

Implementasi Konsultasi Stunting Balita Menggunakan Large Language Models (LLMs)

Tanwir¹, Khasnur Hidjah², Dyah Susilowati³

^{1,2,3}Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Bumigora
Jl. Ismail Marzuki No 22. Cakranegara, Mataram
e-mail: ¹tanwir@universitasbumigora.ac.id, ²khasnur72.h@universitasbumigora.ac.id,
³dyah.bumigora@gmail.com

Abstrak - Stunting pada balita merupakan masalah kesehatan kritis di Indonesia yang memerlukan intervensi berbasis teknologi untuk meningkatkan akses informasi nutrisi. Penelitian ini bertujuan mengembangkan chatbot konsultasi stunting berbasis *Large Language Models* (LLMs) guna menyediakan rekomendasi kesehatan yang akurat dan mudah diakses. Metode yang digunakan berupa Model LLaMA 3 di-*fine-tuning* menggunakan dataset Q&A spesifik stunting berisi 7.642 entri, kemudian dievaluasi dengan matrik ROUGE untuk mengukur kesesuaian semantik respons. Hasil menunjukkan model Stunting mencapai skor ROUGE-1 (72,24%), ROUGE-2 (64,54%), ROUGE-L (70,42%), dan ROUGE-Lsum (70,96%), secara signifikan melampaui model baseline seperti LLaMA3, Deepseek-R1, dan Mistral. Chatbot diimplementasikan dalam aplikasi web berbasis *cloud* dengan arsitektur terdistribusi, dilengkapi enkripsi SSL dan HTTPS untuk menjamin keamanan data. Sistem ini memungkinkan interaksi real-time antara pengguna dan model LLMs melalui antarmuka berbasis *Gradio*. Temuan penelitian mengonfirmasi potensi LLMs dalam menyederhanakan layanan kesehatan preventif, khususnya di daerah dengan sumber daya terbatas.

Kata Kunci: Stunting balita, *Large Language Models* (LLMs), LLaMA 3, Matrik ROUGE, Chatbot konsultasi kesehatan.

PENDAHULUAN

Stunting pada balita merupakan masalah kesehatan masyarakat yang signifikan di Indonesia, ditandai dengan gangguan pertumbuhan dan perkembangan akibat malnutrisi kronis [1]. Laporan *United Nations International Children's Emergency Fund (UNICEF)* menunjukkan 148 juta anak balita mengalami stunting [2]. Di Indonesia angka prevalensi anak balita yang terkena stunting sebesar 21.6% [3] sedangkan Nusa Tenggara Barat (NTB) sendiri berkisar pada angka 18.34% [4]. Stunting ditandai dengan gangguan pertumbuhan dan perkembangan akibat kekurangan gizi kronis yang terjadi dalam jangka waktu panjang, terutama selama 1.000 hari pertama kehidupan [5]. Kondisi ini negatif pada perkembangan fisik dan kognitif, meningkatkan risiko gangguan perkembangan dan penyakit metabolik di masa dewasa [6]. Berdasarkan data Badan Survei Status Gizi Balita Indonesia (SSGBI), prevalensi stunting di Indonesia masih tergolong tinggi dan memerlukan intervensi komprehensif dan berkelanjutan [7].

Salah satu strategi utama dalam menanggulangi stunting adalah penyediaan layanan konsultasi gizi dan kesehatan anak yang mudah diakses, murah, dan informatif. Teknologi chatbot berbasis *Artificial Intelligence* (AI) telah muncul sebagai solusi inovatif untuk menjembatani

kesenjangan layanan, terutama dalam layanan pelanggan [8]. Chatbot dapat memberikan informasi dan saran kepada pengguna secara real-time, tanpa harus bertatap muka dengan tenaga Kesehatan [9].

Beberapa penelitian sebelumnya telah mengembangkan chatbot dalam konteks konsultasi kesehatan menggunakan berbagai pendekatan teknologi, seperti *framework* Rasa, yang mengandalkan pemetaan intent, entitas, dan dialog berbasis pola (*rule-based*) pada kasus chatbot Stunting [10] dan chatbot Diet [11], model Pose-based *Convolutional Neural Network* (P-CNN) untuk klasifikasi percakapan dan pemahaman konteks berbasis urutan kalimat pada kasus chatbot layanan pelanggan [12]. serta pendekatan menggunakan *Neural Language Processing* (NLP) pada kasus chatbot Covid-19 [13]. Namun, pendekatan ini memiliki keterbatasan dalam memahami konteks percakapan yang kompleks, Tidak fleksibel dalam menangani pertanyaan di luar skenario yang telah ditentukan, serta ketergantungan pada data pelatihan yang terbatas.

