

# IMPLEMENTASI NLP (NATURAL LANGUAGE PROCESSING) CHATBOT DESKRIPSI GUA DI KABUPATEN MAROS

Fadly Shabir<sup>\*1</sup>, Billy Eden Wiliam Asrul<sup>2</sup>

<sup>1</sup>Politeknik Negeri Media Kreatif, Jakarta, Indonesia

<sup>2</sup>Universitas Handayani Makassar, Makassar, Indonesia

Penulis korespondensi: Fadly Shabir (fadlyshabir@polimedia.ac.id)

*Kabupaten Maros dikenal dengan kawasan Karst Maros-Pangkep yang menyimpan gua-gua bernilai sejarah, budaya, dan geologi tinggi. Namun, informasi tentang gua-gua tersebut masih tersebar dan belum terintegrasi dalam satu platform yang mudah diakses. Penelitian ini merancang dan mengimplementasikan chatbot berbasis Natural Language Processing (NLP) untuk memberikan deskripsi gua secara otomatis dan interaktif. Sistem menggunakan model DistilBERT untuk pemrosesan bahasa alami dan Supabase sebagai basis data berbasis cloud. Tiga gua yang menjadi fokus penelitian adalah Gua Kharisma, Gua Patta, dan Gua Leang Sampeang Dua. Dataset mencakup pertanyaan dan jawaban dengan lima kriteria: sejarah, ornamen, ukuran, lokasi, dan ciri khas gua. Pengujian black box menunjukkan sistem berfungsi sesuai rancangan dengan respons cepat, akurat, dan relevan.*

**Kata Kunci:** Natural Language Processing, chatbot, DistilBERT, gua, Maros, sistem pendukung keputusan.

## I. PENDAHULUAN

Gua merupakan salah satu bentang alam yang memiliki nilai penting dari sisi geologi, arkeologi, sejarah, dan ekologi. Kabupaten Maros, Sulawesi Selatan, dikenal karena kawasan Karst Maros-Pangkep yang merupakan salah satu kawasan karst terbesar di dunia [1]. Gua-gua di Maros memiliki karakteristik unik, mulai dari stalaktit dan stalagmit, hingga lukisan prasejarah yang menjadi bukti kehidupan manusia purba.

Permasalahan utama yang dihadapi adalah keterbatasan informasi deskriptif yang tersedia bagi masyarakat umum maupun peneliti. Informasi gua masih tersebar dalam berbagai bentuk — laporan tertulis, catatan lapangan, atau publikasi akademik — yang belum terintegrasi dalam satu sistem mudah diakses.

Solusi yang diusulkan adalah sistem chatbot berbasis Natural Language Processing (NLP) yang mampu memberikan deskripsi gua secara otomatis dan interaktif. NLP merupakan cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan manusia melalui bahasa alami [2]. Chatbot berbasis NLP telah banyak diaplikasikan pada berbagai domain layanan informasi, termasuk layanan wisata dan akademik [3], [4], [5]. Tiga gua yang menjadi fokus penelitian adalah Gua Kharisma,

Gua Patta, dan Gua Leang Sampeang Dua di Kabupaten Maros.

Penelitian ini mengimplementasikan model DistilBERT dan platform Supabase sebagai penyimpanan data berbasis cloud, dengan tujuan meningkatkan aksesibilitas informasi wisata gua dan mendukung pengembangan pariwisata berbasis teknologi di Kabupaten Maros. Penggunaan chatbot berbasis NLP untuk informasi pariwisata telah terbukti mampu meningkatkan efisiensi layanan informasi wisata secara signifikan [6].

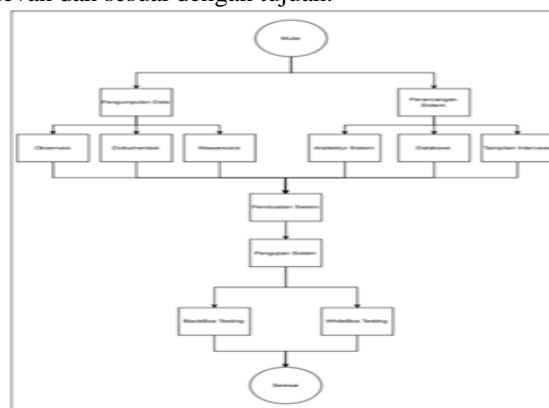
## II. METODE PENELITIAN

### A. Metode Pengembangan Sistem

Pengembangan sistem dilakukan menggunakan pendekatan prototyping yang terdiri dari tahapan: pengumpulan data, perancangan sistem, pembuatan sistem, dan pengujian sistem [7]. Pengumpulan data dilakukan melalui observasi langsung ke tiga gua di Kabupaten Maros, wawancara dengan pengelola wisata, serta dokumentasi dari literatur dan portal pemerintah.

