

# Integrasi Natural Language Processing (NLP) dalam Pengembangan dan Optimalisasi Data Warehouse: Sebuah Systematic Literature Review

Ririn Syahfitri<sup>1</sup>, Widodo Sihotang<sup>2</sup>, Haifah Ilmi<sup>3</sup>, Rais Affaruq Zunnurain<sup>4</sup>

<sup>1,2,3,4</sup> Fakultas Teknologi dan Bisnis, Program Studi Bisnis Digital  
Universitas Putra Abadi Langkat. Indonesia

## ARTICLE INFO

### Article history:

Received: Jan 2, 2026  
Revised: Jan 15, 2026  
Accepted: Jan 24, 2026

### Keywords:

Data Warehouse;  
Natural Language Processing;  
Semantic ETL;  
Systematic Literature Review;  
Text-to-SQL.

## ABSTRAK

Perkembangan big data dan meningkatnya kebutuhan akses data yang cepat serta akurat mendorong integrasi teknologi Natural Language Processing (NLP) dalam arsitektur data warehouse modern. Penelitian ini bertujuan untuk memetakan metode NLP yang digunakan dalam pengembangan dan optimalisasi data warehouse serta mengevaluasi dampaknya terhadap kualitas data, kinerja query, dan efisiensi proses ETL. Metode yang digunakan adalah Systematic Literature Review berbasis protokol PRISMA dengan sumber data dari Scopus, IEEE Xplore, SpringerLink, ScienceDirect, dan ACM Digital Library, mencakup publikasi tahun 2014–2024. Hasil kajian menunjukkan bahwa NLP memainkan peran penting dalam otomatisasi query, semantic ETL, ekstraksi metadata, text-to-SQL, dan pengelolaan data tidak terstruktur. Integrasi NLP terbukti meningkatkan relevansi hasil pencarian, mempercepat pemrosesan data, serta memperluas fungsi data warehouse menjadi sistem analitik berbasis pemahaman bahasa alami. Penelitian ini memberikan kontribusi berupa pemetaan komprehensif teknologi NLP untuk data warehouse, identifikasi tren penelitian, serta gap yang masih terbuka. Rekomendasi penelitian selanjutnya mencakup eksperimen implementasi pada lingkungan industri, pengembangan dataset bahasa Indonesia untuk text-to-SQL, dan integrasi NLP dengan platform analitik berbasis cloud.

*This is an open access article under the CC BY-NC license.*



## Corresponding Author:

Ririn Syahfitri  
Fakultas Teknologi dan Bisnis, Program Studi Bisnis Digital  
Universitas Putra Abadi Langkat. Indonesia  
Jl. Letjen R. Soeprapto No.10, Sumatera Utara 20814. Indonesia  
Email: ririn12@gmail.com

## 1. PENDAHULUAN

Perkembangan big data mendorong organisasi untuk mengelola data dalam skala besar secara lebih efektif melalui penerapan data warehouse. Namun, implementasi data warehouse masih menghadapi berbagai tantangan seperti kompleksitas query, proses ETL yang rumit, integrasi data tidak terstruktur, serta tuntutan akses data yang cepat dan akurat (Iskandar et al., 2024). Seiring dengan kemajuan Artificial Intelligence, khususnya Natural Language Processing (NLP), berbagai pendekatan mulai dimanfaatkan untuk meningkatkan kinerja data warehouse, antara lain melalui otomatisasi query, semantic search, ekstraksi metadata, text-to-SQL, dan peningkatan kualitas data. Oleh karena itu, diperlukan kajian yang memetakan bagaimana teknologi NLP diintegrasikan secara optimal dalam arsitektur data warehouse modern (Setiawan et al., 2023).

Meskipun integrasi NLP dalam data warehouse menunjukkan potensi yang signifikan, hingga kini belum tersedia pemetaan komprehensif terkait pola dan strategi integrasi yang digunakan. Selain itu, belum terdapat kejelasan mengenai metode NLP yang paling efektif dalam

meningkatkan performa, kualitas data, dan efisiensi proses pada data warehouse. Rumusan masalah ini menjadi dasar bagi perlunya kajian sistematis yang mendalam (Andriyani et al., 2024).

Kajian yang secara khusus menggabungkan aspek teknis data warehouse dengan perkembangan NLP terbaru, termasuk pendekatan berbasis transformer, masih sangat terbatas. Literatur yang ada cenderung membahas keduanya secara terpisah, sehingga belum tersedia tinjauan sistematis yang menyoroti tren penelitian, tantangan implementasi, serta arah pengembangan teknologi di masa depan. Keterbatasan tersebut menunjukkan adanya kesenjangan penelitian yang perlu diisi melalui studi literatur yang lebih terstruktur (Adekamwa et al., 2024).

Penelitian ini bertujuan untuk: 1. mengidentifikasi metode NLP yang telah diterapkan dalam pengembangan dan optimalisasi data warehouse; 2. menganalisis model integrasi NLP seperti text-to-SQL, semantic ETL, knowledge graph, serta otomatisasi query; 3. mengevaluasi efektivitas integrasi NLP terhadap performa dan kualitas data dalam data warehouse; 4. serta mengidentifikasi kesenjangan penelitian dan peluang pengembangan di masa mendatang.

