Jurnal Ilmu Komputer
Volume 1 Nomor 1 2026
ISSN (Online) XXXX-XXXX
DOI: https://doi.org/

Perancangan Aplikasi Percakapan AI Berbasis Web Menggunakan Speech-to-Text, Text-to-Speech, dan Voicevox untuk Melatih Percakapan dalam Bahasa Jepang

Farran Yaphta Quinton[1], Yurico Widia Pranata[1] , Muhammad Pandhu Palenjawi[1*] , Rafly Febriansyah[1], Sherly Auliya Putry[1], Lidia Pratiwi[1], Esron Rikardo Nainggolan[2]

[1] Prodi Teknologi Informasi, Fakultas Tehnik dan Informatika, Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98 Senen, Jakarta Pusat, Indonesia

[2]Prodi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
Jl. Raya Jatiwaringin No.2, Jakarta Timur, Indonesia

e-mail korespondensi: 17230752@bsi.ac.id

Abstrak - Proses pembelajaran bahasa asing kerap menghadapi kendala pada tahap latihan percakapan, khususnya ketika pembelajar bahasa asing tidak memiliki teman untuk belajar yang dapat diajak berlatih secara rutin, pelafalan yang ala kadarnya, dan tingkat kepercayaan diri yang rendah saat melakukan percakapan secara langsung. Kondisi tersebut menunjukkan pembelajar bahasa asing membutuhkan media yang memungkinkan pembelajar berlatih secara mandiri dalam lingkungan yang nyaman dan fleksibel. Penelitian ini bertujuan untuk mengembangkan aplikasi percakapan berbasis kecerdasan buatan yang memanfaatkan suara sebagai sarana pendukung pembelajaran bahasa Jepang. Metode yang digunakan meliputi teknologi *Speech-to-Text* untuk mengubah suara pengguna menjadi teks, pemrosesan bahasa menggunakan API ChatGPT sebagai pusat pengolahan percakapan, serta pemanfaatan teknologi *Text-to-Speech* berbasis VOICEVOX untuk menghasilkan umpan balik percakapan yang terasa hidup. Aplikasi ini dikembangkan dalam bentuk situs web menggunakan JavaScript yang terintegrasikan dengan layanan API sehingga mampu memproses data suara dan teks secara cepat. Hasil yang diharapkan dari penelitian ini adalah tersedianya sistem percakapan AI yang mampu memberikan jawaban dalam bentuk teks dan suara yang cocok dan pas, serta menciptakan pengalaman belajar yang lebih interaktif dan alami, sehingga dapat menjadi solusi yang tepat bagi pembelajar bahasa Jepang dalam meningkatkan kemampuan percakapan secara mandiri.

Kata Kunci : Kecerdasan buatan; Speech-to-text; Text-to-speech; Percakapan berbasis suara; Voicevox

Abstract *- The process of learning a foreign language often faces obstacles during the conversation practice stage, especially when foreign language learners do not have friends to practice with regularly, have mediocre pronunciation, and have low confidence in direct conversations. These conditions indicate that foreign language learners need media that allows them to practice independently in a comfortable and flexible environment. This research aims to develop an artificial intelligence-based conversational application that utilizes voice as a supporting tool for Japanese language learning. The methods used include Speech-to-Text technology to convert the user's voice into text, language processing using the ChatGPT API as a conversation processing center, and the use of Text-to-Speech technology based on VOICEVOX to produce lively conversational feedback. This application is developed as a website using JavaScript integrated with API services so that it can process voice and text data quickly. The expected result of this research is the availability of an AI conversational system that can provide answers in the form of text and voice that match and fit, and create a more interactive and natural learning experience, so that it can be the right solution for Japanese language learners to improve their conversational skills independently.*

Keywords : artificial intelligence; speech-to-text; text-to-speech; voice-based conversation; Voicevox

1. Pendahuluan

Keterampilan berbicara (*speaking*) merupakan salah satu aspek fundamental yang perlu dikuasai dalam pembelajaran bahasa asing, termasuk bahasa Jepang, selain penguasaan kosakata dan tata bahasa. Dalam konteks pembelajaran bahasa Jepang di perguruan tinggi, kemampuan berbicara menjadi kompetensi penting karena pembelajar tidak hanya dituntut memahami struktur linguistik, tetapi juga mampu mengungkapkan gagasan secara jelas, runtut, dan efektif melalui interaksi lisan (Adriana et al., 2025). Pada praktiknya, banyak pembelajar mengalami kendala dalam merangkai kata-kata saat berbicara karena tidak bisa berlatih sendirian dan

membutuhkan teman untuk berbicara, serta kurangnya percaya diri dalam berlatih. Akibatnya, proses pembelajaran hanya menerapkan pemahaman teori dibandingkan latihan percakapan yang interaktif dan praktisis. Karena itu, pemanfaatan kecerdasan buatan dalam pembelajaran bahasa Jepang mampu memberikan umpan balik secara cepat dan menghadirkan simulasi komunikasi yang sesuai dengan cara orang Jepang berbicara, sehingga lebih mendukung pembelajaran mandiri dan kesiapan berkomunikasi (Rita Agustina Karnawati et al., 2025).

