

Implementasi AI Voice Assistant Berbasis ESP32 Pada Wearable Smart Cultural Mask Untuk Edukasi Budaya Indonesia

Galih Enggal Jati^{1*}, Naufal Danadyaksa², Ilham Nugroho³, Maulana Shultan Mubayyin⁴, Rudi Susanto⁵

¹Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

^{1*}250103074@mhs.udb.ac.id (penulis korespondensi)

²Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

²250103127@mhs.udb.ac.id

³Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

³250103122@mhs.udb.ac.id

⁴Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

⁴250103123@mhs.udb.ac.id

⁵Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

⁵rudi_susanto@udb.ac.id

Abstrak— Perkembangan teknologi digital memberikan peluang besar dalam mendukung pelestarian dan penyebaran informasi budaya Indonesia kepada masyarakat, khususnya generasi muda. Namun, media edukasi budaya yang masih banyak menggunakan metode konvensional menyebabkan proses penyampaian informasi kurang interaktif dan belum mampu mengikuti perkembangan teknologi saat ini. Penelitian ini bertujuan untuk mengembangkan wearable Smart Cultural Mask berbasis Internet of Things (IoT) yang mengintegrasikan teknologi Artificial Intelligence (AI) Voice Assistant sebagai media edukasi budaya Indonesia. Metode penelitian yang digunakan adalah Research and Development (R&D) dengan tahapan analisis kebutuhan, perancangan sistem, implementasi prototype, dan pengujian sistem. Smart Cultural Mask dikembangkan menggunakan ESP32 sebagai mikrokontroler utama, modul mikrofon digital INMP441 sebagai input suara, serta modul penguat audio MAX98357A dan speaker sebagai keluaran suara. Sistem bekerja dengan menerima perintah suara pengguna, kemudian data suara dikirim melalui koneksi WiFi menuju layanan AI Voice Assistant untuk diproses dan menghasilkan respons suara yang diteruskan kembali melalui speaker. Hasil implementasi menunjukkan bahwa prototype Smart Cultural Mask berhasil direalisasikan dan mampu menjalankan fungsi utama sistem, yaitu menerima input suara, melakukan komunikasi dengan layanan AI Voice Assistant, serta menghasilkan keluaran suara sebagai respons kepada pengguna. Berdasarkan pengujian Black Box Testing, seluruh fungsi sistem berhasil berjalan sesuai dengan rancangan, dengan waktu respons utama pada proses komunikasi AI sekitar 10 detik yang dipengaruhi oleh kondisi koneksi internet. Penelitian ini berfokus pada implementasi dan pengujian fungsional sistem, sehingga evaluasi terhadap performa AI Voice Assistant maupun efektivitas Smart Cultural Mask sebagai media edukasi budaya belum dilakukan dan menjadi ruang lingkup penelitian selanjutnya.

Kata kunci— Artificial Intelligence, ESP32, Internet of Things, Voice Assistant, Budaya Indonesia.

Abstract— The development of digital technology provides significant opportunities to support the preservation and dissemination of Indonesian cultural information to society, especially among younger generations. However, conventional cultural education media often results in less interactive learning experiences and has not fully adapted to current technological developments. This study aims to develop a wearable Smart Cultural Mask based on the Internet of Things (IoT) integrated with Artificial Intelligence (AI) Voice Assistant technology as an interactive Indonesian cultural education medium. The research method used is Research and Development (R&D), consisting of requirement analysis, system design, prototype implementation, and system testing stages. The Smart Cultural Mask was developed using ESP32 as the main microcontroller, an INMP441 digital microphone module as the voice input device, and a MAX98357A audio amplifier module with a speaker as the audio output device. The system receives voice commands from users, transmits voice data through a WiFi connection to the AI Voice Assistant service for processing, and produces voice responses through the speaker. The implementation results indicate that the Smart Cultural Mask prototype was successfully developed and was able to perform its main functions, including receiving voice input, communicating with the AI Voice Assistant service, and delivering voice responses through the speaker. Based on Black Box Testing, all system functions operated according to the design, with an average response time of approximately 10 seconds during AI communication, depending on internet connectivity. This study focuses on system implementation and functional testing; therefore, the performance of the AI Voice Assistant and the effectiveness of the Smart Cultural Mask as a cultural education medium have not yet been evaluated and remain opportunities for future research.

Keywords— Artificial Intelligence, ESP32, Internet of Things, Voice Assistant, Indonesian Culture.

I. PENDAHULUAN

Indonesia merupakan negara dengan kekayaan budaya yang sangat beragam, mulai dari seni pertunjukan, pakaian tradisional, bahasa daerah, hingga berbagai bentuk warisan budaya benda dan tak benda. Keberagaman tersebut menjadi identitas nasional yang memiliki nilai sejarah, sosial, dan filosofis bagi masyarakat Indonesia [1]. Perkembangan teknologi informasi dan perubahan pola konsumsi informasi masyarakat menyebabkan tantangan baru dalam proses pengenalan dan pelestarian budaya, terutama dalam menarik minat generasi muda terhadap budaya lokal [2].

