BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terdahulu mengenai penulisan tugas akhir telah dilakukan oleh banyak peneliti dan akademisi di berbagai disiplin ilmu. Beberapa temuan umum yang ditemukan dalam penelitian.

Tabel 2.1 Penelitian Terkait

No	Topik	Metode	Hasil	Ref
1.	Aplikasi Chatbot Berbasis Dialogflow dan NLP untuk Pelayanan Informasi Akademik pada Fakultas Ilmu Komputer Universitas Tomakaka	Mengintegrasikan kapabilitas Natural Language Processing (NLP) ke dalam arsitektur penanganan kueri akademik.	Membuktikan secara empiris bahwa pemahaman bahasa alami adalah standar teknologi fundamental untuk mengeskalasi kualitas dan daya tanggap layanan informasi interaktif.	(Munawirah et al., 2025)
2.	Chatbot Layanan Akademik Menggunakan KNearest Neighbor	Mengotomatisasi sistem layanan administratif menggunakan algoritma klasifikasi pada platform digital operasional kampus.	Mengesahkan urgensi otomatisasi antarmuka layanan publik sebagai solusi komputasional untuk mengeliminasi inefisiensi yang diakibatkan oleh keterbatasan sumber daya staf akademik.	(Nugraha & Sebastian, 2021)
3.	Pengadaan Denah Kampus untuk Meningkatkan Kemudahan Navigasi	Pengamatan langsung mengenai kesulitan pengunjung dan mahasiswa dalam mencari lokasi gedung.	Membuktikan bahwa panduan denah kampus yang mudah diakses sangat dibutuhkan. Penelitian ini menjadi dasar alasan kuat	(Nurdiansah et al., 2024)

https://lib.mercubuana.ac.id

			mengapa fitur pencarian rute dibuat pada proyek ini.	
4.	UE-NER-2025: A GPT-Based Approach to Multi-Lingual Named Entity Recognition on Urdu and English	Pengkajian metode dasar mengenai bagaimana pengenalan entitas bernama (NER) memproses bahasa.	Memberikan dasar teori yang kuat bahwa fungsi utama NER adalah mengenali serta mengelompokkan kata dari kalimat yang tidak beraturan agar bisa diproses lebih lanjut oleh komputer.	(Ahmad et al., 2025)
5.	Penerapan Algoritma A-Star sebagai Metode Pathfinding pada Game Lokasi Logika Cerdas untuk Non-Player Character	Penggunaan algoritma A-Star (A*) untuk mencari rute terpendek dengan menghindari rintangan.	Menunjukkan bahwa A-Star sangat akurat dalam mencari jalan tercepat antar dua titik. Metode ini diterapkan langsung pada proyek untuk menghitung rute di denah kampus berdasarkan titik persimpangan.	(Gibran et al., 2026)
6.	Inkonsistensi Bahasa Ujaran Navigasi Google Maps: Benarkah AI Belum Sepenuhnya Memahami Bahasa Manusia?	Penilaian terhadap kejelasan suara panduan arah (TTS) pada aplikasi peta navigasi.	Menjelaskan bahwa suara sistem sangat membantu pengguna mencari jalan tanpa harus terus melihat layar. Pada proyek ini, konsep tersebut diterapkan agar asisten dapat berbicara sambil menunjukkan rute.	(Khoirunnisa et al., 2025)

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id

7.	Integrasi Model LLM Ollama dan OpenStreetMap API pada BI untuk Rekomendasi Lokasi	Menjalankan program kecerdasan buatan secara langsung di komputer perangkat keras lokal (Edge Computing).	Membuktikan bahwa menjalankan model AI di komputer lokal membuat sistem lebih cepat merespons dan bebas dari gangguan internet. Ini menjadi dasar sistem proyek yang dijalankan secara lokal (tanpa komputasi awan).	(Fadhil Ahmad et al., 2025)
8.	Multi-task Learning for Named Entity Recognition and Intent Classification in Natural Language Understanding Applications	Mengusulkan skema komputasi hibrida (Multitask Learning) untuk menyatukan pengenalan entitas dan klasifikasi niat pengguna (Intent Classification).	Mendukung arsitektur klasifikasi intent pada sistem yang bertujuan untuk mitigasi fenomena halusinasi AI, memastikan determinisme respons pada instruksi penjadwalan maupun rute.	(Perdana & Adikara, 2025)
9.	Customized Named Entity Recognition Using Bert for the Social Learning Management System Platform CourseNetworking	Penggunaan NER khusus untuk mengambil data penting dari sebuah susunan kalimat.	Menyimpulkan bahwa NER adalah cara yang tepat untuk mengenali nama orang atau lokasi dari kalimat biasa. Pada proyek ini, NER dipakai untuk menemukan nama dosen dan ruangan.	(Padmanandam et al., 2024)
10.	Pengembangan Sistem Kasir Berbasis Suara dengan Fuzzy	Penerapan algoritma Fuzzy Matching untuk mencocokkan kemiripan teks.	Jurnal ini menunjukkan bahwa metode Fuzzy Matching dapat menangani kesalahan	(Meudea et al., 2025)

