

TUGAS AKHIR - EC234801

PEMBANGKITAN NOTIFIKASI UCAPAN PADA ROBOT SERVICE MENGGUNAKAN LLM BERBASIS SENSOR MULTIMODAL

Aryaduta Putra Perkasa

NRP 5024201077

Dosen Pembimbing

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.

NIP 19580916 198601 1 001

Muhtadin, S.T., M.T.

NIP 19810609 200912 1 003

Program Studi Strata 1 (S1) Teknik Komputer

Departemen Teknik Komputer

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya

2024



TUGAS AKHIR - EC234801

PEMBANGKITAN NOTIFIKASI UCAPAN PADA ROBOT SERVICE MENGGUNAKAN LLM BERBASIS SENSOR MULTIMODAL

Aryaduta Putra Perkasa

NRP 5024201077

Dosen Pembimbing

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.

NIP 19580916 198601 1 001

Muhtadin, S.T., M.T.

NIP 19810609 200912 1 003

Program Studi Strata 1 (S1) Teknik Komputer

Departemen Teknik Komputer

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya

2024

[Halaman ini sengaja dikosongkan]



FINAL PROJECT - EC234801

***SPEECH NOTIFICATION GENERATION FOR SERVICE
ROBOTS USING MULTIMODAL SENSOR-BASED LLM***

Aryaduta Putra Perkasa

NRP 5024201077

Advisor

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.

NIP 19580916 198601 1 001

Muhtadin, S.T., M.T.

NIP 19810609 200912 1 003

Undergraduate Study Program of Computer Engineering

Department of Computer Engineering

Faculty of Intelligent Electrical and Informatics Technology

Sepuluh Nopember Institute of Technology

Surabaya

2024

[Halaman ini sengaja dikosongkan]

LEMBAR PENGESAHAN

PEMBANGKITAN NOTIFIKASI UCAPAN PADA ROBOT SERVICE MENGUNAKAN LLM BERBASIS SENSOR MULTIMODAL

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Teknik pada
Program Studi S-1 Teknik Komputer
Departemen Teknik Komputer
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh: **Aryaduta Putra Perkasa**
NRP. 5024201077

Disetujui oleh Tim Penguji Tugas Akhir:

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
NIP: 19580916 198601 1 001

(Pembimbing I)

.....

Muhtadin, S.T., M.T.
NIP: 19810609 200912 1 003

(Pembimbing II)

.....

Eko Pramunanto, S.T., M.T..
NIP: 19661203199412 1 001

(Penguji I)

.....

Ir. Hany Boedinugroho, M.T..
NIP: 19610706 198701 1 001

(Penguji II)

.....

Ahmad Zaini, S.T., M.Sc..
NIP: 197504192 00212 1 003

(Penguji III)

.....

Mengetahui,
Kepala Departemen Teknik Komputer FTEIC - ITS

Dr. Supeno Mardi Susiki Nugroho, S.T., M.T..
NIP. 197003131 99512 1 0011

SURABAYA
Juni, 2024

[Halaman ini sengaja dikosongkan]

APPROVAL SHEET

SPEECH NOTIFICATION GENERATION FOR SERVICE ROBOTS USING MULTIMODAL SENSOR-BASED LLM

FINAL PROJECT

Submitted to fulfill one of the requirements
for obtaining a degree Bachelor of Engineering at
Undergraduate Study Program of Computer Engineering
Department of Computer Engineering
Faculty of Intelligent Electrical and Informatics Technology
Sepuluh Nopember Institute of Technology

By: **Aryaduta Putra Perkasa**
NRP. 5024201077

Approved by Final Project Examiner Team:

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
NIP: 19580916 198601 1 001

(Advisor I)

.....

Muhtadin, S.T., M.T.
NIP: 19810609 200912 1 003

(Co-Advisor II)

.....

Eko Pramunanto, S.T., M.T..
NIP: 19661203199412 1 001

(Examiner I)

.....

Ir. Hany Boedinugroho, M.T..
NIP: 19610706 198701 1 001

(Examiner II)

.....

Ahmad Zaini, S.T., M.Sc..
NIP: 197504192 00212 1 003

(Examiner III)

.....

Acknowledged,
Head of Computer Engineering Department F-ELECTICS - ITS

Dr. Supeno Mardi Susiki Nugroho, S.T., M.T..
NIP. 197003131 99512 1 0011

SURABAYA
June, 2024

[Halaman ini sengaja dikosongkan]

PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama Mahasiswa / NRP : Aryaduta Putra Perkasa / 5024201077
Departemen : Teknik Komputer
Dosen Pembimbing / NIP : Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng. / 19580916
198601 1 001

Dengan ini menyatakan bahwa Tugas Akhir dengan judul "PEMBANGKITAN NOTIFIKASI UCAPAN PADA ROBOT SERVICE MENGGUNAKAN LLM BERBASIS SENSOR MULTIMODAL " adalah hasil karya sendiri, berfsifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

Surabaya, June 2024

Mengetahui
Dosen Pembimbing

Mahasiswa

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
NIP. 19580916 198601 1 001

Aryaduta Putra Perkasa
NRP. 5024201077

[Halaman ini sengaja dikosongkan]

STATEMENT OF ORIGINALITY

The undersigned below:

Name of student / NRP : Aryaduta Putra Perkasa / 5024201077
Department : Computer Engineering
Advisor / NIP : Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng. / 19580916
198601 1 001

Hereby declared that the Final Project with the title of "*SPEECH NOTIFICATION GENERATION FOR SERVICE ROBOTS USING MULTIMODAL SENSOR-BASED LLM*" is the result of my own work, is original, and is written by following the rules of scientific writing.

If in future there is a discrepancy with this statement, then I am willing to accept sanctions in accordance with provisions that apply at Sepuluh Nopember Institute of Technology.

Surabaya, June 2024

Acknowledged
Advisor

Student

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
NIP. 19580916 198601 1 001

Aryaduta Putra Perkasa
NRP. 5024201077

[Halaman ini sengaja dikosongkan]

ABSTRAK

Nama Mahasiswa : Aryaduta Putra Perkasa
Judul Tugas Akhir : PEMBANGKITAN NOTIFIKASI UCAPAN PADA ROBOT SERVICE
MENGGUNAKAN LLM BERBASIS SENSOR MULTIMODAL
Pembimbing : 1. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
2. Muhtadin, S.T., M.T.

Penelitian ini bertujuan untuk mengembangkan sistem pembangkitan notifikasi ucapan pada robot service menggunakan Language Model (LLM) berbasis sensor multimodal. Dengan memanfaatkan teknologi sensor yang mampu menangkap beragam input, seperti suara, gerakan, dan visual, robot service dapat secara efektif berkomunikasi dengan pengguna melalui notifikasi ucapan yang lebih kontekstual dan responsif. LLM digunakan sebagai inti pengolahan bahasa untuk meningkatkan kemampuan pemahaman dan generasi respons pada robot. Sistem ini diharapkan dapat meningkatkan interaksi manusia-mesin dengan memungkinkan robot memberikan notifikasi dengan lebih intuitif dan adaptif, menciptakan pengalaman pengguna yang lebih memuaskan. Metodologi penelitian melibatkan pengembangan dan integrasi teknologi sensor multimodal, implementasi LLM, serta evaluasi kinerja sistem melalui uji coba simulasi dan pengujian langsung. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan teknologi interaksi manusia-mesin yang lebih canggih dan efisien.

Kata Kunci: Robot *Service*, Notifikasi Ucapan, *LLM*, Sensor Multimodal, Interaksi Manusia-Robot, Pemrosesan Bahasa Alami.

[Halaman ini sengaja dikosongkan]

ABSTRACT

Name : Aryaduta Putra Perkasa
Title : *SPEECH NOTIFICATION GENERATION FOR SERVICE ROBOTS USING MULTIMODAL SENSOR-BASED LLM*
Advisors : 1. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
2. Muhtadin, S.T., M.T.

This study aims to develop a speech notification generation system for service robots using a multimodal sensor-based Language Model (LLM). By leveraging sensor technology capable of capturing diverse inputs, such as sound, motion, and visual cues, robot services can effectively communicate with users through more contextual and responsive speech notifications. The LLM serves as the core language processor to enhance language understanding and response generation in the robot. The system is expected to improve human-machine interaction by enabling robots to provide notifications in a more intuitive and adaptive manner, creating a more satisfying user experience. The research methodology involves the development and integration of multimodal sensor technology, the implementation of LLM, and the evaluation of system performance through simulation tests and direct testing. The outcomes of this research are anticipated to contribute to the advancement of sophisticated and efficient human-robot interaction technology.