Pelestarian budaya pada era digital tidak hanya dilakukan melalui dokumentasi, tetapi juga melalui pengembangan media penyampaian informasi yang lebih adaptif terhadap perkembangan teknologi. Pemanfaatan teknologi informasi dalam digitalisasi budaya dapat memperluas akses masyarakat terhadap informasi budaya serta membantu proses edukasi melalui media digital yang lebih interaktif. Penggunaan teknologi dalam pembelajaran juga mampu meningkatkan keterlibatan pengguna karena memberikan pengalaman belajar yang lebih menarik dibandingkan metode konvensional [3].

Salah satu teknologi yang memiliki potensi besar dalam pengembangan media edukasi interaktif adalah Internet of Things (IoT). IoT memungkinkan perangkat fisik melakukan pengumpulan data, pemrosesan informasi, dan komunikasi melalui jaringan internet sehingga dapat menghasilkan sistem yang lebih cerdas dan terhubung [4]. Dalam pengembangan perangkat IoT, mikrokontroler ESP32 banyak digunakan karena memiliki kemampuan pemrosesan yang baik, dukungan konektivitas WiFi dan Bluetooth, serta konsumsi daya yang relatif rendah sehingga sesuai untuk implementasi perangkat portabel dan *embedded system* [5].

Perkembangan IoT semakin berkembang dengan adanya integrasi teknologi Artificial Intelligence (AI). Salah satu penerapan AI yang banyak digunakan adalah teknologi Voice Assistant yang menggabungkan metode

pengenalan suara (*speech recognition*) dan pemrosesan bahasa alami (*Natural Language Processing/NLP*) untuk memungkinkan komunikasi antara manusia dan perangkat secara lebih alami. Teknologi tersebut telah diterapkan pada berbagai bidang seperti otomasi rumah, layanan informasi, dan sistem pembelajaran berbasis interaksi suara [6].

Konsep wearable technology menjadi salah satu bidang yang berkembang dalam pembuatan perangkat pintar. Wearable device memungkinkan teknologi elektronik digunakan secara langsung oleh manusia sehingga memberikan pengalaman interaksi yang lebih personal dan praktis. Implementasi wearable technology pada bidang edukasi dapat memberikan pendekatan baru dalam penyampaian informasi karena perangkat dapat digunakan sebagai media pembelajaran yang lebih dekat dengan aktivitas pengguna [7].

Penggabungan konsep wearable technology dengan unsur budaya Indonesia memberikan peluang untuk menciptakan inovasi baru dalam bidang edukasi budaya. Topeng tradisional yang selama ini dikenal sebagai bagian dari seni pertunjukan dan simbol budaya dapat dikembangkan menjadi perangkat interaktif dengan menambahkan kemampuan komunikasi berbasis AI. Dengan pendekatan tersebut, topeng tidak hanya berfungsi sebagai objek budaya, tetapi juga sebagai media penyampaian informasi yang mampu memberikan pengalaman belajar secara langsung.

Beberapa penelitian sebelumnya telah mengembangkan sistem berbasis ESP32 untuk pengolahan suara dan perangkat IoT interaktif. Penelitian terkait sistem pengenalan perintah suara berbasis ESP32 menunjukkan bahwa perangkat tersebut mampu digunakan sebagai pusat kendali dalam sistem berbasis suara dengan komunikasi jaringan. Penelitian lain juga menunjukkan bahwa integrasi AI pada perangkat interaktif dapat meningkatkan kemampuan sistem dalam memahami masukan pengguna dan memberikan respons secara otomatis. Namun, sebagian besar penelitian tersebut masih berfokus pada otomasi perangkat dan belum banyak mengarah pada

pengembangan wearable device berbasis budaya Indonesia [8].

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan Implementasi AI Voice Assistant Berbasis ESP32 Pada Wearable Smart Cultural Mask Untuk Edukasi Budaya Indonesia. Sistem yang dirancang menggabungkan teknologi IoT, AI Voice Assistant, dan konsep wearable device dalam bentuk topeng budaya interaktif. Penelitian ini bertujuan untuk menghasilkan rancangan perangkat yang mampu memberikan informasi budaya Indonesia melalui interaksi suara sehingga dapat menjadi alternatif media edukasi budaya yang lebih inovatif.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode Research and Development (R&D) yang bertujuan untuk mengembangkan rancangan prototype Smart Cultural Mask berbasis ESP32 dengan integrasi AI Voice Assistant sebagai media edukasi budaya Indonesia. Metode R&D dipilih karena penelitian ini berfokus pada proses pengembangan suatu produk teknologi melalui beberapa tahapan, yaitu analisis kebutuhan, perancangan sistem, implementasi prototype, dan evaluasi terhadap sistem yang dikembangkan [9]. Tahapan dalam metode ini meliputi :