https://lib.mercubuana.ac.id

	Matching di Koperasi Harapan Mulya Kediri		ketik (typo) dengan baik. Pada proyek ini, metode tersebut dipakai sebagai jalur cepat (fast lane) untuk mengenali nama ruangan atau nama dosen.	
11.	Assessing Geospatial Accuracy in Mapping Applications: A Focus on Google Earth	Membandingkan perbedaan titik koordinat dan panjang garis dari peta satelit Google Earth dengan gambar hasil pengukuran asli bangunan (*As-Built Drawing*) di lapangan menggunakan perhitungan statistik.	Penelitian ini membuktikan bahwa meskipun titik koordinat bumi bisa sedikit bergeser karena pengaruh pembagian zona wilayah satelit, ukuran jarak panjang yang diambil lewat peta satelit terbukti tetap stabil dan tidak berbeda jauh dengan hasil ukur manual di lapangan. Temuan ini menjadi dasar yang kuat bagi penulis untuk mengambil data jarak jalan (dalam meter) menggunakan fitur pengukur jarak di Google Maps tanpa harus mengukurnya secara manual dengan meteran fisik.	(Alqahtani, 2024)
12.	Application of A* algorithm in intelligent vehicle path planning	Membandingkan cara kerja dan kecepatan pencarian jalan terpendek antara algoritma Dijkstra, BFS, dan	Mengonfirmasi bahwa algoritma A* adalah metode terbaik karena berhasil menggabungkan	(Wang dkk., 2022)

https://lib.mercubuana.ac.id

		A-Star (A*) menggunakan peta simulasi komputer yang dipenuhi rintangan.	kelebihan algoritma Dijkstra (yang menjamin pencarian rute terpendek) dengan kelebihan BFS (yang membuat pencarian menjadi lebih cepat menggunakan fungsi perkiraan jarak). Teori ini mendukung keputusan penulis menggunakan algoritma A* untuk menghitung jalan terdekat dari posisi awal pengguna menuju gedung tujuan di kampus.	
13.	Interacting with educational chatbots: A systematic review	Tinjauan pustaka mengenai penerapan asisten virtual berupa teks dan suara di berbagai lingkungan pendidikan.	Menyatakan bahwa asisten virtual terbukti mengurangi beban tugas staf kampus dan memberikan kemudahan layanan yang praktis bagi warga kampus.	(Kuhail et al., 2023)
14.	Implementasi Chatbot sebagai Virtual Assistant Penerimaan Mahasiswa Baru pada Universitas Bumigora	Pembuatan asisten virtual otomatis untuk tanya-jawab layanan publik.	Menemukan bahwa asisten otomatis sangat menghemat waktu dan dapat melayani 24 jam. Hal ini mendukung alasan pembuatan asisten mandiri (kiosk) di lingkungan kampus pada proyek ini.	(Amrullah et al., 2022)

