

# PENGEMBANGAN DAN IMPLEMENTASI VOICEBOT MULTIBAHASA DENGAN TEKNOLOGI RETRIEVAL-AUGMENTED GENERATION UNTUK ROBOT TOUR GUIDE

Winston Franchity Salim, Indar Sugiarto, Handy Wicaksono  
Program Studi Teknik Elektro, Universitas Kristen Petra, Surabaya, Jawa Timur  
E-Mail: winstonfsalim123@gmail.com, indi@petra.ac.id, handy@petra.ac.id

**Abstrak** – Pada tahun 2022, PT MRT dan Universitas Gunadarma membuat inovasi robot pintar yang digunakan untuk robot patroli dan robot tour guide di sekitar lingkungan MRT Jakarta. Penelitian ini bertujuan untuk mengembangkan inovasi tersebut dengan mengimplementasikan sistem voicebot pada robot tour guide dengan memanfaatkan teknologi speech-to-text (STT), text-to-speech (TTS), dan retrieval-augmented generation (RAG) yang dimana nantinya penelitian ini dapat mengubah sistem robot yang dulunya masih menggunakan manusia dalam pengoperasiannya menjadi sistem robot yang berjalan secara otomatis tetapi masih bisa memberikan informasi dengan cara yang interaktif. Penelitian ini akan menggunakan reComputer j202 sebagai hardware utama dari pengolahan suara secara local serta menggunakan protokol komunikasi MQTT untuk berintegrasi dengan perangkat lainnya seperti ROS dan juga Raspberry Pi. Hasil dari pengujian menunjukkan bahwa sistem berhasil memanfaatkan teknologi speech-to-text (STT), text-to-speech (TTS), dan retrieval-augmented generation (RAG) yang memiliki waktu respons yang lumayan cepat dan mampu beradaptasi terhadap pertanyaan yang diberikan dikarenakan memanfaatkan teknologi RAG. Penelitian ini juga berhasil mengontrol perangkat lain dengan menggunakan protokol komunikasi MQTT.

**Kata Kunci** – robot tour guide, voicebot, retrieval-augmented generation (RAG), speech-to-text (STT), text-to-speech (TTS), MQTT

## I. PENDAHULUAN

Di Indonesia perkembangan teknologi sudah semakin maju, Indonesia sudah mulai melibatkan kecerdasan buatan (AI) untuk perkembangan negara Indonesia. Menurut KORIKA, organisasi kolaborasi riset dan inovasi AI nasional, pemerintah telah merilis Strategi Nasional Artificial Intelligence 2045 (Stratnas AI) untuk memperkuat infrastruktur digital di sektor-sektor utama seperti kesehatan, birokrasi, pendidikan, keamanan pangan, dan kota pintar. Strategi ini dirancang untuk mengoptimalkan layanan publik dengan memanfaatkan kecerdasan buatan (AI) dalam otomatisasi pelayanan dan analisis data, yang dapat mempercepat proses administratif dan meningkatkan kualitas layanan kepada masyarakat [1]. Pada tahun 2022 lalu, PT MRT di Jakarta bekerja sama dengan universitas Gunadarma untuk membuat sebuah inovasi robot pintar yang digunakan sebagai robot tour guide di sekitar lingkungan MRT. Inovasi robot tersebut dapat melakukan interaksi dua arah yang terhubung langsung dengan petugas di stasiun MRT dengan tujuan supaya para pengunjung di sekitar MRT dapat bertanya dan berinteraksi secara langsung dengan petugas tanpa harus pergi ke ruangan informasi [2]. Namun

peran robot tersebut masih harus bergantung pada manusia, dimana petugas MRT Jakarta harus selalu ada bagi pengunjung yang mau berinteraksi maupun bertanya.



Gambar 1. Inovasi robot pintar milik MRT Jakarta [3].

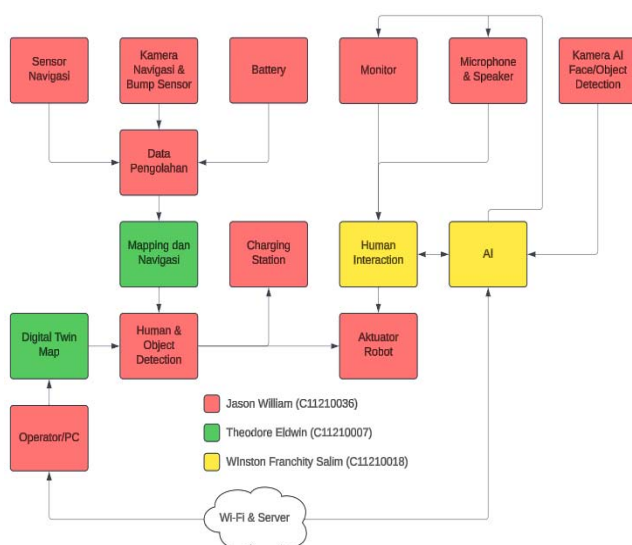
Dengan kemajuan teknologi kecerdasan buatan yang sangat pesat, interaksi antar robot dan juga manusia dapat ditingkatkan menjadi lebih efisien, interaktif, responsive serta tidak memerlukan bantuan manusia. Beberapa penelitian menunjukkan bahwa sistem voicebot dapat meningkatkan interaktifitas pada robot dan efisiensi sistem sehingga memungkinkan robot untuk berinteraksi dengan pengunjung secara mandiri tanpa memerlukan bantuan manusia lagi. Pada penelitian yang menguji performa sistem voicebot, disebutkan bahwa voicebot mampu untuk memberikan informasi kepada pengunjung dengan baik. Penelitian tersebut juga melakukan uji coba dalam tingkat kebisingan tertentu dan hasilnya menunjukkan sistem voicebot masih dapat merespon dengan baik [4]. Terdapat juga penelitian voicebot berbasis website yang menggunakan teknologi Closed Domain Question Answering System dalam pengambilan informasi [5]. Hal ini membuat sistem voicebot menjadi semakin interaktif dan juga informatif dikarenakan jawaban yang diberikan bersumber langsung dari dokumen yang telah disediakan. Namun teknologi Closed Domain Question Answering System memiliki keterbatasan dan hanya cocok digunakan untuk satu topik yang spesifik. Selain itu interaksi robot tidak hanya bergantung pada sistem voicebot melainkan juga pada pergerakan robot yang bisa adaptif, fleksibel dan otomatis. Penerapan robot operating system (ROS) pada robot dapat membuat pergerakan robot menjadi otomatis berdasarkan

perintah dari sistem. Selain mampu memberikan informasi dan berinteraksi dengan manusia, sistem voicebot juga harus mampu berinteraksi dengan pergerakan robot sehingga membuat robot menjadi lebih interaktif.