A. Analisis Kebutuhan Sistem

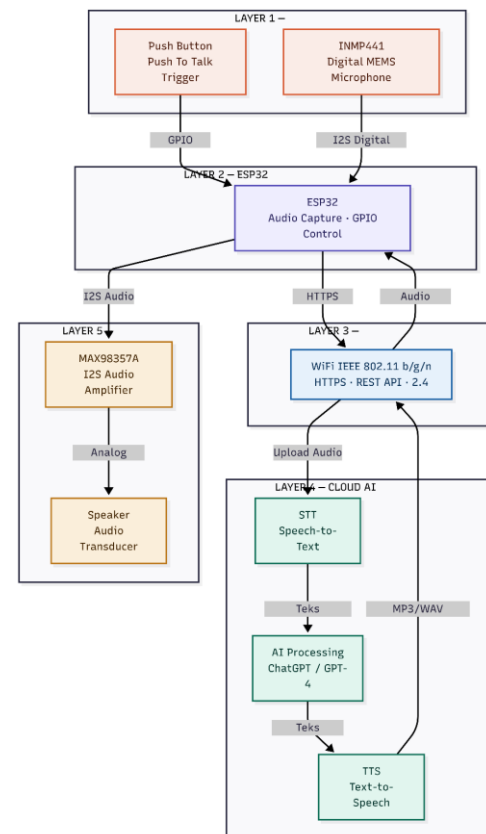
Analisis kebutuhan dilakukan untuk mengidentifikasi komponen hardware dan software yang diperlukan dalam mewujudkan Smart Cultural Mask. Kebutuhan utama sistem meliputi kemampuan akuisisi suara berkualitas tinggi, konektivitas internet yang stabil, mekanisme pengiriman data audio ke server AI melalui protokol HTTPS, serta kemampuan memutar respons audio yang diterima dari server.

B. Perancangan Arsitektur Sistem

Tahap Arsitektur Smart Cultural Mask dirancang dalam lima lapisan yang bekerja secara berurutan. Lapisan pertama adalah lapisan input, terdiri dari tombol PTT dan

mikrofon INMP441. Lapisan kedua adalah unit pemroses ESP32. Lapisan ketiga adalah antarmuka jaringan WiFi IEEE 802.11 b/g/n dengan protokol HTTPS. Lapisan keempat adalah Cloud AI yang menangani pipeline STT-LLM-TTS. Lapisan kelima adalah output audio melalui amplifier MAX98357A dan speaker 3W [10].

1. Diagram Blok Sistem



Gambar 1. Diagram Blok Sistem Smart Cultural Mask

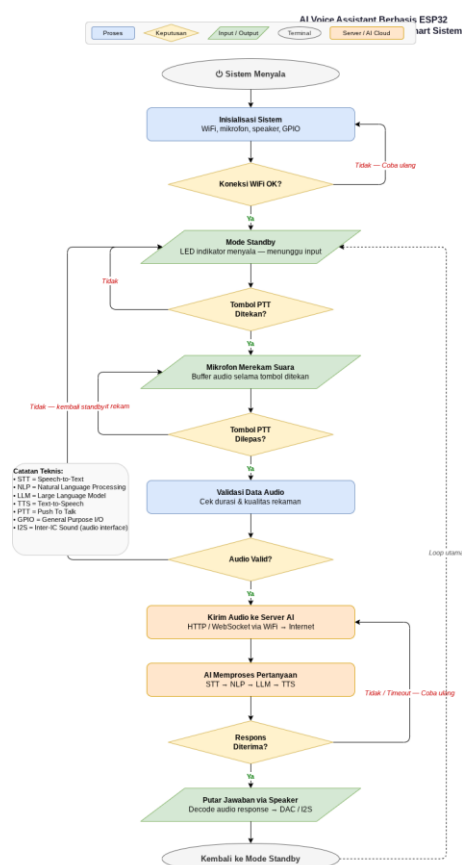
Gambar 1 menunjukkan rancangan arsitektur Smart Cultural Mask, sistem diawali dengan proses penerimaan masukan suara dari pengguna melalui tombol aktivasi (*push button*) dan modul mikrofon digital INMP441. Data suara yang diperoleh kemudian diproses oleh ESP32 sebagai pusat kendali sistem.

ESP32 berperan dalam melakukan pengolahan awal data serta mengatur komunikasi dengan layanan *cloud AI* melalui koneksi WiFi menggunakan protokol HTTPS. Data suara yang dikirimkan akan diproses melalui sistem *Speech-to-Text* (STT) untuk

mengubah suara menjadi teks, kemudian diproses oleh AI untuk menghasilkan respons yang sesuai.

Hasil pemrosesan AI selanjutnya dikonversi kembali menjadi suara melalui proses *Text-to-Speech* (TTS). Data audio yang dihasilkan kemudian dikirimkan kembali menuju ESP32 dan diteruskan melalui modul penguat audio MAX98357A untuk memperkuat sinyal sebelum menghasilkan keluaran suara melalui speaker.

2. Flowcart Sistem



Gambar 2. Flowcart Sistem

Gambar 2 menunjukkan alur kerja sistem Smart Cultural Mask mulai dari proses inialisasi perangkat hingga menghasilkan keluaran suara kepada pengguna. Proses diawali ketika sistem dinyalakan, kemudian ESP32 melakukan inialisasi terhadap komponen yang digunakan, seperti koneksi WiFi, modul mikrofon, speaker, dan konfigurasi GPIO.

Setelah proses inialisasi berhasil, sistem melakukan pengecekan koneksi jaringan dan

masuk ke mode *standby* untuk menunggu masukan dari pengguna. Ketika tombol PTT (*Push To Talk*) ditekan, modul mikrofon mulai melakukan perekaman suara. Data audio yang diperoleh kemudian divalidasi untuk memastikan kualitas rekaman sebelum dikirimkan menuju server AI melalui koneksi internet.

