

## Analisis Bibliometrik Tren Penelitian *Natural language processing* pada *Chatbot* dari Tahun 2019-2023

M. Salman Alfharezi <sup>1</sup>, Dani Marcelo Septiano Kuntardi <sup>2</sup>

<sup>1,2</sup> Program Studi Ilmu Perpustakaan, Fakultas Ilmu Budaya, Universitas Diponegoro  
Jl. Prof. Soedarto, SH, Kampus Undip Tembalang, Semarang, Indonesia  
E-mail: <sup>1</sup>salmanalfharezii@gmail.com, <sup>2</sup>danimarcelo45@gmail.com

### Abstract

This research presents the results of an analysis on the development of research in natural language processing for the implementation of artificial intelligence in the form of chatbots. The research conducted a systematic mapping of relevant scientific publications from 2019 to 2023. Its primary objectives are to identify dominant study fields, explore the most intriguing topics, and analyze shifts in research trends over this time period. The findings highlight that the dominant study fields in research conducted between 2019 and 2023 include computer science, engineering, and mathematics. This emphasis underscores a primary focus on technical development and aspects of artificial intelligence technology. Additionally, the research identifies significant interest in the application of natural language processing within chatbots, particularly concerning participants, datasets, and their relation to COVID-19. These findings indicate that the interaction between humans and artificial intelligence through chatbots, as well as the effects of the COVID-19 pandemic, remain compelling topics for researchers. The study emphasizes the crucial role of high-quality and representative datasets in the development of effective and accurate natural language processing models. Furthermore, by tracking the keywords that determine research focus year by year, the research reveals shifts in trends and interests. This comprehensive overview provides valuable insights into the evolution of research in natural language processing for the implementation of artificial intelligence in chatbots. Such information is valuable for researchers, practitioners, and developers interested in continuously enhancing natural language processing systems and harnessing artificial intelligence within chatbot contexts.

**Keywords:** *Natural Language Processing; Artificial Intelligence; Chatbot.*

### Abstrak

Penelitian ini menyajikan hasil analisis tentang perkembangan penelitian *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot*. Melalui pemetaan sistematis terhadap publikasi ilmiah terkait selama periode 2019 sampai 2023. Penelitian ini bertujuan untuk mengidentifikasi bidang studi yang dominan, topik yang paling menarik minat, serta pergeseran tren penelitian dari waktu ke waktu. Hasil penelitian menunjukkan bahwa bidang studi yang paling dominan dalam penelitian 2019-2023 adalah *computer science*, *engineering*, dan *mathematics*. Hal ini menunjukkan fokus utama pada pengembangan teknis dan aspek-aspek teknologi *artificial intelligence*. Selanjutnya, penelitian juga menemukan minat yang signifikan dalam penggunaan *natural language processing* pada *chatbot* yang terkait dengan *participant*, *dataset*, dan *covid*. Hal ini mengindikasikan bahwa interaksi antara manusia dan *artificial intelligence* menggunakan *chatbot* serta dampak pandemi covid-19, menjadi topik menarik bagi para peneliti. Penelitian ini juga menekankan pentingnya *dataset* yang berkualitas dan representatif dalam pengembangan model *natural language processing* yang efektif dan akurat. Selain itu, melalui pemetaan kata kunci yang menentukan fokus penelitian dari tahun ke tahun, terlihat pergeseran tren dan minat yang ada, dengan demikian penelitian ini memberikan gambaran yang komprehensif tentang perkembangan penelitian *natural language processing* pada

implementasi *artificial intelligence* dalam bentuk *chatbot* selama periode 2019 sampai 2023. Informasi ini dapat berguna bagi peneliti, praktisi, dan pengembang yang tertarik untuk terus memperbaiki sistem *natural language processing* dan memanfaatkan *artificial intelligence* dalam konteks *chatbot*.

**Kata Kunci:** *Natural Language Processing; Artificial Intelligence; Chatbot.*

## PENDAHULUAN

Perkembangan teknologi informasi saat ini berkembang dengan begitu pesat. Teknologi Informasi merupakan teknologi yang memproses data untuk menghasilkan informasi berkualitas, relevan, akurat, dan tepat waktu yang digunakan dalam berbagai konteks seperti pribadi, bisnis, dan pemerintahan, serta menjadi dasar penting dalam pengambilan keputusan (Wardiana, 2002). Berkat kemampuan manusia dalam inovasi serta kreatifitasnya telah muncul inovasi teknologi informasi, salah satunya *artificial intelligence*. AI (*artificial intelligence*) merupakan kecerdasan buatan yang meniru perilaku manusia (Pakpahan, 2021). Secara sederhana, *artificial intelligence* memiliki kemampuan untuk meniru kecerdasan manusia dalam menyelesaikan berbagai tugas dan pekerjaan. Bahkan, *artificial intelligence* memiliki potensi untuk mengambil alih peran manusia dalam beberapa jenis pekerjaan. Perkembangan *artificial intelligence* sudah berkembang tidak hanya sekedar kecerdasan buatan. Baru-baru ini penggunaan Chat GPT menjadi sesuatu yang disorot dalam implementasi *artificial intelligence* dalam bentuk *chatbot*.