Kemajuan terbaru di bidang pemrosesan bahasa alami *Natural Language Processing* (NLP), khususnya dengan hadirnya *Large Language Models* (LLMs) seperti GPT, BERT, dan LLaMA, telah memungkinkan pengembangan sistem percakapan yang lebih fleksibel, adaptif, dan mendekati interaksi

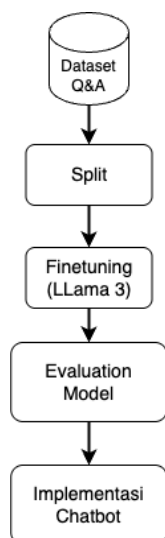


manusia [14][15]. LLMs adalah jenis model pembelajaran mesin yang mempelajari pola statistik dalam teks, memungkinkan untuk memprediksi kata-kata berikutnya dalam urutan [16]. Model ini dapat bersifat umum atau spesifik tugas dan telah menunjukkan potensi di berbagai aplikasi, terutama dalam sains dan kedokteran. LLMs dapat memproses beragam jenis data, termasuk teks ilmiah, catatan medis elektronik, dan database besar urutan biologis, menjadikannya alat yang berharga untuk memahami dan bernalar di berbagai modalitas [16]. Namun demikian, hingga saat ini belum banyak penelitian yang mengimplementasikan integrasi LLMs dalam konteks konsultasi stunting pada balita. Mayoritas studi masih berfokus pada pendekatan berbasis aturan (*rule-based*) atau sistem pembelajaran mesin konvensional tanpa kemampuan pengambilan informasi dinamis.

Penelitian ini bertujuan untuk mengeksplorasi penerapan model chatbot konsultasi stunting balita menggunakan LLMs. Sistem ini dirancang untuk memberikan saran nutrisi dan kesehatan yang akurat, kontekstual, dan sesuai dengan referensi terpercaya seperti pedoman dari Kementerian Kesehatan, WHO. Kontribusi dari penelitian ini adalah pengembangan sebuah aplikasi yang berfungsi untuk memberikan informasi terkait stunting pada balita kepada masyarakat secara informatif dan mudah diakses.

METODE PENELITIAN

Penelitian ini menggunakan LLMs untuk tujuan membantu Masyarakat dalam memberikan informasi terkait saran nutrisi kesehatan pada balita terkait stunting. Pada gambar 1 merupakan alur penelitian model chatbot konsultasi stunting yang terdiri dari Dataset Q&A, Split data, Finetuning (LLama 3), Evaluasi Model dan Implementasi chatbot dalam bentuk aplikasi website.



Gambar 1. Alur penelitian model chatbot stunting

1. Dataset

Dataset dalam bentuk pertanyaan dan jawaban (Q&A) yang bersumber dari buku dan jurnal terkait stunting yang berjumlah 7.642. Sumber dari huggingface dataset dengan link <https://huggingface.co/datasets/kodetr/stunting-qa-2025>

Table 1. Dataset Q&A stunting

No	Pertanyaan	Jawaban
1	Apa saja ciri-ciri anak yang mengalami stunting?	Anak-anak yang mengalami stunting cenderung pendek, kerdil, dan kurus, tetapi tubuhnya tetap proporsional.
2	Kapan tindakan pencegahan untuk mengurangi kejadian stunting harus dimulai?	Tindakan pencegahan untuk mengurangi kejadian stunting harus dimulai sebelum kelahiran melalui perawatan perinatal dan gizi ibu.
....
7.641	Kapan pengukuran lingkaran kepala dilakukan pada anak usia 0-12 bulan?	Pengukuran lingkaran kepala dilakukan setiap 3 bulan pada anak usia 0-12 bulan.
7.642	Apa saja prioritas program pencegahan stunting di Indonesia?	Ibu hamil dan anak umur 0-2 tahun atau rumah tangga 1000 hpk.

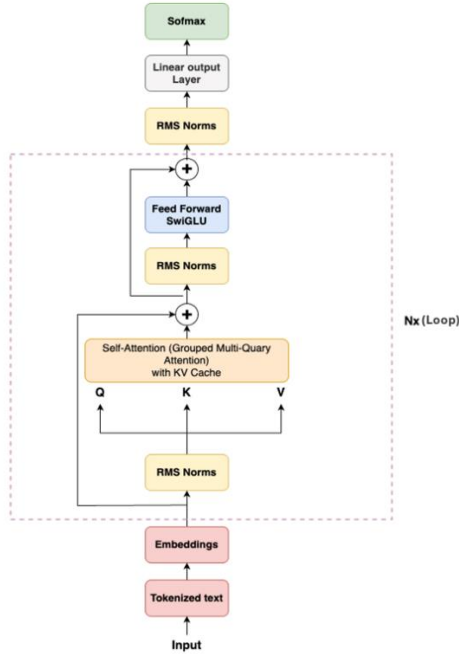
2. Split Data

Setelah itu, dilakukan ekstraksi informasi untuk mengidentifikasi data penting, yang kemudian dibagi menjadi 90% untuk pelatihan dan 10% untuk pengujian, dengan pertimbangan bahwa rasio ini tetap representatif dalam mengevaluasi performa model, terutama untuk dataset yang besar dan kompleks [17].