Pada sisi AI, data suara diproses melalui tahapan *Speech-to-Text* (STT), pemrosesan bahasa alami (*Natural Language Processing/NLP*), dan *Text-to-Speech* (TTS) untuk menghasilkan respons suara. Respons tersebut kemudian dikirim kembali ke perangkat dan diputar melalui speaker melalui proses konversi audio menggunakan antarmuka I2S.

Setelah respons berhasil diputar, sistem kembali ke mode *standby* untuk menunggu perintah berikutnya. Apabila terjadi kegagalan koneksi, data audio tidak valid, atau respons AI tidak diterima, sistem akan melakukan proses pengulangan sesuai alur yang telah dirancang.

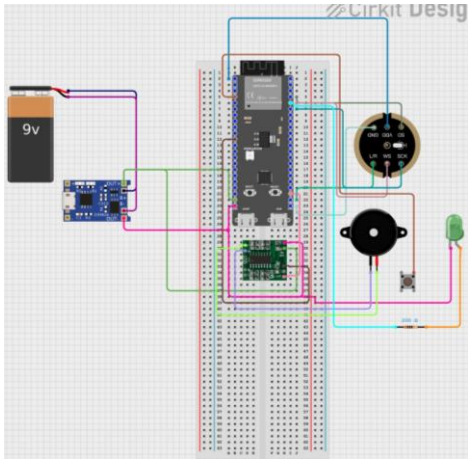
Berdasarkan alur sistem pada Gambar 4, implementasi program dilakukan untuk merealisasikan proses kerja perangkat sesuai rancangan [11]. Program pada ESP32 dibuat untuk mengatur komunikasi antar komponen, koneksi jaringan, pembacaan input, serta pengolahan keluaran audio.

3. Desain Pengkabelan

Tabel 1. Konfigurasi Pin Komponen Smart Cultural

No	Nama Komponen	Mask	Pin Modul	Pin ESP32
1.	ESP32		-	-
2.	INMP441	VDD GND SCK WS SD		3.3V GND GPIO 14 GPIO 15
3.	MAX98357A	LRC BCLK DIN VCC GND		GPIO 4 GPIO 5 GPIO 6 3.3 V GND
4.	Speaker 3v	(+) (-)		Output MAX98357A GND
5.	Baterai	(+) (-)		B+ TP4056 B- TP4056
6.	Push Button	Output GND		
7.	TP4056 Charging Module	OUT + OUT -		Input Power Sistem GND
8.	Saklar On/Off	Input		Jalur Power

Konfigurasi tersebut memungkinkan komunikasi antara modul input suara, sistem pengolahan, dan perangkat keluaran audio. Rangkaian pengkabelan direalisasikan berdasarkan konfigurasi pin yang telah ditentukan.



Gambar 3. Desain Pengkabelan Smart Cultural Mask

Gambar 3 menunjukkan rancangan pengkabelan komponen pada system Smart Cultural Mask. Rangkaian tersebut menggambarkan hubungan antara ESP32 sebagai pusat kendali dengan komponen pendukung lainnya, seperti modul mikrofon INMP441, modul audio MAX98357A, sumber daya, serta komponen pendukung sistem.

Konfigurasi pengkabelan dilakukan berdasarkan kebutuhan komunikasi antar perangkat agar proses penerimaan suara, pengolahan data, dan keluaran audio dapat berjalan sesuai dengan rancangan sistem.

C. Spesifikasi Perangkat Keras

Komponen Perangkat keras yang digunakan dalam pengembangan Smart Cultural Mask dipilih berdasarkan kebutuhan sistem serta kemampuan integrasi antar komponen. Komponen utama perangkat keras yang digunakan ditunjukkan pada Tabel 2.

Tabel 2. Komponen Perangkat Keras

No	Nama Komponen	Fungsi
1.	ESP32	Mikrokontroler utama & Wifi
2.	INMP441	Input suara pengguna
3.	MAX98357A	Penguat suara speaker
4.	Speaker 3v	Output suara AI
5.	Baterai	Sumber daya portable
6.	Push Button	Tombol push-to-talk
7.	TP4056 Charging Module	Charging module
8.	Topeng Budaya	Media Wearable
9.	Saklar On/Off	Menghidupkan/Mematikan alat

Pemilihan ESP32 dilakukan karena memiliki kemampuan pemrosesan, konektivitas WiFi dan Bluetooth, serta mendukung pengembangan perangkat berbasis IoT [5]. Modul mikrofon INMP441 digunakan karena mampu menghasilkan data suara digital yang sesuai untuk proses pengolahan suara. Komponen audio berupa PAM8403 dan speaker digunakan untuk mendukung sistem keluaran suara agar informasi dapat diterima oleh pengguna.

D. Perancangan Perangkat Lunak

Perancangan perangkat lunak dilakukan untuk mengatur komunikasi antara ESP32-S3, perangkat masukan suara, dan layanan AI Voice Assistant berbasis cloud. Perangkat lunak dikembangkan menggunakan Visual Studio Code dengan ekstensi PlatformIO IDE menggunakan bahasa pemrograman C/C++. Sistem dirancang untuk menjelaskan fungsi akuisisi suara, encoding data audio, transmisi data melalui protocol HTTPS, serta pemutaran respons audio yang diterima dari server.