Aplikasi tersebut dibangun atas dasar pendekatan *natural language processing* yang merupakan salah satu bidang keilmuan dalam kecerdasan buatan yang mempelajari mengenai interaksi antara komputer dan bahasa alami yang digunakan oleh manusia. Model komputasi semacam ini memiliki manfaat dalam mempermudah komunikasi antara manusia dan komputer, terutama dalam konteks pencarian informasi (Maitri & Sutopo, 2019). Berkembangnya teknologi memberikan proses kegiatan manusia menjadi lebih mudah dari sebelumnya. Terdapat banyak pilihan produk *artificial intelligence*, *chatbot* menjadi salah satu favorit orang-orang dalam menggunakan *artificial intelligence* seperti Chat GPT dan lain sejenisnya. Perkembangan *artificial intelligence* seperti *chatbot* telah mengalami pertumbuhan yang sangat masif setelah pandemi Covid-19. Seiring dengan perubahan dalam gaya hidup dan kebiasaan komunikasi yang dipengaruhi oleh pandemi, *artificial intelligence* telah menjadi semakin penting dalam menyediakan solusi dan layanan yang mudah diakses secara digital (Kuppuswamy, et al., 2023). Teknologi ini telah digunakan secara luas dalam berbagai sektor, termasuk kesehatan, bisnis digital, dan banyak lagi.

Penelitian sebelumnya yang berjudul “Research Trends in the International Literature on Natural Language Processing, 2000-2019 – A Bibliometric Study” ditulis oleh Roberto E Lopez-Martinez dan Gerardo Sierra dari Journal of Scientometrics Research tahun 2020 yang menyajikan tinjauan menggunakan metode bibliometrik terhadap keadaan penelitian di bidang *natural language processing*, yang mencakup metode teknik semantik, teknik statistik, atau kombinasi keduanya dalam rentang tahun 2000 sampai 2019. Hasilnya menunjukkan bahwa, dalam beberapa tahun terakhir *natural language processing* sering berinteraksi dengan subjek yang terkait dengan *artificial intelligence*, seperti *deep*

*learning, machine learning* dan *neural networks* (Lopez-Martinez & Sierra, 2020). Penelitian ini menggunakan metode bibliometrik untuk melihat sebaran keadaan penelitian internasional tentang *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot* dengan meninjau dari penelitian yang bersumber dari Scopus dengan mengidentifikasi dan melihat area penelitian yang mana saja yang banyak dikaji selama rentang tahun 2019 sampai 2023 setelah pandemi Covid-19.