3. Finetuning

Model Transformer LLaMa 3 digunakan dalam penelitian pembuatan chatbot pencegahan stunting, model Llama 3 menunjukkan performa yang sangat kompetitif, bahkan menyamai atau melampaui model lain seperti GPT-4 pada berbagai benchmark, termasuk pemahaman bahasa, reasoning, dan question answering [18]. Model Llama 3 ditingkatkan melalui fine-tuning dengan dataset stunting. Fine-tuning menyesuaikan model terlatih agar lebih relevan. Dalam Llama 3, teks ditokenisasi menjadi unit kecil (token), lalu dikonversi ke vector melalui

embedding untuk memahami hubungan semantik. Representasi ini dinormalisasi dengan Root Mean Square Normalization (RMS Norms) guna menjaga stabilitas pelatihan dan mencegah exploding atau vanishing gradients [18].



Gambar 2. Arsitektur Tranformer Llama 3

Pada tahap inti, Model menggunakan Self-Attention untuk memahami konteks global teks. Varian Grouped Multi-Query Attention meningkatkan efisiensi dengan mengelompokkan query, terutama pada urutan panjang. KV Cache mempercepat inferensi dengan menyimpan Key (K) dan Value (V), menghindari perhitungan ulang di setiap Langkah, didefinisikan sesuai persamaan (1).

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Setelah proses *Self-Attention*, hasilnya diproses oleh lapisan *Feed Forward* yang menggunakan fungsi aktivasi *Switch-Gated Linear Unit* (SwiGLU). Fungsi SwiGLU memberikan keseimbangan antara linearitas dan non-linearitas, meningkatkan efisiensi komputasi tanpa mengorbankan performa model. Dalam arsitektur ini, output dari *Self-Attention* dan *Feed Forward* saling ditambahkan melalui *Residual Connection*, sebuah teknik yang menjaga informasi dari lapisan sebelumnya tetap terjaga sehingga tidak hilang selama propagasi melalui jaringan, didefinisikan sesuai persamaan (2).

$$FFD(z) = \max(0, zW_1 + b_1)W_2 + b_2 \quad (2)$$

Proses ini diulang N kali agar model lebih memahami hubungan antar token. Hasil akhirnya diteruskan ke Linear Output Layer, yang mengonversinya menjadi distribusi probabilitas melalui Softmax. Fungsi ini memastikan total probabilitas seluruh kelas menjadi 1, memungkinkan prediksi akurat, defined asin Eq (3).

$$P(y) = \text{softmax}(z) = \frac{\exp(z_j)}{\sum_{k=1}^{|V|} \exp(z_k)} \quad (3)$$

4. Evaluasi Model

Evaluasi terhadap LLMs, dalam penelitian ini menggunakan matrik ROUGE untuk mengukur kesamaan semantik antara teks prediksi. Recall-Oriented Understudy for Gisting Evaluation (ROUGE) merupakan metode evaluasi otomatis yang sering digunakan untuk menilai kualitas ringkasan yang dihasilkan oleh mesin dengan membandingkannya terhadap ringkasan referensi yang dibuat oleh manusia. ROUGE menghitung jumlah overlap kata atau n-gram antara ringkasan yang dihasilkan sistem dengan ringkasan referensi, defined asin Eq (5), (6) dan (7) [19].

$$ROUNGE - 1 = \frac{\text{Number of matching unigrams}}{\text{Total number of nigrams in the reference}} \quad (5)$$

$$ROUNGE - 2 = \frac{\text{Number of matching bigrams}}{\text{Total number of bigrams in the reference}} \quad (6)$$

$$ROUNGE - L = \frac{LCS(\text{Result text}, \text{References})}{\text{Reference Length}} \quad (7)$$

(i) ROUGE-1: Mengukur sejauh mana teks hasil (*generated text*) mencakup kata-kata individu (disebut *unigram*) dari teks referensi, ROUGE-2: Matrik ini menggunakan bigram (urutan dua kata) untuk membandingkan teks hasil model dengan teks referensi, ROUGE-L: Menggunakan konsep Longest Common Subsequence (LCS) untuk mengevaluasi kualitas teks yang dihasilkan dibandingkan dengan teks referensi, Jumlah unigram yang cocok: Total kata yang muncul di teks hasil dan juga ada di teks referensi, Jumlah total unigram dalam referensi: Total semua kata yang ada dalam teks referensi tanpa pengulangan, Jumlah bigram yang cocok: Jumlah bigram yang sama persis antara teks yang dihasilkan oleh model dengan teks referensi, Jumlah total bigram dalam referensi: Total jumlah bigram yang ada dalam teks referensi, LCS (Teks hasil, Referensi): Panjang subsekuensi terpanjang yang sama (Longest Common Subsequence) antara teks yang dihasilkan

dan teks referensi, Panjang Referensi: Jumlah kata dalam teks referensi.