1. Arsitektur Pipeline AI Voice Assistant

Sistem mengimplementasikan pipeline pemrosesan tiga tahap yang berjalan secara sekuensial, meliputi tahap Speech-to-Text (STT), inferensi Large Language Model (LLM), dan Text-to-Speech (TTS). Seluruh proses komputasi dijalankan pada infrastruktur cloud, sementara ESP32-S3 berfungsi sebagai edge device yang bertugas melakukan akuisisi audio, encoding, transmisi HTTPS, dan pemutaran audio balasan

2. Tahap STT dan LLM: Google Gemini 2.5 Flash

Pada penelitian ini, tahap STT dan LLM tidak diimplementasikan sebagai dua layanan terpisah, melainkan menggunakan satu model multimodal tunggal yaitu Google Gemini 2.5 Flash [13]. Model tersebut mampu memproses masukan audio sekaligus menghasilkan respons teks dalam satu API call, sehingga mengeliminasi latensi tambahan yang timbul apabila tahap transkripsi dan inferensi bahasa dijalankan secara terpisah.

Data suara direkam oleh mikrofon INMP441 dengan spesifikasi mono, frekuensi sampling 16.000 Hz, resolusi 16-bit PCM, dan durasi perekaman tiga detik. Data tersebut dikemas dalam format WAV kemudian dikonversi ke representasi Base64 sebelum dikirimkan ke endpoint Google Generative Language API v1beta melalui protokol HTTPS POST [13]. Ukuran payload yang dikirimkan berkisar ± 128 KB. Respon yang diterima berupa teks jawaban dalam bahasa Indonesia yang diekstrak dari field "candidates[0].content.parts[0].text" pada struktur JSON balasan.

Spesifikasi teknis tahap ini dirangkum sebagai berikut:

- *Model*: gemini-2.5-flash (Google Generative Language API v1beta) [13]
 - *Endpoint*:
generativelanguage.googleapis.com/v1beta/models/gemini-2.5-flash:generateContent
 - *Format masukan*: WAV Base64, mono, 16.000Hz, 16-bit PCM
 - *Format keluaran*: Teks JSON (bahasa Indonesia)
 - *Protokol Komunikasi*: HTTPS REST API, timeout 60 detik
3. Tahap TTS: ElevenLabs eleven_multilingual_v2

Teks Jawaban yang diperoleh dari tahapan sebelumnya dikirimkan ke layanan Text-to-Speech berbasis jaringan saraf tiruan dari ElevenLabs untuk dikonversi

menjadi audio sintetis berbahasa Indonesia [14].

Format keluaran yang diminta adalah pcm_24000, yaitu raw PCM S16LE (*Signed 16-bit Little Endian*), mono, frekuensi sampling 24.000 Hz, tanpa header container. Format ini dipilih karena data dapat langsung diputar melalui antarmuka I2S pada ESP32-S3 tanpa melakukan proses decoding tambahan di sisi perangkat, sehingga mengurangi beban komputasi edge device. Ukuran data audio yang diterima berkisar antara 160.000 hingga 280.000 byte, setara dengan durasi audio ± 3 hingga 6 detik.

Spesifikasi teknis tahap ini dirangkum sebagai berikut:

- *Model*: eleven_multilingual_v2 (ElevenLabs TTS API v1)[14]
 - *Endpoint*:
api.elevenlabs.io/v1/text-to-speech/{voice_id}
 - *Format masukan*: Teks JSON (bahasa Indonesia)
 - *Format keluaran*: PCM S16LE, mono, 24.000 Hz
 - *Protokol komunikasi*: HTTPS REST API, timeout 30 detik
4. Tahap Pemutaran Audio:ESP32-S3 dan MAX98357A

Data PCM yang diterima dari ElevenLabs disimpan sementara pada PSRAM ESP32-S3 kemudian dialirkan secara bertahap (*streaming*) melalui antarmuka I2S (I2S_NUM_1) menuju amplifier Class-D MAX98357A. Karena data PCM dari ElevenLabs bersifat mono, setiap sampel diduplikasi ke kanal kiri dan kanan sebelum ditulis ke buffer I2S yang dikonfigurasi dalam mode stereo pada frekuensi sampling 24.000 Hz, resolusi 16-bit.

E. Pengujian Sistem

Tahap akhir penelitian dilakukan dengan pengujian fungsionalitas sistem untuk mengetahui keberhasilan prototype Smart Cultural Mask dalam menjalankan fungsi yang telah dirancang. Pengujian dilakukan dengan

memberikan masukan pada sistem dan mengamati keluaran yang dihasilkan [12].

Parameter pengujian meliputi aktivasi perangkat, koneksi ESP32 dengan jaringan, penerimaan input suara, komunikasi dengan AI Voice Assistant, keluaran audio, serta waktu respons sistem. Hasil pengujian digunakan untuk mengevaluasi kesesuaian kinerja sistem dengan rancangan yang telah dibuat.

III. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

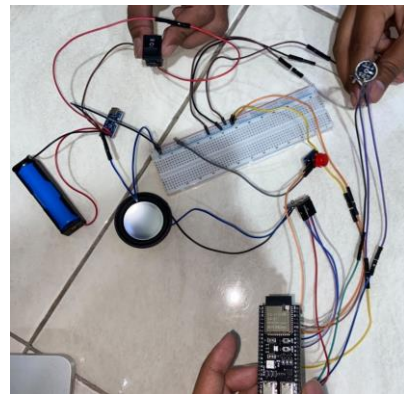
Hasil implementasi berupa prototype Smart Cultural Mask berbasis ESP32 dengan integrasi AI Voice Assistant sebagai media edukasi budaya Indonesia. Sistem ini menggabungkan konsep wearable technology, Internet of Things (IoT), dan kecerdasan buatan untuk memberikan pengalaman interaksi budaya berbasis suara.

Sistem yang dikembangkan bekerja dengan menerima masukan suara dari pengguna melalui modul mikrofon, kemudian data suara diproses oleh ESP32 dan dikomunikasikan dengan AI Voice Assistant melalui jaringan internet. Respons yang dihasilkan oleh AI selanjutnya diteruskan melalui speaker sebagai keluaran suara kepada pengguna.

B. Implementasi Perangkat Keras

Implementasi perangkat keras dilakukan berdasarkan rancangan sistem yang telah dibuat pada tahap metodologi penelitian. Pada tahap ini seluruh komponen elektronik direalisasikan menjadi sebuah prototype Smart Cultural Mask yang dapat digunakan sebagai perangkat wearable berbasis IoT dan AI Voice Assistant. Komponen utama yang digunakan terdiri dari ESP32 sebagai pusat kendali sistem, modul mikrofon digital INMP441 sebagai input suara, modul MAX98357A sebagai pengolah dan penguat sinyal audio, speaker sebagai keluaran suara, serta baterai dan modul charging sebagai sumber daya perangkat.

1. Realisasi Rangkaian Elektronik



Gambar 4. Realisasi Rangkaian Pengkabelan

Gambar 4 menunjukkan realisasi sebelum dilakukan pemasangan pada media topeng, seluruh komponen elektronik terlebih dahulu dirangkai dan dilakukan pengujian awal untuk memastikan setiap perangkat dapat bekerja sesuai fungsinya. Proses perakitan dilakukan dengan menghubungkan ESP32 dengan modul input suara, modul audio, sistem catu daya, serta perangkat pendukung lainnya sesuai konfigurasi yang telah ditentukan.

Hasil perakitan menunjukkan bahwa komunikasi antar komponen dapat berjalan dengan baik. ESP32 mampu menerima data suara dari modul mikrofon dan meneruskan data untuk diproses oleh sistem. Selain itu, jalur keluaran audio melalui modul MAX98357A dan speaker juga berhasil bekerja sesuai dengan kebutuhan sistem.

2. Integrasi Komponen Pada Topeng



Gambar 5. Integrasi Perangkat Keras

Gambar 5 menunjukkan hasil integrasi seluruh komponen elektronik pada media topeng budaya. Proses pemasangan dilakukan setelah rangkaian elektronik berhasil diuji untuk memastikan perangkat dapat bekerja dengan baik ketika digunakan sebagai wearable device.

Penempatan komponen dilakukan dengan mempertimbangkan fungsi dan kenyamanan penggunaan. ESP32 berperan sebagai pusat kendali sistem, modul mikrofon digunakan sebagai input suara pengguna, sedangkan speaker digunakan sebagai keluaran respons suara dari AI Voice Assistant.

3. Implementasi Smart Cultural Mask sebagai Wearable Device



Gambar 6. Penggunaan Smart Cultural Mask

Gambar 6 menunjukkan hasil akhir prototype Smart Cultural Mask dari sisi depan dan belakang. Tampak depan memperlihatkan bentuk desain topeng budaya yang digunakan sebagai media wearable, sedangkan tampak belakang menunjukkan posisi pemasangan komponen elektronik seperti ESP32, modul mikrofon, speaker, baterai, dan rangkaian pendukung lainnya.

Hasil implementasi menunjukkan bahwa seluruh komponen telah berhasil diintegrasikan ke dalam media topeng sehingga dapat digunakan sebagai perangkat wearable.

Penempatan komponen dilakukan dengan mempertimbangkan fungsi sistem serta kenyamanan pengguna saat menggunakan Smart Cultural Mask.

C. Implementasi Perangkat Lunak

Pengembangan program dilakukan menggunakan Visual Studio Code dengan ekstensi PlatformIO IDE. Beberapa library digunakan untuk mendukung kebutuhan sistem, seperti komunikasi WiFi, konfigurasi komunikasi I2S, dan pengolahan data audio.

```

1 #include <WiFi.h>
2 #include <driver/i2s.h>
3
4 #define BUTTON_PIN 4
5
6 void setup() {
7   Serial.begin(115200);
8
9   pinMode(BUTTON_PIN, INPUT_PULLUP);
10
11   // Inisialisasi koneksi WiFi
12   WiFi.begin(ssid, password);
13
14   // Inisialisasi modul audio I2S
15   setupI2S();
16 }
17
18 void loop() {
19
20   if (digitalRead(BUTTON_PIN) == LOW) {
21
22     // Proses perekaman suara
23     recordAudio();
24
25     // Pengiriman data ke AI
26     sendToAI();
27
28     // Pemutaran respons suara
29     playResponse();
30   }
31 }

