P-ISSN : 2580-4316
https://doi.org/10.37817/ikraith-informatika.v10i2 E-ISSN : 2654-8054

Tinjauan Arsitektur Transformer Dan Penerapannya Pada Pemrosesan Bahasa Alami

Marcel Alezandro Sihombing[1], Irene Lestaria Sinaga[2], Octav Kornelius Hutagaol[3], Judea Tirta Jordan Simamora[4], Alex Septama Sihite[5]
[1)4]Program Studi Ilmu Komputer, Fakultas Ilmu Pengetahuan Alam dan Matematika, Universitas HKBP Nommensen Pematangsiantar, Pematangsiantar, Indonesia

E-mail: [1]marcelalezandro25@gmail.com, [2]irenesinaga924@gmail.com, [3]korneliushutagaol82@gmail.com, [4]judeasimamora@gmail.com, [5]alexsihite08@gmail.com

ABSTRAK

Perkembangan pesat dalam bidang pemrosesan bahasa alami (PBA) tidak terlepas dari kemunculan arsitektur Transformer yang pertama kali diperkenalkan pada tahun 2017. Arsitektur ini mengandalkan mekanisme self-attention yang memungkinkan model memproses seluruh token dalam satu waktu, berbeda dengan pendekatan sekuensial pada RNN atau LSTM. Artikel ini menyajikan tinjauan literatur mengenai arsitektur Transformer, komponen utamanya seperti multi-head attention, positional encoding, dan feed-forward network, serta berbagai penerapannya dalam tugas-tugas PBA seperti terjemahan mesin, analisis sentimen, dan pembuatan teks. Selain itu, dibahas pula kelebihan dan keterbatasan Transformer dibandingkan arsitektur sebelumnya, serta tantangan seperti kebutuhan komputasi yang tinggi dan konsumsi memori yang besar. Tinjauan ini diharapkan dapat memberikan pemahaman dasar yang komprehensif mengenai Transformer bagi peneliti dan praktisi di bidang machine learning.

Kata kunci : *Transformer, self-attention, pemrosesan bahasa alami, neural network, deep learning*

ABSTRACT

The rapid development in the field of natural language processing (NLP) cannot be separated from the emergence of the Transformer architecture first introduced in 2017. This architecture relies on a self-attention mechanism that allows the model to process all tokens at once, unlike the sequential approach in RNN or LSTM. This article presents a literature review of the Transformer architecture, its main components such as multi-head attention, positional encoding, and feed-forward network, as well as its various applications in NLP tasks such as machine translation, sentiment analysis, and text generation. In addition, the advantages and limitations of the Transformer compared to previous architectures are discussed, along with challenges such as high computational requirements and large memory consumption. This review is expected to provide a comprehensive basic understanding of the Transformer for researchers and practitioners in the field of machine learning.

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Keyword : *Transformer, self-attention, natural language processing, neural network, deep learning*

1. PENDAHULUAN

Perkembangan kecerdasan buatan, khususnya machine learning, telah mengubah cara komputer memahami bahasa manusia. Pemrosesan bahasa alami (natural language processing/NLP) mengalami kemajuan pesat, dengan tugas seperti terjemahan mesin, analisis sentimen, peringkasan teks, dan pembuatan konten otomatis kini dapat dilakukan dengan akurasi mendekati manusia(Vaswani, 2017).

Sebelum Transformer, pendekatan dominan di NLP adalah recurrent neural network (RNN) dan long short-term memory (LSTM). Keduanya memproses data secara sekuensial, sehingga sulit menangkap ketergantungan jarak jauh dan tidak dapat diparalelkan efisien(Cascade-correlation & Chunking, 1997). Pelatihan model ini lambat, meskipun modifikasi seperti bidirectional LSTM dan attention mechanism telah dicoba(Bahdanau et al., 2015).

Terobosan besar terjadi tahun 2017 ketika Google Brain memperkenalkan Transformer melalui makalah "Attention Is All You Need". Transformer mengandalkan self-attention yang memungkinkan pemrosesan paralel dan menangkap hubungan antarkata lebih baik (Vaswani, 2017). Pendekatan ini tidak hanya mempercepat proses pelatihan karena dapat diparalelkan, tetapi juga mampu menangkap hubungan antarkata dalam kalimat dengan lebih baik, terlepas dari jarak antar kata tersebut.