## METODE PENELITIAN

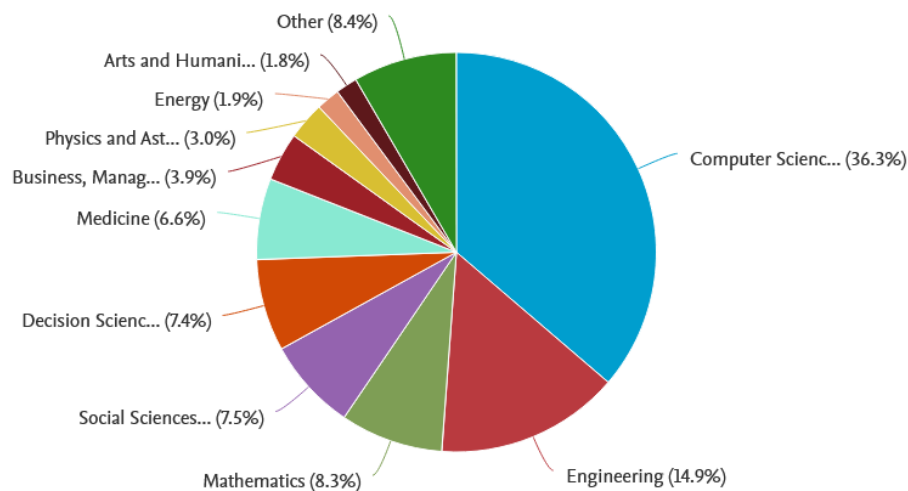
Metode penelitian yang akan digunakan dalam penelitian ini mencakup dua pendekatan utama: *citation network analysis* dan *text mining*. *Citation network analysis* merupakan metode yang digunakan untuk menganalisis jaringan kutipan antara publikasi ilmiah. Jaringan kutipan terdiri dari simpul yang merepresentasikan artikel penelitian atau penulis, serta tautan yang menghubungkan kutipan antara penelitian-penelitian tersebut. Jaringan kutipan ini memiliki sifat *acyclic*, artinya tidak terdapat putaran tertutup dalam jaringan, namun dapat berkembang dari waktu ke waktu dengan adanya publikasi baru (Leicht, et al., 2007). Hubungan kutipan dalam jaringan ini sering mencerminkan korelasi semantik antara dua penelitian (Börner, et al., 2003). Analisis jaringan kutipan dapat digunakan untuk menyelidiki hubungan antara publikasi ilmiah, mengidentifikasi perkembangan ilmu pengetahuan, dan mengungkap inovasi teknologi. Diasumsikan bahwa penelitian yang dikutip memiliki minat yang sama atau berbagi topik yang serupa, dengan menyelidiki struktur dalam jaringan kutipan dapat memahami interkoneksi antara karya ilmiah dan struktur pengetahuan dalam berbagai disiplin ilmu. Pada penelitian kali ini VOSviewer akan dijadikan alat untuk melakukan *citation network analysis*.

Sedangkan *text mining* merupakan pendekatan yang digunakan untuk mengekstraksi informasi dan merangkum konten dari berbagai dokumen tertulis dan kumpulan teks seperti situs web, surat kabar, buku, email, dan artikel. Pendekatan ini melibatkan penerapan pemrosesan bahasa alami untuk mengubah teks yang tidak terstruktur dalam dokumen menjadi format data terstruktur, sehingga memungkinkan identifikasi fakta dan hubungan (Aggarwal & Zhai, 2012). *Text mining* digunakan untuk mengekstraksi istilah kunci dari penelitian untuk mengklasifikasikan topik utama dalam suatu penelitian, dengan menggabungkan kedua pendekatan ini, penelitian ini akan meninjau hubungan antara publikasi ilmiah melalui analisis jaringan kutipan, serta melakukan analisis teks untuk mengekstraksi informasi penting dan meringkas konten dari dokumen-dokumen yang relevan.

## HASIL DAN PEMBAHASAN

Proses tinjauan penelitian ini terdiri dari tiga fase. Pertama, mengumpulkan data dari database yang bersumber dari Scopus. Kedua, catatan bibliografi data diambil untuk membuat dataset untuk bidang penelitian yang kemudian digunakan untuk melihat hasil visualisasi di VOSviewer. Ketiga, memanfaatkan *text mining* untuk mengekstraksi dan mendistribusi kata kunci informasi dari setiap tahun. Koleksi yang diambil meliputi

artikel jurnal, makalah konferensi, buku dan bab buku. Kombinasi kata kunci yang dipakai adalah “Chatterbot OR Chatbot” sebagai string pencarian untuk pencarian topik dan membatasi pengumpulan literatur artikel yang diterbitkan selama periode 2019–2023 (data diambil pada 24 Mei 2023). Ini berarti bahwa semua publikasi yang mengandung istilah-istilah ini dalam daftar abstrak, judul, atau kata kunci diambil dan dimasukkan dalam kumpulan data awal penelitian. Kemudian diambil catatan bibliografi lengkap dari publikasi ini untuk membuat kumpulan data untuk analisis lebih lanjut.