```

Gambar 7. Potongan Program Sistem

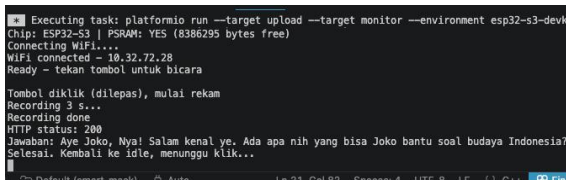
Potongan program pada Gambar 7 menunjukkan bagian utama kode yang digunakan untuk melakukan konfigurasi awal sistem, menghubungkan ESP32 dengan jaringan, serta mengatur komunikasi dengan perangkat audio. Program tersebut menjadi penghubung antara rancangan perangkat keras dengan proses kerja sistem yang telah ditunjukkan pada flowchart.

```

* Executing task: platformio run --target upload --target monitor
Leaving...
Hard resetting via RTS pin...
--- Terminal on /dev/cu.usbmodem14101 | 115200 8-N-1
--- Available filters and text transformations: debug, default, dir
file, nocontrol, printable, send_enter, time
--- More details at https://bit.ly/pio-monitor-filters
--- Quit: Ctrl+C | Menu: Ctrl+T | Help: Ctrl+T followed by Ctrl+H
Chip: ESP32-S3 | PSRAM: YES (8386295 bytes free)
Connecting WiFi....
WiFi connected - 10.32.72.28
Ready - tekan tombol untuk bicara

```

Gambar 8. Hasil Koneksi ESP32 dengan Jaringan WIFI



Gambar 9. Proses Perekaman Suara melalui Serial Monitor

Gambar 9 menunjukkan proses pengambilan input suara pada sistem Smart Cultural Mask. Setelah perangkat berhasil terhubung ke jaringan, sistem berada pada mode standby dan menunggu perintah pengguna. Ketika tombol ditekan, sistem mengaktifkan proses perekaman melalui modul mikrofon INMP441 selama durasi yang telah ditentukan. Setelah proses perekaman selesai, serial monitor menampilkan status *recording done* yang menunjukkan bahwa data suara berhasil diperoleh dan siap untuk diproses lebih lanjut.

D. Pengujian Fungsionalitas Sistem

Tahap akhir penelitian dilakukan melalui pengujian fungsionalitas sistem untuk mengetahui kesesuaian kinerja Smart Cultural Mask terhadap rancangan yang telah dibuat. Pengujian dilakukan dengan mengamati kerja setiap bagian sistem, mulai dari proses aktivasi perangkat hingga menghasilkan keluaran suara. Berikut adalah hasil pengujian fungsionalitas yang ditunjukkan pada tabel 3 berikut.

Tabel 3. Hasil Pengujian Fungsionalitas Sistem Smart Cultural Mask

No	Pengujian	Skenario pengujian	Hasil pengukuran	Hasil
1.	Aktivasi Sitem	Menghidupkan perangkat Smart Cultural Mask menggunakan saklar daya	Sistem berhasil melakukan inialisasi perangkat dalam 2 detik	Berhasil
2.	Koneksi Jaringan	Menghubungkan ESP32 dengan jaringan internet	ESP32 berhasil terhubung dengan jaringan Wifi dalam 2 detik	Berhasil
3.	Tombol Push To Talk	Menekan tombol untuk memulai proses interaksi suara	Tombol berhasil memberikan perintah aktivasi perekaman dalam 3 detik	Berhasil
4.	Input Suara	Memberikan perintah suara melalui modul mikrofon INMP441	Sistem berhasil menerima dan merekam suara pengguna dalam waktu 3 detik	Berhasil

5.	Pengiriman Data Suara	Mengirimkan data audio hasil rekaman menuju layanan AI Voice Assistant	Data suara berhasil dikirim dalam 5 detik, bergantung koneksi internet	Berhasil
6.	Pemrosesan AI Voice Assistant	Memberikan pertanyaan mengenai budaya Indonesia dan menunggu respon AI	Respon AI berhasil diterima dengan waktu respons 10 detik, bergantung koneksi internet	Berhasil
7.	Output Audio	Memutar hasil respons AI melalui speaker menggunakan modul MAX98357A	Suara respons berhasil keluar melalui speaker dalam 5 detik, bergantung kondisi internet	Berhasil
8.	Mode Standby Sistem	Memberikan perintah baru setelah proses interaksi selesai	Sistem secara otomatis kembali ke kondisi awal (<i>standby</i>) setelah proses pemrosesan selesai dan siap menerima perintah berikutnya	Berhasil

Berdasarkan hasil pengujian Black Box Testing, Smart Cultural Mask telah mampu menjalankan seluruh fungsi utama sesuai dengan rancangan sistem. Pengujian menunjukkan bahwa perangkat dapat melakukan aktivasi, koneksi WiFi, perekaman suara, pengiriman data menuju layanan AI Voice Assistant, menerima respons, serta menghasilkan keluaran suara melalui speaker. Namun, pengujian dalam penelitian ini masih terbatas pada aspek fungsional sistem sehingga belum mencakup evaluasi performa AI Voice Assistant, seperti tingkat keberhasilan pengenalan suara, akurasi jawaban terhadap pertanyaan budaya Indonesia, maupun efektivitas perangkat sebagai media edukasi budaya melalui pengujian kepada pengguna.