Di sisi lain, tidak semua pembelajar memiliki lingkungan belajar yang mendukung untuk melatih kemampuan berbicara secara berkelanjutan. Berbagai hambatan masih sering muncul, seperti keterbatasan kosakata, rasa kurang percaya diri, serta kecemasan berlebihan ketika harus berbicara atau menyampaikan gagasan secara lisan. Hambatan ini juga dipengaruhi oleh perbedaan struktur bahasa yang digunakan dalam kehidupan sehari-hari antara bahasa Indonesia dan bahasa Jepang, sehingga menyulitkan pembelajar dalam mengekspresikan ide secara tepat (Intan Bestari et al., 2025). Kondisi tersebut menunjukkan bahwa latihan percakapan memerlukan media yang mampu memfasilitasi proses komunikasi secara bertahap dan nyaman. Dalam komunikasi dua arah berbasis suara, teknologi sintesis suara dinilai perlu menghadirkan variasi serta ekspresi yang lebih kaya agar interaksi dapat berlangsung secara lebih lancar dan alami (Homma Yukinori et al., 2023). Oleh karena itu, keberadaan media pembelajaran alternatif menjadi penting untuk mendukung latihan percakapan yang fleksibel dan mandiri.

Pemanfaatan teknologi kecerdasan buatan dalam sistem pembelajaran adaptif membuka peluang besar bagi penyediaan media pembelajaran yang responsif dirancang untuk menyesuaikan materi, metode, serta kecepatan belajar berdasarkan karakteristik dan kebutuhan masing-masing pembelajar. Pendekatan ini memungkinkan proses pembelajaran menjadi lebih responsif dan personal, sehingga mampu meningkatkan efektivitas pembelajaran serta keterlibatan peserta didik dalam mencapai tujuan belajar yang optimal (Respati, 2025). Sistem-sistem berbasis AI, seperti platform pemrosesan bahasa alami, pengenal ucapan otomatis, dan pengajar virtual, mampu menciptakan lingkungan belajar yang interaktif dan menarik sekaligus memberikan umpan balik secara langsung. Melalui mekanisme tersebut, pembelajar dapat berlatih berinteraksi dengan berbagai topik dan konteks komunikasi baik dalam situasi formal maupun informal, sehingga meningkatkan efektivitas proses pembelajaran bahasa secara keseluruhan (Katonáné Gyönyörű, 2025). Paparan terhadap beragam skenario percakapan tersebut membantu pembelajar menyesuaikan penggunaan bahasa secara lebih fleksibel serta melatih kemampuan berpikir kritis dalam menyampaikan pendapat dan menjelaskan gagasan. Proses ini berkontribusi terhadap peningkatan kompetensi komunikatif secara menyeluruh, termasuk kemampuan berargumentasi dan berkomunikasi secara efektif dalam berbagai situasi nyata (Novita Pratiwi et al., 2024). Dengan demikian, sistem percakapan AI berpotensi menjadi media pembelajaran yang interaktif dan dapat digunakan tanpa ketergantungan pada mitra percakapan secara langsung.

Berdasarkan permasalahan dan peluang tersebut, penelitian ini bertujuan untuk merancang dan membangun aplikasi *Friend Talk AI*, yaitu sistem percakapan berbasis kecerdasan buatan yang memungkinkan pengguna berinteraksi melalui suara maupun teks. Aplikasi ini diharapkan mampu memberikan respons secara alami dalam bentuk teks dan suara sehingga dapat dimanfaatkan sebagai sarana latihan percakapan bahasa Jepang secara mandiri, interaktif, dan mudah diakses.

2. Metode Penelitian

Metodologi penelitian ini disusun sebagai kerangka kerja terstruktur yang digunakan untuk menjelaskan tahapan perancangan, pengembangan, dan pengujian aplikasi percakapan berbasis kecerdasan buatan yang diusulkan. Pendekatan yang diterapkan berorientasi pada pengembangan sistem (*system development*), dengan tujuan menghasilkan aplikasi yang dapat beroperasi secara stabil dan sesuai dengan kebutuhan pembelajaran bahasa Jepang. Pengembangan sistem difokuskan pada penyediaan fitur latihan percakapan berbasis suara, sehingga aplikasi yang dibangun tidak hanya berfungsi sebagai media interaksi, tetapi juga mampu mendukung pembelajar dalam melatih keterampilan berbicara secara lebih kontekstual dan interaktif (Sisephaputra, 2023).