6. Implementasi Chatbot

Pada tahap implementasi, model chatbot yang telah melalui proses fine-tuning dan evaluasi kemudian diterapkan ke dalam bentuk aplikasi berbasis website yang dirancang untuk memberikan informasi seputar stunting pada balita kepada masyarakat. Aplikasi ini memiliki antarmuka yang sederhana dan interaktif, memungkinkan pengguna untuk mengajukan pertanyaan secara langsung. Pertanyaan tersebut diproses oleh backend yang telah diintegrasikan dengan model LLaMA 3 hasil fine-tuning, dan jawaban yang relevan akan ditampilkan kepada pengguna. Komunikasi antara frontend dan backend dilakukan melalui *Application Programming Interface* (API).

HASIL DAN PEMBAHASAN

Dataset Q&A, di mana dari total 7.642 data yang tersedia, sebanyak 90% atau 6.878 data digunakan sebagai data pelatihan (*training*), sementara 10% sisanya atau 764 data digunakan sebagai data validasi (*validation*). Pembagian ini bertujuan agar model dapat dilatih menggunakan data pelatihan, sedangkan data validasi digunakan untuk memantau performa model selama proses pelatihan guna menghindari overfitting dan membantu dalam proses seperti pemilihan model terbaik.

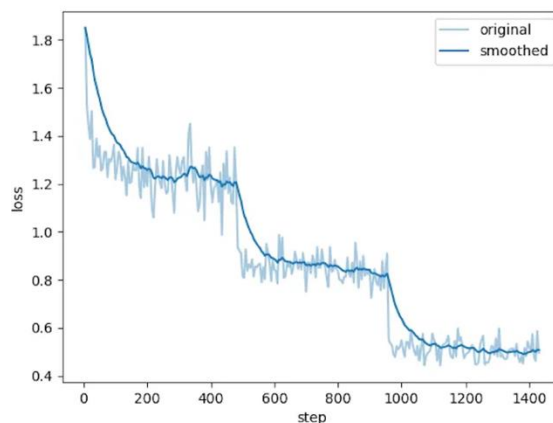
Model LLMs Stunting dilatih menggunakan GPU Nvidia L4 dengan VRAM 24GB, kecepatan TFLOPs sebesar 121. Table 2 menunjukkan hasil matrix pelatihan.

Tabel 2. Matrik Pelatihan

Matrix	Value
<i>Epoch</i>	3
<i>Num input tokens seen</i>	19.009.976
<i>Total flos</i>	799.440.666
<i>Train loss</i>	0.872
<i>Train runtime</i>	1:06:36
<i>Train samples persecond</i>	5.737
<i>Train steps persecond</i>	0.358

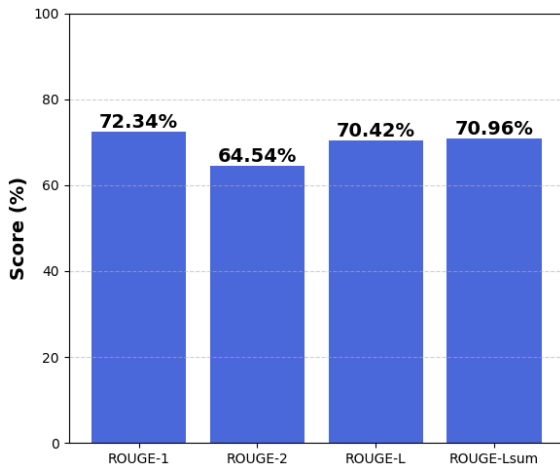
Hasil pelatihan model menunjukkan bahwa proses training dilakukan selama 2,9987 *epoch* dengan jumlah token input yang diproses sebanyak 19.009.976. Total operasi *floating point* (FLOPs) yang dicapai adalah sebesar 799.440.666 GFLOPs, dengan nilai *train loss* sebesar 0,872, yang mengindikasikan bahwa model telah berhasil belajar dari data dengan performa yang cukup baik. Waktu yang

