowledge graph-enhanced PLMs (KGPLMs). It highlights the rapid evolution of the field, with numerous new technologies being introduced, indicating a need for a comprehensive survey that focuses specifically on KGPLMs.

4."ChatGPT v.s. Media Bias: A Comparative Study of GPT-3.5 and Fine-tuned Language Models" Zehao Wen1, Rabih Younes2, arXiv:2306.11489v2 [cs.CL] 30 Jan 2022. The paper reviews existing research on media bias detection, highlighting the transition from manual to automated methods. It emphasizes the capabilities of large language models (LLMs), particularly GPT-based models, in various tasks, including media bias detection. It discusses the challenges faced by current automated approaches, such as the reliance on labeled datasets and the inherent biases in the labeling process, which complicate the detection of media bias.

5."Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text" TIANCHENG TANG, XINHUAI TANG, AND TIANYI YUAN, date of publication October 12, 2020, date of current version November 3, 2020. Digital Object Identifier 10.1109/ACCESS.2020.3030468. The paper discusses the evolution of sentiment analysis, highlighting that most research has concentrated on binary or ternary sentiment analysis in monolingual texts. However, it notes a shift towards multilingual and multi-dimensional sentiment analysis, particularly in code-switching contexts. The introduction of BERT has significantly improved sentiment analysis performance. Previous studies have utilized BERT for various tasks, including aspect-based sentiment analysis and emotion detection in code-switching text, but there is limited research specifically targeting BERT's application in this area.

6."Consistency Analysis of ChatGPT, "Myeongjun Erik Jang1 Thomas Lukasiewicz2,1, arXiv:2303.06273v3 [cs.CL] 14 Nov 2023. The paper discusses various types of consistency in language models, particularly focusing on semantic consistency, which is crucial for reliable NLP tasks. It highlights that models should maintain consistent predictions across semantically equivalent contexts. It reviews previous works that have explored different consistency types, such as symmetric consistency, which requires models to be invariant to input order, and negation consistency, where predictions should differ for negated inputs.

7."Evaluating the Performance of ChatGPT for Spam Email Detection", Shijing Si, Yuwei Wu1 arXiv:2402.15537v2 [cs.CL] 19 Jun 2024. The paper evaluates ChatGPT's performance in spam email detection, comparing it with traditional machine learning methods like naive Bayes, SVM, LR, DNN, and BERT classifiers. It highlights the advancements in natural language processing and the potential of large language models in various tasks, including spam identification.

8."Efficient argument classification with compact language models and chatgpt-4 refinements", Marcin Pietron, Rafał Olszowski, MIT Massachusetts, Jakub Gomułka *arXiv:2403.15473v1 [cs.CL] 20 Mar 2024*. The paper focuses on argument mining (AM), which involves identifying and extracting argumentative components and their relationships. It highlights the shift from traditional methods to deep learning models, particularly the use of BERT and ChatGPT-4 for argument classification.

9."(Chat)GPT vs BERT Dawn of Justice for Semantic Change Detection", Francesco Periti, Haim Dubossarsky, Nina Tahmasebi, arXiv:2401.14040v3 [cs.CL] 29 Apr 2024. The paper evaluates the performance of (Chat)GPT and BERT in the context of semantic change detection, specifically focusing on two diachronic extensions of the Word-in-Context (WiC) task: Tempo WiC and Histo WiC. This is the first study to investigate (Chat)GPT for historical linguistic purposes, challenging previous evaluations that reported low accuracy under single setups.

10."Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data George Manias1", Argyro Mavrogiorgou1, Athanasios Kiourtis1, Chrysostomos Symvoulidis1, Dimosthenis Kyriazis1, arXiv:2401.14040v3 [cs.CL] 12 Jun 2020. The paper discusses the evolution of argument classification (AC) within argument mining (AM), highlighting the transition from traditional methods to deep learning approaches. Early works utilized hierarchical structures, while recent studies have shifted towards non-hierarchical frameworks, particularly with the introduction of Transformer-based models like BERT.

