

Transformers in Machine Learning: Literature Review

Thoyyibah T^{1*}, Wasis Haryono¹, Achmad Udin Zailani¹, Yan Mitha Djaksana¹, Neny Rosmawarni², Nunik Destria Arianti³

¹ Faculty of Computer Science, Informatics Engineering Study Program, Universitas Pamulang, Banten, Indonesia.

² Faculty of Computer Science, Informatics Engineering Study Program, Universitas Pembangunan Nasional Veteran, Jakarta, Indonesia.

³ Faculty of Computer Engineering and Design, Information System Study Program, Nusa Putra University, Bandung, Indonesia.

Received: July 15, 2023

Revised: August 20, 2023

Accepted: September 25, 2023

Published: September 30, 2023

Corresponding Author:

Thoyyibah T

doseno01116@npam.ac.id

DOI: [10.29303/jppipa.v9i9.5040](https://doi.org/10.29303/jppipa.v9i9.5040)

© 2023 The Authors. This open access article is distributed under a (CC-BY License)



Abstract: In this study, the researcher presents an approach regarding methods in Transformer Machine Learning. Initially, transformers are neural network architectures that are considered as inputs. Transformers are widely used in various studies with various objects. The transformer is one of the deep learning architectures that can be modified. Transformers are also mechanisms that study contextual relationships between words. Transformers are used for text compression in readings. Transformers are used to recognize chemical images with an accuracy rate of 96%. Transformers are used to detect a person's emotions. Transformer to detect emotions in social media conversations, for example, on Facebook with happy, sad, and angry categories. Figure 1 illustrates the encoder and decoder process through the input process and produces output. the purpose of this study is to only review literature from various journals that discuss transformers. This explanation is also done by presenting the subject or dataset, data analysis method, year, and accuracy achieved. By using the methods presented, researchers can conclude results in search of the highest accuracy and opportunities for further research.

Keywords: Accuracy; Machine learning; Transformer

Introduction

At first, the transformer is a neural network architecture that is considered as input (Reinauer et al., 2021; Grechishnikova, 2021; Ramos-Pérez et al., 2021; Moutik et al., 2023; Röder, 2023; Lin et al., 2022; Abed et al., 2023). In principle, a transformer is needed to solve sequential problems such as sentences, which is an artificial neural network architecture (Luitse et al., 2021; Taye, 2023; Yang & Wang, 2020). The transformer is in charge of connecting the encoder and decoder to the text in stages (Singla et al., 2020; Alqudsi et al., 2019; Liu & Chen, 2022) input and output are interconnected via context vectors. The autoencoder is capable of converting input into output. Besides being used in Natural Language Process transformers, it is also used in computer vision (Ghojogh et al., 2020; He et al., 2023). Transformer is a concept in Natural language Processing (NLP) in Deep Learning (Singla et al., 2020).

Transformer is part of NLP, an open-source library (Singla et al., 2020). The Natural Language process uses many words that are processed through transformers (Singla et al., 2020; Khurana et al., 2023; Caucheteux & King, 2022). Several studies use transformers, namely the application of machine learning using the transformer method in the health sector (Alqudsi et al., 2019; Arshed et al., 2023; Sanmarchi et al., 2023). Transformers used to examine people's views through politics on social media apply BERT, BERT, LSTM, Support Vector Machine, Decision Trees, Naïve Bayes, and Electra. In this study, the highest accuracy value used the Electra model with a value of 70% (Öztürk et al., 2022). The use of TurnGPT transformer model is used in spoken dialogue (Ekstedt et al., 2020). Research that applies transformers to musical chord recognition using the bi-directional Transformer for chord recognition (BTC) method (Park et al., 2019). Transformer using block model (Yan Xiao A et al., 2020). Transformer imaging using 3D on the plane (Dong Yang et al., 2021).

How to Cite:

T, T., Haryono, W., Zailani, A. U., Djaksana, Y. M., Rosmawarni, N., & Arianti, N. D. (2023). Transformers in Machine Learning: Literature Review. *Jurnal Penelitian Pendidikan IPA*, 9(9), 604–610. <https://doi.org/10.29303/jppipa.v9i9.5040>

The transformer used for sentiment analysis uses the BERT model (Durairaj et al., 2021; Kokab et al., 2022; Prottasha et al., 2022).