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id

15.	Implementasi Chatbot Menggunakan Framework Langchain Berbasis Llm Gpt	Penggunaan Large Language Model (LLM) untuk membangun sistem tanya jawab.	Menunjukkan kemampuan LLM dalam memahami kalimat secara utuh dan panjang. Pada proyek ini, LLM digunakan khusus untuk menjawab pertanyaan pengguna yang bersifat rumit atau panjang.	(Muhajir et al., 2025)
16.	Penggunaan Chatbot pada Sistem Informasi Buku Berbasis Web Menggunakan Metode Natural Language Processing (NLP)	Penerapan pra-pemrosesan teks (Text Preprocessing) sebelum teks dimasukkan ke pemodelan NLP.	Menjelaskan bahwa pembersihan teks (seperti menghapus simbol) sangat penting untuk meningkatkan keakuratan sistem. Pada proyek ini, langkah tersebut dipakai sebelum teks diproses oleh mesin.	(Hamdani et al., 2026)
17.	Optimalisasi Korpus Bahasa Indonesia untuk Peningkatan Kualitas Text-to-Speech Berbasis Diphone Concatenation	Pengujian kualitas suara dari mesin Text-to-Speech (TTS) berdasarkan tingkat kejelasan dan kealamian nada.	Menyimpulkan bahwa suara mesin harus jelas dan terdengar alami. Pada proyek ini, temuan ini menjadi dasar mengapa suara asisten virtual harus disesuaikan agar pengguna nyaman mendengar jawabannya.	(Amiruddin & Furqan, 2025)
18.	Wav2wav: Wave-toWave Voice Conversion	Mengkaji arsitektur transformasi Voice Conversion yang difokuskan pada pelestarian susunan fonetik sekaligus	Menguatkan landasan teoritis bahwa implementasi Retrievalbased Voice Conversion (RVC)	(Jeong et al., 2024)

https://lib.mercubuana.ac.id

		memanipulasi timbre profil vokal.	sanggup mereplika karakteristik vokal target secara presisi, mendongkrak	
19.	Hybrid TF-IDF and Embedding Model for Improving Similarity and Clustering Accuracy	Melakukan komputasi derajat kemiripan teks matematis dengan memanfaatkan metrik pengukuran kemiripan vektor (Cosine Similarity).	Mendasari kalkulasi semantik bahwa kedekatan spasial antar-vektor (embedding) dalam dimensi ruang merepresentasikan ekuivalensi komputasional dari kesamaan makna antarinstruksi.	(Ero Wahyu Pratomo, 2026)
20.	A Stacking Ensemble Based on Lexicon and Machine Learning Methods for the Sentiment Analysis of Tweets	Mengaplikasikan algoritma deteksi sentimen berbasis leksikon (Lexicon-based) untuk mengekstraksi polaritas afektif pada data tekstual.	Menyediakan kerangka komputasional untuk mengklasifikasikan sentimen respons (positif, negatif, netral) yang difungsikan sebagai parameter aktuasi dinamis penggerak emosi avatar 2D.	(Malebary & Abulfaraj, 2024)

2.2 Teori Utama

Dalam teori utama ini berisi konsep-konsep fundamental dalam membangun asisten virtual kampus agar hasilnya menjadi lebih bagus dan dapat berjalan lancar

2.2.1 *Natural Language Processing* (NLP)

Natural Language Processing (NLP) merupakan cabang fundamental Kecerdasan Buatan (AI) yang menjembatani interaksi kognitif antara komputer dan bahasa alami manusia. Pada ekosistem

https://lib.mercubuana.ac.id

asisten virtual modern berbasis *Named Entity Recognition* (NER) dan *Large Language Models* (LLMs), prapemrosesan teks tradisional yang agresif, seperti *stopword removal* dan *stemming*, tidak lagi relevan karena berisiko mendegradasi struktur sintaksis kalimat. Sebagai contoh, reduksi preposisi spasial (seperti "di" atau "ke") dapat mengaburkan makna arah yang krusial bagi ekstraksi entitas lokasi. Oleh karena itu, pendekatan modern mempertahankan tata bahasa utuh dengan mengonversi sekuens kalimat menjadi representasi vektor padat (*word embeddings*), sehingga sistem mampu mengekstrak makna kontekstual secara holistik selayaknya kognisi manusia (Perdana & Adikara, 2025).

Pada penelitian ini, teknik-teknik NLP yang diterapkan meliputi *text preprocessing* untuk normalisasi teks transkripsi, *Named Entity Recognition* berbasis *fuzzy string matching* untuk pengenalan entitas kampus, serta *intent classification* berbasis pola kata kunci untuk menentukan jenis pertanyaan pengguna. Tantangan utama dalam penerapan NLP pada sistem ini terletak pada sifat input yang berasal dari transkripsi suara, di mana hasil transkripsi sering kali mengandung kesalahan ejaan, kata yang terputus, atau struktur kalimat yang tidak lengkap sehingga memerlukan mekanisme pencocokan yang toleran terhadap variasi tersebut.