11. Analyzing Sentiments Regarding ChatGPT Using Novel BERT: A Machine Learning Approach Sudheesh R, MuhammadMujahid2, Furqan Rustam, Rahman Shafique 4 Mónica Gracia Villar 6,7,8, Julién Brito Ballester 6,9,10, Isabel de la Torre Diez, Venkata Chunduri 5 andImranAshraf4, Neural Computing and Applications (2023) 35:21415–21431. The paper discusses various studies that have analyzed sentiments on social media platforms, highlighting the importance of such analyses in understanding public opinions. For instance, Lee et al. utilized social media data to investigate the Taliban's control over Afghanistan and racism in the workplace, achieving notable accuracy in sentiment classification.

12."ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems", Yunjian Qiu, Yan Jin,ELSVIER, Intelligent Systems with Applications 21 (2024) 200308. The paper discusses the emergence of Large Language Models (LLMs) like ChatGPT and their applications across various domains, particularly in enhancing human work efficiency and aiding in tasks such as scientific writing and concept explanation. It highlights the need for integrating domain-specific knowledge into LLMs to improve their reliability and effectiveness in specialized fields like engineering design.

13."Evaluation of BERT and ChatGPT models in inference, paraphrase and similarity tasks", Radmir Kim1, Anton Kotsenko1, Aleksandr Andreev1, Anastasiia Bazanova1, Dmitry Aladin1, David Todua1, Aleksei Marushchenko1, and E3S Web of Conferences 515, 03016 (2024) TT21C-2024. The paper explores the application of BERT and ChatGPT models specifically in mechanical engineering, focusing on their capabilities in natural language processing (NLP) tasks such as analyzing technical documentation and diagnosing malfunctions. It discusses the architectural features of both models, highlighting their strengths and weaknesses in the context of machine learning and natural language understanding.

14."Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data", George Manias1, Argyro Mavrogiorgou1, Athanasios Kiourtis1, Chrysostomos Symvoulidis1, Dimosthenis Kyriazis1. The paper discusses the rapid growth of telemedicine, especially during the COVID-19 pandemic, highlighting the need for efficient medical dialogue summarization algorithms to manage lengthy conversations effectively. It compares the performance of various models, including BERTSUM, BART, and ChatGPT, in generating medical dialogue summaries, emphasizing the shift towards large language models (LLMs) like GPT-3.5.

15."Exploring the potential of ChatGPT in medical dialogue summarization: a study on consistency with human preferences", Yong Liu1, Sheng gen Ju1 and Junfeng Wang, Liu et al. BMC Medical Informatics and Decision Making https://doi.org/10.1186/s12911-024-02481-8 . The paper discusses the need for sophisticated multilingual approaches in natural language processing (NLP) due to the challenges posed by multilingual and multicultural societies. It highlights ongoing efforts by major AI research teams to develop comprehensive multilingual systems. It reviews the utilization of multilingual sentence embeddings and classifiers based on pre-trained models and transfer learning, which are essential for effective multilingual text categorization and sentiment analysis.

Methodology

It is basically an approach by the BERT model. So, knowing it works in the case of ChatGPT, the multi-phased approach which it employed was to find out how well the model of BERT or any other of ChatGPT compared to any NLP related task pertaining to mechanical engineering. This approach can be divided into four categories, namely, task selection, model evaluation, performance evaluation, and bottlenecks identification, for describing the merits and demerits of one of the two models that exist in the system in an organized manner.

1. Task selection and classification

- Applicable NLP activities: Logical Inference Text Similarity Paraphrasing
- Group tasks by semantic dimension to understand how the model behaves with complex texts in the field of engineering.

2. Model Architecture and Evaluation

- Compare Architectures:
- BERT: transformer architecture with deep contextual understanding.
- ChatGPT: a conversation having large generative capacity, with strengths and weaknesses. Strengths: BERT is contextual, but for chat GPT- it has very high generative capacity, and the weakness is having a shallow conversational model.

3. Performance Comparison

- Accuracy, precision, recall, and finally the F1 score.
- Zero-Shot Vs. Fine-tuning: both zero shot, and with the fine tuning over the task specific training on the performance of every model.