Metode penelitian yang digunakan mengombinasikan pendekatan deskriptif dan eksperimental. Pendekatan deskriptif diterapkan untuk menggambarkan karakteristik sistem, struktur arsitektur, serta alur kerja aplikasi, sedangkan pendekatan eksperimental digunakan untuk mengevaluasi fungsi sistem melalui pengujian langsung terhadap fitur yang dikembangkan. Pendekatan serupa juga banyak diterapkan dalam pengembangan sistem berbasis web, di mana kerangka pengembangan sistem informasi, seperti *System Development Life Cycle* (SDLC), dimanfaatkan untuk memberikan struktur yang jelas dalam proses perancangan hingga implementasi, sehingga memudahkan proses evaluasi kinerja sistem (Dewa & Azizah, 2024). Selain itu, pengembangan aplikasi yang mengintegrasikan layanan kecerdasan buatan umumnya mengacu pada model pengembangan perangkat lunak yang disertai tahapan pengujian fungsional guna memastikan sistem berjalan sesuai dengan kebutuhan pengguna dan tujuan penelitian (Uun Hariyanti et al., 2025; Ade Nur Hidayatulloh et al., 2024).

Sebagai dasar perancangan sistem, studi literatur dilakukan untuk mengkaji penelitian-penelitian terdahulu yang relevan dengan pengembangan aplikasi percakapan berbasis AI dan teknologi suara. Proses ini kemudian dilengkapi dengan pengamatan terhadap berbagai video tutorial teknis yang berkaitan dengan implementasi sistem percakapan berbasis web dan integrasi layanan AI. Hasil dari tahapan ini digunakan sebagai acuan dalam merancang arsitektur aplikasi Friend Talk AI, termasuk penentuan modul-modul utama yang diperlukan, seperti pengenal ucapan otomatis (*automatic speech recognition*), pemahaman bahasa alami (*natural language*

understanding), pengelolaan dialog, serta integrasi layanan eksternal. Alur proses percakapan dirancang agar input suara pengguna dapat diterima, dikonversi menjadi teks, diproses secara semantik, dan dikembalikan sebagai respons yang sesuai secara waktu nyata. Perancangan alur dan komponen inti ini bertujuan untuk memastikan sistem tidak hanya berfungsi secara teknis, tetapi juga efektif dalam mendukung pembelajaran bahasa melalui interaksi suara yang responsif (Sriram, 2025).

Arsitektur sistem yang dikembangkan bersifat berbasis web dan terdiri atas beberapa modul yang saling terintegrasi. Antarmuka pengguna berperan sebagai media utama interaksi antara pengguna dan sistem, yang memungkinkan pengguna memberikan masukan dalam bentuk suara maupun teks serta menerima respons dari aplikasi. Ketika pengguna mengirimkan input suara, sistem terlebih dahulu memproses sinyal audio tersebut melalui modul *speech-to-text* (STT), yang merupakan implementasi dari teknologi *automatic speech recognition* (ASR). Pada tahap ini, sinyal suara ditangkap dan diubah menjadi representasi digital, kemudian ditranskripsikan ke dalam bentuk teks. Proses transkripsi ini memiliki peran penting karena hasil teks yang diperoleh akan menjadi dasar bagi tahapan pemrosesan bahasa selanjutnya dalam sistem kecerdasan buatan (Billings & McDonnell, 2025).

Teks hasil konversi dari modul STT selanjutnya diteruskan ke API ChatGPT yang berfungsi sebagai pusat pemrosesan bahasa alami. Pada tahap ini, sistem menganalisis masukan pengguna dengan mempertimbangkan konteks percakapan yang sedang berlangsung, instruksi sistem, serta karakter percakapan yang dipilih. API ChatGPT bertugas mengolah teks input tersebut secara semantik sehingga respons yang dihasilkan tidak hanya relevan terhadap pertanyaan pengguna, tetapi juga koheren dan sesuai dengan konteks dialog. Pendekatan ini memungkinkan sistem menghasilkan jawaban yang menyerupai pola komunikasi alami antara manusia dan asisten virtual berbasis kecerdasan buatan (Saputra & Harefa, 2025).

Setelah respons teks dihasilkan, sistem melanjutkan proses ke modul *text-to-speech* (TTS) berbasis VOICEVOX. Modul ini bertanggung jawab untuk mengonversi teks bahasa Jepang menjadi keluaran suara. VOICEVOX dipilih karena kemampuannya dalam menghasilkan suara sintetis dengan tingkat kealamian yang baik, sebagaimana ditunjukkan dalam evaluasi objektif kualitas suara pada penelitian sebelumnya (Mizumoto et al., 2025). Sebelum proses sintesis suara dilakukan, teks respons terlebih dahulu diproses untuk menghilangkan elemen pemformatan yang tidak diperlukan, sehingga kualitas audio yang dihasilkan tetap terjaga. Keluaran suara kemudian dikirimkan kembali ke antarmuka pengguna dan diputar sebagai respons sistem.