Machine translation uses the concept of transformers (Qiang Wang et al., 2019). Transformers were used in the best food review using BERT (Pasaribu et al., 2020). Transformers are used to assess tourist visits (Irfan et al., 2021; Wu et al., 2023; Bohatyrewicz et al., 2019). Sentiment analysis on film reviews using transformers (Putri et al., 2022; Bacco et al., 2021). Transformer introduces BERT and GPT with encoder and decoder (Ghojogh, 2020; Paaß & Giesselbach, 2023).

Transformers are used for text compression in readings (Li et al., 2021). Transformers are used to recognize chemical images with an accuracy rate of 96% (Rajan et al., 2021). Transformers are used to detect a person's emotions (Zhong et al., 2019; Graterol et al., 2021; Ghosh et al., 2023). Transformer to detect emotions in social media conversations, for example on Facebook with categories of happy, sad, and angry (Zhong et al., 2019; Acheampong et al., 2020; Li et al., 2020). Machine learning uses a transformer model with 1000+ transformers to analyze health problems (Wu et al., 2020). Transformers were used to predict the next 10 weeks with influenza case subjects (Rendragraha et al., 2021; Santangelo et al., 2023; Piccialli et al., 2021).

Theory

The transformer model is also used to detect abusive language in Indonesian online news comments. Classification of the text is done with the category of offensive, normal, or non-offensive (Tran et al., 2020). The transformer model is used for energy monitoring using the concept of algorithms in machine learning in the health sector (Valencia et al., 2021). Transformer method to analyze a person's stress (Vaswani et al., 2017). The transformer model is used for the Classification of Music in Indonesia (Thoyyibah et al., 2022). Transformers are also mechanisms that study contextual relationships between words. Figure 1 illustrates the encoder and decoder process through the input process and produces output. In Transformer there are two mechanisms, namely:

Code Generator

Encoder is used to read all text input at once. The encoder consists of a stack of identical layers. Each layer has two sub-layers, namely the self-attention layer and the feed-forward neural network. with a self-attention layer, the encoder can help nodes not only focus on the word they see but also get the semantic context of the word. Each position in the encoder can handle all positions in the previous layer in the encoder.

Decoder

The decoder is used to generate predictive output sequences. The decoder also consists of a stack of identifiable layers. Each layer consists of two sub-layers like those of the encoder, with an additional attention layer between the two layers to help the current node get to the main content it needs to pay attention to by performing multi-head attention on the encoder output. Similar to the encoder, the self-attention layer in the decoder makes each position in the decoder capable of handling all previous and current positions. From the various journals described earlier, researchers are interested in studying transformers. In several aspects, this research is different from previous research. Focuses on critical thinking skills. First, this research focuses on all the articles that have been published from 2017 to 2022; almost all of which are international journals. Second, this research is devoted to further reviewing various articles with critical thinking skills as the main focus being discussed. Third, various parameters are used as the basis for content analysis, through analysis of methods, accuracy, and others.

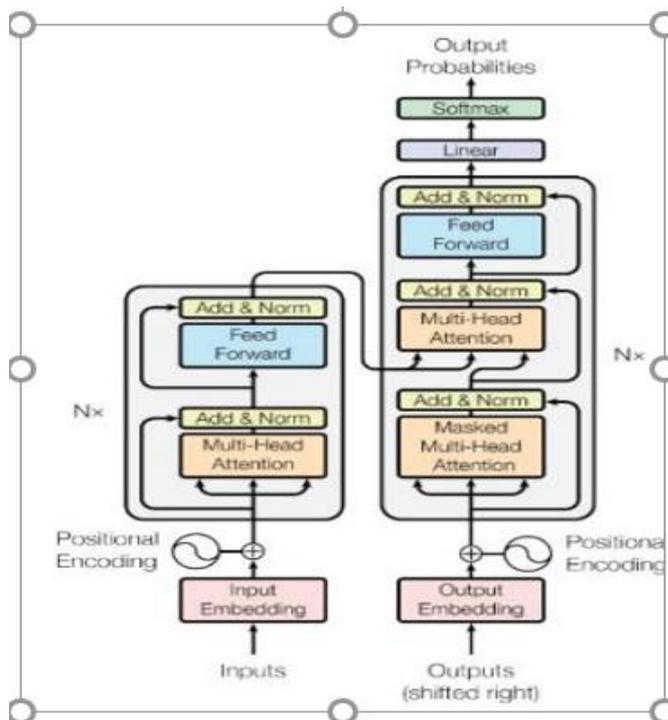


Figure 1. Transformer model

Method

Research Design

This research journal has a principle for analyzing the contents of previous literature reviews, which focuses on the findings of various studies that have been published in various international journals.