Sejak diperkenalkan, Transformer telah menjadi fondasi bagi berbagai model besar (large language models) seperti BERT, GPT, dan T5 yang mendominasi berbagai benchmark NLP(Kenton et al., 2019)(Radford et al., 2018). Keberhasilan Transformer juga meluas ke luar domain NLP, merambah ke bidang visi komputer, pengenalan suara, dan bioinformatika. Hal ini menunjukkan bahwa arsitektur Transformer bukan sekadar inovasi teknis, melainkan sebuah paradigma baru dalam pemodelan data sekuensial.

Meskipun demikian, Transformer bukan tanpa kelemahan. Beberapa tantangan yang masih dihadapi antara lain kebutuhan komputasi dan memori yang sangat besar, terutama saat menangani teks dengan panjang konteks yang ekstensif. Kompleksitas kuadratik dari mekanisme self-attention menjadi kendala dalam penerapannya pada perangkat dengan sumber daya terbatas(Tay & Mar, 2022). Selain itu, isu terkait bias dan etika dalam model berbasis Transformer juga semakin menjadi perhatian di kalangan peneliti(Bender & Mcmillan-major, 2021).

Berdasarkan latar belakang tersebut, artikel ini bertujuan untuk memberikan tinjauan literatur yang komprehensif mengenai arsitektur Transformer, meliputi komponen-komponen utamanya, mekanisme kerja, serta penerapannya dalam

berbagai tugas pemrosesan bahasa alami. Selain itu, artikel ini juga akan membahas kelebihan, keterbatasan, dan tantangan yang masih dihadapi oleh model berbasis Transformer. Tinjauan ini diharapkan dapat menjadi referensi awal bagi mahasiswa dan peneliti yang ingin memahami dasar-dasar Transformer sebelum mendalami implementasi teknisnya.

2. LANDASAN TEORI

Pemrosesan Bahasa Alami (Natural Language Processing)

Pemrosesan bahasa alami atau Pemrosesan bahasa alami atau Natural Language Processing (NLP) merupakan cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia, bertujuan agar komputer dapat memahami, menginterpretasikan, dan menghasilkan bahasa manusia secara bernilai. Tugas umum NLP meliputi terjemahan mesin, analisis sentimen, peringkasan teks, penjawab pertanyaan, dan pembuatan teks otomatis. Tantangan utamanya adalah ambiguitas bahasa, di mana makna kata bergantung pada konteks. Pendekatan awal menggunakan aturan linguistik dan statistika, namun kini pendekatan berbasis machine learning dan neural networks lebih dominan karena mampu menangkap pola kompleks dari data teks besar.

Recurrent Neural Network (RNN)

RNN dirancang untuk memproses data sekuensial dengan mempertahankan informasi dari langkah waktu sebelumnya melalui struktur siklus internal. Namun, RNN kesulitan menangkap ketergantungan jarak jauh (long-range dependencies) dan rentan terhadap masalah gradien hilang atau meledak selama pelatihan. (Cascade-correlation & Chunking, 1997).

Long Short-Term Memory (LSTM)

LSTM hadir sebagai solusi kelemahan RNN dengan mekanisme gerbang yang mengatur aliran informasi, sehingga dapat menyimpan informasi lebih lama. Meski lebih unggul, LSTM tetap memproses data secara sekuensial, menyulitkan paralelisasi dan memperlambat pelatihan pada data besar.

Mekanisme Perhatian (Attention Mechanism)

Mekanisme perhatian pertama kali diperkenalkan untuk meningkatkan kinerja model sequence-to-sequence berbasis RNN dalam terjemahan mesin.(Bahdanau et al., 2015). Ide dasarnya adalah memberikan bobot berbeda pada setiap bagian input saat menghasilkan output, sehingga model dapat "fokus" pada bagian relevan. Pendekatan ini menjadi cikal bakal Transformer yang sepenuhnya mengandalkan mekanisme perhatian tanpa RNN atau CNN.

Arsitektur Transformer

Arsitektur Transformer diperkenalkan oleh tim Google Brain dalam makalah berjudul "Attention Is All You Need" pada tahun 2017(Vaswani, 2017). Berbeda dengan pendekatan sebelumnya, Transformer sepenuhnya mengandalkan mekanisme perhatian diri (self-attention) tanpa menggunakan RNN atau CNN. Hal ini memungkinkan model untuk memproses seluruh token dalam satu

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

waktu secara paralel, yang secara signifikan mempercepat pelatihan dan meningkatkan kemampuan menangkap ketergantungan jarak jauh.