## 2. METODE

Penelitian ini menggunakan desain Systematic Literature Review (SLR) yang mengacu pada pedoman PRISMA untuk memastikan proses identifikasi, seleksi, dan analisis literatur dilakukan secara sistematis dan dapat direplikasi. Metode SLR dipilih karena mampu memberikan pemetaan komprehensif mengenai perkembangan penelitian terkait integrasi Natural Language Processing (NLP) dalam data warehouse (Tribuana et al., 2025).

Kriteria inklusi ditetapkan untuk menjamin relevansi artikel yang dianalisis. Artikel yang dipertimbangkan adalah publikasi ilmiah berjenis empiris, eksperimen, atau model arsitektur yang diterbitkan dalam periode 2014 hingga 2024. Artikel juga harus berfokus pada integrasi NLP dengan sistem data warehouse dan berasal dari jurnal atau prosiding terindeks Scopus, Web of Science (WoS), atau SINTA. Sebaliknya, kriteria eksklusi mencakup artikel non-ilmiah, tinjauan tanpa metode, publikasi non-peer-reviewed, serta penelitian yang menggunakan NLP namun tidak relevan dengan konteks data warehouse. Dengan demikian, hanya artikel yang memenuhi standar akademik dan memiliki kontribusi signifikan yang disertakan dalam analisis (Solehah & Hendriani, 2024).

Literatur dikumpulkan dari lima database ilmiah utama, yaitu Scopus, IEEE Xplore, SpringerLink, ScienceDirect, dan ACM Digital Library, untuk memastikan cakupan pencarian yang lengkap dan representatif (Darmawan Napitupulu et al., n.d.). Pencarian dilakukan menggunakan kombinasi kata kunci seperti "Natural Language Processing", "NLP integration", "data warehouse optimization", "text-to-SQL", "semantic ETL", dan "automated data management", yang dirancang untuk menangkap berbagai penelitian yang relevan dengan topik integrasi NLP dalam data warehouse.

Proses penyaringan mengikuti tahapan PRISMA, dimulai dari tahap identification untuk mengumpulkan semua artikel dari database, kemudian screening dengan menghapus duplikasi serta menilai relevansi berdasarkan judul dan abstrak. Selanjutnya, pada tahap eligibility, artikel dianalisis secara penuh untuk memastikan kesesuaian dengan kriteria inklusi. Pada tahap akhir, yaitu included studies, hanya artikel yang berkualitas dan relevan yang dipilih untuk dianalisis lebih lanjut (Utami et al., 2021).

Setiap artikel yang lolos evaluasi kemudian melalui proses ekstraksi data, di mana informasi penting dikumpulkan meliputi nama penulis, tahun publikasi, metode NLP yang digunakan, arsitektur data warehouse yang diusulkan, tujuan integrasi, hasil kinerja yang dicapai, tantangan implementasi, serta peluang penelitian masa depan. Untuk memastikan kualitas penelitian yang ditinjau, dilakukan penilaian kualitas menggunakan dua instrumen, yaitu CASP (Critical Appraisal Skills Programme) dan JBI Critical Appraisal Checklist. Kedua instrumen tersebut membantu menilai integritas metodologis, validitas hasil, dan transparansi pelaporan dari setiap artikel sehingga hanya penelitian yang kredibel dan berkualitas yang menjadi dasar temuan SLR ini (Widiana et al., 2025).

## 3. HASIL DAN PEMBAHASAN

### Gambaran Umum Studi yang Terpilih

Bagian ini menyajikan gambaran umum mengenai studi-studi yang terpilih setelah melalui proses seleksi berbasis prosedur PRISMA. Berdasarkan hasil pencarian dan penyaringan pada lima database utama, diperoleh sejumlah artikel yang memenuhi kriteria inklusi dan relevan dengan fokus penelitian mengenai integrasi Natural Language Processing (NLP) dalam data warehouse (Wahid & Kususiyanah, 2024).

Artikel yang terpilih berasal dari rentang publikasi tahun 2014 hingga 2024, menunjukkan perkembangan penelitian yang cukup konsisten selama satu dekade terakhir. Secara umum, terlihat adanya tren peningkatan jumlah penelitian terutama sejak tahun 2018, seiring dengan munculnya model NLP berbasis deep learning dan transformer yang mulai banyak diadopsi dalam berbagai aplikasi data warehouse, termasuk text-to-SQL, semantic enrichment, metadata extraction, dan otomatisasi pengelolaan data. Peningkatan ini menunjukkan bahwa integrasi NLP menjadi perhatian penting dalam pengembangan arsitektur data warehouse modern. Selain itu, mayoritas studi yang terpilih merupakan penelitian empiris dan eksperimen yang menekankan evaluasi performa sistem, sedangkan sebagian lainnya berfokus pada pengembangan arsitektur konseptual maupun framework integrasi. Gambaran ini memberikan dasar yang kuat untuk analisis lebih lanjut terkait metode, model integrasi, manfaat, serta tantangan implementasi NLP dalam data warehouse (Mulyana et al., 2025).