Gambar 1. Hasil Pemetaan Bidang Studi

Sumber: Hasil Perhitungan VOSviewer

Pada hasil pemetaan tren bidang studi terdapat 10 bidang studi yang paling mendominasi (Gambar 1) dengan tema penelitian *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot*, dari 7785 hasil pencarian yaitu, *computer science* sebanyak 2825 hasil (36,3%), *engineering* sebanyak 1159 hasil (14,9%), *mathematics* sebanyak 648 hasil (8,35), *social sciences* sebanyak 582 hasil (7,5%), *decision sciences* sebanyak 579 hasil (7,4%), *medicine* sebanyak 511 hasil (6,6%), *business, management and accounting* sebanyak 306 hasil (3,9%), *physics and astronomy* sebanyak 236 hasil (3%), *energy* sebanyak 149 (1,9%), dan *arts and humanities* sebanyak 137 hasil (1,8%). Untuk bidang studi lainnya yaitu sebanyak (8,4%) hasil terdiri dari *psychology* sebanyak 108 hasil, *materials science* sebanyak 98 hasil, *health professions* sebanyak 84 hasil, *biochemistry, genetics and molecular biology* sebanyak 56 hasil, *environmental science* sebanyak 54 hasil, *chemical engineering* sebanyak 48 hasil, *neuroscience* sebanyak 30 hasil, *nursing* sebanyak 26 hasil, *agricultural and biological sciences* sebanyak 24 hasil, *chemistry* sebanyak 22 hasil, *multidisciplinary* sebanyak 19 hasil, *earth and planetary sciences* sebanyak 12 hasil, *immunology and microbiology* sebanyak 9 hasil, *pharmacology, toxicology and pharmaceuticals* sebanyak 6 hasil, *dentistry* sebanyak 1 hasil, dan *veterinary* sebanyak 1 hasil. Berdasarkan hasil pemetaan tersebut bidang studi *computer science*, *engineering*, dan *mathematics*





*agents* sebanyak 76 hasil, *human computer interaction* sebanyak 59 hasil, *machine learning* sebanyak 54 hasil, *learning systems* sebanyak 50 hasil, *deep learning* sebanyak 43 hasil, dan *students* sebanyak 42 hasil. Selanjutnya tahun 2020 terdapat 10 kata kunci yang paling mendominasi yaitu, *chatbot or chatbots* sebanyak 622 hasil, *artificial intelligence* sebanyak 156 hasil, *natural language processing or Natural language processing* sebanyak 139 hasil, *natural language processing systems* sebanyak 98 hasil, *conversational agents* sebanyak 89 hasil, *human* sebanyak 58 hasil, *machine learning* sebanyak 53 hasil, *human computer interaction* sebanyak 48 hasil, *natural languages* sebanyak 47 hasil, dan *user interfaces* sebanyak 47 hasil.

Untuk tahun 2021 terdapat 10 kata kunci yang paling mendominasi yaitu, *chatbot or chatbots* sebanyak 954 hasil, *artificial intelligence* sebanyak 225 hasil, *natural language processing systems* sebanyak 130 hasil, *natural language processing* sebanyak 110 hasil, *conversational agents* sebanyak 91 hasil, *machine learning* sebanyak 88 hasil, *human* sebanyak 74 hasil, *students* sebanyak 72 hasil, *deep learning* sebanyak 67 hasil, dan *article* sebanyak 55 hasil. Selanjutnya tahun 2022 terdapat 10 kata kunci yang paling mendominasi yaitu, *chatbot or chatbots* sebanyak 1206 hasil, *artificial intelligence* sebanyak 223 hasil, *natural language processing systems* sebanyak 183 hasil, *natural language processing* sebanyak 141 hasil, *natural language* sebanyak 130 hasil, *machine learning* sebanyak 111 hasil, *conversational agents* sebanyak 107 hasil, *language processing* sebanyak 95 hasil, *human* sebanyak 91 hasil dan *deep learning* sebanyak 84 hasil. Terakhir tahun 2023 terdapat 10 kata kunci yang paling mendominasi yaitu, *chatbot or chatbots* sebanyak 387 hasil, *artificial intelligence* sebanyak 111 hasil, *human or humans* sebanyak 96 hasil, *natural language processing* sebanyak 51 hasil, *natural language processing systems* sebanyak 42 hasil, *natural language* sebanyak 35 hasil, *conversational agents* sebanyak 33 hasil, *machine learning* sebanyak 33 hasil, *language processing* sebanyak 31 hasil dan Covid-19 sebanyak 29 hasil.

Pada hasil pemetaan menunjukkan perubahan kata kunci yang mendominasi dari tahun ke tahun. Hal ini mencerminkan pergeseran minat penelitian dan perkembangan tren dalam bidang *natural language processing* pada chatbot. Contohnya, kata kunci "*chatbot or chatbots*" menjadi salah satu kata kunci yang dominan dalam semua tahun yang diamati, menunjukkan pentingnya *chatbot* dalam konteks *natural language processing* pada *artificial intelligence*. Sedangkan kata kunci "Covid-19" muncul pada tahun 2023, mengindikasikan adanya pengaruh pandemi terhadap penelitian dalam bidang ini. Hal ini menunjukkan upaya peneliti dalam mempelajari penggunaan *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot* dalam menghadapi pandemi dan memahami dampaknya dalam berbagai aspek.