Dalam implementasi aplikasi Friend Talk AI, karakter suara Zundamon dari VOICEVOX digunakan sebagai agen percakapan utama. Pemilihan karakter ini didasarkan pada karakteristik suara yang jelas, artikulasi yang baik, serta intonasi yang ringan dan ekspresif. Selain aspek kualitas suara, Zundamon juga dipilih karena mampu merepresentasikan persona percakapan yang konsisten dan komunikatif. Kehadiran karakter dengan identitas suara dan gaya bicara tertentu membuat interaksi terasa lebih personal dan mendekati percakapan nyata. Hal ini memungkinkan pengguna untuk berlatih mendengarkan serta meniru pelafalan kosakata dan struktur kalimat bahasa Jepang dengan ritme yang alami. Pendekatan berbasis karakter ini dinilai dapat meningkatkan keterlibatan dan motivasi pembelajar, karena interaksi yang dihasilkan tidak bersifat monoton dan lebih menyerupai dialog sehari-hari dibandingkan sistem percakapan generik tanpa persona yang jelas (Rackauckas & Hirschberg, 2025).

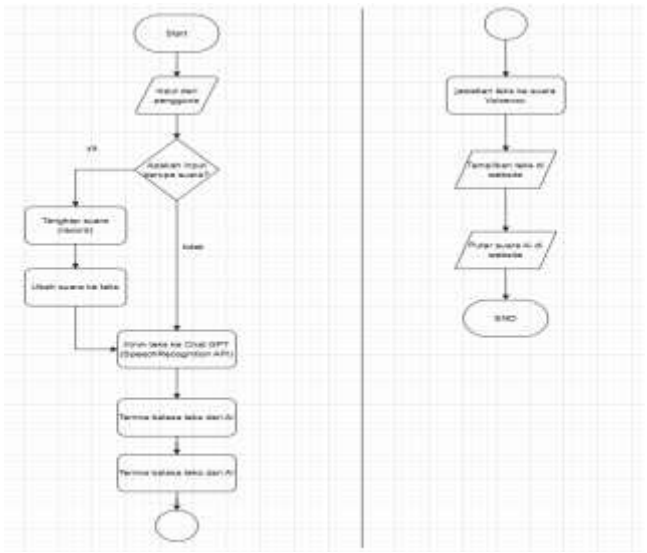
Seluruh modul dan layanan tersebut diintegrasikan dalam satu kesatuan aplikasi berbasis web dengan memanfaatkan bahasa pemrograman JavaScript serta layanan *Application Programming Interface* (API) yang mendukung pemrosesan data secara waktu nyata. Integrasi ini memungkinkan alur interaksi berlangsung secara berkesinambungan, mulai dari penerimaan input pengguna, pemrosesan bahasa alami, hingga penyajian respons dalam bentuk teks dan suara. Pendekatan integrasi serupa juga diterapkan pada sistem percakapan interaktif berbasis web seperti PublicTalk, yang memanfaatkan pemrosesan bahasa alami untuk mengenali maksud pengguna dan menghasilkan respons secara dinamis. Integrasi antarmuka berbasis JavaScript dengan pemanggilan API NLP secara real-time terbukti mampu mendukung penyajian respons yang relevan, kontekstual, dan cepat (Fauzi & Sutabri, 2025). Dengan mengadopsi pendekatan tersebut, sistem yang dikembangkan dalam penelitian ini diharapkan dapat memberikan pengalaman latihan percakapan bahasa Jepang yang interaktif, responsif, dan mendekati kondisi komunikasi nyata.

Algoritma sistem pada penelitian ini dirancang untuk menggambarkan alur interaksi antara pengguna dan sistem secara terstruktur, mulai dari proses input hingga output yang diterima pengguna. Proses ini diawali ketika pengguna memberikan masukan berupa suara (P), yang kemudian diproses oleh sistem melalui mekanisme *speech-to-text* (STT) untuk mengubah input suara menjadi teks (Q). Setelah suara pengguna berhasil dikonversi menjadi teks, hasil tersebut dikirimkan ke *Application Programming Interface* (API) kecerdasan buatan (R), sehingga sistem AI dapat menerima dan memproses pesan yang diberikan. Berdasarkan input tersebut, AI menghasilkan respons dalam bentuk teks (S) sesuai dengan konteks percakapan. Selanjutnya, teks respons yang dihasilkan oleh AI diproses kembali oleh sistem menggunakan teknologi *text-to-speech* (TTS) dengan memanfaatkan VOICEVOX (T) untuk mengubah teks menjadi suara. Pada tahap ini, sistem tidak hanya menampilkan respons dalam bentuk teks pada antarmuka website (U), tetapi juga memutar output suara kepada pengguna (Z), sehingga interaksi terasa lebih alami. Melalui rangkaian proses tersebut, pengguna memperoleh pengalaman percakapan yang interaktif dan mendekati komunikasi manusia secara nyata (V). Mekanisme ini dirancang untuk mendukung tujuan utama sistem, yaitu membantu pengguna belajar dan berlatih percakapan bahasa Jepang secara mandiri dan nyaman, serta mengurangi hambatan psikologis seperti rasa malu dan kurang percaya diri dalam berbicara (W).