E. Keterbatasan Penelitian

Pengujian pada penelitian ini difokuskan pada aspek fungsional sistem menggunakan metode Black Box Testing untuk memastikan setiap fungsi utama Smart Cultural Mask berjalan sesuai dengan rancangan. Oleh karena itu, penelitian ini belum mengevaluasi performa AI Voice Assistant secara kuantitatif, seperti *Word Error Rate* (WER), tingkat keberhasilan pengenalan suara, tingkat keberhasilan pemahaman perintah, maupun akurasi jawaban terhadap pertanyaan budaya Indonesia.

Selain itu, penelitian ini belum melibatkan pengguna dalam proses evaluasi sehingga efektivitas Smart Cultural Mask sebagai media edukasi budaya belum dapat dibuktikan secara empiris. Dengan demikian, penelitian selanjutnya disarankan untuk melakukan evaluasi performa AI serta uji pengguna guna memperoleh bukti empiris mengenai kinerja sistem dan efektivitas perangkat sebagai media edukasi budaya Indonesia.

IV. KESIMPULAN

Berdasarkan Penelitian ini berhasil mengimplementasikan prototype Smart Cultural Mask berbasis ESP32 dengan integrasi AI Voice Assistant yang mampu menerima masukan suara, mengirimkan data melalui jaringan internet, memproses pertanyaan menggunakan layanan AI, serta menghasilkan respons suara melalui speaker. Berdasarkan hasil pengujian menggunakan metode Black Box Testing, seluruh fungsi utama sistem berhasil berjalan sesuai dengan rancangan, meliputi aktivasi perangkat, koneksi WiFi, perekaman suara, pengiriman data, penerimaan respons AI, keluaran audio, dan kemampuan sistem kembali ke kondisi *standby* setelah proses interaksi selesai.

Penelitian ini berfokus pada implementasi dan pengujian fungsional sistem sehingga hasil penelitian belum mencakup evaluasi terhadap performa AI Voice Assistant maupun efektivitas Smart Cultural Mask sebagai media edukasi budaya. Oleh karena itu, evaluasi terhadap kedua aspek tersebut menjadi rekomendasi untuk penelitian selanjutnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan dalam proses penelitian ini. Terima kasih disampaikan kepada dosen pembimbing

yang telah memberikan arahan, masukan, dan bimbingan selama proses pengembangan penelitian. Penulis juga mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam proses perancangan, implementasi, hingga pengujian sistem Smart Cultural Mask berbasis ESP32 dengan integrasi AI Voice Assistant. Dukungan dan kontribusi yang diberikan sangat membantu dalam penyelesaian penelitian ini..

REFERENSI

- [1] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, *Pokok Pikiran Kebudayaan Indonesia*. Jakarta, Indonesia: Kementerian Pendidikan dan Kebudayaan Republik Indonesia, 2018.
- [2] S. S. Bahri, "Digitalisasi kebudayaan sebagai upaya pelestarian budaya Indonesia di era revolusi industri 4.0," *Jurnal Sosial Humaniora*, vol. 12, no. 2, pp. 135–142, 2021.
- [3] R. W. Sari and M. F. Arif, "Pemanfaatan teknologi digital sebagai media pembelajaran interaktif dalam meningkatkan kualitas pendidikan," *Jurnal Teknologi Pendidikan*, vol. 9, no. 2, pp. 85–94, 2021.
- [4] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010, doi: 10.1016/j.comnet.2010.05.010.
- [5] Espressif Systems, *ESP32 Series Datasheet*. Shanghai, China: Espressif Systems, 2023.
- [6] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed. Stanford, CA, USA: Stanford University, 2023.
- [7] S. Young, M. Gasic, B. Thomson, and J. D. Williams, "POMDP-based statistical spoken dialogue systems: A review," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013, doi: 10.1109/JPROC.2012.2225812.
- [8] S. Mann, "Wearable computing: A first step toward personal imaging," *Computer*, vol. 30, no. 2, pp. 25–32, 1997, doi: 10.1109/2.566147.
- [9] Sugiyono, *Metode Penelitian dan Pengembangan (Research and Development/R&D)*. Bandung, Indonesia: Alfabeta, 2019.
- [10] I. Sommerville, *Software Engineering*, 10th ed. Boston, MA, USA: Pearson, 2015.
- [11] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 9th ed. New York, NY, USA: McGraw-Hill Education, 2019.
- [12] M. A. Khan and K. Salah, "IoT security: Review, blockchain solutions, and open challenges," *Future Generation Computer Systems*, vol. 82, pp. 395–411, 2018, doi: 10.1016/j.future.2017.11.022.
- [13] Google, "Gemini API Documentation - Generate Content," Google AI for Developers, 2024. [Online]. Available: <https://ai.google.dev/gemini-api/docs/text-generation>
- [14] ElevenLabs, "Text to Speech - API Reference," ElevenLabs Documentation, 2024. [Online]. Available: <https://elevenlabs.io/docs/api-reference/text-to-speech>