Perhatian Diri (Self-Attention)

Mekanisme perhatian diri adalah inti dari arsitektur Transformer. Sederhananya, perhatian diri memungkinkan setiap kata dalam sebuah kalimat untuk "memperhatikan" kata-kata lain dalam kalimat yang sama untuk menentukan konteksnya. Proses ini bekerja dengan cara menghitung skor kesamaan antara setiap kata dengan kata-kata lainnya. Semakin tinggi skor kesamaan, semakin besar perhatian yang diberikan.

Perhatian Multi-Kepala (Multi-Head Attention)

Transformer tidak hanya menggunakan satu mekanisme perhatian diri, tetapi beberapa mekanisme yang berjalan secara paralel, yang disebut perhatian multi-kepala (multi-head attention). Setiap "kepala" (head) mempelajari aspek hubungan yang berbeda antara kata-kata dalam kalimat. Misalnya, satu kepala mungkin fokus pada hubungan tata bahasa, sementara kepala lain fokus pada hubungan semantik. Hasil dari semua kepala kemudian digabungkan untuk menghasilkan representasi yang lebih kaya dan komprehensif.

Penyandian Posisi (Positional Encoding)

Salah satu kelemahan dari mekanisme perhatian diri adalah ia tidak secara inheren memahami urutan atau posisi kata dalam kalimat. Berbeda dengan RNN yang memproses kata secara berurutan, Transformer memproses semua kata sekaligus. Untuk mengatasi hal ini, Transformer menambahkan informasi posisi ke setiap kata melalui penyandian posisi (positional encoding). Informasi ini memungkinkan model untuk mengetahui posisi relatif dan absolut dari setiap kata dalam urutan, yang sangat penting untuk memahami struktur bahasa.

Jaringan Umpan Maju (Feed-Forward Network)

Setelah melewati lapisan perhatian multi-kepala, output dari setiap kata diproses lebih lanjut oleh jaringan umpan maju (feed-forward network) yang diterapkan secara independen pada setiap posisi. Jaringan ini berfungsi untuk mentransformasi representasi kata ke dalam ruang yang lebih kompleks, menangkap fitur-fitur non-linear yang tidak dapat ditangkap oleh lapisan perhatian saja.

Koneksi Residual dan Normalisasi Lapisan

Untuk mempermudah proses pelatihan dan mencegah masalah gradien yang hilang, Transformer menggunakan koneksi residual (residual connections), yaitu menambahkan input dari lapisan sebelumnya ke output lapisan saat ini. Selain itu, normalisasi lapisan (layer normalization) diterapkan untuk menstabilkan dan mempercepat proses pelatihan dengan menormalkan aktivasi di setiap lapisan.

Model Turunan Transformer

Keberhasilan arsitektur Transformer telah melahirkan

P-ISSN : 2580-4316
https://doi.org/10.37817/ikraith-informatika.v10i2 E-ISSN : 2654-8054

berbagai model besar yang mendominasi bidang NLP. Dua arsitektur paling terkenal adalah BERT dan GPT.

BERT (Bidirectional Encoder Representations from Transformers)

BERT (2018) hanya menggunakan encoder Transformer dan memahami konteks secara bidireksional (kiri dan kanan), unggul dalam klasifikasi teks, penjawab pertanyaan, dan pengenalan entitas bernama (Kenton et al., 2019).

GPT (Generative Pre-trained Transformer)

Berbeda dengan BERT, GPT menggunakan bagian decoder dari Transformer dan bersifat autoregresif, yang berarti model memprediksi kata berikutnya dalam urutan berdasarkan kata-kata sebelumnya. GPT dilatih pada data teks yang sangat besar dan mampu menghasilkan teks yang koheren dan kontekstual. Model GPT, terutama generasi terbarunya, menjadi fondasi bagi berbagai aplikasi AI generatif seperti ChatGPT(Radford et al., 2018).

Kelebihan dan Keterbatasan Transformer

Meskipun sangat sukses, Transformer memiliki kelebihan dan keterbatasan yang perlu dipahami.

Kelebihan

Arsitektur Transformer menawarkan beberapa keunggulan signifikan dibandingkan pendekatan sebelumnya:

1. Paralelisasi: Kemampuan untuk memproses seluruh urutan secara paralel membuat pelatihan Transformer jauh lebih cepat dibandingkan RNN atau LSTM yang bersifat sekuensial.
2. Penangkapan Ketergantungan Jarak Jauh: Mekanisme perhatian diri memungkinkan Transformer menangkap hubungan antara kata-kata yang berjauhan dalam sebuah kalimat dengan lebih baik.
3. Skalabilitas: Transformer dapat dilatih pada dataset yang sangat besar dan menjadi dasar bagi model-model bahasa besar (large language models).