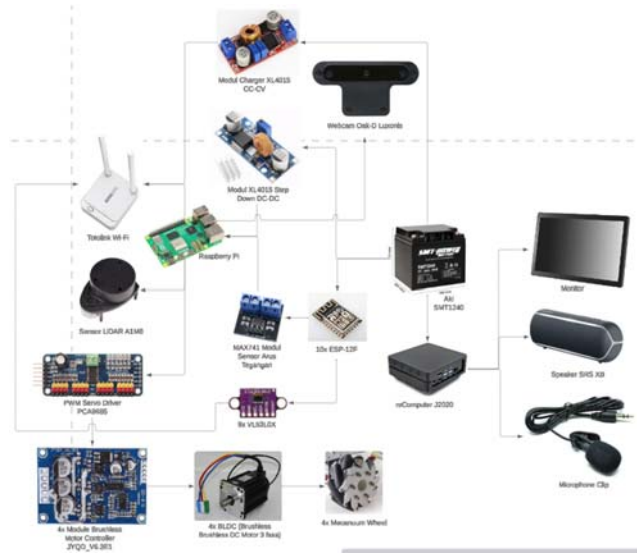
Terdapat penelitian yang memungkinkan robot dikendalikan melalui protokol komunikasi message queuing telemetry transport (MQTT) [6]. Hal ini memungkinkan sistem voicebot dapat terhubung langsung dengan sistem ROS sehingga membuat sistem voicebot tidak hanya memberikan informasi dan berinteraksi dengan manusia tetapi juga dapat memberikan perintah kontrol terhadap robot secara real-time. Maka dari itu pada penelitian ini akan memfokuskan sistem voicebot multibahasa untuk diterapkan pada robot tour guide yang akan dikendalikan menggunakan robot operating system (ROS). Sistem voicebot pada penelitian ini akan menggunakan teknologi speech to text (STT) dan text to speech (TTS) yang bersifat open-source, pendekatan teknologi pengambilan informasi yang berbeda yaitu menggunakan retrieval-augmented generation (RAG), mendukung 3 bahasa utama yaitu Mandarin, Indonesia dan Inggris serta menggunakan protokol komunikasi MQTT untuk menghubungkan sistem voicebot dengan robot operating system (ROS) dalam proses kendali secara real-time.

## II. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk merancang dan mengimplementasi sistem voicebot yang dilengkapi teknologi retrieval-augmented generation (RAG) pada robot tour guide untuk membuat robot lebih interaktif dan otonom. Penelitian ini akan lebih berfokus pada sistem voicebot yang bertujuan untuk meningkatkan interaktifitas robot dengan manusia yang akan dirancang dengan menggunakan reComputer j202 sebagai hardware utama dan menggunakan server untuk proses retrieval-augmented generation (RAG). Sistem voicebot akan berkomunikasi dengan perangkat lain seperti Raspberry pi dan ROS dengan menggunakan protokol komunikasi message queuing telemetry transport (MQTT). Sementara bagian pergerakan robot tour guide yaitu robot operating system (ROS) dan penambahan hardware robot lainnya akan dikerjakan oleh rekan penulis. Secara umum desain sistem dan hardware pada robot tour guide dapat dilihat seperti pada gambar 2 dan gambar 3.



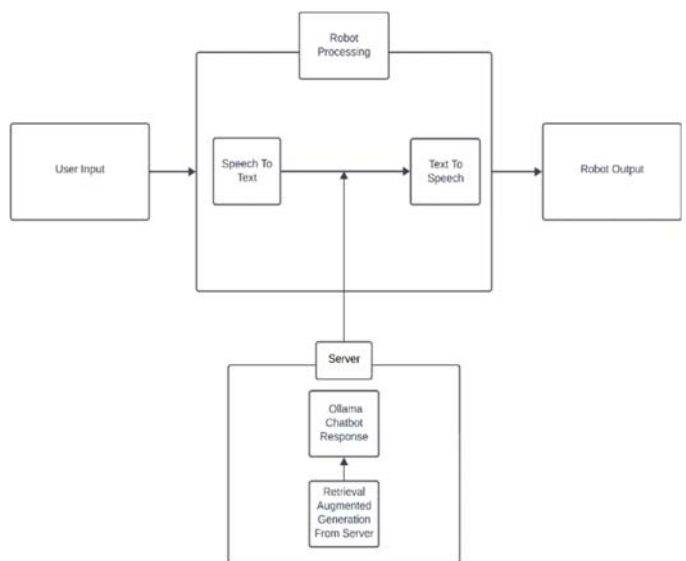
Gambar 2. Desain Umum Sistem Robot Tour Guide.



Gambar 3. Desain Umum Hardware Robot Tour Guide.

### A. Desain Umum Sistem Voicebot

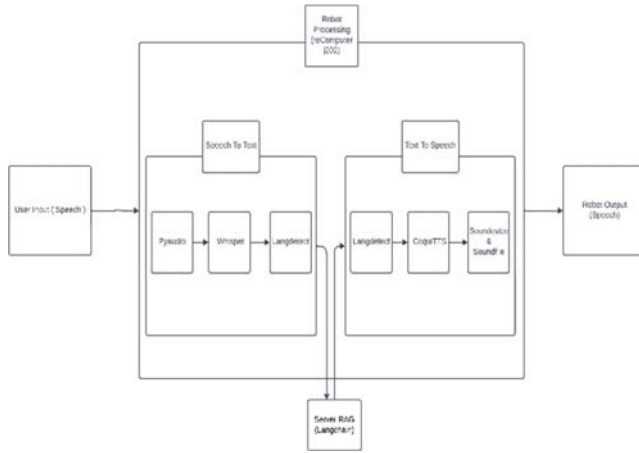
reComputer j202 berfungsi sebagai wadah untuk melakukan pengolahan suara secara lokal dimana di dalam terdapat sistem speech to text (STT) dan juga text to speech (TTS) yang dapat mendeteksi 3 bahasa utama yang akan digunakan yaitu Mandarin, Indonesia dan Inggris. Sistem voicebot akan berjalan saat sistem menerima input suara berupa pertanyaan dalam 3 bahasa utama dari sebuah mikrofon. Setelah itu proses STT akan mengubah input suara tersebut menjadi sebuah teks yang nantinya akan diproses oleh server yang sudah diimplementasikan teknologi RAG untuk dilakukan pengecekan apakah input dari pengguna menanyakan sesuatu yang relevan dengan dokumen yang disediakan oleh RAG. Jika tidak maka jawaban yang akan dikirimkan menggunakan jawaban dari model AI bernama Gemma 3 dari platform Ollama. Setelah mendapatkan jawaban dalam berupa teks, selanjutnya proses TTS akan mengubah teks tersebut menjadi sebuah suara lagi yang akan dikeluarkan melalui speaker ke pengguna sesuai dengan bahasa pada saat memberikan input pertanyaan. Diagram blok sistem dapat dilihat pada gambar 3.



Gambar 4. Blok Diagram Sistem Voicebot.

**B. Library yang Digunakan dalam Sistem Voicebot**

Pada penelitian ini sistem voicebot menggunakan beberapa library open-source. Untuk melakukan proses perekaman suara, library yang digunakan adalah Pyaudio, Proses speech to text (STT) menggunakan library Whisper, proses text to speech (TTS) menggunakan library CoquiTTS yang menggunakan model Tacotron untuk bahasa Mandarin dan VITS untuk bahasa Inggris dan Indonesia, library Langdetect untuk mendeteksi bahasa dari sebuah teks, library Sounddevice dan Soundfile yang digunakan untuk memutar dan membaca file audio secara langsung dan terakhir library Langchain untuk retrieval-augmented generation (RAG). Diagram penggunaan library bisa dilihat pada gambar 5 dan 6.