### B. Tahapan Penelitian

Tahapan penelitian dalam pengembangan sistem pendukung keputusan asisten deskripsi gua di Maros dilakukan secara sistematis agar hasil yang diperoleh relevan dan sesuai dengan tujuan.



Gambar 1. Tahapan Penelitian

#### 1. Mulai

Tahapan ini adalah langkah awal dimulainya penelitian atau pengembangan sistem. Pada tahap ini ditetapkan

tujuan, ruang lingkup, dan arah dari kegiatan yang akan dilakukan. Semua aktivitas selanjutnya berangkat dari titik awal ini sebagai acuan.

## 2. Pengumpulan Data

Pengumpulan data merupakan tahap penting untuk memperoleh informasi dasar yang akan digunakan dalam pembangunan sistem. Data dikumpulkan melalui beberapa metode seperti observasi, dokumentasi, dan wawancara. Tujuannya adalah memahami kebutuhan pengguna dan kondisi nyata di lapangan.

## 3. Observasi

Observasi dilakukan dengan cara mengamati langsung objek penelitian atau lokasi yang relevan. Melalui observasi, peneliti dapat melihat kondisi nyata yang tidak selalu bisa digambarkan dengan data tertulis. Hal ini membantu memperoleh data yang lebih akurat dan faktual.

## 4. Dokumentasi

Metode dokumentasi dilakukan dengan cara mengumpulkan data dari dokumen, catatan, arsip, laporan, atau sumber tertulis lainnya. Dokumentasi memberikan informasi pendukung yang berguna untuk melengkapi hasil observasi maupun wawancara, sehingga data lebih komprehensif.

## 5. Wawancara

Wawancara dilakukan dengan cara bertanya langsung kepada narasumber atau pihak terkait. Tujuannya untuk mendapatkan informasi yang lebih mendalam, baik berupa pengalaman, pendapat, maupun kebutuhan pengguna. Wawancara membantu menemukan detail yang tidak bisa diperoleh melalui observasi dan dokumentasi saja.

## 6. Perancangan Sistem

Setelah data terkumpul, tahap berikutnya adalah merancang sistem. Perancangan mencakup bagaimana sistem akan dibangun, mulai dari arsitektur, desain database, hingga tampilan antarmuka. Tahap ini sangat penting karena menentukan struktur dasar dari sistem yang akan dikembangkan.

## 7. Arsitektur Sistem

Arsitektur sistem adalah rancangan kerangka dasar yang menjelaskan bagaimana komponen sistem saling terhubung. Pada tahap ini ditentukan alur kerja sistem, modul-modul yang dibutuhkan, dan hubungan antarbagian.

## 8. Database

Database dirancang untuk menyimpan data yang diperlukan dalam sistem. Perancangan database mencakup struktur tabel, relasi antar tabel, serta jenis data yang akan disimpan. Database yang baik memungkinkan sistem dapat mengolah data secara cepat, aman, dan terstruktur.

## 9. Tampilan Antarmuka

Tampilan antarmuka (interface) merupakan bagian sistem yang langsung berinteraksi dengan pengguna. Pada tahap ini, dirancang desain yang sederhana, mudah digunakan, dan informatif agar pengguna merasa nyaman saat menggunakan sistem.

## 10. Pembuatan Sistem

Tahap ini merupakan implementasi nyata dari rancangan yang telah dibuat. Pembuatan sistem mencakup proses coding atau pemrograman sesuai dengan desain arsitektur, database, dan tampilan antarmuka yang sudah dirancang sebelumnya.