4. Bottle-necks Identification

- Identify Domain-specific Problems:
- Semantics of complex text ChatGPT
- BERT: Not applicable in the conversational and interactive nature of applications
- Error Analysis: A qualitative long narrative of patterns in error.

5. Development Recommended for a Model

- Hybrid Approach: Inclusion of Expert Systems in Preprocessing Mivar
- Illuminated design to function: Technical practicability for change of imperative engineering changes and input configurations

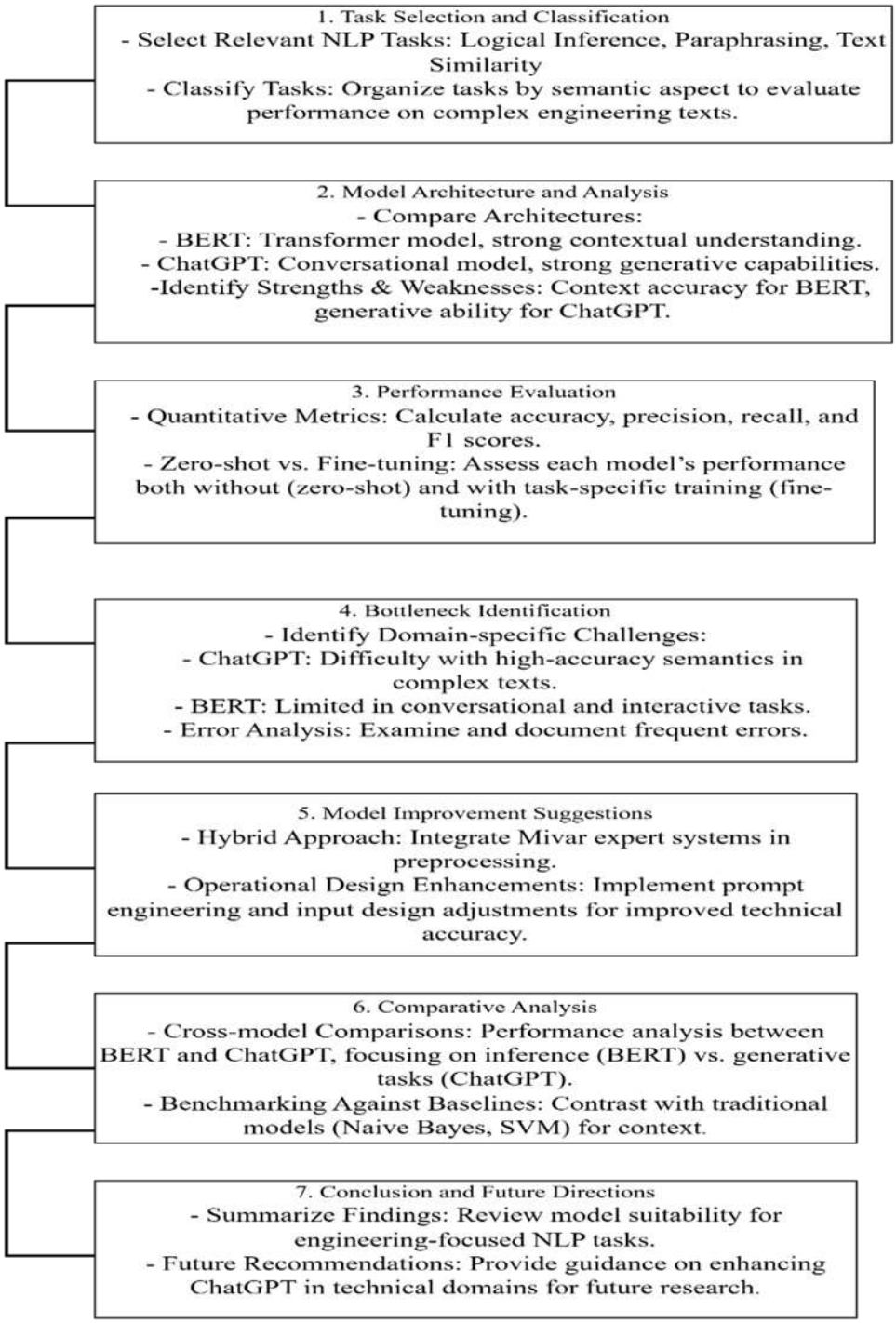
6. Comparison End

- Cross Model Comparison: Conclusion of BERT in relation to capability generation in the case of ChatGPT.
- Benchmark against older systems such as Naive Bayes and SVM.