Gambar 5. Blok Diagram Library yang Digunakan.

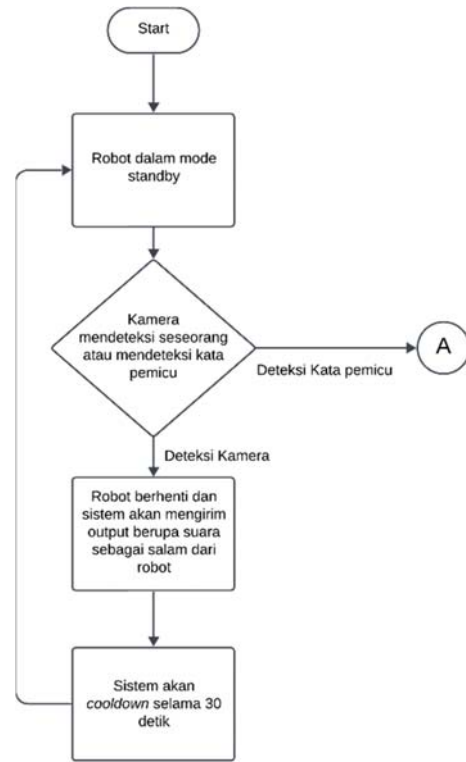
**C. Integrasi Sistem Voicebot pada Robot Tour Guide**

Agar sistem voicebot dapat memberikan pengalaman interaktif maka sistem ini diintegrasikan ke dalam robot tour guide. Sistem voicebot tidak hanya berperan untuk menangani pertanyaan dari pengguna tetapi juga dalam mengontrol pergerakan robot maupun interaksi robot lainnya seperti memberi pesan salam berupa suara. Dalam penelitian ini, sistem voicebot terintegrasi dengan kamera dan juga motor dari robot. Cara kerja sistem voicebot secara keseluruhan dapat dilihat pada gambar 5, 6 dan 7. Pada bagian awal gambar 5, robot yang sedang bergerak akan berada dalam mode standby yang menunggu deteksi awal berupa deteksi kamera dan juga deteksi kata pemicu. Saat Kamera mendeteksi adanya manusia maka robot akan berhenti dan sistem voicebot akan mengirim pesan salam berupa suara. Setelah itu robot akan kembali bergerak dan kamera akan cooldown selama 30 detik.

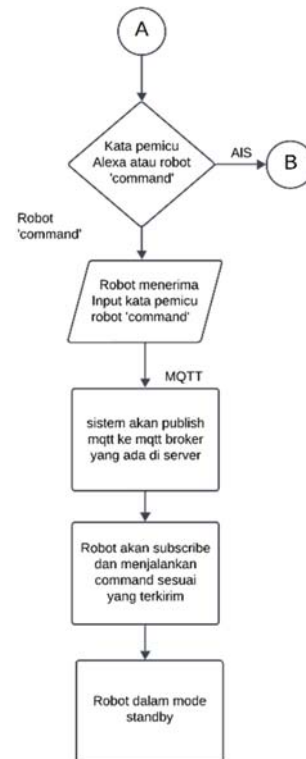
Deteksi kata pemicu pada sistem voicebot dibagi menjadi dua yaitu kata pemicu “robot command” dan kata pemicu “AIS”. Kata pemicu pertama yaitu “robot command” berfungsi untuk mengatur pergerakan robot yang dapat dikontrol melalui suara oleh sistem voicebot. Saat sistem menerima input kata pemicu “robot command” maka sistem melakukan publish data menggunakan protokol komunikasi MQTT yang ada di server. Setelah itu ROS akan menerima data tersebut dengan melakukan subscribe dan akan menjalankan command atau perintah sesuai yang terkirim. Jika sudah maka robot akan kembali berjalan dan kembali dalam mode standby.

Deteksi kata pemicu “AIS” berfungsi untuk menerima pertanyaan dari pengguna. kata “AIS” sendiri merupakan kepanjangan dari “Artificial Intelligence Support” yang berarti sistem pendukung atau bantuan berbasis AI. Saat sistem

voicebot menerima kata pemicu “AIS” maka robot akan berhenti dan sistem voicebot akan melakukan proses perekaman pertanyaan pengguna. Setelah mendapatkan input pertanyaan pengguna maka sistem akan memproses input suara pertanyaan tersebut dan mengubahnya menjadi sebuah teks

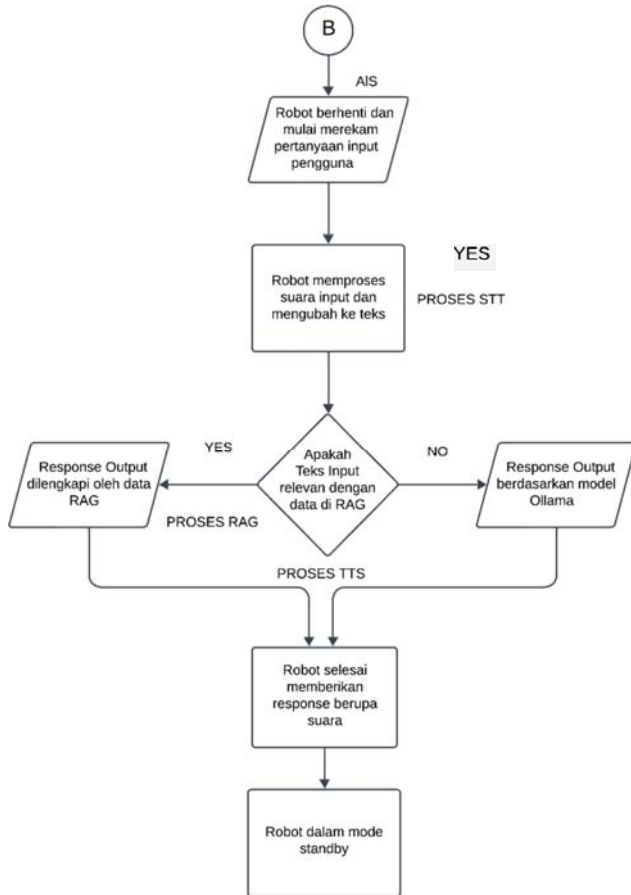


Gambar 7. Flowchart Sistem Voicebot 1.



Gambar 8. Flowchart Sistem Voicebot 2.

dengan menggunakan library STT. Teks tersebut akan dikirimkan ke server RAG untuk dilakukan pengecekan apakah ada dokumen yang relevan dengan pertanyaan tersebut. Jika ada maka jawaban atau response yang diberikan akan dilengkapi dengan data dari dokumen tersebut. Jika tidak maka jawaban atau response yang diberikan akan berdasarkan dari model AI LLM Ollama. Terakhir setelah mendapatkan jawaban atau response dari RAG ataupun Ollama, server akan meneruskan kembali teks hasil jawaban tersebut ke sistem voicebot. Lalu sistem voicebot akan mengubah teks tersebut menjadi output suara lagi dengan menggunakan library TTS. Setelah jawaban dalam bentuk suara sudah selesai diputar maka robot akan kembali berjalan dan berada pada mode standby untuk menerima kata pemicu selanjutnya.