Data Source

The data collected from the analysis in the form of the contents of this literature review comes from international journals about transformers in machine learning. All journals are taken from various sources of international journals. Hence, all the contents of the literature review that reviews transformers in machine learning are collected from each of these journals. The literature review analyzed in this study has been published online, in a total of 23 journals that examine transformers in machine learning. All literature reviews are analyzed in this study to find out what methods are used in transformers in machine learning.

Research Instruments

This research is a literature review of several research journals related to transformers in machine learning. The review was conducted on some of the most recent research efforts utilizing machine learning. Furthermore, this study comes from several literacies and includes problem-solving efforts that are divided into areas from the perspective of each machine learning category with various methods available in the research journal. The data collection process used to examine some of this literature is very useful for finding and obtaining study sources based on previous relevant research. Supporting theories, data, and information as references in the documentation.

There are several main aspects in Table 1 which are presented in this research journal from the literature review of several international journals. These aspects include (A) type of research, (B) year, (C) Dataset, (D) Transformer Model, (E) Accuracy, (F) Field of Research.

Table 1. Aspects and Categories Used for Content

Aspects	Categories
Kind of Research	A.1 Experiment
	A.2 Literature review
Year	B.1 2017
	B.2 2019
	B.3 2020
	B.4 2021
	B.5 2022
Dataset	C.1 Public
	C.2 Private
Transformer	D.1 Transformer
	D.2 BERT
	D.3 GPT

Result and Discussion

The number of article publications shows how high the accuracy value of the research carried out in a certain period is. Referring to the graph shown in the table above, which reviews a transformer method that shows the level of accuracy in each method, shows that there is

an average highest accuracy value of up to 90%, and there is also an average accuracy value of only 10%. In addition, the number of publications each year always appears using different methods. The trend of research using different methods shows that transformers can be implemented in various fields, not referring to only one field.

Type of Research

Some researchers use this type of experimental research. The type of experimental research here is almost all using a public or unpublic database. Besides that, some researchers are more interested in studying transformers in theory, so the authors also use literature reviews for several studies from various sources. Figure 1 is a graph of the journal's analysis of the transformer.

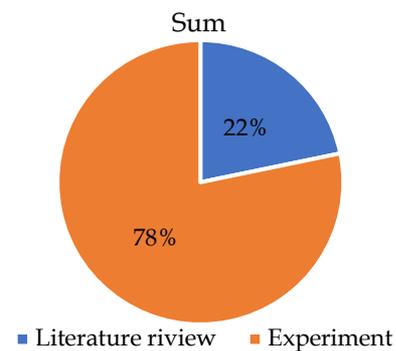


Figure 2. Research type diagram

In Figure 2 there are 2 types of research, namely literature review and experiment. This study examines many journals the largest number of which is 78% in the experimenter and 22% of the total journals studied.

Research Year

Some researchers use this type of experimental research. The type of experimental research here is almost all using a public or unpublic database. Besides that, some researchers are more interested in studying transformers in theory, so the authors also use literature reviews for several studies from various sources.

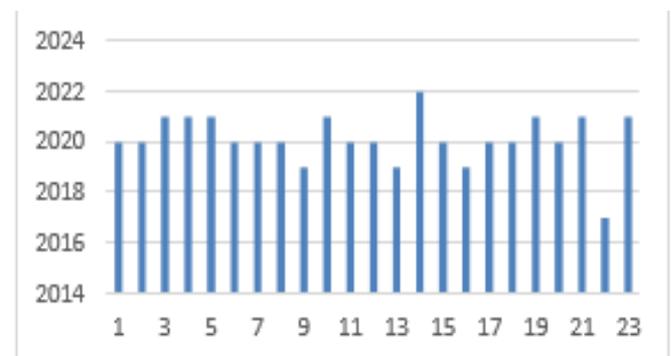


Figure 3. Year of research

The research years in Figure 2 consist of 2017, 2019, 2020, 2021, and 2022. The most researched year was 2020, namely 11, 2021, 6, 2019, 3, 2017, 1, 2022, 1. Of the five years, there was 2018 when there was no research, so this is an opportunity for transformer research in 2018.