### **Metode NLP yang Digunakan dalam Studi**

Berdasarkan analisis terhadap studi-studi yang terpilih, ditemukan bahwa terdapat beberapa pendekatan utama Natural Language Processing (NLP) yang digunakan dalam pengembangan dan optimalisasi data warehouse. Salah satu kategori yang paling menonjol adalah NLP berbasis transformer, yang menunjukkan peningkatan signifikan dalam lima tahun terakhir (Deva Pratama & Rijati, 2024). Model seperti BERT, RoBERTa, dan GPT serta varian transformer lainnya digunakan terutama untuk tugas text-to-SQL, semantic parsing, dan ekstraksi informasi tingkat lanjut, karena kemampuannya dalam memahami konteks bahasa secara mendalam.

Selain itu, sejumlah studi masih menggunakan pendekatan rule-based NLP, terutama untuk kebutuhan spesifik seperti ekstraksi metadata terstruktur, validasi data berbasis aturan linguistik, serta otomatisasi query yang membutuhkan konsistensi sintaksis. Pendekatan ini banyak digunakan pada lingkungan data warehouse tradisional yang memiliki struktur data relatif stabil (Erdisna et al., 2025).

Selanjutnya, ditemukan pula penggunaan NLP berbasis machine learning, di mana algoritma seperti Naïve Bayes, SVM, dan Random Forest dimanfaatkan untuk klasifikasi teks, identifikasi entitas, atau pendeteksian anomali yang berkaitan dengan kualitas data. Metode ini umum digunakan pada studi-studi awal sebelum teknologi deep learning mendominasi (Solahudin, 2024).

Kategori lainnya yang semakin berkembang adalah model deep learning untuk text-to-SQL, yang berfokus pada penerjemahan bahasa alami ke dalam query SQL secara otomatis. Model neural semantic parsing, encoder-decoder networks, hingga arsitektur sequence-to-sequence berbasis attention banyak digunakan untuk meningkatkan akurasi dan fleksibilitas dalam menghasilkan query yang kompleks. Metode ini memberikan kontribusi besar dalam mengurangi kompleksitas interaksi pengguna dengan data warehouse serta meningkatkan kemampuan sistem dalam memahami kebutuhan pengguna tanpa memerlukan keahlian teknis SQL (Raharjo, 2024).

Secara keseluruhan, variasi metode NLP ini menunjukkan bahwa integrasi NLP dalam data warehouse berkembang secara multidimensional, mulai dari pendekatan berbasis aturan hingga model transformer yang lebih maju, sesuai dengan kebutuhan dan kompleksitas arsitektur sistem pada masing-masing penelitian (R. C. Tarumingkeng, 2024).

### **Model Integrasi NLP yang Ditemukan dalam Literatur**

Hasil penelusuran terhadap studi-studi yang terpilih menunjukkan bahwa integrasi Natural Language Processing (NLP) dalam data warehouse dilakukan melalui berbagai model dan pendekatan arsitektural. Salah satu model yang paling banyak dibahas adalah Model ETL Semantik, yang memanfaatkan teknik NLP untuk meningkatkan tahap Extract–Transform–Load melalui ekstraksi entitas, analisis semantik, dan penyelarasan data tidak terstruktur agar sesuai dengan skema data warehouse. Pendekatan ini memungkinkan proses integrasi data menjadi lebih akurat, adaptif, dan mampu menangani kompleksitas data heterogen (Selviani et al., 2025).

Selain itu, banyak penelitian mengembangkan Query Recommendation System, yaitu sistem yang memanfaatkan pemrosesan bahasa alami untuk memberikan saran query berdasarkan pola penggunaan, konteks pencarian, dan kebutuhan informasi pengguna. Sistem ini bertujuan mengurangi beban pengguna dalam menyusun query kompleks serta meningkatkan efisiensi interaksi dengan data warehouse.

Model integrasi lainnya adalah Automated Metadata Annotation, yang menggunakan NLP untuk melakukan anotasi, klasifikasi, dan pemetaan metadata secara otomatis. Pendekatan ini sangat bermanfaat dalam meningkatkan kualitas dokumentasi data, memperkuat struktur katalog

data, serta mempermudah proses penelusuran informasi dalam lingkungan data warehouse (Estede et al., 2025).

Studi-studi terbaru juga banyak mengembangkan Conversational Query Interface, yaitu antarmuka tanya-jawab berbasis percakapan (chat-based) yang memungkinkan pengguna melakukan query ke data warehouse menggunakan bahasa alami. Integrasi ini biasanya memanfaatkan model deep learning atau transformer untuk menerjemahkan pertanyaan pengguna ke dalam bentuk SQL yang relevan, sehingga menurunkan hambatan teknis bagi pengguna non-teknis (R. C. Tarumingkeng, n.d.).