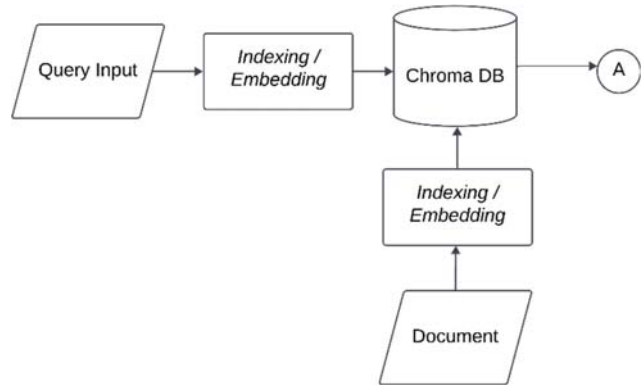


Gambar 9. Flowchart Sistem Voicebot 3

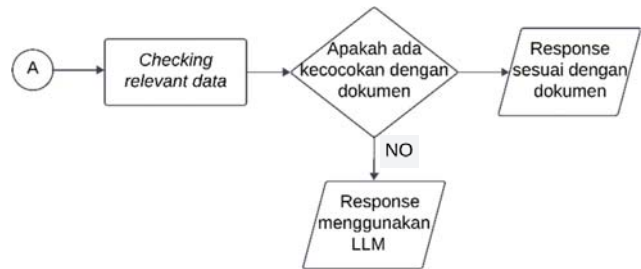
**D. Sistem Retrieval-Augmented Generation (RAG)**

Sistem retrieval-augmented generation menggunakan library Langchain. Cara kerja dari sistem RAG adalah saat sistem mendapatkan teks input, maka akan dilakukan proses Indexing atau Embedding dimana pada proses tersebut mengubah teks input menjadi sebuah vektor numerik yang disimpan dalam database bernama Chroma. Selain itu, dokumen yang tersedia pada sistem RAG juga melalui proses yang sama yaitu Indexing atau Embedding. Namun dikarenakan dokumen yang bertugas sebagai informasi kemungkinan memiliki jumlah kata yang sangat banyak sehingga dokumen tersebut akan dibagi menjadi sebuah pecahan teks yang disebut Chunk. Chunk tersebut yang akan diubah menjadi vektor dan disimpan ke dalam database Chroma. Setelah itu vektor input dan vektor dokumen berdasarkan tiap chunk akan dilihat kecocokan yang

ada. Jika terdapat kecocokan maka RAG akan memberikan jawaban sesuai dengan kecocokan tersebut. Jika tidak terdapat kecocokan maka jawaban yang diberikan menggunakan model AI LLM. Dalam penelitian ini model LLM yang digunakan adalah Gemma 3.



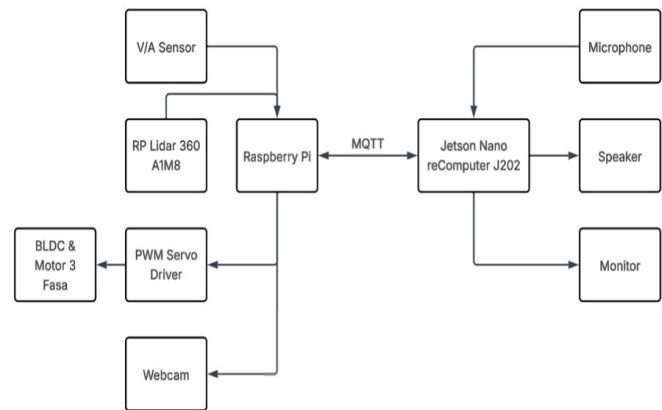
Gambar 9. Flowchart Sistem RAG 1.



Gambar 10. Flowchart Sistem RAG 2.

**E. Komunikasi Sistem Voicebot dengan Perangkat Lain**

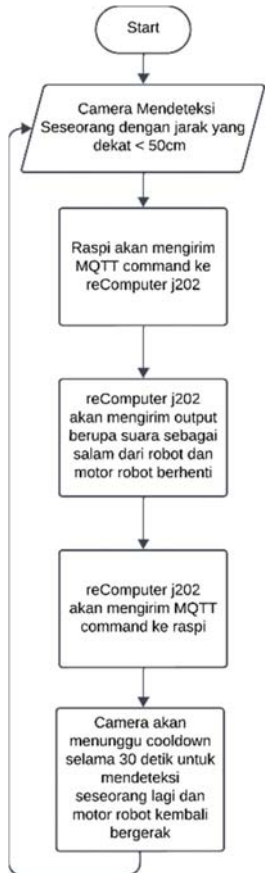
Sistem voicebot perlu diintegrasikan dengan perangkat keras robot lainnya seperti Raspberry Pi yang dilengkapi sebuah kamera untuk mendeteksi manusia. Selain itu, sistem voicebot juga harus mampu terintegrasi dengan sistem ROS agar sistem voicebot ini bisa mengontrol pergerakan robot sehingga menciptakan interaksi antar sistem secara keseluruhan. Protokol komunikasi yang digunakan adalah MQTT. Diagram Blok sistem komunikasi menggunakan MQTT dapat dilihat pada gambar 5.



Gambar 11. Diagram Blok Komunikasi Sistem Voicebot dengan Perangkat Keras Robot Lainnya.

1. Komunikasi Sistem Voicebot dengan Kamera Luxonis OAK-D

Sistem voicebot perlu terintegrasi dengan kamera yang berfungsi sebagai pendeteksi manusia agar tidak saling bertabrakan. Kamera OAK-D menggunakan bantuan model dari mobilenet-SSD yang berisi dataset tentang person detection. Sistem voicebot akan saling mengirim data sebagai bentuk komunikasi dengan kamera. Saat kamera mendeteksi adanya manusia maka kamera akan mengirimkan data kepada reComputer j202 sebagai perangkat sistem voicebot melalui bantuan protokol komunikasi MQTT. Hal ini akan membuat sistem voicebot berhenti sejenak dalam menerima input pertanyaan dari pengguna dan akan mengirimkan pesan salam berupa suara. Selain itu robot juga akan berhenti saat kamera mendeteksi manusia. Setelah pesan selesai diputar reComputer j202 akan mengirimkan data kembali ke kamera untuk memberikan informasi agar robot bisa kembali berjalan dan kamera akan cooldown selama 30 detik sebelum bisa mendeteksi manusia kembali. Hal ini berfungsi agar kamera tidak terus menerus mendeteksi manusia dan robot tidak terus menerus mengirim pesan salam kepada pengunjung.



Gambar 12. Flowchart Komunikasi Kamera dalam Raspberry Pi dengan Sistem Voicebot dalam reComputer j202

2. Komunikasi Sistem Voicebot dengan ROS

Sistem voicebot juga perlu terintegrasi dalam pergerakan robot yang dimana pergerakan robot dalam penelitian ini menggunakan ROS. Cara kerja komunikasi antara sistem voicebot dengan ROS dapat dilihat pada gambar 8. Untuk penelitian ini command atau perintah yang tersedia adalah berhenti/stop, jalan/start, antarkan ke waypoint/ go to

waypoint, dan Go to Lift. List command yang disediakan untuk perintah robot bisa dilihat pada tabel 1.