Tabel 1. Fakta dan Bentuk Logika Sistem

Simbol	Fakta (Premis)	Arti
P	Pengguna berbicara (input suara)	Ada suara dari user
X	Pengguna memakai teks (input teks)	User mengetik pembicaraannya
Q	Sistem mengubah suara menjadi teks (STT)	Speech-to-text aktif
R	Teks dikirim ke API AI	AI menerima pesan
S	API AI memberi balasan dalam bentuk teks	AI menghasilkan teks jawaban
T	Sistem mengubah teks jawaban jadi suara menggunakan VOICEVOX (TTS)	Text-to-speech aktif
U	Website menampilkan teks	Output teks tampil ke pengguna
Z	Website memutar suara	Output suara tampil ke pengguna
V	Pengguna mendapat pengalaman interaktif & alami	Tujuan sistem tercapai
W	Pengguna bisa belajar bahasa Jepang dengan nyaman	Masalah kesendirian dan malu teratasi

Alur kerja sistem ini kemudian divisualisasikan dalam bentuk diagram alir (flowchart) sebagaimana ditunjukkan pada Gambar 1, yang menggambarkan keterkaitan setiap proses secara berurutan dari input hingga output sistem.



Gambar 1. Flowchart Proses Kerja Frined Talk AI

3. Hasil dan Pembahasan

Hasil pengembangan aplikasi *Friend Talk AI* memperlihatkan bahwa sistem percakapan berbasis kecerdasan buatan dengan dukungan input dan output suara dapat direalisasikan secara fungsional sebagai media latihan berbicara bahasa Jepang. Aplikasi yang dibangun tidak hanya berfokus pada kemampuan menghasilkan respons teks, tetapi juga menekankan integrasi antara pemrosesan bahasa alami, konversi suara ke teks, serta sintesis suara bahasa Jepang dalam satu alur interaksi yang berkesinambungan. Melalui implementasi ini, sistem mampu memberikan pengalaman percakapan yang lebih interaktif dan mendekati komunikasi lisan sehari-hari.

Secara teknis, *Friend Talk AI* dikembangkan menggunakan arsitektur berbasis web dengan pemisahan peran antara komponen antarmuka pengguna dan komponen pemrosesan sistem. Pendekatan ini dipilih agar pengelolaan sistem menjadi lebih terstruktur serta memudahkan proses pengembangan dan pengujian. Komponen frontend berfungsi sebagai media interaksi pengguna dengan sistem, termasuk menampilkan tampilan visual, menyediakan fitur pemilihan karakter percakapan, serta menerima masukan dalam bentuk teks maupun suara. Antarmuka dirancang secara sederhana agar pengguna dapat langsung memahami alur penggunaan aplikasi tanpa memerlukan penyesuaian teknis yang rumit.

Sementara itu, komponen backend berperan sebagai pusat pengolahan logika sistem dan kecerdasan buatan. Masukan pengguna yang diperoleh dari frontend, baik berupa teks langsung maupun hasil konversi suara,

diteruskan ke model *large language model* (LLM) melalui layanan API. Model ini bertugas memproses masukan tersebut dan menghasilkan respons percakapan yang sesuai dengan konteks interaksi. Respons teks yang dihasilkan kemudian diproses lebih lanjut untuk menghasilkan keluaran suara bahasa Jepang melalui engine *text-to-speech* VOICEVOX. Hasil akhir berupa teks dan audio dikirim kembali ke frontend untuk ditampilkan dan diputar secara bersamaan. Gambaran struktur sistem dan peran masing-masing komponen disajikan pada Tabel 2.

Tabel 2. Struktur Utama Proyek Chat-AI

Komponen	Deskripsi
Backend (NestJS)	Menangani API chat, pemrosesan AI, dan TTS
Frontend (React)	Antarmuka pengguna dan pengelolaan interaksi
OpenRouter API	Penyedia layanan LLM
VoiceVox	*Engine text-to-speech* bahasa Jepang

Cara kerja pembentukan respons percakapan dalam sistem ini menganut pendekatan berbasis karakter, di mana setiap karakter memiliki gaya bicara dan nuansa bahasa yang berbeda. Proses pembangkitan respons teks oleh sistem kecerdasan buatan dirumuskan dalam Persamaan (1) sebagai berikut:

Persamaan (1):

Response = Format(LLM(System Prompt + User Prompt + Character Trait))

Persamaan (1) menjelaskan bahwa jawaban yang dihasilkan sistem tidak hanya bergantung pada masukan pengguna, tetapi juga dipengaruhi oleh instruksi sistem dan karakteristik kepribadian karakter percakapan. *System prompt* berfungsi sebagai arahan utama yang mengatur perilaku sistem, seperti penggunaan bahasa Jepang, gaya bahasa yang santai dan natural, serta aturan tambahan seperti dukungan *furigana* dan struktur kalimat yang sesuai konteks pembelajaran. Instruksi ini memastikan bahwa respons yang dihasilkan tetap konsisten dengan tujuan sistem sebagai media latihan percakapan bahasa Jepang.