Selain itu, ditemukan pula model Data Quality Improvement Using NLP, yang memanfaatkan algoritma NLP untuk mendeteksi inkonsistensi, duplikasi, kesalahan label, dan ketidaklengkapan data. Pendekatan ini membantu meningkatkan kualitas data sebelum atau setelah data dimasukkan ke dalam data warehouse (Fitrah, 2025).

Model terakhir yang muncul cukup signifikan dalam literatur adalah Knowledge Graph Enrichment, yaitu pemanfaatan NLP untuk membangun atau memperkaya knowledge graph sebagai representasi semantik yang membantu meningkatkan analisis data, pemetaan relasi antar entitas, serta mendukung pencarian semantik dalam data warehouse. Model ini menunjukkan potensi besar dalam memberikan pemahaman kontekstual yang lebih luas terhadap data yang tersimpan.

Secara keseluruhan, variasi model integrasi tersebut menunjukkan bahwa NLP memberikan kontribusi yang luas dan beragam pada arsitektur data warehouse modern, mulai dari proses ingest data hingga pemanfaatan data oleh pengguna akhir (Manuaba et al., 2025).

### **Dampak NLP terhadap Optimalisasi Data Warehouse**

Integrasi Natural Language Processing (NLP) dalam data warehouse memberikan berbagai dampak signifikan terhadap peningkatan performa dan efisiensi sistem. Salah satu dampak yang paling terlihat adalah reduksi waktu query, di mana penerapan model text-to-SQL dan antarmuka berbasis bahasa alami memungkinkan pengguna mengeksekusi query dengan lebih cepat tanpa memerlukan kemampuan teknis SQL. Hal ini mengurangi waktu pemrosesan dan mempercepat proses pengambilan keputusan. Selain itu, NLP berkontribusi pada peningkatan relevansi hasil pencarian melalui mekanisme semantic search yang mampu memahami konteks bahasa, sinonim, serta hubungan antar entitas data. Dengan demikian, data warehouse dapat menghasilkan informasi yang lebih akurat dan sesuai kebutuhan pengguna (Anisa, 2025).

Dampak penting lainnya adalah peningkatan kualitas data, khususnya melalui teknik ekstraksi entitas, deteksi kesalahan linguistik, identifikasi duplikasi, dan penyelarasan semantik. Pendekatan ini mampu mengurangi inkonsistensi serta meningkatkan integritas data dalam seluruh siklus hidup data warehouse. NLP juga memperkuat automasi proses ETL, di mana teknik pemrosesan teks membantu mengotomatiskan proses ekstraksi dan transformasi data tidak terstruktur menjadi bentuk terstruktur yang siap dimuat ke data warehouse. Hal ini tidak hanya mempercepat proses ETL, tetapi juga mengurangi kesalahan manual dan meningkatkan skalabilitas sistem (Suriansyah et al., 2025).

Selain itu, NLP memberikan kontribusi signifikan terhadap efisiensi pengelolaan data tidak terstruktur, yang kini menjadi bagian penting dari ekosistem big data. Melalui teknik seperti named entity recognition, topic modeling, dan semantic labeling, data tidak terstruktur seperti dokumen, email, atau laporan dapat diinterpretasikan, dipetakan, dan diintegrasikan secara efektif ke dalam arsitektur data warehouse. Integrasi ini menjadikan data warehouse lebih adaptif dan mampu menangani berbagai jenis data secara lebih komprehensif (Anam, 2025).

### **Tantangan yang Teridentifikasi**

Integrasi Natural Language Processing (NLP) ke dalam data warehouse tidak terlepas dari berbagai tantangan teknis maupun konseptual yang ditemukan dalam sejumlah penelitian. Salah satu tantangan utama adalah kebutuhan komputasi yang tinggi, terutama ketika menggunakan model berbasis deep learning atau transformer yang memerlukan sumber daya CPU dan GPU yang besar untuk proses training maupun inferensi. Hal ini menjadi kendala khususnya pada organisasi dengan infrastruktur terbatas atau yang belum mengadopsi arsitektur komputasi berbasis cloud (TRIANA et al., 2026).

Selain itu, banyak studi menyoroti bahwa NLP modern membutuhkan data training dalam jumlah besar agar mampu menghasilkan performa yang akurat dan stabil. Tantangan muncul ketika data pelatihan tidak tersedia dalam domain spesifik atau tidak mencerminkan struktur data

warehouse yang kompleks. Keterbatasan data berkualitas dapat menurunkan performa model dan mengurangi efektivitas integrasi NLP dalam sistem (Jumarto et al., 2025).

Tantangan berikutnya adalah kompleksitas integrasi arsitektur, di mana penggabungan modul NLP dengan pipeline ETL, query engine, dan komponen data warehouse lainnya memerlukan desain sistem yang matang. Integrasi ini melibatkan penyesuaian antarmuka, alur data, skema metadata, serta mekanisme validasi yang harus berjalan secara konsisten. Kompleksitas ini sering menjadi hambatan dalam implementasi di lingkungan industri yang memiliki sistem legacy atau arsitektur yang tidak fleksibel (Rusdi, 2025).