Keywords: Service Robot, Speech Notification, LLM, Multimodal Sensor, HUMAN-Robot Interaction, Natural Language Processing.

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Puji dan syukur kehadiran Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque.

Penelitian ini disusun dalam rangka Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Keluarga, Ibu, Bapak dan Saudara tercinta yang telah Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero.
2. Bapak Nikola Tesla, S.T., M.T., selaku Quisque ullamcorper placerat ipsum. Cras nibh.
3. Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl.

Akhir kata, semoga Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus.

Surabaya, Juni 2024

Aryaduta Putra Perkasa

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

ABSTRAK	i
ABSTRACT	iii
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Manfaat	2
2 TINJAUAN PUSTAKA	3
2.1 Interaksi Robot-Manusia	3
2.2 Robot <i>Service</i>	3
2.3 Dobot Magician	4
2.4 Model Bahasa Besar	5
2.4.1 Gemma	6
2.4.2 <i>Low Rank Adaptation</i>	7
2.5 Multimodalitas	8
2.6 OpenCV	9
3 DESAIN DAN IMPLEMENTASI	11
3.1 Deskripsi Sistem	11
3.2 Metode	11
3.3 Dataset	12
3.4 <i>Fine Tuning</i>	14

3.5	Evaluasi Metrik	14
4	PENGUJIAN DAN ANALISIS	17
4.1	Pengujian Model Bahasa	17
4.2	Metrik Pengujian	19
	DAFTAR PUSTAKA	21

DAFTAR GAMBAR

2.1	Dobot Magician (“Dobot US”, 2024).	4
2.2	Kinematika (Islam et al., 2019).	5
2.3	Arsitektur Transformer (Vaswani et al., 2023).	6
2.4	Logo Gemma (“Gemma Open Model”, 2024).	6
2.5	Ukuran Parameter Gemma (“Gemma Open Model”, 2024).	7
2.6	Benchmark Gemma 7B dan Mistral 7B (“Gemma Open Model”, 2024).	7
2.7	Mekanisme kerja LoRA (“Gemma Open Model”, 2024).	8
2.8	Konsep Multimodal (“Dobot US”, 2024).	8
2.9	Logo OpenCV (“OpenCV”, 2024).	9
3.1	Alur sistem	11
3.2	Metode	12
3.3	Kriteria Dataset	13
3.4	Format Dataset	13
3.5	Tuning	14
3.6	Metrik <i>Exact Match</i>	14
4.1	Uji Reasoning	17
4.2	Hasil JSON Reasoning	17
4.3	Uji Chain of Command	17
4.4	Hasil JSON Chain of Command	18
4.5	Uji Direct Command	18
4.6	Hasil JSON Direct Command	18
4.7	Uji Impossible Task	18
4.8	Tuning	19

[Halaman ini sengaja dikosongkan]

DAFTAR TABEL

2.1	Denavit-Hartenberg Parameters	5
-----	-----------------------------------------	---

[Halaman ini sengaja dikosongkan]

BAB I

PENDAHULUAN

1.1 Latar Belakang

Robot adalah sebuah mesin yang dapat diprogram untuk melakukan berbagai gerakan dan tugas secara otomatis (Singh & Banga, 2022). Perkembangannya didorong oleh kemajuan teknologi di bidang elektronika, mekanika, dan kecerdasan buatan. Teknologi ini memungkinkan robot untuk memiliki kemampuan untuk melakukan tugas yang lebih kompleks. Saat ini, robot telah digunakan di berbagai bidang, mulai dari industri manufaktur, kesehatan, hingga hiburan.

Robot *service* adalah jenis robot yang dirancang dan diprogram untuk menjalankan berbagai tugas yang bermanfaat bagi manusia (Gonzalez-Aguirre et al., 2021). Kemampuannya untuk diprogram dan menjalankan tugas secara mandiri menjadikannya ideal untuk membantu manusia dalam berbagai aspek kehidupan. Salah satu robot *service* yang cukup terkenal adalah robot penyedot debu. Robot ini sangat populer di ranah konsumen rumah tangga karena kemampuannya membersihkan lantai secara otomatis. Contohnya adalah robot Roomba yang diproduksi oleh perusahaan iRobot (Elara et al., 2014). Robot ini menggunakan sensor dan algoritma navigasi membersihkan ruangan secara efisien.

Interaksi manusia-robot adalah disiplin ilmu yang bertujuan untuk memahami, merancang, dan menilai interaksi sistem robotik untuk digunakan bersama dengan manusia (Mohebbi, 2020). Meningkatnya penggunaan robot pada skala konsumen umum mendorong kebutuhan untuk merancang robot yang dapat berkolaborasi secara efektif dengan manusia. Ada berbagai metode untuk berinteraksi dengan robot, salah satunya adalah melalui sentuhan fisik, seperti menggunakan tombol atau layar sentuh. Penggunaan dengan tombol sentuh memerlukan pengetahuan dasar pengguna dalam pengoperasian robot. Jika pengguna tidak terbiasa dengan kontrol pengoperasian robot, dibutuhkan pelatihan khusus untuk dapat berinteraksi dengan robot *input* (Braga et al., 2023). Metode lain melibatkan sensor untuk mengumpulkan informasi lingkungan, seperti sensor gerakan atau kamera, yang memungkinkan robot menyesuaikan respons sesuai kebutuhan.

Seiring dengan perkembangan teknologi kecerdasan buatan, interaksi dengan robot semakin beragam dengan perkembangan pemrosesan bahasa alami (*Natural Language Processing*). Model bahasa besar (*Large Language Model*) menjadi salah satu tren terpopuler di bidang kecerdasan buatan (*Artificial Intelligence*) dalam beberapa tahun terakhir. Model bahasa besar adalah model yang mampu memproses dan menghasilkan teks (Naveed et al., 2024). Model ini mampu menghasilkan teks, menerjemahkan bahasa, menulis berbagai jenis teks, dan menjawab pertanyaan dengan cara yang informatif. Sebagai contoh, salah satu model bahasa besar yang populer adalah GPT (*Generative Pre-trained Transformer*). Dengan perkembangan ini, robot dapat memahami perintah bahasa alami dan merespon dengan cara yang menyerupai interaksi manusia.

Meskipun demikian, dalam penerapan model bahasa untuk interaksi robot-manusia, ada beberapa hal yang harus dipertimbangkan. Model bahasa yang telah dilatih secara umum sering kali tidak dapat memahami dan merespon sesuai dengan informasi kondisi, kemampuan, dan

lingkungan nyata robot. Pemahaman terhadap informasi ini sangat penting mengingat robot bergerak di lingkungan nyata yang memiliki berbagai batasan seperti batasan lokasi atau aksi yang dapat dilakukan. Salah satu hal yang dapat dilakukan untuk meningkatkan performa model bahasa adalah dengan melakukan (*fine tuning*) dimana di mana model yang sudah ada dilatih kembali dengan menggunakan data khusus yang relevan dengan konteks robot. Proses (*fine tuning*) telah terbukti dapat meningkatkan performa model bahasa terhadap kondisi yang sebelumnya tidak terlihat (Chung et al., 2022). Informasi lingkungan nyata robot dapat diambil dengan integrasi sensor multimodal, seperti kamera, juga menjadi bagian penting dalam meningkatkan pemahaman robot terhadap kondisi lingkungan nyata.

Dalam upaya untuk meningkatkan pengalaman interaksi robot-manusia, penulis menggunakan model bahasa besar berbasis sensor multimodal untuk menghasilkan sistem respon dari robot service yang lebih alami dan sesuai dengan konteks lingkungan robot. Melalui penelitian ini, diharapkan dapat memberikan kontribusi signifikan pada kemajuan bidang interaksi robot-manusia.

1.2 Permasalahan

Bagaimana menerapkan integrasi model bahasa besar dan konsep multimodalitas melalui sensor-sensor dalam pengembangan respon robot *service* untuk meningkatkan kemampuan adaptif, respons kontekstual, dan menciptakan pengalaman interaksi yang lebih personal dan efektif antara robot dan manusia?

1.3 Tujuan

Tujuan penelitian ini adalah untuk mengembangkan sistem interaksi robot-manusia dengan model bahasa besar (LLM) dan sensor multimodal pada respon robot service dengan fokus pada peningkatan kemampuan adaptif, respons kontekstual, serta penciptaan pengalaman interaksi yang lebih alami dan efektif.