Tabel 1. List dan Fungsi Command/Perintah yang Disediakan

Command	Fungsi
Stop / berhenti	Untuk memberhentikan robot
Start / jalan	Untuk memulai kembali proses roaming robot
Antarkan ke waypoint Q/P/I	Untuk membuat robot pergi ke checkpoint yang sudah ditentukan
Go To waypoint Q/P/I	Untuk membuat robot pergi ke checkpoint yang sudah ditentukan
Go to Lift	Untuk membuat robot pergi ke checkpoint yang sudah ditentukan

III. HASIL DAN PEMBAHASAN

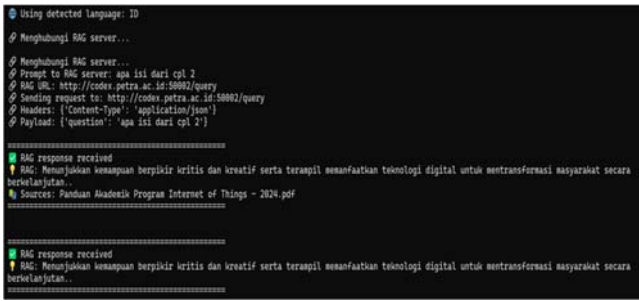
A. Pengujian Informasi RAG

Tujuan dari pengujian ini adalah untuk mengetes apakah sistem yang menggunakan teknologi RAG dapat memberikan informasi yang relevan sesuai dengan dokumen yang diberikan atau tidak. Untuk mengetes keakuratan RAG sistem akan diberikan beberapa dokumen mengenai informasi tertentu. Pada percobaan pertama diberikan data mengenai capaian pembelajaran lulusan. Data yang diberikan dapat dilihat pada tabel 2.

Tabel 2. Informasi CPL

No	Kode CPL	Deskripsi CPL
1	CPL1	Menunjukkan sikap konsisten berpegang pada nilai-nilai Kristiani dalam menjalankan panggilan hidup di kampus dan masyarakat.
2	CPL2	Menunjukkan kemampuan berpikir kritis dan kreatif serta terampil memanfaatkan teknologi digital untuk mentransformasi masyarakat secara berkelanjutan.
3	CPL3	Menunjukkan kemampuan bersinergi sebagai warga negara dan dunia dalam pengabdian sesuai panggilan hidup.
4	CPL4	Mampu memahami teori keteknikan secara komprehensif untuk memahami perkembangan dunia teknik elektro.
5	CPL5	Mampu membuat desain teknis dan mewujudkan rancangan dengan mengadopsi kearifan lokal.
6	CPL6	Memahami konsep eksperimen ilmiah dan melakukannya sesuai kaidah keilmuan, serta mampu melakukan analisis hasil eksperimen dengan akurat dan mendalam.

Pertanyaan yang diberikan pada sistem RAG berupa “apa isi dari CPL yang ke 2”. Pertanyaan tersebut dikirim melalui proses STT lalu masuk ke dalam server RAG. Pada percobaan ini RAG berhasil memberikan jawaban yang sesuai dengan dokumen yang diberikan seperti pada gambar 13. Percobaan selanjutnya masih menggunakan pertanyaan yang sama tetapi dengan bahasa lain yaitu bahasa inggris. Hasil yang didapat juga sesuai dengan dokumen tetapi dalam bahasa inggris seperti pada gambar 14.



Gambar 13. Percobaan Pertama RAG.



Gambar 14. Percobaan Kedua RAG

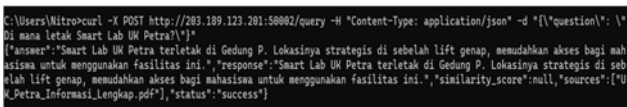
Selanjutnya karena penelitian ini akan berfokus pada robot tour guide maka dokumen akan disesuaikan dengan pertanyaan umum tour guide yang di fokuskan pada lingkungan Universitas Kristen Petra khususnya di gedung P seperti pada gambar 15. Pertanyaan yang sesuai dengan lingkungan Universitas Kristen Petra dan akan dicoba adalah “dimanakah letak Smart Lab UK Petra” dan “apa saja makanan yang ada di kantin UK Petra”

Salah satu fasilitas unggulan yang terdapat di Gedung P adalah **Smart Lab**, sebuah ruang laboratorium modern yang dirancang untuk mendukung kegiatan praktikum berbasis Internet of Things (IoT), dan sistem otomatis. Smart Lab ini terletak secara strategis di **sebelah lift genap**, memudahkan akses bagi mahasiswa dan dosen yang ingin melakukan eksperimen atau riset di bidang teknologi cerdas.

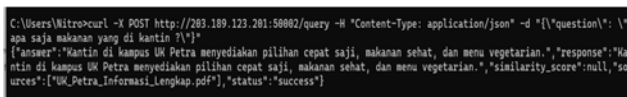
Selain fasilitas akademik, Gedung P juga dilengkapi dengan **kantin mahasiswa** yang terletak di area lantai dasar dan sangat populer di kalangan civitas akademika. Kantin ini menyediakan berbagai pilihan makanan, dan minuman, mulai dari masakan nusantara seperti nasi ayam, soto ayam, mie ayam, hingga makanan cepat saji, makanan sehat, dan pilihan vegetarian. Suasana kantin dirancang nyaman dan bersih, dilengkapi dengan area duduk yang luas, koneksi Wi-Fi gratis, serta area diskusi informal. Tak jarang mahasiswa memanfaatkan ruang kantin ini tidak hanya untuk makan, tetapi juga untuk berdiskusi kelompok, mengerjakan tugas bersama, atau sekadar bersantai sambil menunggu kelas berikutnya.

Dengan keberadaan laboratorium teknologi tinggi seperti Smart Lab dan fasilitas penunjang seperti kantin yang nyaman.

Gambar 15. Informasi yang difokuskan pada lingkungan UK Petra



Gambar 16. Percobaan Ketiga RAG.



Gambar 17. Percobaan Keempat RAG.

Dari semua percobaan yang dilakukan hasil menunjukkan bahwa sistem RAG yang dirancang sudah bisa mengambil informasi relevan dari dokumen yang disediakan. Jawaban dan pertanyaan yang dicoba dapat dilihat pada tabel 3.

Tabel 3. Pengujian RAG

Pertanyaan	Jawaban RAG
Apa isi dari CPL 2	Menunjukkan kemampuan berpikir kritis dan kreatif serta terampil memanfaatkan teknologi digital untuk mentransformasi masyarakat secara berkelanjutan
What is the content of CPL 2	CPL 2 shows the ability to think critically and creatively and skilled in utilizing digital technology to transform the community in a sustainable manner.
Dimanakah letak Smart Lab UK Petra	Smart lab UK Petra terletak di gedung P. Lokasinya strategis di sebelah lift genap, memudahkan akses bagi mahasiswa untuk menggunakan fasilitas ini.
Apa saja makanan yang ada di kantin UK Petra	Kantin di kampus UK Petra menyediakan pilihan cepat saji, makanan sehat, dan menu vegetarian.

Percobaan terakhir adalah Ketika RAG tidak menemukan informasi dalam dokumen maka seharusnya sistem tersebut menggunakan Ollama sebagai pemberi jawaban.



Gambar 18. Percobaan Kelima RAG.