## 11. Pengujian Sistem

Setelah sistem selesai dibuat, dilakukan pengujian untuk memastikan bahwa sistem berjalan sesuai dengan kebutuhan dan tidak terdapat kesalahan yang signifikan. Pengujian bertujuan memvalidasi fungsi sistem dan menilai kualitas kinerjanya.

## 12. BlackBox Testing

*BlackBox Testing* adalah metode pengujian yang berfokus pada fungsionalitas sistem tanpa melihat kode program. Penguji hanya memberikan input dan memeriksa output untuk memastikan hasilnya sesuai dengan yang diharapkan.

## 13. WhiteBox Testing

*WhiteBox Testing* adalah metode pengujian yang melihat ke dalam struktur kode program. Tujuannya adalah untuk memastikan bahwa alur logika, algoritma, dan struktur internal program sudah benar dan tidak ada bug tersembunyi.

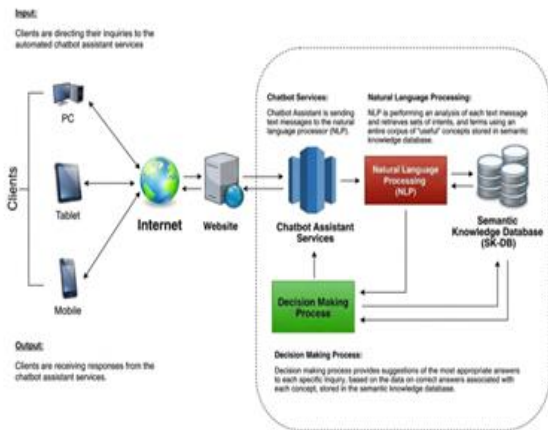
## 14. Selesai

Tahapan terakhir adalah selesai, yaitu saat semua proses dari pengumpulan data, perancangan, pembuatan, hingga pengujian telah dilakukan. Pada tahap ini sistem dianggap siap digunakan sesuai dengan kebutuhan pengguna.

## C. Arsitektur Sistem

Sistem dibangun dengan arsitektur client-server berbasis web [8]. Komponen utama meliputi: (1) Antarmuka pengguna berbasis web menggunakan PHP dan JavaScript; (2) Modul NLP berbasis DistilBERT dari HuggingFace Transformers; (3) Basis data Supabase berbasis cloud; dan (4) Chatbot Assistant Services sebagai penghubung antarmuka dan modul NLP.

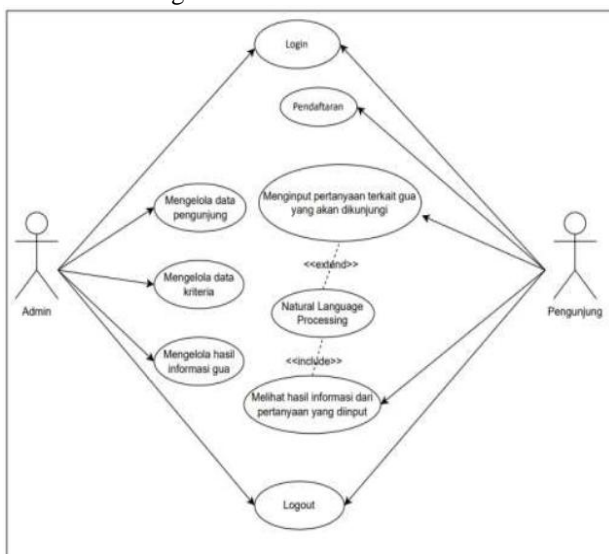
Alur kerja sistem dimulai dari pengguna mengirimkan pertanyaan melalui antarmuka web. Pertanyaan diteruskan ke modul NLP yang melakukan preprocessing, pencocokan dengan dataset, dan jika tidak ditemukan padanan, model DistilBERT digunakan untuk menghasilkan jawaban berbasis konteks.