2.2.2 *Named Entity Recognition* (NER)

Named Entity Recognition (NER) adalah teknik dalam NLP yang bertujuan untuk mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks ke dalam kategori yang telah ditentukan, seperti nama orang, organisasi, lokasi, atau entitas domain-spesifik lainnya. Dalam pendekatan konvensional, NER dapat diimplementasikan melalui metode berbasis aturan (*rule-based*), metode statistik, atau metode berbasis pembelajaran mesin (*machine learning*). Model NER bertugas memindai serangkaian kata di dalam teks dan mengklasifikasikannya ke dalam kelompok kelas entitas (*entity class*) yang telah didefinisikan secara baku, seperti kelas entitas tata letak lokasi/ruangan (*Location*), individu (*Person*), dan institusi atau organisasi (*Organization*) (Padmanandam et al., 2024). Pada

https://lib.mercubuana.ac.id

penelitian ini, digunakan pendekatan *Hybrid NER* yang mengombinasikan *fuzzy string matching* sebagai jalur cepat (*fast lane*) dengan pencarian semantik berbasis *vector database* sebagai jalur cadangan (*slow lane*). Pendekatan hybrid ini dirancang untuk memaksimalkan akurasi pengenalan entitas sekaligus meminimalkan latensi respons.

Mekanisme *fast lane* pada sistem ini bekerja dengan cara membentuk *n-gram* (kombinasi satu hingga lima kata) dari teks input, kemudian mencocokkannya terhadap basis data alias entitas menggunakan algoritma *fuzzy matching* dengan library RapidFuzz. Sistem menerapkan *dynamic threshold* yang menyesuaikan ambang batas kemiripan berdasarkan panjang teks kandidat, sehingga dapat meminimalkan kesalahan pencocokan pada teks pendek. Pada penelitian (Meudea et al., 2025), yang membuktikan bahwa kombinasi pengenalan suara (*Speech-to-Text*) dengan *fuzzy string matching* berbasis RapidFuzz mampu menekan *Word Error Rate* (WER) dan mereduksi kesalahan identifikasi entitas lisan secara signifikan.

2.2.3 Algoritma A-Star (A*)

Algoritma A* merupakan algoritma pencarian jalur terpendek yang menggabungkan keunggulan dari algoritma Dijkstra yang menjamin penemuan solusi optimal, dengan pendekatan *Breadth First Search* (BFS) yang menggunakan fungsi heuristik untuk mengarahkan pencarian (Wang et al., 2022). Secara formal, algoritma A* mengevaluasi setiap simpul berdasarkan fungsi

$$f(n) = g(n) + h(n)$$

di mana g(n) merepresentasikan biaya aktual dari simpul awal ke simpul n, dan h(n) merepresentasikan estimasi biaya dari simpul n ke simpul tujuan menggunakan fungsi heuristik. Selama fungsi heuristik yang digunakan bersifat *admissible* (tidak pernah melebih-lebihkan biaya sebenarnya), algoritma ini dijamin akan selalu menemukan rute navigasi yang paling efisien.

https://lib.mercubuana.ac.id

Pada penelitian ini, algoritma A* diterapkan untuk pencarian rute navigasi di dalam lingkungan kampus. Graf navigasi direpresentasikan dalam format JSON yang memuat simpul-simpul (*nodes*) dengan koordinat spasial pada denah kampus, serta sisi-sisi (*edges*) yang merepresentasikan jalur penghubung antar-lokasi. Untuk mendapatkan metrik jarak yang realistis, penentuan bobot pada setiap sisi diukur secara langsung menggunakan fitur *measure distance* dari layanan Google Maps. Pemanfaatan platform pemetaan digital ini dinilai sangat layak secara akademis karena terbukti memiliki keandalan yang konsisten dan akurasi geospasial yang memadai untuk mendukung perancangan rute navigasi maupun pemodelan tata ruang (Alqahtani, 2024). Fungsi heuristik yang digunakan sebagai pendukung navigasi adalah jarak *Euclidean*, yang memenuhi syarat *admissible* karena jarak garis lurus selalu lebih kecil atau sama dengan jarak jalur sebenarnya. Jika dibandingkan dengan algoritma *Breadth First Search* (BFS) maupun Dijkstra yang mengeksplorasi seluruh simpul secara merata penggunaan algoritma A* pada graf spasial ini jauh lebih superior. Fungsi heuristik pada algoritma A* secara efektif memandu arah eksplorasi menuju simpul tujuan, sehingga meminimalkan penumpukan ekspansi simpul (*expansion nodes*) dan menghasilkan efisiensi pencarian komputasi yang jauh lebih tinggi (Wang et al., 2022).