B. Pengujian Output TTS

Tujuan dari pengujian ini adalah untuk mengetahui seberapa cepat processing time suara tiap bahasa yang dihasilkan oleh library CoquiTTS serta menggunakan pertanyaan yang sama tiap bahasa.

1. Bahasa Inggris

Tabel 3. Processing Time Bahasa Inggris

Percobaan	Pertanyaan	Processing time
1	What is the content of CPL 1	4.951490640640259
2		3.111818790435791
3		3.0446267127990723
4		3.001958131790161

Pada percobaan TTS bahasa Inggris menggunakan pertanyaan "what is the content of cpl 1", Hasil menunjukkan bahwa semakin pertanyaan tersebut ditanyakan berulang ulang maka semakin cepat TTS memberikan output suara. Dapat dilihat pada percobaan pertama waktu yang dibutuhkan untuk menghasilkan dan memutar output suara adalah 4.95 detik lalu dilakukan percobaan berulang ulang hingga percobaan terakhir hanya membutuhkan waktu 3 detik untuk memutar output suara tersebut.

2. Bahasa Indonesia

Tabel 4. Processing Time Bahasa Indonesia

Percobaan	Pertanyaan	Processing time
1	Apa isi dari CPL 1	7.170746326446533
2		8.012677431106567
3		5.766978740692139
4		5.748572111129761

Pada percobaan TTS bahasa Indonesia menggunakan pertanyaan yang sama dengan bahasa Inggris yaitu "apa isi dari CPL 1", Hasil juga menunjukkan bahwa semakin pertanyaan tersebut ditanyakan berulang ulang maka semakin cepat TTS memberikan output suara. Dapat dilihat pada percobaan pertama waktu yang dibutuhkan untuk menghasilkan dan memutar output suara adalah 7.17 detik lalu dilakukan percobaan berulang ulang hingga percobaan terakhir hanya membutuhkan waktu 5.74 detik untuk memutar output suara tersebut.

3. Bahasa Mandarin

Tabel 5. Processing Time Bahasa Mandarin

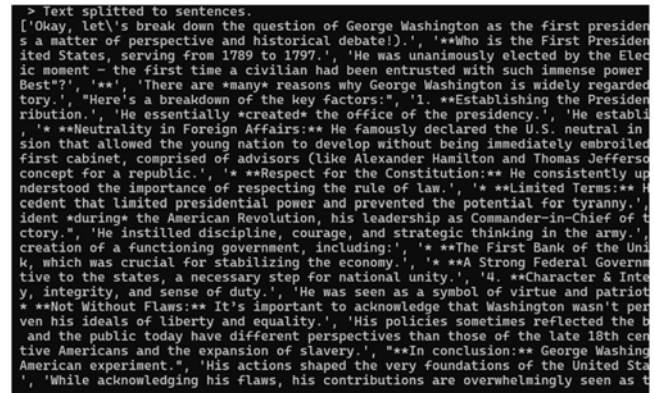
Percobaan	Pertanyaan	Processing time
1	你好吗?	11.116130828857422
2		5.294870376586914
3		4.6804726123809814
4		4.910390853881836

Pada percobaan TTS bahasa Mandarin menggunakan pertanyaan "你好吗?", Hasil pada bahasa Mandarin juga menunjukkan bahwa semakin pertanyaan tersebut ditanyakan berulang ulang maka semakin cepat TTS memberikan output suara.

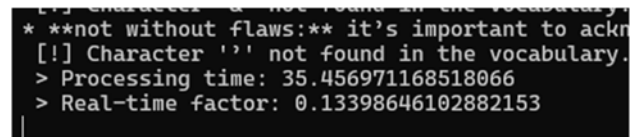
4. Pengujian Berdasarkan Panjang Response

Pengujian terakhir adalah dengan melihat apakah processing time TTS berpengaruh dengan panjang sebuah jawaban atau response yang diberikan. Percobaan pertama dilakukan dengan melihat waktu yang dihasilkan saat mendapat jawaban yang panjang. Sedangkan pada percobaan kedua dilakukan pada

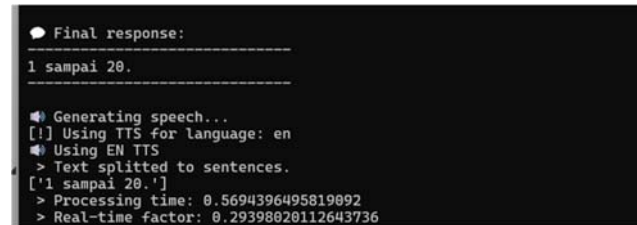
jawaban yang pendek. Hasil menunjukkan bahwa waktu yang dihasilkan saat jawaban panjang lebih lama daripada jawaban pendek seperti yang ada pada gambar 21 saat response panjang menghabiskan waktu sebanyak 35 detik dan pada gambar 22 saat response pendek hanya menghabiskan waktu sebanyak 0.5 detik.



Gambar 19. Percobaan Response Panjang.



Gambar 20. Hasil Processing Time Response panjang.



Gambar 21. Percobaan Response Pendek.

Hasil dari pengujian ini menunjukkan bahwa kecepatan processing time berdasarkan panjangnya sebuah jawaban atau response serta model TTS bahasa Inggris lebih cepat dibandingkan bahasa lain dikarenakan model bahasa Inggris dilatih dan menggunakan dataset yang lebih besar daripada bahasa yang lain seperti Indonesia dan juga Mandarin. Pengujian kecepatan performa processing time ini berdasarkan dari GPU reComputer j202 yang mempunyai 384-core NVIDIA Volta GPU dan RAM 8gb.

C. Pengujian Output STT

Tujuan dari pengujian ini adalah untuk mengetahui keakuratan STT dari library Whisper. Percobaan akan dilakukan dalam bahasa Mandarin, Inggris dan Indonesia. Percobaan akan dilakukan sebanyak mungkin dan akan menampilkan pertanyaan yang diucapkan dan juga hasil transkrip dari library Whisper.

1. Bahasa Indonesia

Pada pengujian dalam bahasa Indonesia masih ada beberapa yang tidak sesuai dengan pertanyaan yang diucapkan tetapi kebanyakan sudah benar.

Tabel 6. Deteksi Bahasa Indonesia

Percobaan	Pertanyaan	Hasil Transcribe	Deteksi Bahasa
1	"Apa isi dari CPL yang ke 2"	"Apa isi dari JPL y2"	indonesia
2	"Apa isi dari CPL yang ke delapan"	Apa isi dari CPL yang ke-8"	indonesia
3	"siapakah presiden pertama di indonesia"	"siapakah presiden pertama di indonesia"	indonesia
4	"Apa yang terjadi pada tambang nikel di raja ampat"	"Apa yang terjadi pada tambang nikel di raja ampat"	indonesia
5	"bagaimana cara untuk melakukan pertambahan"	"bagaimana cara untuk melakukan pertambahan"	indonesia
6	"Apa isi dari CPL yang ke tujuh"	"Apa sih dari JPL yang ke tujuh"	indonesia
7	"dimana tempat TPS di surabaya"	"Di mana tempat TPS disurabaya"	indonesia

2. Bahasa Inggris

Pada pengujian dalam bahasa Inggris semua pertanyaan berhasil di transkrip dengan benar.