Keterbatasan

Meskipun sangat sukses, Transformer juga memiliki beberapa keterbatasan:

1. Kompleksitas Komputasi: Mekanisme perhatian diri memiliki kompleksitas kuadratik terhadap panjang urutan ($O(N^2)$), yang berarti semakin panjang teks yang diproses, semakin besar kebutuhan komputasi dan memori yang dibutuhkan. Hal ini menjadi kendala saat menangani dokumen yang sangat panjang.
2. Sumber Daya yang Besar: Melatih model Transformer dari awal membutuhkan

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

infrastruktur komputasi yang sangat besar (GPU/TPU) dan waktu yang lama, sehingga tidak semua peneliti atau praktisi dapat melakukannya.

3. Keterbatasan Konteks: Meskipun lebih baik dari RNN, Transformer tetap memiliki batasan dalam panjang konteks yang dapat diproses secara efektif karena keterbatasan memori dan kompleksitas komputasi.

3. METODOLOGI

Metodologi yang digunakan dalam penelitian ini adalah studi literatur (literature review) dengan pendekatan sistematis untuk mengumpulkan, menganalisis, dan mensintesis berbagai sumber ilmiah terkait arsitektur Transformer dan penerapannya dalam pemrosesan bahasa alami. Penelitian ini tidak melibatkan pengumpulan data primer atau pengujian model, melainkan berfokus pada kajian mendalam terhadap teori, konsep, dan temuan dari penelitian-penelitian sebelumnya.

Strategi Pencarian Literatur

Pencarian literatur dilakukan dengan menggunakan kata kunci utama (keywords) yang relevan dengan topik penelitian. Kata kunci yang digunakan dalam proses pencarian adalah kombinasi dari istilah-istilah berikut:

a. "Transformer architecture"
b. "self-attention mechanism"
c. "natural language processing"
d. "BERT"
e. "GPT"
f. "large language models"
g. "neural network for NLP"
h. "attention is all you need"

Sumber Database

Sumber literatur diambil dari beberapa database ilmiah terpercaya yang memiliki kredibilitas tinggi dalam bidang ilmu komputer dan kecerdasan buatan, antara lain:

1. Google Scholar - untuk pencarian literatur secara umum
2. IEEE Xplore - untuk publikasi di bidang teknik dan komputer
3. ACM Digital Library - untuk publikasi di bidang komputasi
4. ScienceDirect - untuk jurnal-jurnal ilmiah terakreditasi
5. arXiv.org - untuk makalah pracetak (preprints) terbaru

Kriteria Inklusi dan Eksklusi

Untuk memastikan kualitas dan relevansi literatur yang digunakan, diterapkan kriteria inklusi dan eksklusi sebagai berikut:

Kriteria Inklusi:

1. Artikel atau makalah yang membahas arsitektur Transformer atau turunannya
2. Publikasi dalam 10 tahun terakhir (2015–2025)

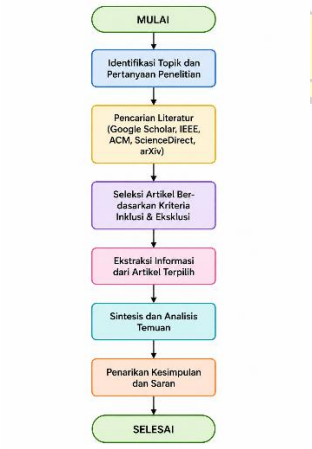
3. Artikel yang diterbitkan dalam jurnal terakreditasi atau prosiding konferensi internasional
4. Makalah yang memiliki kontribusi teoretis atau konseptual yang jelas
5. Artikel yang ditulis dalam bahasa Inggris atau Indonesia

Kriteria Eksklusi:

1. Artikel yang hanya membahas implementasi teknis tanpa penjelasan konseptual
2. Publikasi yang tidak memiliki referensi atau sumber yang jelas
3. Artikel yang tidak relevan dengan topik pemrosesan bahasa alami
4. Makalah yang tidak tersedia secara full-text

Diagram Alur Penelitian

Proses penelitian ini mengikuti alur sistematis yang digambarkan pada diagram alur berikut. Diagram ini menunjukkan tahapan yang dilalui mulai dari identifikasi topik hingga penarikan kesimpulan.