1.4 Batasan Masalah

Adapun beberapa batasan dalam pembahasan tugas akhir ini yaitu, sebagai berikut:

1. Sensor multimodal yang digunakan adalah kamera.
2. Robot yang digunakan adalah robot lengan Dobot Magician.

1.5 Manfaat

Penelitian ini diharapkan memberikan manfaat yang signifikan pada pengembangan teknologi Robot Service dan interaksi robot-manusia. Integrasi Model Bahasa Besar (LLM) dan konsep multimodalitas melalui sensor-sensor bertujuan untuk meningkatkan kemampuan adaptif, respons kontekstual, dan menciptakan pengalaman interaksi yang lebih personal antara robot dan manusia. Dengan peningkatan kemampuan komunikasi robot service, diharapkan memberikan respon yang lebih efisien, efektif, dan sesuai dengan kebutuhan pengguna.

BAB II

TINJAUAN PUSTAKA

Pada penelitian sebelumnya berjudul "Language Models for Human-Robot Interaction" dengan tujuan untuk menunjukkan potensi dan batasan dari model bahasa besar dalam konteks interaksi manusia-robot secara langsung (Islam et al., 2019). Dalam penelitian ini, model bahasa besar OpenAI GPT-3 berhasil diintegrasikan dengan robot Pepper dan Nao untuk menciptakan sistem dialog verbal terbuka. Saat pengujian, model yang digunakan adalah GPT-3 Davinci, dan hasil implementasi ini disajikan memberikan kesempatan untuk partisipasi manusia untuk terlibat dalam percakapan terbuka dengan robot Pepper dan Nao, memungkinkan mereka untuk memahami batasan dari model bahasa besar dalam konteks interaksi manusia-robot. Meskipun berhasil di tahap pengujian, perlu dicatat bahwa robot menggunakan model yang belum terintegrasi sepenuhnya dengan robot dan tidak memiliki informasi umum mengenai keadaan robot ataupun lingkungan sekitarnya.

2.1 Interaksi Robot-Manusia

Interaksi Manusia-Robot (HRI) adalah disiplin ilmu yang mempelajari tentang dinamika interaksi antara manusia dan robot (Vasconez et al., 2019). Disiplin ini memfokuskan pada pemahaman, perancangan, dan evaluasi cara manusia dan robot berinteraksi di berbagai konteks dengan manusia. Tujuan utamanya adalah untuk membuat interaksi antara manusia dan robot semakin lancar dan terasa alami, sehingga mengurangi waktu yang dibutuhkan bagi pengguna untuk beradaptasi. Dengan perkembangan teknologi yang semakin maju dan penetrasi robot yang semakin luas dalam berbagai aspek kehidupan manusia, sangat penting untuk memberikan prioritas pada pengalaman pengguna dan memastikan bahwa setiap interaksi tersebut mudah dipahami, efisien, dan, yang terpenting, berfokus pada kebutuhan manusia (Chikwendu et al., 2023). Dalam interaksi robot-manusia, sangat penting untuk menganalisis dan mengoptimalkan aspek-aspek seperti antarmuka pengguna, perilaku robot, dan dampak yang mungkin timbul selama interaksi tersebut dengan tujuan menciptakan interaksi yang efektif, intuitif, dan mendukung kerjasama harmonis antara manusia dan robot dalam berbagai konteks. Pemahaman bahasa manusia oleh robot, baik melalui pemrosesan bahasa alami maupun pengenalan suara, memungkinkan interaksi yang lebih kompleks dan alami.

Di sektor industri, interaksi robot-manusia telah menciptakan lingkungan kerja yang dinamis. Robot kolaboratif adalah jenis robot yang dapat bekerja bersama manusia dengan aman, menghadirkan peluang untuk meningkatkan produktivitas dan efisiensi. Sementara itu, di bidang layanan dan kesehatan, robot memberikan dukungan dalam pemantauan kesehatan dan terapi, membawa dampak positif dalam pelayanan dan kesejahteraan manusia.

2.2 Robot Service

Robot *Service* adalah robot yang dirancang untuk memberikan layanan kepada manusia (Lee, 2021). Robot service memiliki di berbagai bidang, seperti rumah tangga dan komersial. Dalam mendukung interaksi dengan manusia, Robot *service* dapat menerapkan berbagai

metode komunikasi, termasuk melalui ucapan, tindakan, dan antarmuka yang dirancang intuitif. Sensor-sensor seperti sensor suara, kamera, dan sensor gerak membekali Robot *service* dengan kemampuan responsif terhadap perubahan lingkungan sekitar, menciptakan pengalaman interaksi yang dinamis.

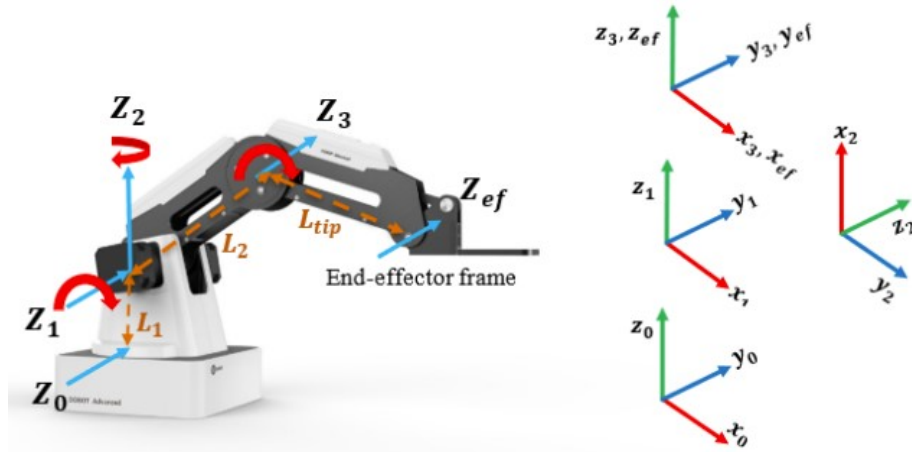
Aplikasi yang luas mencakup berbagai sektor, seperti layanan pelanggan, di mana robot *service* dapat memberikan informasi, bantuan, atau menerima pesanan. Sebagai asisten pribadi, kemampuannya membantu dalam manajemen jadwal dan memberikan informasi cuaca memperluas perannya di kehidupan sehari-hari. Di sektor kesehatan, robot *service* dapat memberikan dukungan atau mengingatkan jadwal obat, sementara di bidang pendidikan, menjadi penyajian materi pembelajaran yang interaktif. Robot *service* dapat membantu dalam perkembangan teknologi untuk memperbaiki kualitas layanan dan membentuk interaksi manusia dengan mesin menjadi lebih responsif dan dinamis.

2.3 Dobot Magician



Gambar 2.1: Dobot Magician (“Dobot US”, 2024).

Dobot Magician adalah robot lengan kolaboratif yang memiliki 4 poros. Robot ini memiliki tingkat akurasi yang tinggi, yakni sekitar 0,2 mm, dan menggunakan chip industri STM32 sebagai pengontrolnya. Untuk mengontrol robot ini, tersedia *API* dalam bahasa pemrograman *Python* yang dapat digunakan oleh pengembang untuk mengontrol robot. Robot ini juga mempunyai Aplikasi Dobot Lab yang menawarkan fitur pengontrol berbasis *Graphical User Interface* untuk mempermudah pengguna tanpa latar belakang dalam pemrograman untuk mengontrol robot. Robot ini juga memiliki berbagai macam pilihan *end effector* yang dapat digunakan untuk mengatur fungsionalitas robot. *End effector* yang tersedia meliputi pencapit (*claw*) untuk menjepit objek, penghisap (*suction cup*) yang dapat digunakan untuk menyedot objek, pena untuk melakukan penulisan atau penggambaran, dan ada juga aksesoris yang dapat digunakan untuk melakukan pencetakan 3d. Untuk aksesoris yang digunakan diluar *end effector* terdapat aksesoris citra *vision* yang dapat digunakan untuk pengolahan citra dan juga *conveyor belt* untuk memindahkan barang.



Gambar 2.2: Kinematika (Islam et al., 2019).