Datasets

This study in Figure 4, uses public datasets and private datasets. Public datasets are datasets that are easy to search for. Private datasets are hard-to-find datasets with certain approvals. Private datasets are usually used in the health sector.

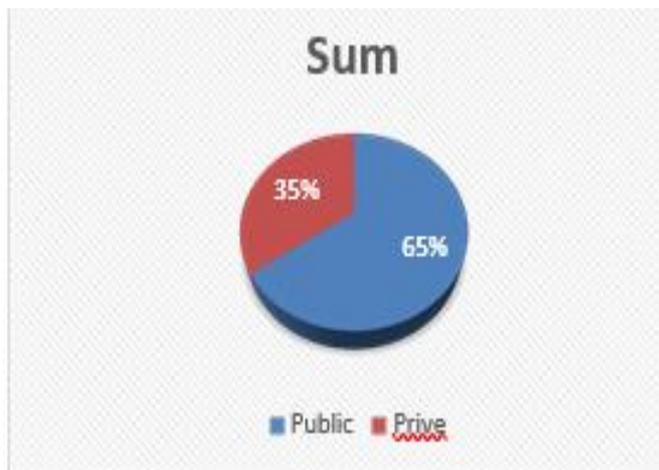


Figure 4. Dataset types

Research Methods

In this study, there are several methods used in making the research method consisting of transformers, BERT, and GPT. BERT and GPT are variants of the transformer itself. The use of the transformer method is in Figure 5. Dominates more than the others, this is an opportunity to explore the BERT and GPT methods. Of the 23 papers I read, 16 used transformers, 6 used BERT, and 1 used GPT.

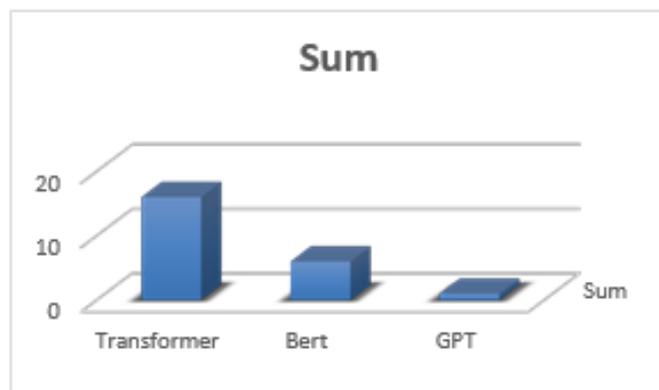


Figure 5. Transformer model

Level of Accuracy

The accuracy rate in Figure 6 averages above 40%. Some of the studies analyzed did not mention their accuracy, only told the process of collecting data or comparing methods. A total of 10 of the 23 journals analyzed reported the accuracy achieved. 1 journal 54% [20]. There is 1 that reaches 99.1% [1].