## SIMPULAN

Berdasarkan hasil penelitian yang disajikan dapat disimpulkan beberapa poin yang relevan yaitu, bidang studi yang dominan dalam penelitian *natural language processing* pada *chatbot* adalah *computer science*, *engineering*, dan *mathematics*. Hal ini menunjukkan fokus utama dalam pengembangan dan pemahaman aspek-aspek teknis dari *artificial intelligence*. Selanjutnya, hasil penelitian juga menunjukkan minat yang signifikan dalam

penggunaan *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot* yang terkait dengan *participant*, *dataset*, dan *covid*. Hal ini mengindikasikan bahwa *natural language processing* dalam analisis data terkait memahami interaksi antara manusia dan sistem *artificial intelligence* yang menggunakan *chatbot* serta pandemi Covid-19 telah menjadi topik yang menarik bagi para peneliti. Selain itu, Penelitian tersebut juga menekankan pentingnya *dataset* yang berkualitas dan representatif dalam pengembangan model *natural language processing* yang efektif dan akurat. Terakhir dari hasil pemetaan kata kunci yang menentukan fokus penelitian yang berbeda setiap tahunnya, mencerminkan pergeseran tren dan minat dalam *natural language processing* pada implementasi *artificial intelligence* dalam bentuk *chatbot* dari tahun ke tahun.

## DAFTAR PUSTAKA

- Aggarwal, C. C. & Zhai, C. (2012). *Mining Text Data*. Springer.
- Barakhnin, V. B. et al. (2018). "The Automatic Processing of the Texts in Natural Language, Some Bibliometric Indicators of the Current State of this Research Area". *Journal of Physics: Conference Series*, 117(1), p. 012001.
- Börner, K., Chen, C. & Boyack, K. W. (2003). "Visualizing Knowledge Domains". *Annual Review of Information Science and Technology*, Volume 37, p. 179–255.
- Chen, X. et al. (2018). "A Bibliometric Review of *Natural language processing* Empowered Mobile Computing". *Wireless Communications And Mobile Computing*. 2018
- Chollet, F. (2017). *Deep Learning With Python*. Manning Publications.
- Hays, D. G. (1967). *Introduction to Computational Linguistics, Mathematical Linguistics and Automatic Language Processing*. American Elsevier Publishing Co.
- Kuppuswamy, P. et al. (2023). "A Study Examining the Effectiveness of *Artificial intelligence* in Predicting Human Behavior Changes in Their Lifestyle of Post Covid Exposure". *Journal of Survey in Fisheries Sciences*, 10(1S), pp. 3449-3463.
- Leicht, E. A., Clarkson, G., Shedden, K. & Newman, M. E., (2007). "Large-Scale Structure of Time Evolving Citation Networks". *European Physical Journal B*, 59(1), p. 75–83.
- Lopez-Martinez, R. E. & Sierra, G. (2020). "Research Trends in the International Literature on Natural Language Processing, 2000-2019 – A Bibliometric Study". *Journal of Scientometric Research*, 9(3), pp. 310-318.
- Maitri, A. L. & Sutopo, J. (2019). *Rancang Bangun Chatbot sebagai Pusat Informasi Lembaga Kursus dan Pelatihan Menggunakan Pendekatan Natural Language Processing*. Universitas Teknologi Yogyakarta.
- Manning, C. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press Ltd.
- Mäntylä, M. V., Graziotin, D. & Kuutila, M. (2018). "The Evolution of Sentiment Analysis: A Review of Research Topics, Venues and Top Cited Papers". *Computer Science Review*, Volume 27, pp. 16-33.



- Pakpahan, R. (2021). "Analisis Pengaruh Implementasi *Artificial intelligence* dalam Kehidupan Manusia". *JISICOM (Journal of Information System, Informatics and Computing)*, 5(2), pp. 506-513.
- Pustejovsky, J. & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning*. O'Reilly.
- Wang, X. Y. et al. (2019). "Tracking Knowledge Evolution, Hotspots and Future Directions of Emerging Technologies in Cancers Research: A Bibliometrics Review". *Journal of Cancer*, 10(12), pp. 2643-53.
- Wardiana, W., (2002). *Perkembangan Teknologi Informasi di Indonesia*. Universitas Komputer Indonesia.
- Zu, X. & Yu, D. (2019). "Bibliometrics Analysis On Big Data Research (2009–2018)". *Journal of Data, Information and Management*, 1(1), pp. 3-15.