Terdapat beberapa pendekatan dalam implementasi NER, yaitu: (1) pendekatan berbasis aturan (*rule-based*), (2) pendekatan berbasis model *machine learning* seperti CRF dan BiLSTM-CRF, serta (3) pendekatan berbasis pengetahuan (*knowledge-based*) yang memanfaatkan kamus untuk mencocokkan entitas (Kusumaningtyas et al., 2023) . Pada penelitian ini, NER diimplementasikan menggunakan pendekatan berbasis pengetahuan dengan teknik *fuzzy string matching*. Pendekatan ini dipilih karena domain entitas bersifat tertutup (*closed-domain*), di mana seluruh entitas kampus telah terdefinisi dalam basis data alias

https://lib.mercubuana.ac.id

2.2.4 Asisten Virtual Kampus

Asisten virtual merupakan program komputer berbasis kecerdasan buatan yang dirancang untuk meniru interaksi manusia melalui teks maupun suara demi membantu penyelesaian tugas serta mempermudah penyebaran informasi. Dalam dunia pendidikan, kehadiran teknologi ini terbukti mampu meningkatkan keterlibatan pengguna, memberikan bantuan personal yang tepat, serta membantu efisiensi tugas para pengajar (Kuhail et al., 2023). Pada lingkungan kampus modern, teknologi ini menjadi sangat penting sebagai layanan mandiri untuk menjembatani hambatan dalam mengakses informasi internal kampus bagi mahasiswa baru maupun tamu melalui interaksi komputer yang mudah menyesuaikan diri dengan kebutuhan pengguna.

Pada penelitian ini mengembangkan sebuah asisten virtual melalui rangkaian proses pengolahan data yang terpadu. Rekaman suara dari pengguna akan diubah menjadi teks oleh modul *Speech-to-Text*, untuk kemudian diperiksa oleh sistem pencari entitas agar maksud dari pertanyaan tersebut dapat dikenali. Hasil pengenalan ini secara dinamis akan menentukan arah jawaban sistem: melakukan penarikan data nyata secara langsung, menghitung rute peta lingkungan kampus menggunakan algoritma A*, atau meneruskan pertanyaan yang rumit kepada *Large Language Model* (LLM). Hasil pengolahan ini kemudian ditampilkan melalui karakter dua dimensi yang dapat bergerak dan mengeluarkan suara. Pengalihan tugas-tugas berulang kepada asisten virtual mandiri ini terbukti sangat efektif untuk mengurangi beban kerja bagian administrasi kampus, menyediakan layanan informasi yang siap sedia selama 24 jam penuh, serta mengatasi masalah antrean pelayanan pada jam-jam sibuk perkuruan tinggi (Amrullah et al., 2022).

2.3 Teori Pendukung

Teori pendukung memaparkan landasan ilmu dan spesifikasi teknologi yang merakit, mendayagunakan, dan mendukung kelancaran operasional teori utama, khususnya dalam menyesuaikan batasan spesifikasi komputasi lokal.

https://lib.mercubuana.ac.id

2.3.1 *Large Language Model* (LLM)

Large Language Model (LLM) merupakan model kecerdasan buatan pengolah bahasa berskala besar yang dilatih menggunakan basis data teks dalam jumlah yang sangat masif. Model bahasa generatif seperti keluarga GPT telah membuktikan kapabilitas yang tinggi dalam memahami konteks percakapan, mengikuti arahan spesifik, serta menghasilkan teks jawaban yang runtut dan sesuai dengan perintah yang diberikan (Muhajir et al., 2025).