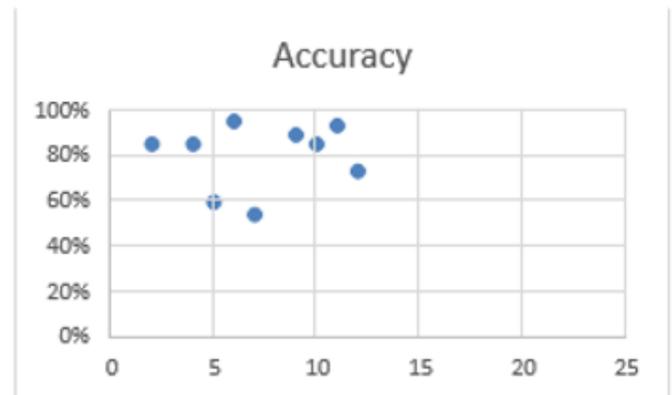


Figure 6. Accuracy level

Field of Data Analysis

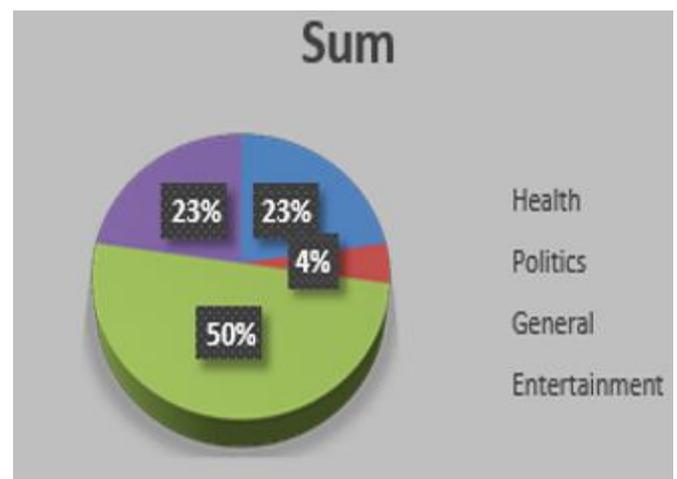


Figure 7. Field of analysis

Conclusion

Some of the conclusions from this research consist of: Most types of research are in experiments. this is because it is easy to find data on this type of research. research opportunities, Of the five years, there was 2018 when there was no research, so it became an opportunity for transformer research in 2018, The type of dataset used uses more public data than private data. The transformer method in e r arch analyzed dominates the transformer itself, the field of transformer research dominates in the general field.

Acknowledgments

Thanks to all parties who have supported the implementation of this research. I hope this research can be useful.

Author Contributions

Conceptualization, T.T., W. H., A. U. Z., Y. M. D., N. R., N. D. A.; methodology, T.T.; validation, W. H. and A. U. Z.; formal analysis, Y. M. D; investigation, N. R., and N. D. A.; formal analysis, T.T.; investigation, W. H and A. U. Z.; resources, Y. M. D. and N. R.; data curation, N. D. A; writing – original draft preparation, T.T, and W. H.; writing – review and editing, A. U. Z.: visualization, Y. M. D and N. R.; supervision, N. D. A.; project administration, T.T; funding acquisition, W. H., and A. U. Z. All authors have read and agreed to the published version of the manuscript.

Funding

This research was independently funded by researchers.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Abed, M., Imteaz, M. A., Ahmed, A. N., & Huang, Y. F. (2023). A novel application of transformer neural network (TNN) for estimating pan evaporation rate. *Applied Water Science*, 13(2), 31. <https://doi.org/10.1007/s13201-022-01834-w>
- Acheampong, F. A., Wenyu, C., & Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7). <https://doi.org/10.1002/eng2.12189>
- Alqudsi, A., & El-Hag, A. (2019). Application of machine learning in transformer health index prediction. *Energies*, 12(14), 2694. <https://doi.org/10.3390/en12142694>
- Arshed, M. A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M. (2023). Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models. *Information*, 14(7), 415. <https://doi.org/10.3390/info14070415>
- Durairaj, A. K., & Chinnalagu, A. (2021). Transformer based Contextual Model for Sentiment Analysis of Customer Reviews: A Fine-tuned BERT. *International Journal of Advanced Computer Science and Applications*, 12(11). Retrieved from https://thesai.org/Downloads/Volume12No11/Paper_53-Transformer_based_Contextual_Model_for_Sentiment_Analysis.pdf
- Bacco, L., Cimino, A., Dell'Orletta, F., & Merone, M. (2021). Explainable Sentiment Analysis: A Hierarchical Transformer-Based Extractive Summarization Approach. *Electronics*, 10(18), 2195. <https://doi.org/10.3390/electronics10182195>
- Bohatyrewicz, P., Płowucha, J., & Subocz, J. (2019). Condition Assessment of Power Transformers Based on Health Index Value. *Applied Sciences*, 9(22), 4877. <https://doi.org/10.3390/app9224877>
- Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, 5(1), 134. <https://doi.org/10.1038/s42003-022-03036-1>
- Putri, C. A. (2020). Analisis sentimen review film berbahasa Inggris dengan pendekatan bidirectional encoder representations from transformers. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 6(2), 181-193. <https://doi.org/10.35957/jatisi.v6i2.206>
- Pasaribu, D. J. M., Kusriani, K., & Sudarmawan, S. (2020). Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan Bert Embedding. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 9-20. <http://dx.doi.org/10.35585/inspir.v10i1.2568>
- Yang, D., Myronenko, A., Wang, X., Xu, Z., Roth, H. R., & Xu, D. (2021). T-AutoML: Automated machine learning for lesion segmentation using transformers in 3d medical imaging. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3962-3974). Retrieved from https://openaccess.thecvf.com/content/ICCV2021/papers/Yang_T-AutoML_Automated_Machine_Learning_for_Lesion_Segmentation_Using_Transformers_in_ICCV_2021_paper.pdf
- Ekstedt, E., & Skantze, G. (2020). *Turngpt: a transformer-based language model for predicting turn-taking in spoken dialog*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.findings-emnlp.268>
- Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2023). VAD-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes. *Information Processing & Management*, 60(2), 103234. <https://doi.org/10.1016/j.ipm.2022.103234>
- Ghojogh, B., & Ghodsi, A. (2020). *Attention mechanism, transformers, BERT, and GPT: tutorial and survey*. <https://doi.org/10.31219/osf.io/m6gcn>
- Graterol, W., Diaz-Amado, J., Cardinale, Y., Dongo, I., Lopes-Silva, E., & Santos-Libarino, C. (2021). Emotion Detection for Social Robots Based on NLP Transformers and an Emotion Ontology. *Sensors*, 21(4), 1322. <https://doi.org/10.3390/s21041322>
- Grechishnikova, D. (2021). Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Scientific Reports*,