Tabel 6. Deteksi Bahasa Inggris

Percobaan	Pertanyaan	Hasil Transcribe	Deteksi Bahasa
1	"What is the content of CPL 2"	"What is the content of CPL 2"	English
2	"Who is the first america president"	"Who is the first american president"	English
3	"Who is the second indonesia president"	"Who is the second indonesia president"	English
4	"what is the minimum score for graduating"	"what is the minimum score for graduating"	English
5	"what is CPL 3"	"what is CPL 3"	English
6	"Where is the bathroom"	"Where is the bathroom"	English
7	"What is your name"	"What is your name"	English

3. Bahasa Mandarin

Pada pengujian dalam bahasa Mandarin semua pertanyaan berhasil di transkrip dan dirubah menjadi goresan mandarin dengan benar.

Tabel 6. Deteksi Bahasa Mandarin

Percobaan	Pertanyaan	Hasil Transcribe	Deteksi Bahasa
1	Ni hao ma	你好吗?	Mandarin
2	Wo ai ni	我爱你	Mandarin
3	Wo yao chi fan	不要吃饭	Mandarin
4	Zhe shi shen me	这是什么	Mandarin

D. Pengujian Kamera

1. Pengujian Deteksi Manusia

Cara kerja dari pendeteksian manusia adalah ketika kamera mendeteksi manusia dengan jarak kurang dari 50 cm. Pada saat jarak kurang dari 50 cm maka tidak terjadi apa apa seperti pada gambar 23. Namun saat kamera mendeteksi manusia dengan jarak kurang dari 50 cm maka sistem kamera pada gambar 12 akan dijalankan dan tampilan akan menjadi seperti pada gambar 24.

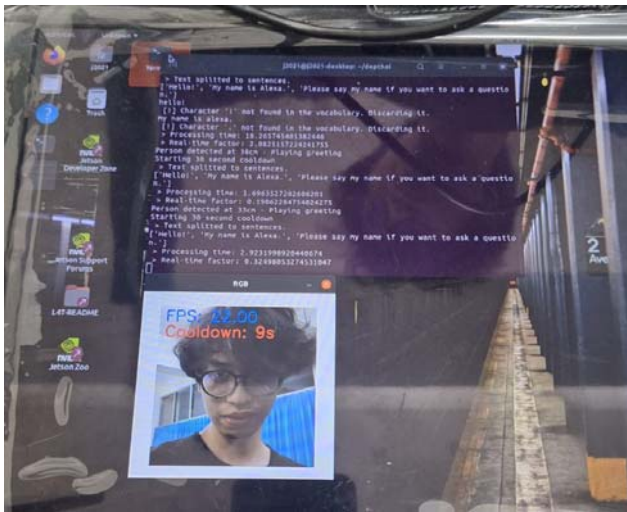


Gambar 22. Kamera Tidak Mendeteksi.



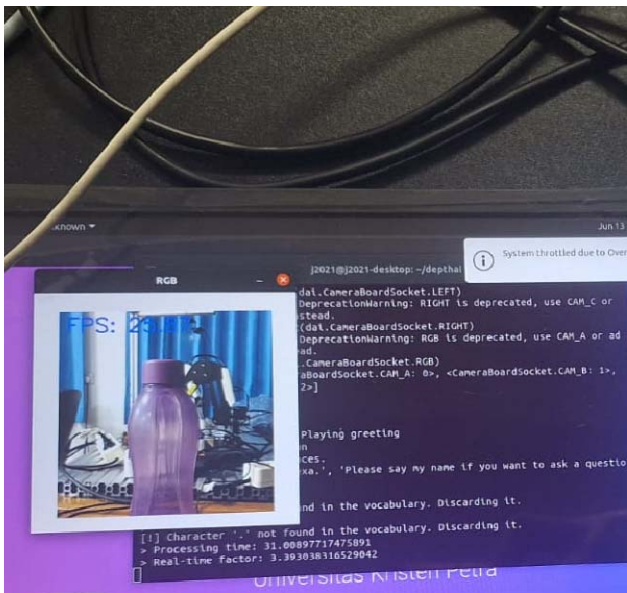
Gambar 23. Kamera Mendeteksi.

Dilakukan juga pengujian dengan berbeda manusia, sistem masih bisa mendeteksi sesuai dengan percobaan sebelumnya.



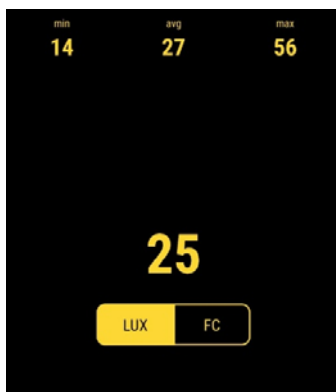
Gambar 24. Percobaan Berbeda Manusia.

Terakhir melakukan percobaan deteksi objek lain selain manusia. Pada percobaan ini digunakan objek botol minum untuk menguji keakuratan model AI kamera dan hasilnya kamera tidak mendeteksi apa apa.



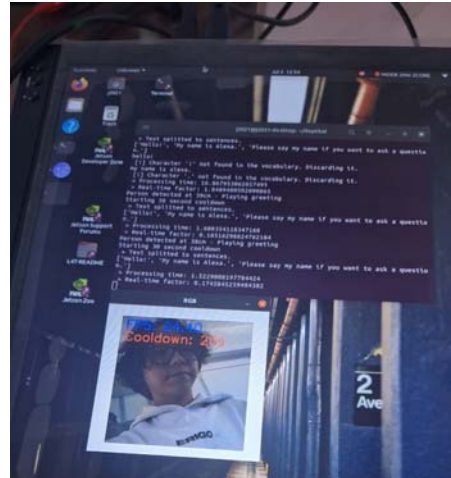
Gambar 25. Percobaan Beda Objek

2. Pengujian Deteksi dengan berbagai pencahayaan

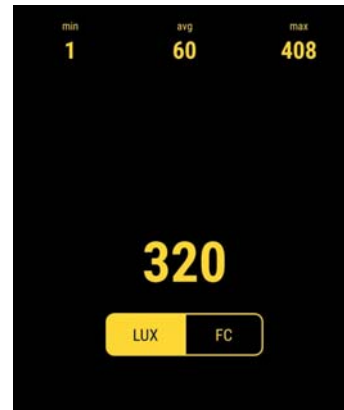


Gambar 26. Kondisi Cahaya Saat Ruangn Tanpa Lampu

Pengujian selanjutnya adalah untuk mengetahui kualitas kamera dalam mendeteksi objek dengan berbagai pencahayaan. Percobaan pertama adalah saat kondisi ruangan tanpa lampu dan cahaya yang terdeteksi hanya 25 lumen. Kamera masih dapat mendeteksi adanya manusia. Percobaan kedua adalah saat kondisi ruangan dengan lampu dan Cahaya yang terdeteksi 320 lumen. Kamera juga berhasil mendeteksi adanya manusia.



Gambar 27. Kamera Mendeteksi Dalam Keadaan Ruangn Tanpa Lampu



Gambar 28. Kondisi Cahaya Saat Ruangn dengan Lampu



Gambar 29. Kamera Mendeteksi Dalam Keadaan Ruangn dengan Lampu.