dibutuhkan untuk menyelesaikan pelatihan mencapai 1 jam 6 menit 36 detik, dengan rata-rata kecepatan pelatihan sebesar 5,737 sampel per detik dan 0,358 langkah pelatihan per detik. Gambar 3 menunjukkan grafik penurunan nilai loss terhadap jumlah langkah (*step*) selama proses pelatihan model. Kurva biru muda mewakili nilai loss asli yang bersifat fluktuatif, sedangkan kurva biru tua menunjukkan versi yang telah dihaluskan (*smoothed*) untuk memberikan representasi tren yang lebih jelas. Dari grafik ini terlihat bahwa nilai loss menurun secara konsisten seiring bertambahnya jumlah langkah, yang mengindikasikan bahwa model mengalami konvergensi dan pembelajaran berjalan efektif. Penurunan bertahap yang signifikan pada beberapa titik juga mengindikasikan adanya fase-fase pelatihan yang memungkinkan model memperbaiki kesalahan secara bertahap.



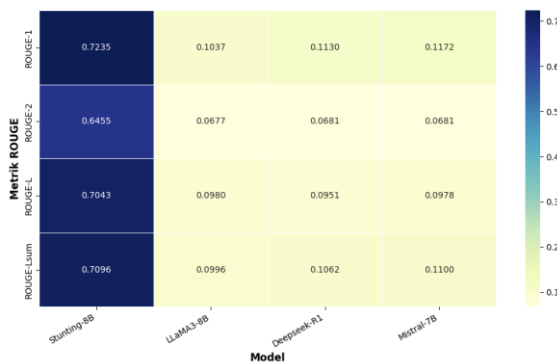
Gambar 3. Stunting model training loss graph

Pengujian dilakukan menggunakan 50 sampel data uji secara subjektif yang melibatkan seorang profesional dalam bidang Kesehatan terkait stunting pada balita untuk mengidentifikasi respon yang dihasilkan. Gambar 3. Menunjukkan hasil evaluasi model chatbot untuk konsultasi pencegahan stunting menggunakan matrik ROUGE menunjukkan kinerja yang sangat baik. Model memperoleh skor ROUGE-1 sebesar 72.24%, yang mengindikasikan bahwa sebagian besar kata-kata penting dari jawaban referensi berhasil direproduksi. Skor ROUGE-2 sebesar 64.54% menunjukkan bahwa model mampu mempertahankan hubungan antar kata secara berurutan (bigram) dengan baik. Selain itu, skor ROUGE-L dan ROUGE-Lsum masing-masing sebesar 70.42% dan 70.96% menunjukkan bahwa model juga mampu menghasilkan urutan kata dan ringkasan yang mendekati struktur serta isi jawaban referensi. Secara keseluruhan, hasil ini menegaskan bahwa model memiliki performa yang tinggi dalam menghasilkan respons yang relevan dan sesuai dengan konteks konsultasi stunting.



Gambar 4. Evaluasi menggunakan ROUGE pada model chatbot stunting

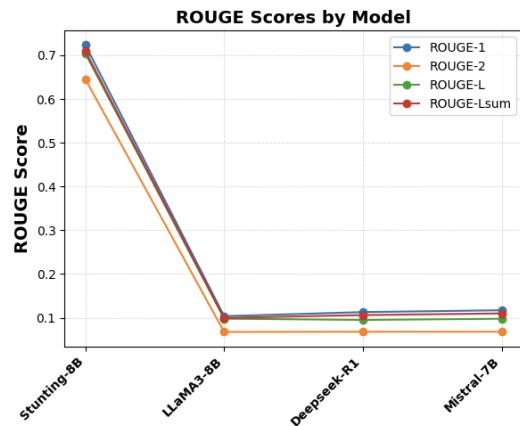
Gambar 5 menunjukkan perbandingan skor ROUGE dari empat model LLM, yaitu Stunting-8B, LLaMA3, Deepseek-R1, dan Mistral, terhadap empat matrik evaluasi: ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-Lsum. Model Stunting-8B secara konsisten menunjukkan performa terbaik, dengan skor ROUGE-1 (0.7235), ROUGE-2 (0.6455), ROUGE-L (0.7043), dan ROUGE-Lsum (0.7096), yang secara signifikan lebih tinggi dibandingkan model lainnya. Sebaliknya, model baseline seperti LLaMA3, Deepseek-R1, dan Mistral menunjukkan skor ROUGE yang jauh lebih rendah pada keempat matrik, mengindikasikan bahwa model Stunting-8B memiliki kemampuan yang lebih unggul dalam menghasilkan ringkasan yang relevan dan sesuai konteks, khususnya dalam domain pencegahan stunting.