Link i	Link Twist α_{i-1}	Link Length a_{i-1}	Link Offset d_i	Joint Angle θ_i
1	0	0	d	θ_1
2	$\frac{\pi}{2}$	0	0	θ_2
3	0	a_1	0	θ_3
4	0	a_2	0	θ_4

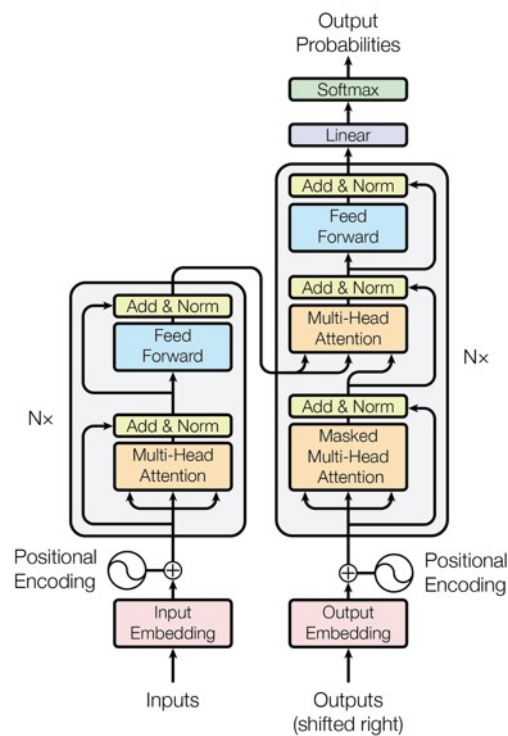
Tabel 2.1: Denavit-Hartenberg Parameters

Berikut ini adalah kinematika dan parameter Denavit-Hartenberg robot lengan Dobot Magician. Robot ini mempunyai empat *joint*. *Joint* pertama terletak di *base*, *joint* kedua terletak di lengan bagian belakang, *joint* ketiga terletak di lengan bagian depan, dan *joint* keempat terletak dekat *end effector*.

2.4 Model Bahasa Besar

Model bahasa besar adalah model yang dirancang untuk dapat memahami dan menghasilkan teks (Wang et al., 2023). Model ini menggunakan teknik *deep learning* dan pemrosesan bahasa alami untuk mengenali pola dalam teks, dan kemudian menggunakan pola tersebut untuk menghasilkan teks baru. Untuk melakukan training dalam model bahasa besar, diperlukan data teks dengan jumlah sangat besar dan kualitasnya akan mempengaruhi performa model bahasa (Liu et al., 2024).

Perkembangan model bahasa besar tidak lepas dengan perkembangan arsitektur transformer. Arsitektur transformer adalah salah satu jenis arsitektur dalam *neural network* yang populer digunakan untuk pemrosesan teks. Model ini menggunakan mekanisme *attention* untuk menangkap hubungan jarak jauh antara kata-kata, sehingga memungkinkan model untuk memahami konteks teks. Pada arsitektur transformer, *attention* adalah proses yang memungkinkan model untuk memfokuskan pada kata-kata tertentu dalam teks (Vaswani et al., 2023). Proses ini dilakukan dengan menghitung skor *attention* untuk setiap kata dalam teks. Skor *attention* ini kemudian digunakan untuk menentukan seberapa besar pengaruh kata tersebut terhadap kata yang sedang diproses.



Gambar 2.3: Arsitektur Transformer (Vaswani et al., 2023).

Model GPT *Generative Pre-training Transformer* dan PaLM *Pathways Language Model* adalah dua model bahasa besar populer yang dikembangkan oleh OpenAI dan Google. Kedua model ini menggunakan mekanisme *attention* untuk menangkap hubungan jarak jauh antara kata-kata, sehingga memungkinkan model bahasa besar untuk memahami konteks teks.

2.4.1 Gemma



Gambar 2.4: Logo Gemma (“Gemma Open Model”, 2024).

Gemma adalah sebuah model bahasa besar yang bersifat *open source* dan ringan yang dikembangkan oleh Google DeepMind. Model-model ini dirancang untuk memahami dan menghasilkan teks dengan kemampuan yang serupa dengan manusia, serta dapat digunakan

untuk berbagai aplikasi dalam bidang pemrosesan bahasa alami, seperti memahami teks, menghasilkan teks, dan penerjemahan. Gemma merupakan model teks ke teks yang berarti mempunyai input teks dan menghasilkan output teks.

Model	Embedding Parameters	Non-embedding Parameters
2B	524,550,144	1,981,884,416
7B	786,825,216	7,751,248,896

Gambar 2.5: Ukuran Parameter Gemma (“Gemma Open Model”, 2024).

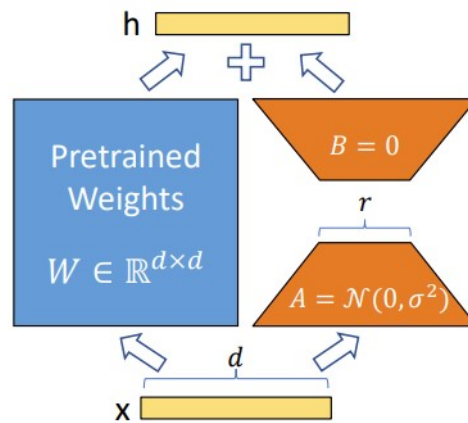
	Mistral	Gemma
Benchmark	7B	7B
ARC-c	60.0	61.9
HellaSwag	83.3	82.2
MMLU	64.2	64.6
TruthfulQA	42.2	44.8
Winogrande	78.4	79.0
GSM8K	37.8	50.9
Average	61.0	63.8

Gambar 2.6: Benchmark Gemma 7B dan Mistral 7B (“Gemma Open Model”, 2024).

Gemma hadir dengan dua jenis ukuran parameter, 2b dan 7b. Jenis ukuran ini merepresentasikan jumlah paramater yang dilatih dalam model bahasa. Model Gemma 7b memiliki arsitektur yang lebih kompleks, cocok digunakan untuk menangani pemrosesan bahasa yang memerlukan pemahaman konteks yang kompleks dan penanganan tugas NLP yang lebih mendalam. Di sisi lain, Gemma 2b dirancang untuk melakukan pemrosesan bahasa pada perangkat dengan sumber daya komputasi yang lebih kecil, memastikan kinerja yang optimal dalam lingkungan yang memiliki keterbatasan sumber daya komputasi. Berdasarkan hasil benchmark, model Gemma memiliki rata - rata performa terbaik dibandingkan model lain dengan ukuran paramater sejenis.

2.4.2 Low Rank Adaptation

Low Rank Adaptation atau disingkat LoRA merupakan teknik untuk melakukan *fine tuning* yang digunakan model bahasa besar untuk menyesuaikan dengan data spesifik yang baru tanpa harus melatih ulang seluruh model. Teknik ini melibatkan pengurangan dimensi dari representasi internal model, yang dapat dilakukan dengan menggunakan metode faktorisasi matriks (*Low Rank Matrix*). Dengan mengurangi kompleksitas model, low rank adaptation memungkinkan penyesuaian yang cepat dan efisien terhadap data baru, sambil tetap mempertahankan sebagian besar pengetahuan yang telah diperoleh oleh model asli.

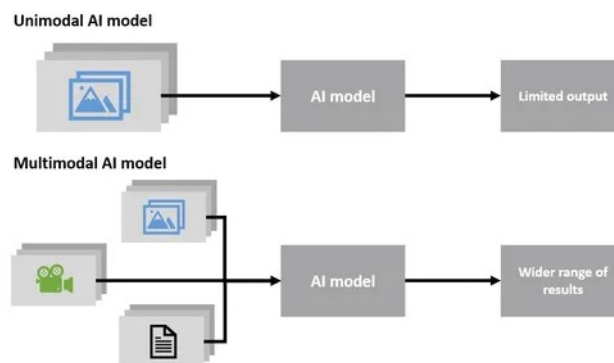


Gambar 2.7: Mekanisme kerja LoRA (“Gemma Open Model”, 2024).

Dengan menggunakan proses ini, memungkinkan model untuk memperoleh pemahaman yang lebih mendalam terhadap data spesifik tertentu dengan menghilangkan komponen yang kurang relevan atau umum dari representasi internalnya. Kemudian, model yang telah disesuaikan dapat digunakan untuk tugas-tugas khusus dalam domain tersebut dengan kinerja yang lebih baik daripada model yang belum disesuaikan. Keunggulan utama LoRA terletak pada efisiensi dalam penyesuaian model: sebuah model yang telah dilatih sebelumnya dapat dibagi dan digunakan untuk membangun banyak modul LoRA kecil untuk tugas-tugas yang berbeda.