Sementara itu, *user prompt* merepresentasikan masukan langsung dari pengguna, baik yang berasal dari input teks maupun hasil konversi suara ke teks. Masukan ini menjadi konteks utama percakapan yang akan diproses oleh *large language model*. Komponen *character trait* melengkapi proses dengan menambahkan atribut kepribadian karakter, seperti gaya bicara, nuansa ekspresi, dan karakteristik bahasa tertentu, sehingga setiap respons memiliki ciri khas sesuai karakter yang dipilih.

Ketiga komponen tersebut diproses secara bersamaan oleh *large language model* untuk menghasilkan respons teks yang kontekstual dan relevan. Tahap *formatting* pada keluaran model bertujuan untuk menyesuaikan struktur respons agar siap ditampilkan pada antarmuka sistem, termasuk penyesuaian bahasa dan konsistensi percakapan. Implementasi teknis dari mekanisme yang dijelaskan dalam Persamaan (1) ditunjukkan pada gambar 2, yang memperlihatkan proses pemanggilan model AI serta penyusunan prompt pada sisi backend aplikasi.

Gambar 2. Implementasi pembentukan respons percakapan berbasis LLM

.

Respons teks yang telah dihasilkan selanjutnya diproses untuk menghasilkan keluaran suara melalui teknologi *text-to-speech*. Proses konversi teks menjadi audio dalam sistem ini dapat dirumuskan sebagai berikut:

Persamaan (2):

Audio = VoiceVox(StripHTML(Response))

Persamaan (2) menggambarkan mekanisme konversi teks respons menjadi audio menggunakan engine VOICEVOX. Sebelum proses sintesis suara dilakukan, teks respons terlebih dahulu melalui tahap *preprocessing* berupa pembersihan elemen pemformatan. Proses *StripHTML* bertujuan untuk menghilangkan tag HTML, *ruby tags*, serta simbol tambahan yang digunakan untuk tampilan visual, sehingga teks yang dikirimkan ke engine *text-to-speech* berada dalam bentuk yang bersih dan siap diproses.

Tahap pembersihan ini penting untuk menjaga kualitas suara yang dihasilkan, karena elemen pemformatan dapat mengganggu proses sintesis audio. Setelah teks berada dalam format yang sesuai, sistem mengirimkan data tersebut ke VOICEVOX dengan parameter suara dan karakter yang telah ditentukan sebelumnya. Engine VOICEVOX kemudian menghasilkan audio bahasa Jepang dengan pelafalan dan intonasi yang menyerupai penutur asli.

Keluaran audio yang dihasilkan dikirim kembali ke antarmuka pengguna dan diputar secara bersamaan dengan tampilan teks respons. Dengan mekanisme ini, pengguna memperoleh umpan balik dalam bentuk visual dan audio secara simultan. Implementasi teknis dari proses yang dijelaskan dalam Persamaan (2) ditunjukkan pada gambar 3, yang memperlihatkan tahapan pembersihan teks dan pemanggilan layanan sintesis suara VOICEVOX.

```
// Synthesize voice via local VoiceVox engine if available
async synthesizeVoice(text: string, speaker?: number) {
  try {
    if (!speaker) speaker = 1; // default speaker id

    // Extract only the main Japanese text (without furigana/translations)
    let plainText = text;

    // Remove furigana and readings from ruby tags
    plainText = plainText.replace(/<ruby>([^<]+)<rt>.*?<\/rt>(<rt>.*?<\/rt>)*<\/ruby>/g, '$1');

    // Remove any remaining HTML tags
    plainText = plainText.replace(/<[^>]*>/g, '');

    // Clean up any extra whitespace
    plainText = plainText.trim();

    // 1) get audio_query
    const audioQueryResp = await axios.post(
      `http://127.0.0.1:50021/audio_query?text=${encodeURIComponent(plainText)}&speaker=${speaker}`,
    );

    // 2) post synthesis
    const synthResp = await axios.post(
      `http://127.0.0.1:50021/synthesis?speaker=${speaker}`,
      audioQueryResp.data,
      { responseType: 'arraybuffer' },
    );

    const buffer = synthResp.data as ArrayBuffer;
    const base64 = Buffer.from(buffer).toString('base64');
    return { data: base64, mime: 'audio/wav' };
  } catch (e) {
    console.error(
      'Voice synthesis failed:',
      e instanceof Error ? e.message : 'Unknown error',
    );
    return null;
  }
 }
}
```