Gambar 5. Perbandingan skor ROUGE dari empat model LLMs

Gambar 6 menunjukkan perbandingan skor ROUGE antara empat model LLM, yaitu Stunting-8B, LLaMA3, Deepseek-R1, dan Mistral, berdasarkan matrik ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-Lsum. Model Stunting secara jelas unggul dengan skor ROUGE tertinggi di semua matrik, menunjukkan kemampuannya dalam

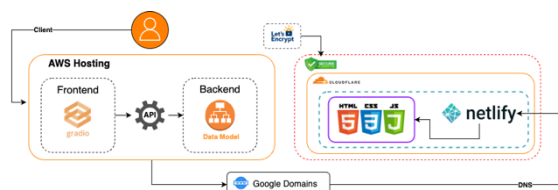
menghasilkan ringkasan teks yang paling relevan dan sesuai konteks. Sebaliknya, LLaMA3, Deepseek-R1, dan Mistral menunjukkan performa yang jauh lebih rendah dan relatif setara, dengan semua skor ROUGE berada di kisaran 0.6 sampai 0.7. Hasil ini menekankan efektivitas model Stunting dalam menangani pencegahan stunting.



Gambar 6. Perbandingan ROUGE antara model LLMs

Implementasi layanan konsultasi dan tanya jawab terkait pencegahan stunting. Kami menggunakan pendekatan arsitektur terdistribusi, yang membagi sistem menjadi modul-modul independen meliputi *front-end*, *back-end*, dan *server* komputasi model. Setiap modul dirancang dengan tanggung jawab spesifik, memungkinkan fleksibilitas dalam peningkatan sistem serta penggantian modul secara efisien. Interaksi antar modul dilakukan melalui antarmuka API, yang dirancang agar setiap modul dapat beroperasi secara mandiri tanpa ketergantungan langsung pada fungsi internal modul lainnya.

Dalam aspek keamanan, pendekatan desain kami mengintegrasikan mekanisme perlindungan yang komprehensif. Sistem modular ini menggunakan enkripsi pada tingkat produk, dengan API umum yang terenkripsi untuk memastikan komunikasi antar komponen berjalan secara aman. Selain itu, implementasi arsitektur sistem berbasis *HyperText Transfer Protocol Secure* (HTTPS) disertai dengan enkripsi *server cloud* menggunakan protokol *Secure Sockets Layer* (SSL) untuk meningkatkan perlindungan data. Pendekatan terdistribusi dan modular ini, yang dipadukan dengan prioritas tinggi pada keamanan, dirancang untuk mengatasi tantangan dalam penyelenggaraan layanan konsultasi daring berbasis AI skala besar. Desain ini memberikan fleksibilitas, skalabilitas, serta keandalan sistem yang mendukung keberhasilan implementasi proyek secara keseluruhan.

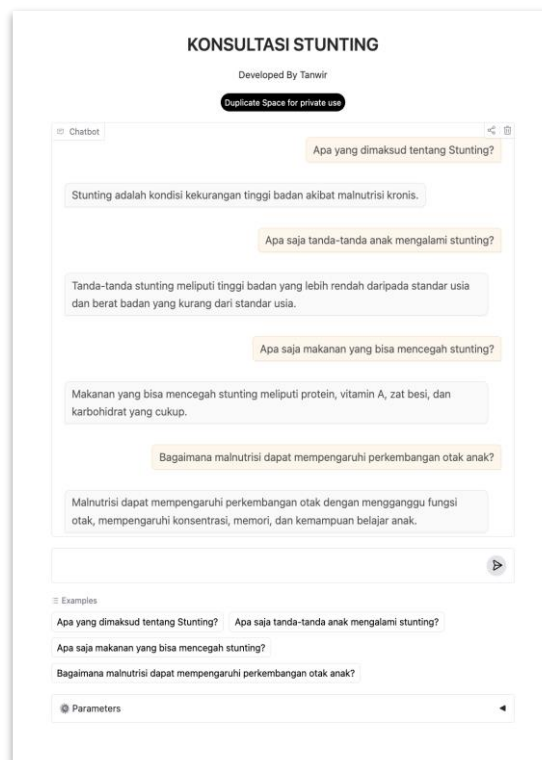


Gambar 7. Online Web front-end architecture

Diagram pada gambar 7 mengilustrasikan arsitektur sistem web yang kami kembangkan secara independent untuk situs berbasis *cloud*. Pada tingkat pengguna, klien berinteraksi dengan aplikasi melalui antarmuka *frontend* yang dibangun menggunakan *Gradio*, sebuah library Python yang dirancang untuk menciptakan antarmuka pengguna yang interaktif dan mudah diakses. *Frontend* ini berkomunikasi dengan *backend* melalui API, yang menghubungkan logika bisnis dan pemrosesan data yang diatur dalam model. Sistem *backend* ini dijalankan pada platform *cloud Amazon Web Services (AWS)*, yang memberikan fleksibilitas, skalabilitas, dan keberlanjutan operasional.