2.5 Multimodalitas

Sistem interaksi manusia-komputer multimodal terdiri dari penggunaan berbagai saluran input dan output (Jia et al., 2020). Pendekatan multimodal ini mengintegrasikan beberapa mode komunikasi, seperti teks, suara, gambar, dan gerakan, untuk menciptakan pengalaman interaktif yang lebih kaya. Dengan memanfaatkan keberagaman saluran ini, sistem dapat mengoptimalkan cara pengguna berinteraksi dengan perangkat komputer. Pengguna dapat memberikan input melalui berbagai cara, seperti mengetik, berbicara, atau menggunakan gerakan tubuh, sedangkan output dapat disampaikan melalui teks, suara, visual, atau kombinasi dari semuanya. Pendekatan multimodal bertujuan untuk meningkatkan aksesibilitas, efisiensi, dan pengalaman pengguna dalam berbagai konteks, memungkinkan interaksi yang lebih intuitif dan menyeluruh antara manusia dan sistem komputer.



Gambar 2.8: Konsep Multimodal (“Dobot US”, 2024).

Penerapan multimodalitas dapat ditemukan dalam berbagai aplikasi, seperti asisten pintar, perangkat seluler, dan sistem interaktif. Misalnya, dalam asisten pintar, pengguna dapat berinteraksi dengan menggunakan kombinasi suara dan teks untuk memberikan perintah atau mendapatkan informasi. Pada perangkat seluler, layar sentuh, suara, dan gerakan dapat digunakan bersamaan untuk memberikan pengalaman pengguna yang lebih interaktif. Pada perangkat pintar rumah, sistem multimodal dapat mengukur suhu ruangan dan jumlah orang di dalam ruangan untuk mengatur suhu pendingin ruangan. Sistem multimodal juga membuka peluang untuk meningkatkan aksesibilitas, memungkinkan individu dengan berbagai kebutuhan untuk berinteraksi dengan cara yang paling nyaman. Dengan menggabungkan berbagai modalitas, multimodalitas tidak hanya memperkaya interaksi antara manusia dan teknologi, tetapi juga mendukung pengembangan sistem yang lebih inklusif dan responsif

2.6 OpenCV



Gambar 2.9: Logo OpenCV (“OpenCV”, 2024).

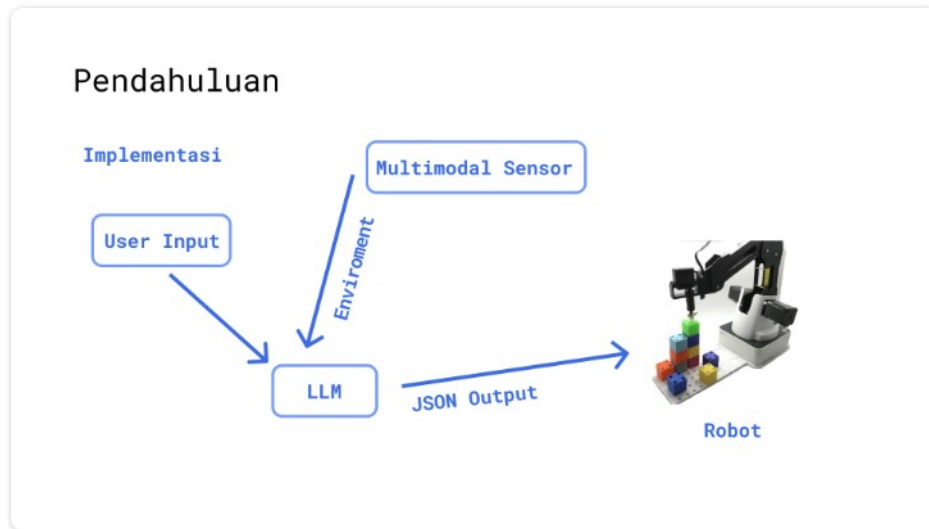
OpenCV (*Open Source Computer Vision Library*) adalah sebuah library bersifat terbuka yang menyediakan berbagai fungsi untuk melakukan pengolahan citra dan video pada komputer (“OpenCV”, 2024). OpenCV telah menjadi salah satu *library* perangkat lunak paling populer dan sangat dipercaya dalam pengolahan citra dan video komputer. Pustaka ini menawarkan beragam alat dan fungsi yang sangat berguna dalam berbagai aplikasi, seperti deteksi objek, pelacakan gerakan, pengenalan pola, dan segmentasi citra.

[Halaman ini sengaja dikosongkan]

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Deskripsi Sistem



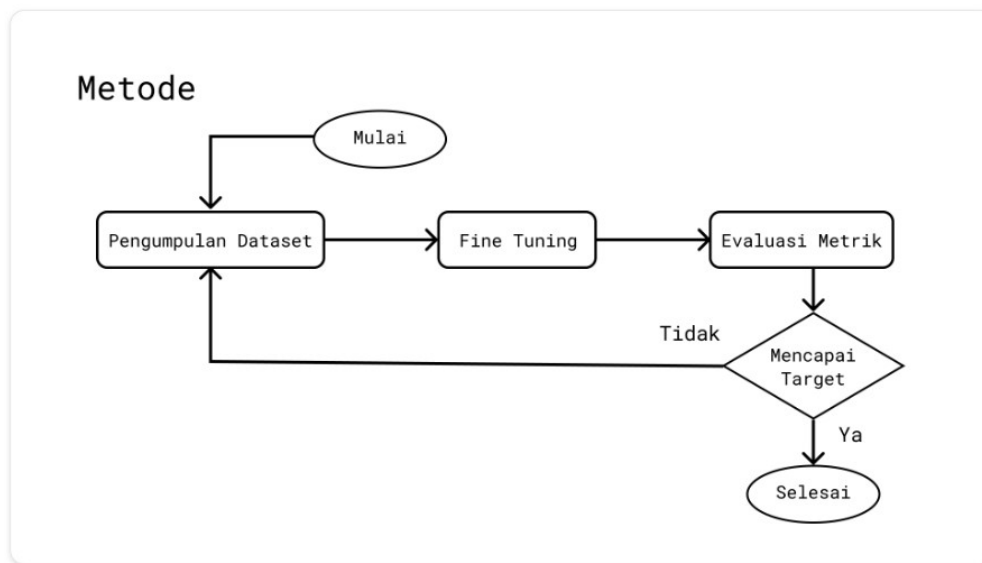
Gambar 3.1: Alur sistem

Metode yang digunakan untuk menyelesaikan tugas akhir ini dapat diuraikan sebagai berikut. Pertama-tama, sistem akan menerima input dari berbagai sumber, termasuk sensor multimodal yang menggunakan kamera dengan OpenCV untuk mendeteksi objek dalam lingkungan sekitar. Informasi kontekstual juga diperoleh, mencakup fungsi yang dapat diakses oleh robot (misalkan fungsi "gerak()") serta tujuan kontrol yang diinginkan oleh pengguna (misalkan "ambil balok merah"). Selanjutnya dengan memanfaatkan teknologi model bahasa besar, sistem memproses seluruh input yang diberikan untuk menghasilkan rencana aksi bagi robot. Pada tahap terakhir, output melibatkan konversi hasil rencana tersebut ke dalam format JSON. Format ini digunakan untuk menyusun serangkaian instruksi yang dapat diinterpretasikan oleh robot. Dengan menggunakan Dobot API Python, informasi tersebut diubah menjadi gerakan Dobot yang dapat dieksekusi oleh robot sesuai dengan aksi yang telah ditetapkan. Dengan mengikuti langkah-langkah ini, sistem mampu menerima instruksi dari pengguna, memahaminya, dan menerjemahkannya menjadi tindakan yang dapat dilakukan oleh robot dalam lingkungan sekitarnya.