Gambar 2. Arsitektur Sistem

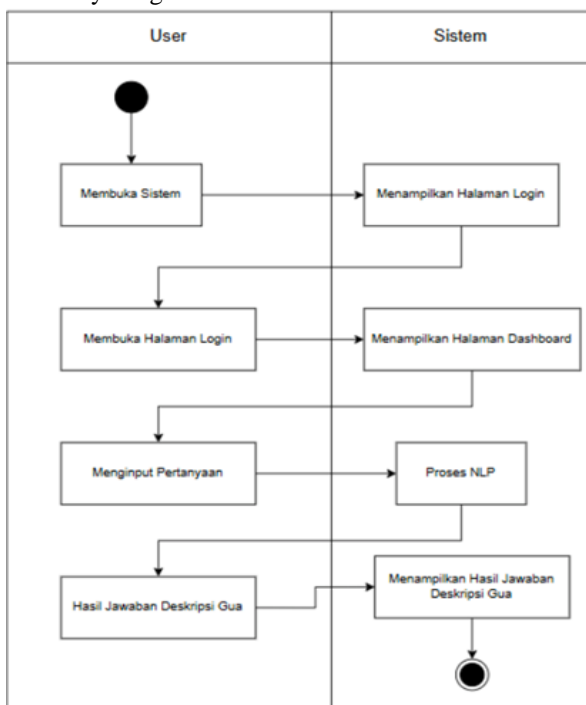
D. Pemodelan Sistem

1. Use Case Diagram



Gambar 3. Use Case Diagram

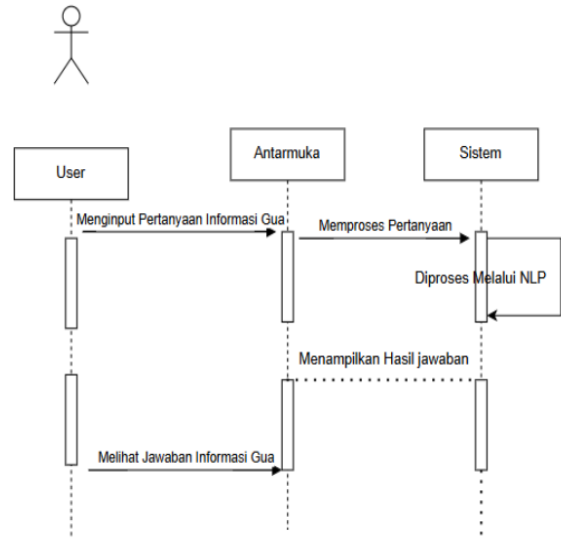
2. Activity Diagram



Gambar 4. Activity Diagram

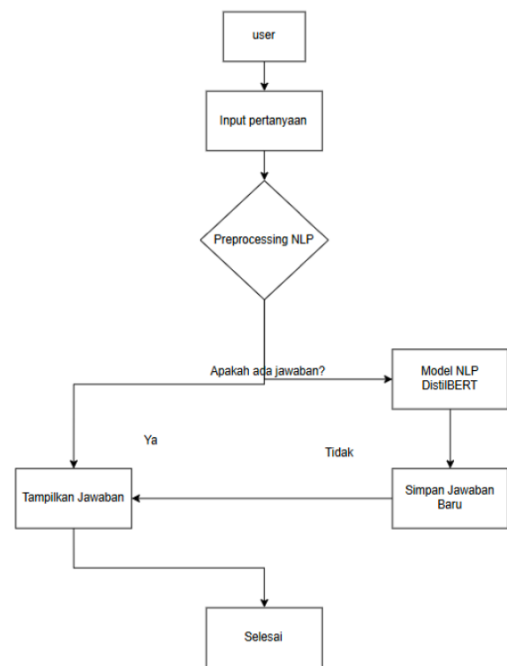
3. Sequence Diagram

Diagram sequence (diagram urutan interaksi) yang menjelaskan alur komunikasi antara User, Antarmuka, dan Sistem dalam aplikasi chatbot informasi gua.



Gambar 5. Sequence Diagram

E. Perancangan Sistem



Gambar 6. Diagram Prototype

Diagram prototype untuk perancangan sistem yang akan dibangun. Pada gambar diatas user menginput 79 pertanyaan yang diproses melalui metode NLP. Jika jawaban ada, maka jawaban akan ditampilkan. Jika tidak ada, model NLP DistilBERT akan bekerja dan menyimpan jawaban yang baru kemudian jawaban ditampilkan.