Dalam sistem asisten virtual kampus ini, mengimplementasikan LLM secara lokal menggunakan model Gemma-4-E2B sebagai komponen utama pemrosesan bahasa. Penggunaan model lokal ini difokuskan untuk menjaga efisiensi beban perangkat keras sekaligus memastikan sistem mampu menjawab pertanyaan-pertanyaan kompleks yang tidak dapat ditangani oleh pangkalan data standar. Model ini diatur untuk menerima instruksi karakter agen, riwayat obrolan terdahulu, serta data dari ekstraksi entitas sebagai informasi tambahan. Metode ini menerapkan konsep *Retrieval-Augmented Generation* (RAG) untuk memasukkan data nyata ke dalam perintah awal (*prompt*), sehingga jawaban model lebih akurat dan terhindar dari halusinasi. Lebih lanjut, kapabilitas komputasi model ini dimanfaatkan untuk terintegrasi secara langsung dengan modul *Speech-to-Text* (STT), memungkinkan pemrosesan rekaman suara menjadi teks transkripsi dalam satu alur inferensi yang terpadu..

2.3.2 *Text Preprocessing & Intent Detection*

Pemrosesan awal teks (*text preprocessing*) adalah tahapan fundamental dalam NLP yang bertujuan untuk membersihkan dan menormalisasi data teks mentah sebelum diproses lebih lanjut. Pada sistem ini, preprocessing mencakup normalisasi untuk menyeragamkan representasi karakter, konversi ke huruf kecil (*lowercasing*) untuk menghilangkan variasi kapitalisasi, penghapusan tanda baca yang tidak relevan, serta normalisasi spasi berlebih. Standardisasi ini secara signifikan mampu mengurangi beban komputasi dan meningkatkan akurasi sistem,

https://lib.mercubuana.ac.id

karena model algoritma dapat langsung memproses struktur teks yang sudah terstruktur dan bersih dari *noise* data (Hamdani et al., 2026)

Intent detection atau deteksi intent merupakan proses mengklasifikasikan maksud atau tujuan di balik pertanyaan pengguna ke dalam kategori-kategori yang telah ditentukan. Dalam penelitian ini, deteksi intent dilakukan menggunakan pendekatan berbasis pencocokan pola kata kunci (*keyword pattern matching*), di mana sistem mencocokkan teks input dengan sekumpulan kata kunci yang diasosiasikan dengan setiap kategori intent. Kategori intent yang didefinisikan meliputi intent faktual (jabatan, akreditasi, gelar), intent jadwal, intent lokasi, serta intent kompleks yang memerlukan pemrosesan lebih lanjut oleh LLM. Pendekatan berbasis pola kata kunci dipilih karena memberikan latensi yang sangat rendah dan determinisme yang tinggi dibandingkan dengan pendekatan berbasis model klasifikasi yang memerlukan inferensi tambahan.

2.3.3 *Speech Processing*

Speech processing mencakup teknologi pemrosesan suara yang meliputi dua arah konversi utama, yaitu *Speech-to-Text* (STT) dan *Text-to-Speech* (TTS). STT merupakan proses konversi sinyal audio ucapan manusia menjadi representasi teks, yang menjadi komponen input utama pada sistem asisten virtual berbasis suara. Sebaliknya, TTS merupakan proses sintesis suara dari teks tertulis, yang memungkinkan sistem memberikan respons dalam bentuk audio yang dapat didengar oleh pengguna. Kualitas keluaran TTS yang ideal ditandai dengan kemampuannya menghasilkan sintesis suara yang dapat dipahami dengan jelas serta memiliki intonasi yang menyerupai pengucapan manusia secara alami (Amiruddin & Furqan, 2025).

Selain STT dan TTS, terdapat teknologi Voice Conversion (VC) yang bertujuan untuk mengubah karakteristik suara dari satu pembicara agar menyerupai suara pembicara lain tanpa mengubah konten linguistik yang diucapkan. Salah satu pendekatan yang berkembang pesat adalah Retrieval-based Voice Conversion (RVC), yang memanfaatkan metode penarikan ciri

https://lib.mercubuana.ac.id

khas suara (timbre) secara efisien (Dong, 2025). Secara mendasar, teknologi ini mengandalkan model HuBERT untuk mengekstrak dan mengenali isi pesan dari sebuah rekaman audio input. Proses ini bertujuan untuk memisahkan antara teks yang diucapkan dengan identitas asli suara pembicara. Dalam pengembangan asisten virtual, penerapan model berbasis HuBERT ini terbukti sangat tangguh untuk mengubah karakter suara dasar menjadi suara baru yang berkualitas tinggi dan selaras dengan identitas tokoh yang dituju (Chung & Nam, 2023).