3.2 Metode

Metode yang diterapkan untuk menyelesaikan tugas akhir ini telah dirancang dengan langkah-langkah yang terperinci untuk memastikan keberhasilan proyek. Tahap pertama adalah pengumpulan dataset, di mana data yang relevan dikumpulkan dalam bentuk pasangan masukan dan keluaran untuk digunakan dalam proses pengembangan model. Selanjutnya, proses finetun-

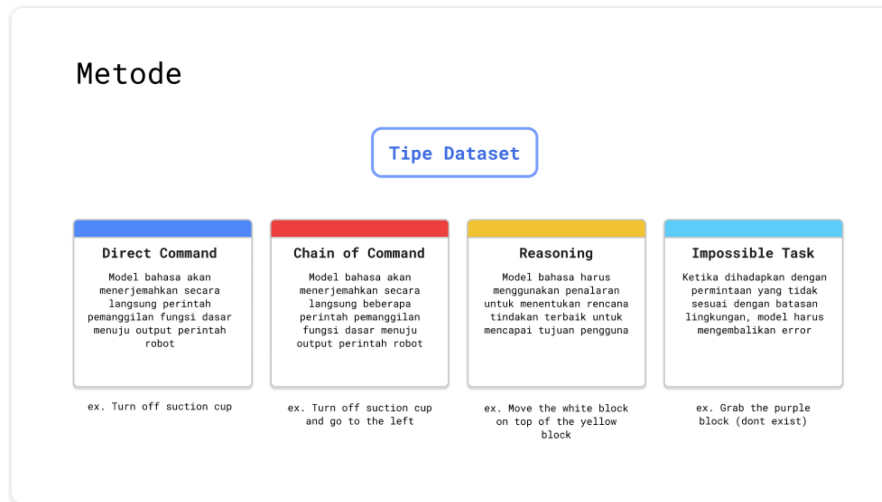
ing dilakukan, di mana model bahasa disesuaikan dengan dataset baru yang telah terkumpul, memastikan bahwa model mampu mengenali pola-pola dalam data yang bersangkutan. Setelah itu, evaluasi metrik menjadi langkah krusial dalam memeriksa performa model yang telah disesuaikan. Dengan menggunakan metrik evaluasi, hasil dari model dievaluasi untuk menentukan seberapa baik model dapat menghasilkan output yang diinginkan. Apabila hasil evaluasi masih belum memenuhi target yang telah ditetapkan, maka dilakukan iterasi kembali pada tahap pengumpulan dataset atau finetuning. Dengan pendekatan ini diharapkan bahwa proyek tugas akhir ini akan menghasilkan model yang optimal dan sesuai dengan kebutuhan yang diinginkan.



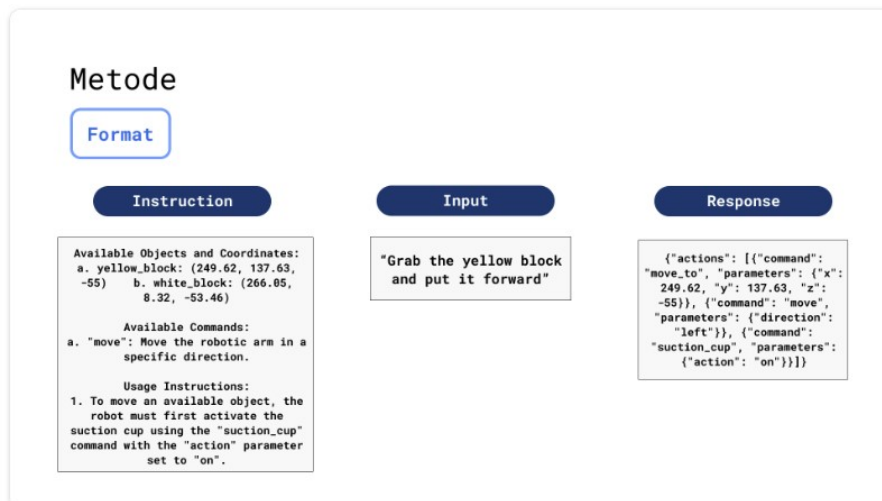
Gambar 3.2: Metode

3.3 Dataset

Dalam pengembangan model untuk pengendalian robot, jenis dataset yang digunakan memiliki peran yang penting dalam memperkuat kemampuan sistem. Pertama, jenis dataset *Direct Command* digunakan untuk melatih model dalam mengakses fungsi dasar robot dengan instruksi langsung. Dataset ini membantu model memahami dan mengeksekusi perintah-perintah dasar dengan akurat. Kedua, dataset *Chain of Command* memberikan latihan pada model untuk mengakses beberapa fungsi dasar robot secara bersamaan, memungkinkan robot untuk melakukan serangkaian tindakan terkait dengan satu instruksi. Selanjutnya, dataset *Reasoning* menekankan pada pelatihan model untuk menentukan rencana tindakan terbaik untuk mencapai tujuan pengguna tanpa adanya instruksi yang eksplisit. Dengan demikian, model diajarkan untuk memahami konteks dan tujuan yang lebih luas dalam mengambil keputusan. Terakhir, kriteria terakhir adalah dataset *Impossible Task*, di mana model diuji dengan permintaan yang tidak sesuai dengan batasan lingkungan. Dalam situasi ini, model harus dapat mengenali batasan tersebut dan mengembalikan error sebagai respons yang tepat.



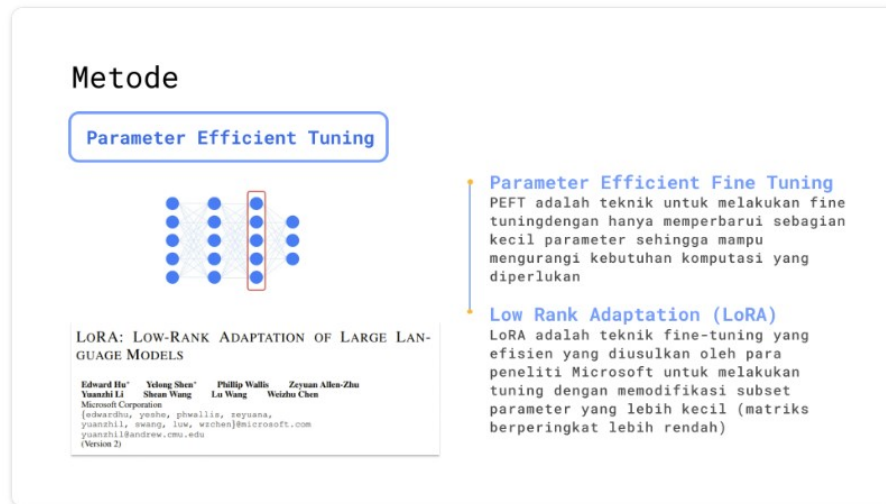
Gambar 3.3: Kriteria Dataset



Gambar 3.4: Format Dataset

Untuk proses fine-tuning, format dataset akan disusun sesuai dengan format Alpaca Stanford, yang memiliki komponen-komponen sebagai berikut. Bagian *Instruction* berisi instruksi yang diperlukan untuk mencapai output yang diinginkan. Dalam penelitian ini, instruksi mencakup kondisi nyata lingkungan seperti objek yang dapat dimanipulasi dan koordinatnya. Selain itu, terdapat pula fungsi dasar yang dapat diakses oleh model bahasa (misalkan fungsi Gerak()) yang digunakan untuk mengendalikan robot. Instruksi penggunaan juga disertakan, yang berisi langkah-langkah yang harus diambil untuk mencapai keluaran yang diharapkan. Bagian *Input* berisi perintah yang diberikan oleh pengguna kepada robot. Contohnya, perintah untuk "mengambil blok merah dan memindahkannya ke depan". Bagian *Output* berisi hasil keluaran berupa rencana aksi robot dalam format JSON. Rencana ini mencakup langkah-langkah yang harus dilakukan oleh robot berdasarkan instruksi yang diberikan dan kondisi lingkungan yang ada. Dengan menyusun dataset dalam format ini, proses fine-tuning dapat dilakukan dengan lebih terstruktur dan memungkinkan model untuk belajar dengan lebih efektif dari interaksi antara pengguna dan robot.

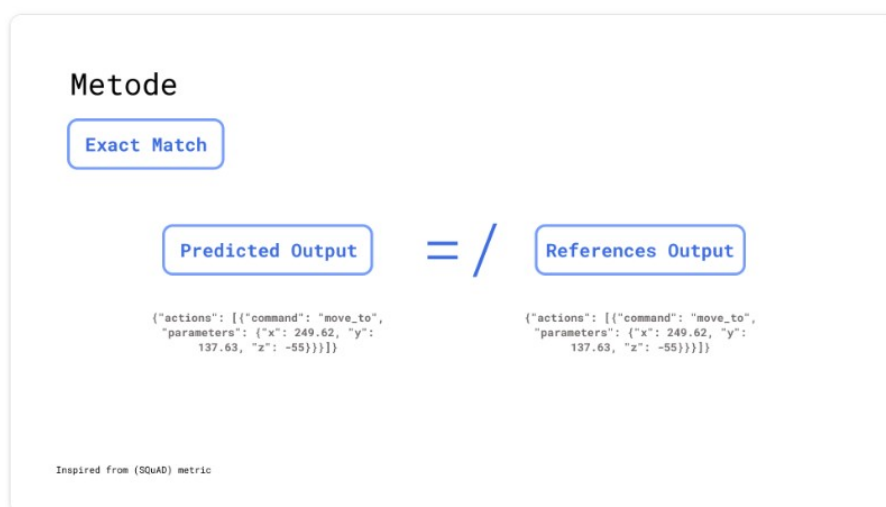
3.4 Fine Tuning



Gambar 3.5: Tuning

Proses *fine tuning* melibatkan penyesuaian model bahasa yang telah ada dengan dataset yang spesifik atau tugas yang diinginkan. Tujuan utamanya adalah untuk meningkatkan kemampuan model bahasa dalam menangani tugas atau lingkungan yang lebih spesifik dengan lebih baik. Dalam penelitian ini, *fine tuning* bertujuan agar model bahasa menghasilkan rencana aksi robot yang optimal berdasarkan instruksi pengguna dan parameter pendukung. Dengan ini, kemampuan robot untuk merespons dan berinteraksi dengan lingkungan sekitarnya dengan lebih efektif. Seiring dengan peningkatan kinerja model, harapannya adalah bahwa robot akan mampu melakukan tugas-tugas yang kompleks dengan lebih baik dan menghasilkan rencana aksi yang lebih tepat sesuai dengan kebutuhan dan preferensi pengguna.