F. Dataset dan Preprocessing NPL

Dataset dibangun dari hasil observasi dan studi literatur, mencakup pertanyaan dan jawaban tentang tiga gua dengan lima kriteria: sejarah terbentuknya gua, ornamen di dalam gua, ukuran gua, lokasi gua, dan ciri

kelas. Dataset disimpan dalam tiga kolom: question, answer, dan context di Supabase.

Preprocessing NLP mencakup dua tahap utama. Pertama, pembersihan teks (text cleaning) yang menghapus karakter non-alfanumerik dan mengubah seluruh huruf menjadi huruf kecil. Kedua, pencocokan pertanyaan (question matching) menggunakan metode string normalization dengan fungsi dataset.find() [9].

Tabel 1. Contoh Dataset Pertanyaan dan Jawaban

No.	Pertanyaan	Jawaban
1	Apa itu Gua Leang-Leang?	Situs prasejarah di Maros
2	Di mana Gua Petta Kere?	Karst Maros-Pangkep, Sulsel
3	Apa keunikan Gua Bulu' Sipong?	Lukisan tangan purba dan karst
4	Apa makna arkeologi Gua Maros?	Lukisan manusia purba tertua dunia
5	Apakah gua terbuka umum?	Ya, Leang-Leang & Rammang-Rammang

### G. Model NPL DistilBERT

Model yang digunakan adalah DistilBERT — varian ringan dari BERT yang dioptimalkan untuk efisiensi dan kecepatan [10] - diimplementasikan melalui pipeline question-answering dari HuggingFace Transformers. Pipeline menerima dua masukan (question dan context) kemudian menghasilkan jawaban beserta skor kepercayaan.

Validasi tipe data diterapkan sebelum data dikirim ke pipeline untuk memastikan question dan context bertipe string. Jika tidak sesuai, sistem merespons dengan kode 400 (Bad Request). Inferensi dijalankan dengan perintah await answerer(question, context). Pendekatan ini sejalan dengan arsitektur model bahasa besar modern seperti RoBERTa [11] dan PaLM [12] yang juga mengandalkan mekanisme attention berbasis transformer untuk pemahaman konteks.

### H. Lokasi Penelitian

Penelitian dilakukan di tiga gua Kabupaten Maros: (1) Gua Patta — 04°58'44,6" LS / 119°40'30,5" BT, Kel. Leang-leang; (2) Gua Leang Sampeang Dua — 04°59'44,8" LS / 119°40'01,3" BT, Kel. Kalabbirang; (3) Gua Kharisma - 5°01'34" LS / 119°44'32" BT, Ds. Labuaja. Penelitian berlangsung Maret–Agustus 2025.

## III. HASIL DAN PEMBAHASAN

### A. Pemodelan Sistem

Sistem dirancang menggunakan diagram UML (Use Case, Activity, dan Sequence Diagram). Use Case Diagram menunjukkan dua aktor: Admin dan Pengunjung. Admin mengelola data pertanyaan-jawaban dan memastikan validitas informasi gua. Pengunjung dapat mendaftar akun, login, menginput pertanyaan, dan melihat hasil.

Activity Diagram menggambarkan alur dari pengguna membuka sistem hingga tampilnya jawaban. Sequence Diagram menjelaskan komunikasi: User → Antarmuka → Sistem NLP → Antarmuka → User.

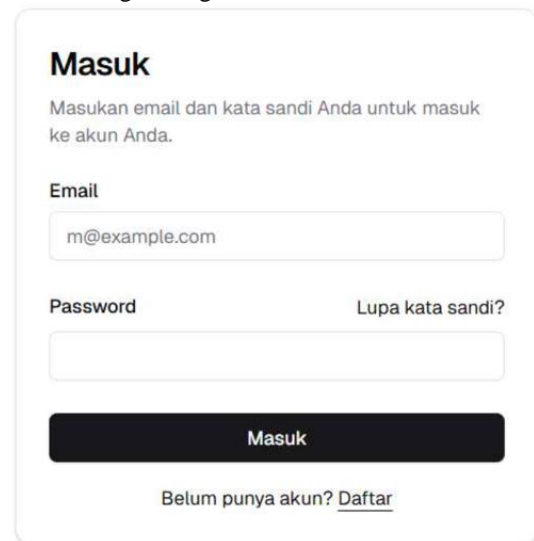
### B. Alur Pemrosesan Prototype

Alur pemrosesan mengikuti diagram prototype: pengguna menginput pertanyaan → preprocessing NLP → cek ketersediaan jawaban di database. Jika ditemukan, jawaban langsung ditampilkan. Jika tidak, DistilBERT memproses dan menghasilkan jawaban baru yang disimpan ke database lalu ditampilkan.