Selain itu, tantangan besar lainnya adalah kebergantungan pada bahasa (multilingual issues). Model NLP yang dilatih pada satu bahasa seringkali tidak bekerja optimal pada bahasa lain, terutama dalam lingkungan organisasi yang menggunakan bahasa lokal atau kombinasi multilingual. Perbedaan struktur bahasa, kosakata, dan ekspresi linguistik menyebabkan model mengalami penurunan akurasi dalam proses text-to-SQL, ekstraksi entitas, maupun pencarian semantik. Tantangan ini menunjukkan perlunya pengembangan model NLP yang lebih adaptif terhadap berbagai bahasa dan domain (Aisyah et al., 2025).

Secara keseluruhan, tantangan-tantangan tersebut menunjukkan bahwa meskipun integrasi NLP dalam data warehouse menawarkan banyak manfaat, implementasinya membutuhkan strategi teknis, infrastruktur, dan sumber daya yang memadai untuk mencapai hasil yang optimal (Rustiayana, Judijanto, Mahendra, et al., 2025).

### **Kesenjangan Penelitian**

Hasil telaah sistematis menunjukkan bahwa terdapat sejumlah kesenjangan penelitian yang masih belum terjawab secara memadai dalam literatur terkait integrasi Natural Language Processing (NLP) dengan data warehouse. Salah satu kesenjangan utama adalah minimnya penelitian yang berfokus pada bahasa Indonesia, baik dalam pengembangan model NLP, evaluasi text-to-SQL, maupun implementasi semantic ETL berbasis bahasa lokal. Kondisi ini menyebabkan keterbatasan adaptasi teknologi NLP dalam lingkungan organisasi di Indonesia, yang secara praktis memerlukan solusi berbasis bahasa lokal untuk memastikan akurasi dan relevansi hasil (Nasution et al., 2025).

Selain itu, banyak studi yang dianalisis belum memberikan evaluasi performa secara kuantitatif yang komprehensif. Sebagian besar penelitian hanya menyajikan hasil eksperimen bersifat kualitatif atau deskriptif tanpa metrik kinerja terstandar seperti akurasi, waktu eksekusi, F1-score, atau latency query. Ketiadaan evaluasi kuantitatif membuat sulit untuk membandingkan efektivitas berbagai model atau menentukan pendekatan NLP yang paling optimal untuk data warehouse (Wihardjo, 2025).

Kesenjangan berikutnya adalah belum adanya standar arsitektur integrasi NLP–Data Warehouse yang disepakati secara luas. Penelitian yang ada menunjukkan keragaman pendekatan arsitektur mulai dari semantic ETL, conversational query engine, hingga knowledge graph enrichment, namun belum ada kerangka baku yang dapat dijadikan acuan dalam implementasi industri. Hal ini menimbulkan tantangan dalam konsistensi desain, interoperabilitas sistem, dan skalabilitas (Ramdhani, 2024).

Selain itu, kajian terkait integrasi NLP dengan teknologi cloud-based data warehouse, seperti Snowflake, Google BigQuery, Amazon Redshift, atau Azure Synapse, masih sangat terbatas. Padahal, platform cloud sudah menjadi standar modern dalam pengelolaan data skala besar. Minimnya perhatian terhadap konteks cloud computing membuat penelitian yang ada kurang relevan dengan kebutuhan organisasi masa kini yang telah beralih ke arsitektur data warehouse berbasis cloud. Kesenjangan ini membuka peluang penelitian untuk mengevaluasi bagaimana NLP dapat diintegrasikan secara efektif dengan ekosistem cloud-native (Wibowo & Kom, 2025).

Secara keseluruhan, kesenjangan penelitian tersebut menunjukkan perlunya pengembangan kajian yang lebih mendalam, terukur, dan relevan dengan konteks penggunaan nyata, sehingga dapat mendorong perkembangan integrasi NLP–data warehouse yang lebih komprehensif di masa mendatang.

## **4. PEMBAHASAN**

### **Interpretation of Findings**

Hasil temuan dari berbagai studi menunjukkan bahwa Natural Language Processing (NLP) memiliki kontribusi signifikan dalam mengoptimalkan berbagai komponen utama pada data warehouse, mulai dari proses ekstraksi data, otomatisasi query, hingga peningkatan aksesibilitas pengguna non-teknis

melalui antarmuka berbasis bahasa alami. Penggunaan model transformer, teknik machine learning, hingga pendekatan text-to-SQL secara konsisten meningkatkan efisiensi pemrosesan data, mempercepat pengambilan keputusan, dan meminimalkan human error dalam manajemen data berskala besar.

Selain itu, integrasi NLP terbukti memperluas fungsi data warehouse dari sekadar sistem penyimpanan dan pelaporan menjadi ekosistem cerdas yang mampu memahami, menafsirkan, dan merespons kebutuhan pengguna melalui bahasa alami. Perubahan ini mendorong data warehouse menuju paradigma human-centered analytics, di mana pengguna tidak lagi harus memahami struktur query yang kompleks. Secara keseluruhan, temuan ini mengindikasikan bahwa integrasi NLP bukan hanya inovasi teknis, tetapi juga transformasi fungsional yang menggeser data warehouse menuju arsitektur yang lebih adaptif, intuitif, dan berbasis pemahaman konteks (I. R. C. Tarumingkeng, n.d.).