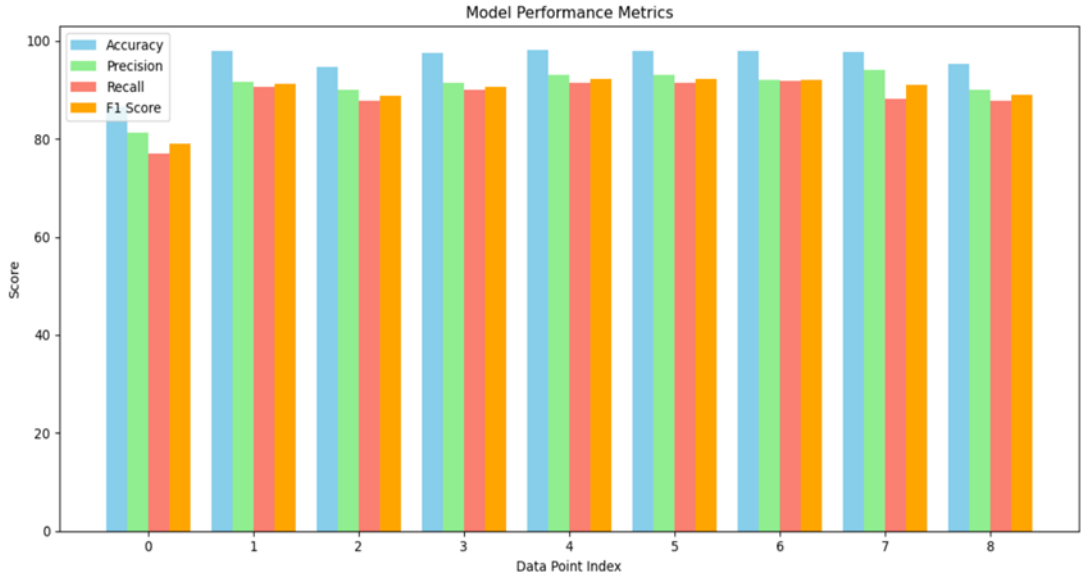
7. Conclusion

- Summary of Results Conclusion: whether this model fits or if it's just a suited variant to engineering-centric NLP tasks.

Workflow



Graph



Results and Analysis

Model Architecture	Pre-train Language	Preprocessing Particularities	Fine-Tuning Dataset	Acc (%)	Pr (%)	Rec (%)	F1 (%)
Baseline	-	n-gram TFIDF POS tags	Off_en_train	86.63	81.19	76.96	79.01
BERT	English	-	Off_en_train	97.90	91.70	90.69	91.19
BERT	English	-	Off_en_train+ OLID+HASOC_en	94.62	89.91	87.75	88.81
mBERT	Multi	-	Off_all_train+ OLID+HASOC_all	97.54	91.42	89.92	90.66
Roberta	English	-	Off_en_train	98.05	93.11	91.43	**92.26**
Roberta	English	-	Off_en_train+ OLID+HASOC_en	98.01	93.03	91.45	92.23
ALBERT	English	-	Off_en_train+	97.96	92.10	91.83	91.96
ALBERT	English	-	Off_en_train+ OLID+HASOC_en	97.79	94.04	88.13	90.99
XLM-Roberta	Multi	-	Off_all_train+ OLID+HASOC_all	95.27	90.10	87.84	88.95

Conclusion

This paper explored the intersection of deep learning, specifically BERT and ChatGPT, in advancing natural language processing for diverse applications. While BERT's bidirectional training provides a nuanced understanding of text, enhancing contextual comprehension for tasks like sentiment analysis and logical inference, ChatGPT excels in generating coherent and contextually appropriate conversational responses. The comparison revealed that while BERT's deep contextual analysis is robust for tasks with complex textual semantics, ChatGPT's strength lies in its generative capabilities, making it better suited for interactive and conversational tasks. The integration of BERT principles in ChatGPT has strengthened conversational AI's ability to deliver accurate, human-like responses, with a notable impact on social media and customer service applications. Future improvements may include hybrid models that combine BERT's detailed contextual processing with ChatGPT's conversational fluency, enabling models with improved performance across diverse NLP tasks.