- 11(1), 321. <https://doi.org/10.1038/s41598-020-79682-4>
- He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., Gao, Y., Wang, Q., Zhang, J., & Shen, D. (2023). Transformers in medical image analysis. *Intelligent Medicine*, 3(1), 59–78. <https://doi.org/10.1016/j.imed.2022.07.002>
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Li, Q., Wu, C., Wang, Z., & Zheng, K. (2020). Hierarchical Transformer Network for Utterance-Level Emotion Recognition. *Applied Sciences*, 10(13), 4447. <https://doi.org/10.3390/app10134447>
- Li, Z., Zhang, Z., Zhao, H., Wang, R., Chen, K., Utiyama, M., & Sumita, E. (2021). Text compression-aided transformer encoding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3840–3857. Retrieved from <https://ieeexplore.ieee.org/abstract/document/9354025>
- Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open*, 3, 111–132. <https://doi.org/10.1016/j.aiopen.2022.10.001>
- Liu, H.-L., & Chen, W.-L. (2022). X-Transformer: A Machine Translation Model Enhanced by the Self-Attention Mechanism. *Applied Sciences*, 12(9), 4502. <https://doi.org/10.3390/app12094502>
- Luitse, D., & Denkena, W. (2021). The great transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society*, 8(2), 20539517211047734. <https://doi.org/10.1177/20539517211047734>
- Moutik, O., Sekkat, H., Tigani, S., Chehri, A., Saadane, R., Tchakoucht, T. A., & Paul, A. (2023). Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data? *Sensors*, 23(2), 734. <https://doi.org/10.3390/s23020734>
- Öztürk, O., & Özcan A. (2022). *Ideology Detection Using Transformer-Based Machine Learning Models*. Retrieved from <https://rb.gy/28t72>
- Paaß, G., & Giesselbach, S. (2023). Improving Pre-trained Language Models. In G. Paaß & S. Giesselbach, *Foundation Models for Natural Language Processing* (pp. 79–159). Springer International Publishing. https://doi.org/10.1007/978-3-031-23190-2_3
- Park, J., Choi, K., Jeon, S., Kim, D., & Park, J. (2019). A Bi-Directional Transformer for Musical Chord. *20th International Society for Music Information Retrieval Conferences* Netherland. <https://doi.org/10.13140/RG.2.2.12303.51362>
- Piccialli, F., Di Cola, V. S., Giampaolo, F., & Cuomo, S. (2021). The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic. *Information Systems Frontiers*, 23(6), 1467–1497. <https://doi.org/10.1007/s10796-021-10131-x>
- Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. *Sensors*, 22(11), 4157. <https://doi.org/10.3390/s22114157>
- Putra, M. F. D. A., & Hidayatullah, A. F. (2021). Tinjauan Literatur: Named Entity Recognition pada Ulasan Wisata. *AUTOMATA*, 2(1). Retrieved from <https://journal.uui.ac.id/AUTOMATA/article/view/17391>
- Rajan, K., Zielesny, A., & Steinbeck, C. (2021). DECIMER 1.0: deep learning for chemical image recognition using transformers. *Journal of Cheminformatics*, 13(1), 1–16. <https://doi.org/10.1186/s13321-021-00538-8>
- Ramos-Pérez, E., Alonso-González, P. J., & Núñez-Velázquez, J. J. (2021). Multi-Transformer: A New Neural Network-Based Architecture for Forecasting S&P Volatility. *Mathematics*, 9(15), 1794. <https://doi.org/10.3390/math9151794>
- Reinauer, R., Caorsi, M., & Berkouk, N. (2021). *Persformer: A transformer architecture for topological machine learning*. <https://doi.org/10.48550/arXiv.2112.15210>
- Rendragraha, A. D., Bijaksana, M. A., & Romadhony, A. (2021). Pendekatan metode Transformers untuk deteksi bahasa kasar dalam komentar berita online Indonesia. *eProceedings of Engineering*, 8(2). Retrieved from https://repository.telkomuniversity.ac.id/pustakaa/files/167651/jurnal_eproc/pendekatan-metode-transformers-untuk-deteksi-bahasa-kasar-dalam-komentar-berita-online-indonesia.pdf
- Röder, L. L. (2023). *Neural Network Architectures for Absorption Spectroscopy*. <https://doi.org/10.21203/rs.3.rs-2832856/v1>
- Sanmarchi, F., Bucci, A., Nuzzolese, A. G., Carullo, G., Toscano, F., Nante, N., & Golinelli, D. (2023). A step-by-step researcher's guide to the use of an AI-based transformer in epidemiology: An exploratory analysis of ChatGPT using the STROBE checklist for observational studies. *Journal of Public Health*. <https://doi.org/10.1007/s10389-023-01936-y>
- Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D., & Cedrone, F. (2023). Machine Learning and Prediction of Infectious Diseases: A Systematic Review. *Machine Learning and Knowledge Extraction*,