### **Implications for Industry**

Temuan penelitian ini memiliki sejumlah implikasi penting bagi industri. Pertama, integrasi NLP dalam data warehouse membuka peluang besar untuk otomatisasi proses bisnis, terutama pada aktivitas yang sebelumnya membutuhkan analisis manual, seperti ekstraksi informasi, pelaporan rutin, hingga penyusunan query yang kompleks. Otomatisasi ini tidak hanya mengurangi beban kerja teknis, tetapi juga meningkatkan konsistensi dan akurasi dalam pengolahan data.

Kedua, penerapan NLP secara langsung berdampak pada peningkatan efisiensi analisis data, karena proses analisis dapat dilakukan lebih cepat melalui antarmuka bahasa alami yang memungkinkan interpretasi data secara instan. Hal ini mengurangi kebutuhan akan keahlian teknis mendalam dalam penulisan query SQL, sekaligus mempercepat proses pengambilan keputusan berbasis data.

Selain itu, integrasi NLP dalam data warehouse memberikan kemudahan akses bagi pengguna non-teknis, seperti manajer, staf operasional, atau pemangku kepentingan yang tidak memiliki latar belakang pemrograman. Dengan kemampuan sistem memahami perintah bahasa sehari-hari, pemanfaatan data menjadi lebih inklusif dan dapat meningkatkan budaya data-driven decision-making di organisasi.

Lebih jauh, potensi implementasi teknologi ini mencakup berbagai sektor, termasuk pendidikan, kesehatan, pemerintahan, dan fintech. Pada sektor pendidikan, NLP dapat membantu analisis data akademik secara otomatis; di bidang kesehatan, teknologi ini dapat meningkatkan integrasi rekam medis; pada pemerintahan, NLP mendukung proses administrasi berbasis data; sementara dalam fintech, sistem berbasis NLP dapat memperkuat keamanan, deteksi fraud, dan pengalaman pengguna (Rustiyana, Judijanto, Aldila, et al., 2025).

### **Implications for Future Research**

Hasil telaah literatur ini membuka berbagai peluang penelitian lanjutan yang dapat dikembangkan di masa mendatang. Pertama, masih diperlukan penelitian yang lebih mendalam mengenai pengembangan model text-to-SQL dalam bahasa Indonesia, mengingat mayoritas studi saat ini berfokus pada bahasa Inggris sehingga belum sepenuhnya mendukung kebutuhan industri dan akademik di Indonesia. Penelitian semacam ini penting untuk menciptakan sistem kueri berbasis bahasa alami yang lebih kontekstual dan relevan dengan karakteristik linguistik lokal.

Kedua, penelitian selanjutnya dapat diarahkan pada integrasi NLP dengan platform Cloud Data Warehouse, seperti Snowflake, BigQuery, dan Azure Synapse. Hingga saat ini, masih sangat sedikit studi yang membahas bagaimana teknologi NLP dapat meningkatkan performa dan skalabilitas data warehouse berbasis cloud, padahal penggunaan solusi cloud semakin dominan di berbagai industri.

Ketiga, terdapat peluang besar untuk mengeksplorasi pemanfaatan Large Language Models (LLM) seperti GPT dalam pemrosesan kueri secara real-time. LLM berpotensi memberikan kemampuan pemahaman konteks yang lebih mendalam dan akurasi yang lebih tinggi dalam menghasilkan query yang kompleks, namun aspek efisiensi komputasi dan keamanan data masih perlu diteliti lebih jauh sebelum dapat diterapkan secara luas.

Terakhir, penelitian lanjutan diperlukan untuk mengembangkan framework standar arsitektur integrasi NLP–Data Warehouse. Saat ini belum terdapat pedoman baku yang dapat dijadikan acuan dalam implementasi sistem tersebut, sehingga adanya framework standar akan membantu peneliti dan praktisi dalam merancang, mengevaluasi, serta mengoptimalkan arsitektur yang konsisten dan dapat direplikasi di berbagai lingkungan organisasi (Liliana & Andry, 2024).

### Limitations of Current SLR

Penelitian Systematic Literature Review ini memiliki beberapa keterbatasan yang perlu dicermati. Pertama, terdapat keterbatasan bahasa publikasi, karena sebagian besar artikel yang dianalisis menggunakan bahasa Inggris, sedangkan publikasi berbahasa Indonesia atau bahasa lain masih sangat terbatas. Hal ini berpotensi menyebabkan bias linguistik dan mengurangi representativitas terhadap konteks penggunaan NLP di lingkungan lokal. Kedua, studi ini dibatasi oleh cakupan database yang digunakan, yaitu Scopus, IEEE Xplore, SpringerLink, ScienceDirect, dan ACM Digital Library. Meskipun database tersebut merupakan sumber ilmiah bereputasi tinggi, masih terdapat kemungkinan bahwa beberapa penelitian relevan tidak terindeks atau tidak muncul dalam hasil pencarian. Ketiga, telaah ini juga terpengaruh oleh perkembangan teknologi NLP yang bergerak sangat cepat, terutama dengan munculnya model-model baru berbasis transformer dan LLM yang mungkin belum sepenuhnya tercakup dalam rentang tahun publikasi yang dianalisis. Akibatnya, beberapa inovasi terbaru kemungkinan belum terwakili dalam hasil SLR ini.