References

1." Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT", Qihuang Zhong, LiangDing, JuhuaLiu, BoDu, DachengTaoarXiv:2302.10198v2 [cs.CL] 2 Mar 2023

2.“Fighting Fire with Fire: Can ChatGPT Detect AI-generated Text?”, Amrita Bhattacharjee, Huan Liu, arXiv:2308.01284v2 [cs.CL] 17 Aug 2023

3.“Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modelling” Linyao Yang, Hongyang Chen, Senior Member, IEEE, Zhao Li, Xiao Ding, Xindong Wu, Fellow, IEEE, arXiv:2306.11489v2 [cs.CL] 30 Jan 2024

4.“ChatGPT v.s. Media Bias: A Comparative Study of GPT-3.5 and Fine-tuned Language Models” Zehao Wen1, Rabih Younes2, arXiv:2306.11489v2 [cs.CL] 30 Jan 2022.

5.“Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text” TIANCHENG TANG, XINHUAI TANG, AND TIANYI YUAN, date of publication October 12, 2020, date of current version November 3, 2020. Digital Object Identifier 10.1109/ACCESS.2020.3030468

6.“Consistency Analysis of ChatGPT” ,Myeongjun Erik Jang1 Thomas Lukasiewicz2,1, arXiv:2303.06273v3 [cs.CL] 14 Nov 2023

7.“Evaluating the Performance of ChatGPT for Spam Email Detection”, Shijing Si, Yuwei Wu1 arXiv:2402.15537v2 [cs.CL] 19 Jun 2024

8.“Efficient argument classification with compact language models and chatgpt-4 refinements”, Marcin Pietron, Rafał Olszowski, MIT Massachusetts, Jakub Gomułka arXiv:2403.15473v1 [cs.CL] 20 Mar 2024

9.“(Chat)GPT vBERT Dawnof Justice for Semantic Change Detection”, Francesco Periti, Haim Dubossarsky, Nina Tahmasebi, arXiv:2401.14040v3 [cs.CL] 29 Apr 2024

10.“Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data George Manias1“,Argyro Mavrogiorgou1,Athanasios Kiourtis1,Chrysostomos Symvoulidis1,Dimosthenis Kyriazis1, arXiv:2401.14040v3 [cs.CL] 12 Jun 2020

11.Analyzing Sentiments Regarding ChatGPT Using Novel BERT: A Machine Learning Approach Sudheesh R, MuhammadMujahid2, Furqan Rustam, Rahman Shafique 4 Mónica Gracia Villar 6,7,8, Julién Brito Ballester 6,9,10, Isabel de la Torre Diez, Venkata Chunduri 5 andImranAshraf4, Neural Computing and Applications (2023) 35:21415–21431.

12.“ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems”, Yunjian Qiu, Yan Jin, ELSVIER, Intelligent Systems with Applications 21 (2024) 200308

13.“Evaluation of BERT and ChatGPT models in inference, paraphrase and similarity tasks”, Radmir Kim1, Anton Kotsenko1, Aleksandr Andreev1, Anastasiia Bazanova1, Dmitry Aladin1, David Todua1, Aleksei Marushchenko1, and Oleg Varlamov1, E3S Web of Conferences 515, 03016 (2024) TT21C-2024

14.“Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data”, George Manias1, Argyro Mavrogiorgou1, Athanasios Kiourtis1, Chrysostomos Symvoulidis1, Dimosthenis Kyriazis1

Neural Computing and Applications (2023) 35:21415–21431

15.“Exploring the potential of ChatGPT in medical dialogue summarization: a study on consistency with human preferences”, Yong Liu1, Shenggen Ju1 and Junfeng Wang, Liu et al. BMC Medical Informatics and Decision Making https://doi.org/10.1186/s12911-024-02481-8