E. Pengujian Komunikasi ROS

Pengujian terakhir adalah pengujian komunikasi antara sistem voicebot dengan ROS. Pengujian ini dilakukan untuk mencoba apakah robot dapat dikontrol menggunakan suara. Pada gambar 30, pengguna mengirimkan perintah berupa “Robot Start” dan sistem mengirimkan data melalui MQTT untuk diterima oleh ROS. ROS menerima seperti pada gambar 31, dimana robot pada simulasi bergerak moving forward dan berada pada mode roaming / primary control.

```

Listening...
Recording for 3 seconds...

Initial transcription to detect language...
Whisper detected language: ID
Using Whisper's language detection: ID
Final detected language: ID
Transcribing in detected language: ID

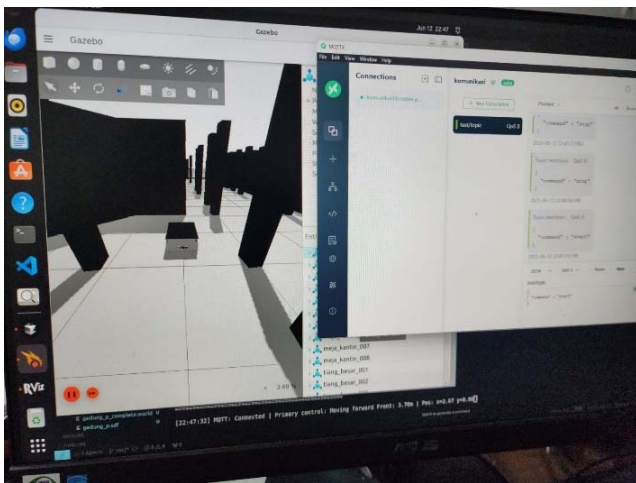
Raw Transcription (ID):
Text: Robot Start

=====
Question detected:
'Robot Start'

=====
Robot command detected!
Executing start command
Sent MQTT command: start

Listening...
Recording for 3 seconds...
    
```

Gambar 30. Sistem Voicebot Mengirim Perintah Start



Gambar 31. ROS Menerima Perintah dan Menjalankan Robot

Pada percobaan kedua sistem mengirimkan data berupa “Robot Stop” untuk menghentikan robot dan ROS menerima perintah tersebut dan menghentikan jalan robot seperti pada gambar 33 dimana robot dalam kondisi paused.

```

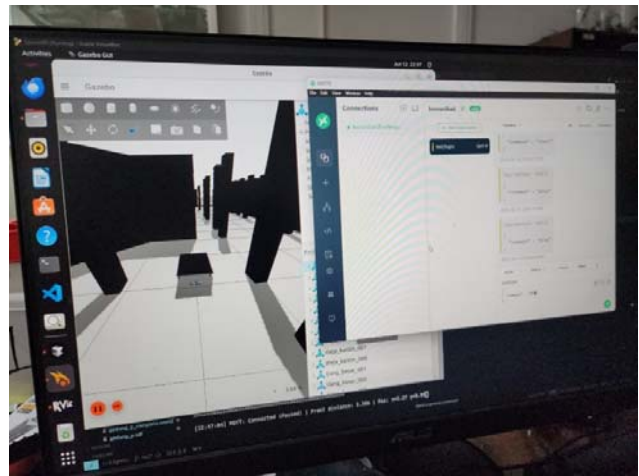
Initial transcription to detect language...
Whisper detected language: ID
Using Whisper's language detection: ID
Final detected language: ID
Transcribing in detected language: ID

Raw Transcription (ID):
Text: Robot, stop.

=====
Question detected:
'Robot, stop.'

=====
Robot command detected!
Executing stop command
Sent MQTT command: stop
    
```

Gambar 32. Sistem Voicebot Mengirim Perintah Stop



Gambar 33. ROS Menerima Perintah dan Memberhentikan Robot

IV. KESIMPULAN

Berdasarkan proses perencanaan, implementasi dan pengujian yang telah dilakukan dapat disimpulkan bahwa sistem ini berhasil memanfaatkan kombinasi teknologi STT, TTS serta RAG untuk membuat sebuah sistem voicebot Multibahasa yang dapat menjawab pertanyaan pengguna dalam 3 bahasa yaitu Inggris, Indonesia dan Mandarin dengan relevan sesuai dengan dokumen yang tersedia. Pengujian library CoquiTTS dan juga Whisper juga berjalan dengan baik. Sistem ini juga berhasil terintegrasi dengan robot dengan memanfaatkan MQTT sebagai protokol komunikasi yang dapat berkomunikasi dengan ROS untuk mengatur pergerakan robot melalui suara dan juga kamera OAK-D yang ditancapkan pada Raspberry Pi untuk mendeteksi manusia. Sistem ini masih dapat dikembangkan lebih lanjut dengan meningkatkan hardware yang lebih baik daripada recomputer j202 sehingga dapat meningkatkan performa dan kecepatan proses sistem serta memungkinkan untuk membuat semua berjalan secara lokal tidak hanya proses pengolahan suara saja sehingga tidak membutuhkan koneksi internet. Selain itu perlu ditambahkan variasi perintah robot untuk mendukung interaksi yang lebih kompleks.

DAFTAR PUSTAKA

- [1] Azhar, M. *How Indonesia’s new national strategy aims to accelerate the growth of Southeast Asia’s largest AI ecosystem*. GovInsider, 2023. available at: <https://govinsider.asia/intl-en/article/how-indonesias-new-national-strategy-aims-to-accelerate-the-growth-southeast-asias-largest-ai-ecosystem>
- [2] Arbi, I. A. *Mengenal "Ropi", robot pintar yang siap melayani penumpang stasiun MRT*. Kompas. 2022, Available at: <https://megapolitan.kompas.com/read/2022/12/30/20125831/mengenal-ropi-robot-pintar-yang-siap-melayani-penumpang-stasiun-mrt-lebak>
- [3] Antara News. *Pengoperasian robot pintar di stasiun MRT Lebak Bulus*. Available at: <https://ramadhan.antaranews.com/foto/3329205/pengoperasian-robot-pintar-di-stasiun-mrt-lebak-bulus>
- [4] Udri, M. H., Yakin, A. N., Hasnah, F. Y., Darlis, D., Rezamela, E., Fitri, I., Fauzi, N. F., & Sunarya, U. "Response speed analysis of interactive voicebot". *Jurnal*

*Teknik Energi Elektrik*, 12(4), 983–994. 2024.  
<http://dx.doi.org/10.26760/elkomika.v12i4.983>

- [5] Athikkal, S., & Jenq, J. “Voice chatbot for hospitality”.  
*Signal & Image Processing: An International Journal (SIPIJ)*, 13(2/3/4), 4–10. 2022.  
<https://doi.org/10.5121/sipij.2022.13401>
- [6] Luna, S. L., Caurin, G. A., & Tronco, M. L. *Usage of MQTT, ROS, and AWS in the manufacturing*. In *International Council of the Aeronautical Sciences (ICAS)*. Available at: [https://www.icas.org/icas\\_archive/ICAS2020/data/papers](https://www.icas.org/icas_archive/ICAS2020/data/papers)