## 5. KESIMPULAN

Integrasi Natural Language Processing (NLP) dalam data warehouse terbukti memberikan dampak signifikan terhadap otomatisasi proses bisnis, pemrosesan data tidak terstruktur, serta peningkatan kualitas dan relevansi query, di mana NLP digunakan secara luas mulai dari tahap ETL, anotasi metadata, semantic search, text-to-SQL, hingga analytical processing. Kajian ini berkontribusi dengan menyediakan pemetaan komprehensif mengenai teknologi NLP yang telah diterapkan dalam konteks data warehouse, sekaligus menyajikan analisis mendalam mengenai tren penelitian, kesenjangan ilmiah, serta arah pengembangan di masa mendatang. Berdasarkan temuan tersebut, penelitian selanjutnya direkomendasikan untuk melakukan eksperimen implementasi NLP pada data warehouse skala industri, mengembangkan dataset bahasa Indonesia khusus untuk task text-to-SQL, serta mengeksplorasi integrasi NLP dengan data lake dan platform analitik berbasis cloud guna meningkatkan skalabilitas dan performa sistem.

## DATAR PUSTAKA

- Adekamwa, A., Mursalim, M., & Indrayanti, I. (2024). Tren penelitian pelayanan publik di Indonesia: Suatu tinjauan sistematis literatur. *Jurnal Administrasi Negara*, 30(3), 240–263.
- Aisyah, S., Eugene, T., & Nurhayati, E. (2025). Tantangan Linguistik dalam Pengimplementasian Big Data Berbahasa Indonesia pada Robot Humanoid: Tinjauan dan Rekomendasi. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 1(1), 9.
- Anam, R. I. (2025). Transformasi Data Warehouse dengan Cloud dalam Era Big Data. *Journal of Technology, Religious Studies, Social and Humanities Science (JTRS)*, 1(1), 48–52.
- Andriyani, W., Purnomo, R., Hendrawan, S. A., Irvani, A. I., Sujarwo, A., Asri, Y. N., Jones, A. H. S., Feta, N. R., Wulandari, I. Y., & Hatma, S. (2024). *Data Sebagai Fondasi Kecerdasan Buatan*. Tohar Media.
- Anisa, P. (2025). Implementasi Data Warehouse untuk Pengolahan Data Akademik Perguruan Tinggi. *Jurnal Komputer Dan Teknik Informatika*, 1(1), 23–30.
- Darmawan Napitupulu, S. T., Kom, M., & CIQnR, C.-B. (n.d.). *Workshop Systematic Literature Review (SLR) Sabtu, 14 Januari 2024*.
- Deva Pratama, A. S., & Rijati, N. (2024). Pengenalan Emosi Ulasan Pelanggan E-Commerce Menggunakan Deep Learning Berbasis Transformer. *Techno. Com*, 23(3).
- Erdisna, E., Murniasih, I., Diantoro, K., Hardini, I. R., Rinaldo, R., Asruddin, A., Ayuningsih, E., Purnomo, R. F., Nurjoko, N., & Fernando, Y. (2025). *Data Warehouse*. CV. Gita Lentera.
- Estede, S., Mohamad Ilham, S. T., Kom, M., Chafid, N., Kom, S., Kom, M., Baan, R. R. S., Kaestria, R., Kom, M., & Aprilia, A. (2025). *KONSEP DASAR DATA WAREHOUSE*. Cendikia Mulia Mandiri.
- Fitrah, R. N. (2025). *SISTEM INTENT DETECTION PADA KELUHAN PASIEN UNTUK REKOMENDASI DOKTER MENGGUNAKAN METODE BIOBERT*. Universitas Islam Sultan Agung Semarang.
- Iskandar, A. P. S., Setiawan, H., Judijanto, L., Mahendra, G. S., Ardi, M., Putri, N. A. R., Handika, I. P. S., Ratih, R., Mandowen, S. A., & Wazaumi, D. D. (2024). *Teknologi big data: Pengantar dan penerapan teknologi big data di berbagai bidang*. PT. Green Pustaka Indonesia.
- Jumarto, S. T., Si, A. M., Didik Kristyanto, S. E., & Feriandy, S. P. (2025). *BUKU ARTIFICIAL INTELLIGENCE DALAM REKRUTMEN SDM MODERN (STRATEGI, SISTEM, DAN*