Gambar 1. Alur Metodologi Studi Literatur

Tahapan Penelitian

Berdasarkan diagram alur, tahapan penelitian meliputi:

1. Identifikasi topik Transformer dan penerapannya di NLP, dengan pertanyaan: bagaimana mekanisme kerja, kelebihan/kekurangan, dan penerapannya;
2. Pencarian literatur pada database terpercaya menggunakan kata kunci yang ditentukan;
3. Seleksi artikel berdasarkan kriteria inklusi/eksklusi;
4. Ekstraksi informasi penting dari artikel terpilih;
5. Sintesis dan analisis temuan;
6. Penarikan kesimpulan dan saran.lanjut.

4. HASIL DAN PEMBAHASAN

Analisis Mekanisme Self-Attention dalam Arsitektur Transformer

Mekanisme self-attention merupakan komponen paling fundamental yang membedakan Transformer dari arsitektur

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

sebelumnya. Berdasarkan tinjauan literatur, self-attention memungkinkan model untuk menghitung hubungan antara setiap pasangan posisi dalam urutan input secara bersamaan, tanpa bergantung pada ketergantungan sekuensial seperti pada RNN atau LSTM(Vaswani, 2017). Hal ini menjadikan Transformer sebagai arsitektur yang sangat efisien dalam menangkap ketergantungan jarak jauh (long-range dependencies) karena setiap token dapat "memperhatikan" token lainnya dalam kalimat yang sama.

Secara konseptual, self-attention mengubah setiap token menjadi tiga vektor: kueri (query), kunci (key), dan nilai (value)(Vaswani, 2017). Kueri berfungsi sebagai pertanyaan dari sebuah token, kunci adalah informasi dari token lain, dan nilai adalah konten aktual token. Perbandingan kueri terhadap semua kunci menentukan bobot kontribusi setiap nilai terhadap representasi akhir token.

Dari hasil analisis, terdapat beberapa keunggulan utama self-attention dibandingkan pendekatan sekuensial. Pertama, self-attention memungkinkan paralelisasi penuh selama proses pelatihan karena seluruh token dalam satu urutan dapat diproses secara bersamaan. Kedua, self-attention memiliki kemampuan superior dalam menangkap hubungan antarkata yang berjauhan dalam sebuah kalimat. Ketiga, fleksibilitas self-attention memungkinkan penerapannya tidak hanya pada teks tetapi juga pada berbagai jenis data lain. Meskipun demikian, self-attention juga memiliki keterbatasan, yaitu kompleksitas komputasi yang bersifat kuadratik terhadap panjang urutan ($O(N^2)$) serta kebutuhan data pelatihan dalam jumlah yang sangat besar(Tay & Mar, 2022).

Perbandingan Transformer dengan Arsitektur Sebelumnya

Berdasarkan tinjauan literatur, Transformer menunjukkan peningkatan kinerja yang signifikan dibandingkan dengan RNN dan LSTM dalam berbagai tugas pemrosesan bahasa alami. Perbedaan mendasar antara Transformer dan arsitektur sebelumnya dapat dirangkum pada Tabel 1.

Tabel 1. Perbandingan arsitektur RNN/LSTM dengan Transformer

Aspek	RNN / LSTM	Transformer
Paralelisasi	Tidak dapat diparalelkan (sekuensial)	Dapat diparalelkan sepenuhnya
Penanganan Ketergantungan Jarak Jauh	Terbatas (masalah gradien hilang)	Sangat baik (self-attention)
Kecepatan Pelatihan	Lambat (proses bertahap)	Cepat (proses paralel)
Skalabilitas	Terbatas	Sangat skalabel
Kebutuhan Komputasi	Rendah hingga sedang	Sangat tinggi

Keunggulan Transformer dalam hal paralelisasi dan penanganan ketergantungan jarak jauh menjadi faktor utama yang mendorong adopsi luas arsitektur ini dalam berbagai aplikasi NLP(Vaswani, 2017). Model-model seperti BERT dan GPT yang dibangun di atas fondasi Transformer

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

berhasil mencapai kinerja terbaik pada berbagai tolok ukur NLP .

Penerapan Transformer dalam Tugas Pemrosesan Bahasa Alami

Terjemahan Mesin (Machine Translation)

Transformer awalnya dirancang untuk terjemahan mesin dan terbukti mengungguli RNN dalam berbagai pasangan bahasa (Vaswani, 2017). Self-attention memungkinkan model menangkap hubungan antarkata dalam kalimat sumber dan target lebih baik, menghasilkan terjemahan lebih akurat, bahkan pada bahasa dengan sumber daya terbatas (Vaswani, 2017).