### C. Implementasi Antarmuka

Sistem memiliki empat halaman utama: (1) Halaman Login dengan autentikasi email; (2) Halaman Admin untuk mengelola dataset; (3) Halaman Dashboard sebagai antarmuka input pertanyaan; (4) Halaman Hasil yang menampilkan jawaban dalam bahasa Indonesia.

#### 1. Halaman Login dengan autentikasi email



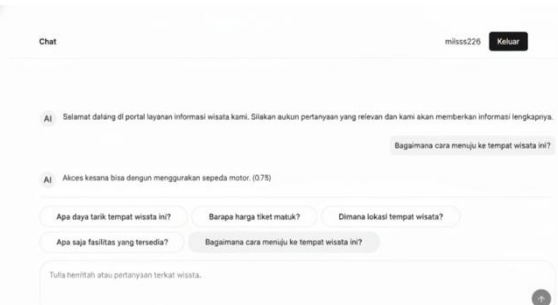
Gambar 7. Halaman Login

#### 2. Halaman Admin



Gambar 8. Halaman Admin

#### 3. Halaman Dashboard dan halaman hasil



Gambar 9. Halaman Dashboard dan Hasil

*D. Implementasi Algoritma NPL*

Implementasi dilakukan dalam lima langkah: (1) Menyiapkan dataset terstruktur (question, answer, context); (2) Preprocessing teks dengan fungsi bersihkan(); (3) Pencocokan pertanyaan menggunakan dataset.find(); (4) Validasi tipe data input; (5) Pemrosesan pipeline DistilBERT untuk inferensi jawaban.

*E. Pengujian Black Box*

Pengujian Black Box dilakukan untuk memvalidasi fungsionalitas sistem dari perspektif pengguna tanpa melihat kode internal. Empat skenario pengujian dijalankan berdasarkan halaman sistem.

Tabel 2. Hasil Pengujian Black Box

No	Form	Hasil Diharapkan	Valid
1	Login	Diarahkan ke dashboard/admin	✓
2	Admin	Input pertanyaan dan jawaban gua	✓
3	Dashboard	Tampil halaman input pertanyaan	✓
4	Hasil	Tampil jawaban informasi gua	✓

Hasil pengujian menunjukkan seluruh skenario valid. Saat jawaban tersedia di database, respons diberikan sangat cepat tanpa inferensi model. Jika tidak ditemukan, DistilBERT menghasilkan respons relevan dari konteks deskriptif gua.

*F. Evaluasi Sistem*

Evaluasi pengguna menunjukkan jawaban sistem jelas, informatif, dan sesuai pertanyaan. Supabase terbukti efektif untuk pengelolaan data real-time. Preprocessing NLP meningkatkan akurasi pencocokan meskipun terdapat variasi penulisan dari pengguna. [MOU1] Evaluasi sistem dilakukan melalui User Acceptance Test (UAT) yang melibatkan 20 responden, terdiri dari 10 mahasiswa Teknik Informatika dan 10 pengunjung wisata gua di Kabupaten Maros. Instrumen pengumpulan data menggunakan kuesioner berbasis skala Likert (1–5) dengan lima indikator: kejelasan jawaban, relevansi informasi, kemudahan penggunaan, kecepatan respons, dan kepuasan keseluruhan. Hasil UAT menunjukkan tingkat kepuasan pengguna sebesar 84,6%, dengan rata-rata skor 4,23 dari skala 5 pada indikator kejelasan jawaban, dan 4,15 pada indikator relevansi informasi. Sebanyak 90% responden menyatakan bahwa jawaban sistem sesuai dengan pertanyaan yang diajukan. Supabase terbukti efektif untuk pengelolaan data real-time, dan preprocessing NLP meningkatkan akurasi pencocokan meskipun terdapat variasi penulisan dari pengguna [13], [14]. Temuan ini konsisten dengan penelitian yang menunjukkan potensi besar model bahasa berbasis AI dalam mendukung pengalaman belajar dan pencarian informasi pengguna secara interaktif [15].