- TRANSFORMASI ORGANISASI*). Penerbit Widina.
- Liliana, L., & Andry, J. F. (2024). *Perancangan Enterprise Architecture Menggunakan Zachman Framework & TOGAF Adm*. Penerbit Andi.
- Manuaba, I. B. K., Mukminna, H., Judijanto, L., Fanani, M. Z., Supartha, I. K. D. G., Junaidi, S., Pasinggi, E. S., Suryadi, D., Purba, D. N., & Nampira, A. A. (2025). *Big Data: Teori dan Penerapan Teknologi Big Data dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- Mulyana, M., Nurendah, Y., & Effendy, M. (2025). *BUSINESS INTELLIGENCE*. Kesatuan Press.
- Nasution, U. H., Syauqi, T. M., Setyawibawa, R. A., Rusdi, R., & Rizky, M. C. (2025). *Manajemen di Era AI "Strategi, Etika, dan Adaptasi Organisasi."* Serasi Media Teknologi.
- Raharjo, B. (2024). *Pergudangan Data (Data Warehousing) Jilid 2*. Penerbit Yayasan Prima Agus Teknik, 1–222.
- Ramdhani, T. I. (2024). *Kajian Arsitektur Berlapis Untuk Pengembangan Aplikasi Dalam Mendukung Pembangunan Aplikasi Di Pemerintahan Provinsi Jawa Barat*. *Creative Research Journal*, 10(01), 17–26.
- Rusdi, M. (2025). *Analisis Platform Aplikasi Low Code dalam Pengembangan Sistem Informasi untuk Industri 4.0 di Indonesia*. Universitas Islam Indonesia.
- Rustiyan, R., Judijanto, L., Aldila, A. S., Suhardi, S., Sudarman, S., Rosadi, M. E., & Rahmawati, E. (2025). *Artificial Intelligence: Pengetahuan dan Pemanfaatannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- Rustiyan, R., Judijanto, L., Mahendra, G. S., Kamil, Z. A., Purba, D. N., Sutoyo, M. N., Hendrayana, I. G., Pasrun, Y. P., & Prayudani, S. (2025). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.
- Selviani, S., Yudhyarta, D. Y., & Susanti, H. (2025). *Pengembangan model integrasi basis data dan sistem manajemen informasi untuk optimalisasi kecerdasan bisnis*. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 6094–6101.
- Setiawan, Z., Irmawati, I., Sepriano, S., Miranda, E., Arifin, N. Y., & Atmaja, K. J. (2023). *DATA WAREHOUSE & BUSINESS INTELLIGENCE: Teori Komprehensif*. PT. Sonpedia Publishing Indonesia.
- Solahudin, W. (2024). *Penerapan deep learning dalam pendidikan di Indonesia: Tantangan dan peluang implementasi teknologi pendidikan*. *Jurnal Ilmu Pendidikan*, 1(1), 48–55.
- Solehah, H. Y., & Hendriani, W. (2024). *PSYSTEMATIC LITERATURE REVIEW (SLR): PERAN ACADEMIC ENGAGEMENT TERHADAP PRESTASI AKADEMIK*. *Jurnal Psikohumanika*, 16(1), 13–29.
- Suriansyah, B., Mz, L. F., Rachman, A. I., & Pratiwi, G. (2025). *Rekonstruksi Arsitektur DataBase untuk Peningkatan Proses Load Data*. *Jurnal Media Informatika*, 6(2), 1455–1460.
- Tarumingkeng, I. R. C. (n.d.). *Meningkatkan Keputusan Bisnis Berbasis Data*.
- Tarumingkeng, R. C. (n.d.). *Machine Learning dalam Data Science: Teknik dan Kasus*.
- Tarumingkeng, R. C. (2024). *Natural Language Processing (NLP)*. RUDYCT E-PRESS, No.
- TRIANA, H., Solahudin, A., & Darmawan, A. R. (2026). *TINJAUAN PERKEMBANGAN KOMPUTASI KUANTUM PADA PLATFORM CLOUD: PELUANG, TANTANGAN, DAN ARAH MASA DEPAN*. *Jurnal Informatika Dan Teknik Elektro Terapan*, 14(1).
- Tribuana, D., Agustan, A. D. H., Halimah, E., & Dianah, K. (2025). *Membangun Taxonomy Riset Big Data Analytics dan Business Intelligence: Systematic Literature Review dalam Konteks Manajemen Informatika*. *Jurnal Teknologi Dan Bisnis Cerdas*, 1(2), 140–154.
- Utami, M. C., Jahar, A. S., & Zulkifli, Z. (2021). *Tinjauan scoping review dan studi kasus*. *Radial*, 9(2), 152–172.
- Wahid, S. H., & Kususiyanah, A. (2024). *Systematic Review Menggunakan Artificial Intelligence (AI) dalam Ilmu Sosial: Systematic Review Using Artificial Intelligence (AI) in Social Science*.
- Wibowo, I. A., & Kom, M. (2025). *Artificial Intelligence (AI) dalam Jaringan Otonom*. yayasan penerbit.
- Widiana, I. G. R., Mahadita, G. W., Samsu, N., & Muzasti, R. A. (2025). *Systematic Review dan Meta Analysis*.
- Wihardjo, E. (2025). *DENGAN SOFTWARE MODERN BAR•*. *Manajemen Data Dengan Software Modern*, 37.