3.5 Evaluasi Metrik



Gambar 3.6: Metrik *Exact Match*

Dalam penelitian ini, metrik *exact match* dipilih sebagai metrik evaluasi. Pendekatan ini menilai kinerja model dengan membandingkan secara langsung antara output prediksi model dengan label atau output yang seharusnya sesuai. Metrik ini dipilih didasarkan pada fakta bahwa output yang dihasilkan oleh model berbentuk JSON, sebuah format yang struktural dan tidak dalam bentuk bahasa alami. Oleh sebab itu, penggunaan metrik *exact match* menjadi relevan karena memungkinkan evaluasi yang langsung terhadap kesesuaian antara output model dengan label yang diharapkan dalam format JSON yang terstruktur. Dengan memastikan kesesuaian ini, kami dapat memastikan bahwa respons yang dihasilkan oleh model sesuai dengan aksi yang diinginkan untuk robot berdasarkan kondisi yang diberikan.

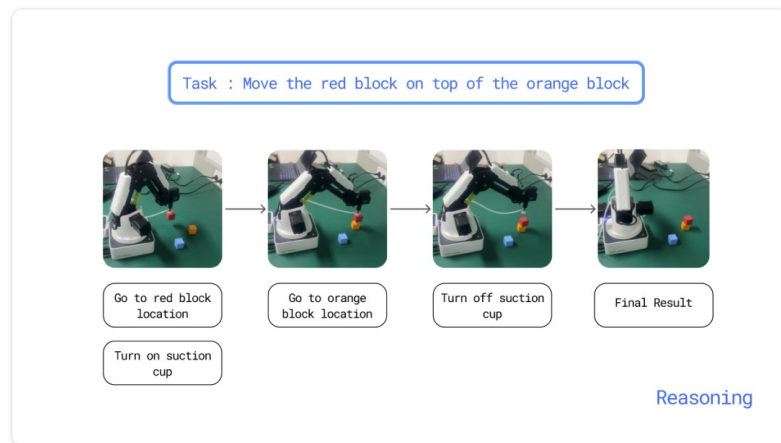
[Halaman ini sengaja dikosongkan]

BAB IV

PENGUJIAN DAN ANALISIS

4.1 Pengujian Model Bahasa

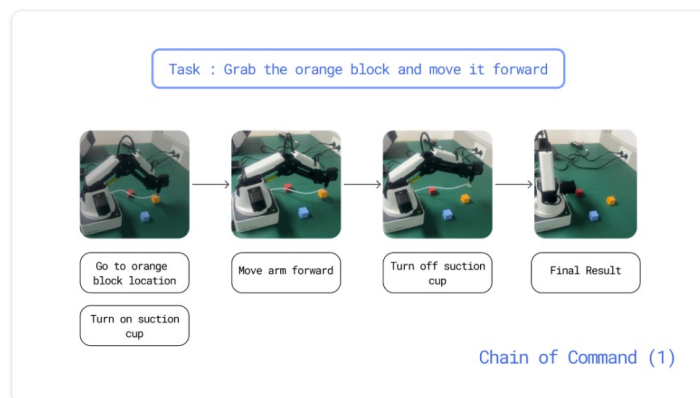
Berikut hasil dari pengujian model bahasa dalam menghasilkan aksi robot



Gambar 4.1: Uji Reasoning

```
{
  "actions": [
    {
      "command": "move_to",
      "parameters": {
        "x": 259.08,
        "y": 13.32,
        "z": -53.46
      }
    },
    {
      "command": "suction_cup",
      "parameters": {
        "action": "on"
      }
    },
    {
      "command": "move_to",
      "parameters": {
        "x": 249.62,
        "y": 137.63,
        "z": -53.46
      }
    },
    {
      "command": "suction_cup",
      "parameters": {
        "action": "off"
      }
    }
  ]
}
```

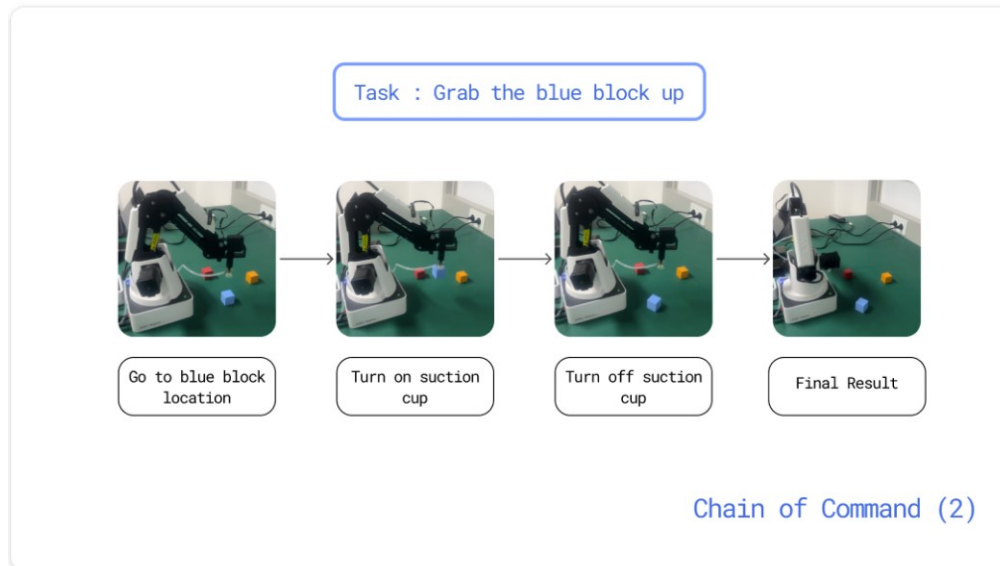
Gambar 4.2: Hasil JSON Reasoning



Gambar 4.3: Uji Chain of Command

```
{
  "actions": [
    {
      "command": "suction_cup",
      "parameters": {
        "action": "on"
      }
    },
    {
      "command": "move_to",
      "parameters": {
        "x": 249.62,
        "y": 137.63,
        "z": -53.46
      }
    },
    {
      "command": "move",
      "parameters": {
        "direction": "forward"
      }
    },
    {
      "command": "suction_cup",
      "parameters": {
        "action": "off"
      }
    }
  ]
}
```

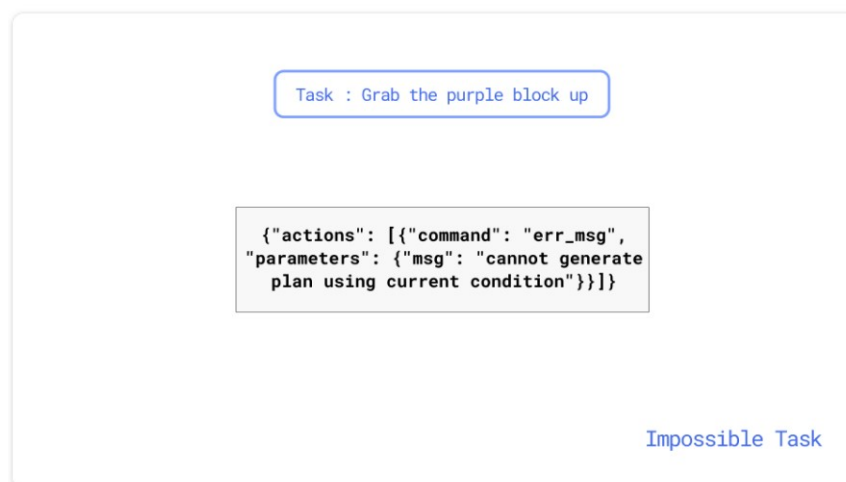
Gambar 4.4: Hasil JSON Chain of Command



Gambar 4.5: Uji Direct Command

```
{
  "actions": [
    {
      "command": "move_to",
      "parameters": {
        "x": 266.05,
        "y": 8.32,
        "z": -53.46
      }
    },
    {
      "command": "suction_cup",
      "parameters": {
        "action": "on"
      }
    }
  ]
}
```