Gambar 3. Implementasi proses sintesis suara menggunakan VOICEVOX

Dari sisi antarmuka, hasil implementasi menunjukkan bahwa aplikasi *Friend Talk AI* dapat dijalankan secara stabil pada lingkungan berbasis web. Antarmuka menyediakan elemen-elemen utama yang mendukung proses percakapan, seperti fitur pemilihan karakter, area masukan teks atau suara, serta indikator pemutaran audio. Tata letak antarmuka dirancang agar alur interaksi mudah dipahami oleh pengguna. Contoh tampilan antarmuka aplikasi ditunjukkan pada Gambar 4.

Gambar 4. Tampilan antarmuka *Friend TalkAI*

Berdasarkan hasil pengujian fungsional, sistem mampu menerima masukan pengguna melalui antarmuka tersebut, memprosesnya menggunakan kecerdasan buatan, serta menghasilkan respons berupa teks dan suara bahasa Jepang secara konsisten. Penyajian respons secara simultan dalam bentuk visual dan audio memberikan pengalaman latihan percakapan yang lebih mendekati kondisi komunikasi nyata. Pengguna tidak hanya membaca hasil percakapan, tetapi juga mendengarkan pelafalan yang benar, sehingga proses latihan keterampilan berbicara bahasa Jepang dapat berlangsung secara lebih efektif.

Secara keseluruhan, hasil implementasi menunjukkan bahwa aplikasi *Friend Talk AI* memiliki potensi sebagai media alternatif untuk latihan percakapan bahasa Jepang secara mandiri. Integrasi antara kecerdasan buatan, teknologi *speech-to-text*, dan *text-to-speech* memungkinkan terjadinya interaksi dua arah berbasis suara yang berjalan secara waktu nyata. Keunggulan sistem ini terletak pada kemampuannya menyediakan mitra percakapan virtual tanpa keterbatasan waktu dan tempat, sehingga pengguna dapat berlatih dengan lebih nyaman tanpa tekanan sosial sebagaimana percakapan langsung dengan manusia. Dengan demikian, sistem percakapan AI berbasis suara yang dikembangkan dalam penelitian ini memberikan kontribusi yang relevan dalam mendukung pembelajaran bahasa Jepang, khususnya pada aspek keterampilan berbicara, serta berpotensi untuk dikembangkan lebih lanjut sebagai media pembelajaran interaktif berbasis kecerdasan buatan.

4. Kesimpulan

Penelitian ini menghasilkan sebuah aplikasi percakapan berbasis kecerdasan buatan bernama *Friend Talk AI* yang dirancang untuk mendukung latihan berbicara bahasa Jepang secara mandiri. Berdasarkan tujuan yang telah dirumuskan pada bagian pendahuluan, sistem yang dikembangkan mampu menjawab permasalahan keterbatasan mitra percakapan dan kesulitan pelafalan melalui pemanfaatan teknologi pemrosesan bahasa alami, *speech-to-text*, serta *text-to-speech* berbasis VOICEVOX. Hasil implementasi dan pengujian fungsional menunjukkan bahwa aplikasi dapat berjalan sesuai dengan rancangan, menerima masukan pengguna, serta menghasilkan respons percakapan dalam bentuk teks dan suara bahasa Jepang secara selaras.

Penerapan mekanisme percakapan berbasis karakter dengan keluaran suara sintetis memberikan pengalaman interaksi yang lebih natural dan personal bagi pengguna. Penyajian respons secara bersamaan dalam bentuk visual dan audio memungkinkan pengguna untuk tidak hanya memahami isi percakapan, tetapi juga mempelajari pelafalan yang tepat. Dengan demikian, aplikasi *Friend Talk AI* dapat dimanfaatkan sebagai media pendukung pembelajaran bahasa Jepang, khususnya untuk melatih keterampilan berbicara dalam suasana yang lebih fleksibel dan nyaman tanpa tekanan interaksi langsung dengan penutur lain.

Meskipun sistem telah berfungsi dengan baik, penelitian ini masih memiliki peluang untuk dikembangkan lebih lanjut. Pengembangan selanjutnya dapat diarahkan pada peningkatan variasi gaya bicara dan ekspresi suara, penambahan konteks percakapan yang lebih beragam sesuai situasi nyata, serta penerapan sistem pada skenario pembelajaran yang lebih luas. Selain itu, penelitian lanjutan disarankan untuk mengevaluasi dampak penggunaan aplikasi terhadap peningkatan kemampuan berbicara bahasa Jepang dalam jangka waktu tertentu. Dengan adanya pengembangan berkelanjutan, sistem percakapan AI berbasis suara diharapkan dapat berkontribusi lebih optimal sebagai media pembelajaran bahasa asing yang adaptif dan inovatif.