2.3.4 *Semantic Search* & *Vector Database*

Semantic search atau pencarian semantik merupakan pendekatan pencarian informasi yang mencocokkan kueri dengan dokumen berdasarkan kesamaan makna. Berbeda dengan pencarian berbasis kata kunci konvensional yang bergantung pada kemunculan kata yang serupa, pencarian semantik merepresentasikan teks dalam bentuk vektor berdimensi tinggi (*embedding*) menggunakan model bahasa yang telah dilatih sebelumnya. Kedekatan makna antara dua teks kemudian diukur menggunakan metrik kemiripan seperti *cosine similarity*, yang menghitung sudut antara dua vektor dalam ruang berdimensi tinggi. Semakin kecil sudut antara dua vektor, semakin tinggi kemiripan semantik antara kedua teks tersebut (Pratomo & Utami, 2025).

Vector database merupakan sistem basis data yang dioptimalkan untuk menyimpan, mengindeks, dan melakukan pencarian pada data berbentuk vektor secara efisien untuk mendukung pengambilan konteks semantik secara presisi (Naquila & Ricafort, 2026). Pada penelitian ini mengimplementasikan *ChromaDB* sebagai vector database yang bertindak sebagai jalur cadangan (slow lane) di dalam arsitektur *Hybrid NER* yang diusulkan. Ketika mekanisme fuzzy matching pada jalur utama tidak berhasil menemukan kecocokan entitas yang memadai, sistem akan melakukan pencarian semantik pada *ChromaDB* untuk menemukan entitas yang relevan berdasarkan kedekatan makna. Hasil pencarian ini kemudian digunakan sebagai metode *Retrieval Augmented Generation* (RAG) yang

https://lib.mercubuana.ac.id

disuntikkan sebagai konteks tambahan ke dalam prompt LLM. Hal ini memastikan model bahasa dapat tetap menghasilkan respons yang akurat meskipun entitas awal tidak terdeteksi melalui pencocokan kata secara langsung.

2.3.5 *Sentiment Analysis* & Live2D

Analisis sentimen merupakan teknik pemrosesan bahasa alami (NLP) yang bertujuan untuk mengidentifikasi dan mengekstraksi informasi subjektif dari teks, khususnya polaritas emosi yang terkandung di dalamnya. Dalam implementasi yang lebih sederhana, metode ini dapat dilakukan melalui pendekatan berbasis kamus kata (*lexicon-based*), di mana sistem mencocokkan kata-kata dalam teks dengan daftar kata yang telah diklasifikasikan ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral (Malebary & Abulfaraj, 2024). Pada sistem asisten virtual, analisis sentimen diterapkan pada teks jawaban dari *Large Language Model* (LLM) untuk mengenali status emosinya. Tujuannya adalah agar sistem bisa memicu perubahan ekspresi pada karakter virtual, sehingga raut wajah yang ditampilkan benar-benar cocok dengan isi jawabannya.

Live2D merupakan teknologi animasi yang memungkinkan ilustrasi karakter dua dimensi (2D) digerakkan secara dinamis tanpa memerlukan pemodelan tiga dimensi (3D). Teknologi ini bekerja dengan cara mendeformasi *mesh* poligon yang dipetakan pada gambar karakter berdasarkan parameter-parameter gerakan yang telah ditentukan, seperti pergerakan mata, dan mulut (Lu et al., 2021). Dalam penelitian ini, Live2D digunakan sebagai antarmuka visual karakter asisten virtual yang mampu menampilkan berbagai ekspresi wajah secara dinamis. Ekspresi yang ditampilkan ditentukan oleh hasil analisis sentimen terhadap teks respons, di mana sentimen positif memicu ekspresi senang (*happy*), sentimen negatif memicu ekspresi sedih (*sad*), dan sentimen netral menampilkan ekspresi diam (*idle*). Integrasi ini menciptakan respons asisten yang interaktif dan sinkron, sehingga secara signifikan meningkatkan kesan alami (naturalitas) dalam berinteraksi dengan pengguna.

https://lib.mercubuana.ac.id