- 5(1), 175–198.
<https://doi.org/10.3390/make5010013>
- Singla, S., & Ramachandra, N. (2020). Comparative analysis of transformer based pre-trained NLP Models. *Int. J. Comput. Sci. Eng*, 8, 40-44.
<https://doi.org/10.26438/ijcse/v8i11.4044>
- Tabinda Kokab, S., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14, 100157.
<https://doi.org/10.1016/j.array.2022.100157>
- Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*, 12(5), 91.
<https://doi.org/10.3390/computers12050091>
- Thoyyibah, T., Abdurachman, E., Heryadi, Y., & Zahra, A. (2022). Transformer Model in Music Mood Classification. *International Journal of Applied Engineering and Technology (London)*, 4(1), 40–44. Retrieved from https://romanpub.com/resources/ijaet_v4-1-2022-08.pdf
- Tran, Q. T., Davies, K., & Roose, L. (2020). Machine learning for assessing the service transformer health using an energy monitor device. *IOSR Journal of Electrical and Electronics Engineering*, 15(6), 1-6. Retrieved from <https://www.iosrjournals.org/iosr-jeec/Papers/Vol15-Issue6/Series-1/A1506010106.pdf>
- Valencia, F., Arcos, H., & Quilumba, F. (2021). Comparison of Machine Learning Algorithms for the Prediction of Mechanical Stress in Three-Phase Power Transformer Winding Conductors. *Journal of Electrical and Computer Engineering*, 2021, 1-9.
<https://doi.org/10.1155/2021/4657696>. 2021
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
- Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019). *Learning deep transformer models for machine translation*.
<https://doi.org/10.48550/arXiv.1906.01787>
- Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). *Deep transformer models for time series forecasting: The influenza prevalence case*.
<https://doi.org/10.48550/arXiv.2001.08317>
- Wu, B., Wang, L., & Zeng, Y.-R. (2023). Interpretable tourism demand forecasting with temporal fusion transformers amid COVID-19. *Applied Intelligence*, 53(11), 14493–14514.
<https://doi.org/10.1007/s10489-022-04254-0>
- Xiao, Y., Jin, Y., Cheng, R., & Hao, K. (2022). Hybrid attention-based transformer block model for distant supervision relation extraction. *Neurocomputing*, 470, 29-39.
<https://doi.org/10.1016/j.neucom.2021.10.037>
- Yang, G. R., & Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*, 107(6), 1048–1070.
<https://doi.org/10.1016/j.neuron.2020.09.005>
- Zhong, P., Wang, D., & Miao, C. (2019). *Knowledge-enriched transformer for emotion detection in textual conversations*.
<https://doi.org/10.48550/arXiv.1909.10681>