Referensi

Andre Farhan Saputra, K. H. (2025). Penerapan metode natural language processing (NLP) dalam implementasi asisten virtual chatbot. *Journal of Research and Publication Innovation, 3*(1), 1–15. https://jurnal.portalpublikasi.id/index.php/JORAPI/article/view/1332/1015

Bestari, I., Basri, M. S., & Nasution, Y. A. (2025). The use of kamishibai as a visual media in tadoku courses to improve Japanese speaking skills for Japanese language students at the faculty of teacher education. *TOFEDU: The Future of Education, 4*(6), 2484–2496. http://journal.tofedu.or.id/index.php/journal/article/view/781

Dewa, S. B. A., & Azizah, N. L. (2024). Perancangan sistem informasi sekolah berbasis web menggunakan metode SDLC. *Indonesian Journal of Applied Technology, 1*(2), 15. https://doi.org/10.47134/ijat.v1i2.3053

Hariyanti, U., Jayadi, D., Afirianto, T., Nurtantyana, R., Zulvarina, P., & Brawijaya, U. (2025). Pengembangan front-end aplikasi MySRE: Tinjauan front-end development of MySRE: A literature review. *Jurnal Teknologi Informasi, 12*(4), 895–902.

Hasibuan, A., & Pawiro, M. A. (2025). Improving Japanese speaking skills during prospective interns: Picture and picture learning model. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching, 9*(1), 60–73. https://doi.org/10.30743/ll.v9i1.10626

Hidayatulloh, A. N., Pane, M. B. W., & others. (2024). Pengembangan aplikasi e-learning menggunakan model rapid application development. *Jurnal Riset Informatika dan Ilmu Komputer, 2*(2), 275–281. http://jurnalmahasiswa.com/index.php/jriin/article/view/1397

Homma, Y., Kanagawa, H., Kobayashi, N., Ijima, Y., & Saito, K. (2023). Expressive text-to-speech synthesis using text chat dataset with speaking style information. *Transactions of the Japanese Society for Artificial Intelligence, 38*(3), 1–12. https://doi.org/10.1527/tjsai.38-3_F-MA7

Karnawati, R. A., Seruni, A. P., & Masrokhah, Y. (n.d.). Pembelajaran bahasa Jepang berbasis AI bagi calon pekerja migran. *Prosiding Seminar Nasional*, 32–37.

Katonáné Gyönyörű, K. I. (2025). Adaptive learning systems and artificial intelligence in language learning. *Gradus, 12*(1), 1–7. https://doi.org/10.47833/2025.1.art.010

M. Erlangga Fauzi, & Sutabri, T. (2025). PublicTalk: Sistem chatbot pintar berbasis natural language processing untuk layanan pemerintahan digital. *Journal Sains Student Research, 3*(2), 426–433. https://doi.org/10.61722/jssr.v3i2.4325

Mizumoto, T., Kojima, A., Fujita, Y., Liu, L., & Sudo, Y. (2025). Is synthetic data truly effective for training speech language models? In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)* (pp. 1808–1812). https://doi.org/10.21437/Interspeech.2025-2693

Pratiwi, N., Efendy, A. G., Rini, H. C., & Ahmed, N. A. (2024). Speaking practice using ChatGPT's voice conversation: A review on potentials and concerns. *Journal of Language, Linguistics, and Instructional Communication, 6*(1), 59–72. https://doi.org/10.35719/jlic.v6i1.149

Rackauckas, Z., & Hirschberg, J. (2025). Animating language practice: Engagement with stylized conversational agents in Japanese learning. *arXiv*. http://arxiv.org/abs/2507.06483

Respati, H. T., & Universitas Veteran Jawa Timur. (2024). Pemanfaatan AI dalam pendidikan: Meningkatkan pembelajaran melalui sistem pembelajaran adaptif. *Jurnal Ilmiah Multidisiplin, 2*(2), 394–400. https://doi.org/10.62017/merdeka

Salsabil, A. D., Rakhmawati, L. A., & Febtwenesty. (2025). From silent learners to confident speakers: The effect of AI voice chat with ChatGPT on EFL speaking skills. *Info: Article, 3*(1), 38–50. CV. Doki Course and Training.

Sisephaputra, B., Ramadhan, A. F., Affifudin, I. F., Noor, N. H., & Hidayatullah, M. A. (2023). Pengembangan aplikasi belajar bahasa Jepang berbasis website. *Jurnal Pendidikan Bahasa Jepang, 16*(1), 50–59.

Venkatesh, S. (2025). Understanding the architecture of voice assistants: A technical deep dive. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11*(2), 587–595. https://ijsrcseit.com/index.php/home/article/view/CSEIT25112398