Analisis Sentimen (Sentiment Analysis)

Dalam analisis sentimen, model berbasis Transformer seperti BERT dan GPT menunjukkan akurasi tinggi dalam memahami emosi, mencapai akurasi hingga 70% (Kenton et al., 2019). Kemampuan self-attention menangkap konteks secara bidireksional memungkinkan pemahaman sentimen lebih akurat pada kalimat kompleks atau ambigu.

Pembuatan Teks (Text Generation)

Transformer decoder-only seperti GPT menjadi fondasi pembuatan teks otomatis. Model GPT generasi terbaru menghasilkan teks koheren dan kontekstual, sering tak terbaca dari teks manusia (Radford et al., 2018). didukung pelatihan data besar dan self-attention yang memahami konteks mendalam.

Aplikasi Lintas Domain

Keberhasilan Transformer meluas ke visi komputer melalui Vision Transformer (ViT), pengenalan suara, dan analisis deret waktu, menunjukkan self-attention memiliki sifat umum untuk berbagai data sekuensial.

Kelebihan dan Keterbatasan Transformer

Kelebihan utama Transformer: kemampuan menangkap konteks panjang, efisiensi pelatihan melalui paralelisasi, skalabilitas pada dataset besar, dan fleksibilitas untuk berbagai tugas dan jenis data (Vaswani, 2017). Keterbatasannya: kompleksitas komputasi kuadratik ($O(N^2)$), kebutuhan data pelatihan sangat besar, ketergantungan pada infrastruktur mahal, isu bias dan etika dari data pelatihan, serta batasan panjang konteks karena keterbatasan memori (Tay & Mar, 2022)(Bender & Mcmillan-major, 2021).

Tantangan dan Penelitian Terkini

Penelitian terkini berfokus pada efisiensi komputasi melalui sparse attention dan model efisien (DistilBERT, ALBERT), adaptasi untuk bahasa sumber daya terbatas termasuk bahasa daerah Indonesia, integrasi multimodal (CLIP, DALL-E), serta upaya mengurangi bias dan memastikan penerapan adil dan etis (Tay & Mar, 2022)(Bender & Mcmillan-major, 2021).

5. KESIMPULAN

Arsitektur Transformer telah mengubah paradigma NLP melalui mekanisme self-attention yang memungkinkan pemrosesan paralel dan penangkapan ketergantungan

jarak jauh lebih efektif daripada RNN dan LSTM. Komponen utama (multi-head attention, positional encoding, feed-forward network) bekerja sinergis menghasilkan representasi bahasa yang kaya dan kontekstual. Transformer melahirkan model turunan seperti BERT dan GPT yang mendominasi berbagai tugas NLP, serta meluas ke visi komputer dan pengenalan suara. Meskipun unggul dalam skalabilitas dan fleksibilitas, Transformer masih menghadapi tantangan kompleksitas komputasi tinggi, kebutuhan data dan infrastruktur besar, serta isu etika bias dalam data pelatihan. Penelitian terkini terus mengatasi keterbatasan ini melalui model lebih efisien dan adaptasi untuk berbagai bahasa, termasuk yang sumber dayanya terbatas.

6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pengampu mata kuliah Machine Learning yang telah memberikan bimbingan dan arahan dalam penyusunan jurnal ini. Ucapan terima kasih juga disampaikan kepada rekan-rekan mahasiswa Program Studi Ilmu Komuangan, Fakultas Ilmu Pengetahuan Alam dan Matematika, Universitas HKBP Nommensen Pematangsiantar yang telah memberikan dukungan dan masukan selama proses penulisan.

DAFTAR PUSTAKA

Bahdanau, D., Cho, K., & Bengio, Y. (2015). *N EURAL M ACHINE T RANSLATION*. 1–15.

Bender, E. M., & Mcmillan-major, A. (2021). On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ? In *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 310, 2021, Virtual Event, Canada* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Cascade-correlation, R., & Chunking, N. S. (1997). *2 PREVIOUS WORK*. *9*(8), 1–32.

Kenton, M. C., Kristina, L., & Devlin, J. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. *Mlm*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2018). *Language Models are Unsupervised Multitask Learners*.

Tay, Y., & Mar, L. G. (2022). *Efficient Transformers : A Survey*. *August 2020*, 1–39.

Vaswani, A. (2017). *Attention Is All You Need*. *Nips*.