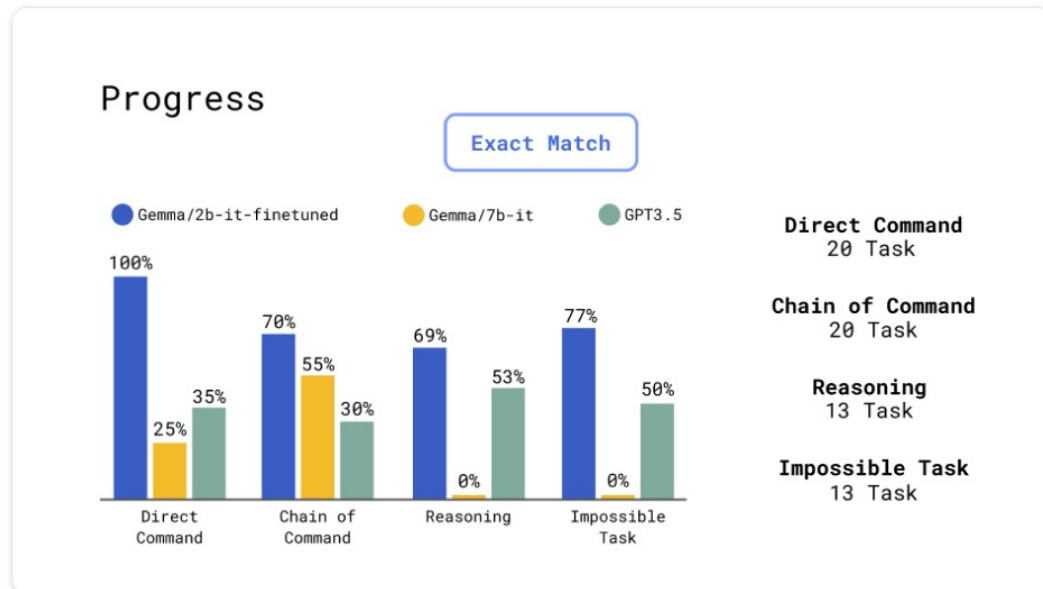
Gambar 4.6: Hasil JSON Direct Command



Gambar 4.7: Uji Impossible Task

4.2 Metrik Pengujian

Berikut hasil dari pengujian metrik dibandingkan dengan model Gemma-7b-it dan juga GPT3.5. Dari hasil tersebut model bahasa yang sudah di finetuned memiliki performa terbaik dibandingkan model yang tidak mengalami finetuned dengan jumlah parameter yang lebih besar. Hal ini dapat disebabkan karena kurangnya adaptasi model bahasa terhadap dataset untuk bagian menghasilkan rencana aksi robotika.



Gambar 4.8: Tuning

Direct command adalah tugas di mana LLM harus mengikuti instruksi yang diberikan secara langsung. Dalam tugas ini, model terbaik adalah model yang sudah di tuned yaitu mencapai 100 persen tingkat kecocokan. Hal ini menunjukkan bahwa semua model LLM mampu memahami dan mengikuti instruksi yang diberikan secara langsung dengan sempurna.

Chain of command adalah tugas di mana LLM harus mengikuti serangkaian instruksi yang diberikan secara berurutan. Dalam tugas ini, Gemma/2b-it-finetuned mencapai skor tertinggi, yaitu 70 persen. Hal ini menunjukkan bahwa model ini mampu memahami dan mengikuti serangkaian instruksi dengan baik. Model Gemma/7b-it dan GPT3.5 mencapai skor yang lebih rendah, yaitu 55 persen dan 30 persen berturut-turut. Hal ini menunjukkan bahwa model-model ini masih memiliki kesulitan dalam memahami dan mengikuti serangkaian instruksi dengan sempurna.

Reasoning adalah tugas di mana LLM harus menggunakan penalarannya untuk menyelesaikan suatu masalah. Dalam tugas ini, Gemma/2b-it-finetuned dan GPT3.5 mencapai skor tinggi, yaitu sebesar 69 persen dan 53 persen. Sedangkan Gemma/7b-it memiliki skor yang sangat rendah. Hal ini menunjukkan bahwa model yang telah tuning mampu menggunakan penalarannya dengan baik untuk menyelesaikan permasalahan yang langkahnya tidak secara eksplisit ditunjukkan oleh input user.

Impossible task adalah tugas yang tidak mungkin diselesaikan oleh LLM karena tidak memenuhi batasan lingkungan atau command. Dalam tugas ini, model Gemma/2b-it-finetuned mencapai skor terbaik yaitu 77 persen disusul oleh GPT3.5 yaitu sebesar 50 persen. Hal ini

menunjukkan bahwa model mampu memahami batasan lingkungan nyata dalam menghasilkan rencana aksi robotik.

DAFTAR PUSTAKA

- Braga, J., Regateiro, F., Stiubiener, I., & Braga, J. (2023). Governance of a dao for facilitating dialogue on human-algorithm interaction and the impact of emerging technologies on society. *International Journal of Advanced Robotic Systems*. <https://doi.org/10.31219/osf.io/wexjr>
- Chikwendu, O. C., Ezeanyim, O., & Igbokwe, N. C. (2023). Human-robot interaction enhancement through ergonomics and human factors: Future directions. *International Journal of Engineering Research and Development*, 19(6), 34–40.
- Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., ... Wei, J. (2022). Scaling instruction-finetuned language models.
- Dobot us. (2024). Retrieved October 5, 2024, from <https://www.dobot.us/product/dobot-magician-educational-version/>
- Elara, M. R., Rojas, N., & Chua, A. (2014). Design principles for robot inclusive spaces: A case study with roomba. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 5593–5599. <https://doi.org/10.1109/ICRA.2014.6907681>
- Gemma open model. (2024). Retrieved February 21, 2024, from <https://blog.google/technology/developers/gemma-open-models/>
- Gonzalez-Aguirre, J. A., Osorio-Oliveros, R., Rodríguez-Hernández, K. L., Lizárraga-Iturralde, J., Morales Menendez, R., Ramírez-Mendoza, R. A., Ramírez-Moreno, M. A., & Lozoya-Santos, J. d. J. (2021). Service robots: Trends and technology. *Applied Sciences*, 11(22), 10702.
- Islam, M. R., Rahman, M., Assad-Uz-Zaman, M., & Rahman, M. (2019). Cartesian trajectory based control of dobot robot.
- Jia, J., He, Y., & Le, H. (2020). A multimodal human-computer interaction system and its application in smart learning environments. *Blended Learning. Education in a Smart Learning Environment*, 12218, 3–14. https://doi.org/10.1007/978-3-030-51968-1_1
- Lee, I. (2021). Service robots: A systematic literature review. *Electronics*, 10(21). <https://doi.org/10.3390/electronics10212658>
- Liu, Y., He, H., Han, T., Zhang, X., Liu, M., Tian, J., Zhang, Y., Wang, J., Gao, X., Zhong, T., Pan, Y., Xu, S., Wu, Z., Liu, Z., Zhang, X., Zhang, S., Hu, X., Zhang, T., Qiang, N., ... Ge, B. (2024). Understanding llms: A comprehensive overview from training to inference.
- Mohebbi, A. (2020). Human-robot interaction in rehabilitation and assistance: A review. *Current Robotics Reports*, 1(3), 131–144.
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A comprehensive overview of large language models.
- Opencv [Accessed: 05 June 2024]. (2024).
- Singh, G., & Banga, V. (2022). Robots and its types for industrial applications. *Materials Today: Proceedings*, 60, 1779–1786.

- Vasconez, J. P., Kantor, G. A., & Auat Cheein, F. A. (2019). Human–robot interaction in agriculture: A survey and current challenges. *Biosystems Engineering*, 179, 35–48. <https://doi.org/https://doi.org/10.1016/j.biosystemseng.2018.12.005>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention is all you need.
- Wang, Z., Dong, W., Tu, K., Yang, Y., Xu, B., Duan, Y., Wei, F., Yang, Z., Jiang, L., Zhou, M., & et al. (2023). A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*.