Untuk pengelolaan domain, sistem ini menggunakan layanan Google Domains, yang memungkinkan pengaturan nama domain dan pengarahannya ke server hosting aplikasi. Selain itu, untuk memastikan keamanan data yang dikirimkan antara klien dan server, sertifikat TLS diterapkan melalui layanan *Let's Encrypt*, yang menyediakan enkripsi koneksi dan melindungi data dari potensi ancaman keamanan. *Cloudflare* berperan penting dalam meningkatkan kinerja aplikasi dengan mengelola distribusi konten, mengurangi latensi, serta memberikan perlindungan terhadap serangan siber seperti *Distributed Denial of Service (DDoS)*. *Cloudflare* juga berfungsi sebagai pengelola *Domain Name System (DNS)*, memastikan bahwa permintaan dari pengguna diarahkan dengan cepat dan tepat ke server yang tepat.

Selain itu, file statis mencakup elemen-elemen dasar seperti HTML, CSS, dan JavaScript, yang merupakan fondasi dari tampilan antarmuka pengguna aplikasi, dihosting di Netlify. Netlify menyediakan hosting untuk file statis dengan optimisasi otomatis, mendukung proses pengembangan yang cepat, serta memastikan kinerja dan keandalan tinggi dalam pengiriman konten. Dengan arsitektur ini, sistem aplikasi berbasis web tidak hanya menyediakan pengalaman pengguna yang responsif dan interaktif, tetapi juga menjamin keamanan data dan operasional yang efisien melalui pemanfaatan teknologi *cloud computing* dan distribusi konten yang mutakhir. Struktur ini sangat cocok untuk aplikasi yang memerlukan skalabilitas tinggi dan ketahanan terhadap beban lalu lintas yang besar, sambil menjaga standar keamanan yang ketat. Berikut gambar 8 menampilkan interface konsultasi stunting.



Gambar 8. Tampilan Interface konsultasi stunting

KESIMPULAN

Penelitian ini berhasil mengimplementasikan chatbot berbasis LLMs untuk konsultasi stunting pada balita, menjawab kebutuhan akses informasi nutrisi yang akurat dan mudah dijangkau di Indonesia. Dengan melakukan *fine-tuning* pada model LLaMA 3 menggunakan dataset khusus Q&A berisi 7.642 entri, model Stunting yang dikembangkan menunjukkan kinerja unggul, mencapai skor ROUGE-1 (72,24%), ROUGE-2 (64,54%), ROUGE-L (70,42%), dan ROUGE-Lsum (70,96%). Hasil ini secara signifikan melampaui model baseline seperti LLaMA3, Deepseek-R1, dan Mistral, menegaskan kemampuannya menghasilkan respons yang relevan secara kontekstual dan akurat secara semantik. Integrasi chatbot ke dalam aplikasi web berbasis *cloud* yang aman dengan fitur HTTPS, enkripsi SSL, dan arsitektur terdistribusi menjamin skalabilitas, aksesibilitas pengguna, serta perlindungan data.

Penelitian ini menggarisbawahi potensi LLMs dalam mengurangi kesenjangan layanan kesehatan, terutama di wilayah dengan sumber daya terbatas. Alat ini menjadi solusi bagi pengasuh dan tenaga kesehatan dalam menangani stunting. Namun, keterbatasan penelitian meliputi ketergantungan pada dataset statis dan perlunya validasi klinis di dunia nyata. Penelitian lanjutan perlu memperluas dukungan multibahasa, mengintegrasikan pembaruan data real-time, serta mengevaluasi keterlibatan

pengguna jangka panjang dan dampak kesehatan. Studi ini menjadi landasan bagi intervensi kesehatan masyarakat berbasis AI, menegaskan peran transformatif LLMs dalam meningkatkan literasi kesehatan komunitas dan layanan preventif.