Model DistilBERT mampu memproses pertanyaan secara efisien baik dari dataset maupun dari konteks deskriptif; (3) Supabase mendukung pengelolaan data real-time dengan performa tinggi; (4) Pengujian black box membuktikan sistem berfungsi sesuai spesifikasi.

Saran untuk penelitian selanjutnya: (1) Perluas dataset dengan pertanyaan lebih bervariasi; (2) Integrasikan media visual seperti peta interaktif, gambar gua, atau video 360 derajat untuk meningkatkan pengalaman pengguna.

DAFTAR PUSTAKA

- [1] M. A. Labib, "Identifikasi Kenampakan Morfologi Gua Banyu dan Sekitarnya di Kecamatan Donomulyo Kabupaten Malang," *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, vol. 5, no. 2, pp. 185–195, Dec. 2021, doi: 10.29408/geodika.v5i2.3323.
- [2] A. Sambolangi, M. Arsyad, V. Arini, T. Karakteristik, G. Kelelawar, and D. Kawasan, "Karakteristik Gua Kelelawar Di Kawasan Karst Rammang-Rammang Maros Tn Bantimurung Bulusaraung," 2021.
- [3] M. Ilyas, T. Khaqiqi, N. H. Harani, and N. A. Id, "Terbit online pada laman web jurnal: <https://ejournalunsam.id/index.php/jicom/> Peningkatan Kinerja Chatbot NLP Asisten: Tinjauan Literatur tentang Metode dan Akurasi dalam Aplikasi Berbasis Percakapan", [Online]. Available: <https://ejournalunsam.id/index.php/jicom/>
- [4] W. Hadinata and L. Stianingsih, "Wira Hadinat a Implementasi Natural Language Processing Implementasi Natural Language Processing pada Chatbot Untuk Helpdesk Informasi Wisata (Studi kasus: Tangerang Raya)." 2024.
- [5] P. B. Wintoro, H. Hermawan, M. A. Muda, and Y. Mulyani, "Implementasi Long Short-Term Memory pada Chatbot Informasi Akademik Teknik Informatika Unila," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 12, no. 1, p. 68, Jun. 2022, doi: 10.36448/expert.v12i1.2593.
- [6] R. Ady Sanjaya and E. Winarno, "Pengembangan Chatbot Informasi Pariwisata di Kabupaten Pati Menggunakan Metode Natural Language Processing Berbasis Dialogflow".
- [7] S. Sulistiyanto, L. Rahmi, N. I. H. Kunio, E. Asoka, and A. S. Wati, "Implementasi Web Development Life Cycle dalam Pembuatan Website Company Profile Ening Production," *Jurnal Penelitian Inovatif*, vol. 4, no. 2, pp. 387–394, Apr. 2024, doi: 10.54082/jupin.324.
- [8] R. Namruddin and M. Author Affiliations, "Innovative Strategies For Integrating Programming Skills In Informatics Engineering Department," 2024. [Online]. Available: <https://ojs.unm.ac.id/vokasiketeknikan>
- [9] D. Oleh and D. Alfian, "Final Project Automatic Summarization System For Medium Articles Using The T5 Algorithm Judul Skirpsi," 2024.
- [10] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Mar. 2020, [Online]. Available: <http://arxiv.org/abs/1910.01108>
- [11] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>
- [12] A. Chowdhery *et al.*, "PaLM: Scaling Language Modeling with Pathways," Oct. 2022, [Online]. Available: <http://arxiv.org/abs/2204.02311>
- [13] A. Holzinger *et al.*, "Towards the Augmented Pathologist: Challenges of Explainable-AI in Digital Pathology," Dec. 2017, [Online]. Available: <http://arxiv.org/abs/1712.06657>
- [14] M. Irsan, "Evaluasi Tingkat Usability Sistem Informasi Akademik (SIKAD) Menggunakan Metode System Usability Scale (SUS)." 2024.
- [15] E. Kasneci *et al.*, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education."

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian disimpulkan: (1) Sistem chatbot NLP berhasil dirancang dan diimplementasikan untuk informasi deskriptif gua di Kabupaten Maros; (2)