REFERENSI

- [1] V. Y. Lameky, "Stunting in Indonesia: Current progress and future directions," *Journal of Healthcare Administration*, vol. 3, no. 1, pp. 82–90, Jun. 2024, doi: 10.33546/joha.3388.
- [2] UNICEF, "Levels and Trends in Child Malnutrition," *UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates*, 2023.
- [3] Q. Rachmah *et al.*, "Peningkatan Pengetahuan Gizi Terkait Makanan Pendamping Asi (Mp-Asi) Melalui Edukasi Dan Hands-On-Activity Pada Kader Dan Non-Kader," *Media Gizi Indonesia*, vol. 17, no. 1SP, pp. 47–52, Dec. 2022, doi: 10.20473/mgi.v17i1sp.47-52.
- [4] Dinas Kesehatan, "Jumlah Balita berdasarkan Status Gizi Kurang, Pendek dan Kurus di Nusa Tenggara Barat." Accessed: Apr. 10, 2025. [Online]. Available: <https://data.ntbprov.go.id/index.php/dataset/9d274132-1480-48ea-aacf-14426b63c7d6/show>
- [5] L. K. Wardani, V. Aulia, M. Hadhikul, and M. Kardila, "Risks of Stunting and Interventions to prevent Stunting," *Journal of Community Engagement in Health*, vol. 6, no. 2, pp. 79–83, Sep. 2023, doi: 10.30994/jceh.v6i2.528.
- [6] E. S. Yanti, R. Imami, and A. Arica, "Raising Awareness of the Impact of Stunting on Toddler Development Through Parental Education," *Amalee: Indonesian Journal of Community Research and Engagement*, vol. 5, no. 2, pp. 751–764, Oct. 2024, doi: 10.37680/amalee.v5i2.5850.
- [7] endang naryono, "STUNTING, THE MAIN THREAT TO HUMAN QUALITY," Aug. 07, 2023. doi: 10.31219/osf.io/2w9yn.
- [8] M. E. Dominguez Tabera *et al.*, "The Application of Artificial Intelligence for the Design of a Virtual Assistant (Chatbot) as an innovative approach in user service: Social Security Center No.102," *International Journal of Combinatorial Optimization Problems and Informatics*, vol. 15, no. 4, pp. 149–162, Nov. 2024, doi: 10.61467/2007.1558.2024.v15i4.484.
- [9] H. G. Cavalcante *et al.*, "DEVELOPING CHATBOTS IN THE FIELD OF HEALTHCARE: A SYSTEMATIC REVIEW," *Revista de Sistemas e Computação*, vol. 12, no. 1, pp. 30–39, 2022, doi: 10.36558/rsc.v12i1.7589.
- [10] W. Hadikurniawati, S. Wijono, D. Manongga, I. Sembiring, and K. D. Hartomo, "Toddler Stunting Consulting Chatbot using Rasa Framework," *Jurnal Rekayasa ElektriKa*, vol. 19, no. 4, Dec. 2023, doi: 10.17529/jre.v19i4.33014.
- [11] M. Arevalillo-Herraez, P. Arnau-Gonzalez, and N. Ramzan, "On Adapting the DIET Architecture and the Rasa Conversational Toolkit for the Sentiment Analysis Task," *IEEE Access*, vol. 10, pp. 107477–107487, 2022, doi: 10.1109/ACCESS.2022.3213061.
- [12] D. Escobar-Grisales, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, "Evaluation of effectiveness in conversations between humans and chatbots using parallel convolutional neural networks with multiple temporal resolutions," *Multimed Tools Appl*, vol. 83, no. 2, pp. 5473–5492, Jan. 2024, doi: 10.1007/s11042-023-14896-y.
- [13] S. Zhou, J. Silvasstar, C. Clark, A. J. Salyers, C. Chavez, and S. S. Bull, "An artificially intelligent, natural language processing chatbot designed to promote COVID-19 vaccination: A proof-of-concept pilot study," *Digit Health*, vol. 9, Jan. 2023, doi: 10.1177/20552076231155679.
- [14] A. Gangavarapu, "LLMs: A Promising New Tool for Improving Healthcare in Low-Resource Nations," in *2023 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, Oct. 2023, pp. 252–255. doi: 10.1109/GHTC56179.2023.10354650.
- [15] A. Goel *et al.*, "LLMs Accelerate Annotation for Medical Information Extraction," Dec. 2023, [Online]. Available: <http://arxiv.org/abs/2312.02296>
- [16] A. Telenti, M. Auli, B. L. Hie, C. Maher, S. Saria, and J. P. A. Ioannidis, "Large language models for science and medicine," *Eur J Clin Invest*, vol. 54, no. 6, Jun. 2024, doi: 10.1111/eci.14183.
- [17] C. Munley, A. Jarmusch, and S. Chandrasekaran, "LLM4VV: Developing LLM-driven testsuite for compiler validation," *Future Generation Computer Systems*, vol. 160, pp. 1–13, Nov. 2024, doi: 10.1016/j.future.2024.05.034.
- [18] C. Curto, D. Giordano, D. G. Indelicato, and V. Patatu, "Can a Llama Be a Watchdog? Exploring Llama 3 and Code Llama for Static Application Security Testing," in *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, IEEE, Sep. 2024, pp. 395–400. doi: 10.1109/CSR61664.2024.10679444.

- [19] S. Kumar, A. Solanki, and N. Jhanjhi, "ROUGE-SS: A New ROUGE Variant for the Evaluation of Text Summarization," *Recent Advances in Computer Science and Communications*, vol. 17, Jun. 2024, doi: 10.2174/0126662558304595240528111535.