

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Mobilitas dan kesadaran spasial merupakan aspek krusial bagi kemandirian individu. Bagi penyandang tunanetra, tantangan dalam mempersepsi lingkungan sekitar secara *real-time* menjadi hambatan signifikan dalam aktivitas sehari-hari. Keterbatasan *visual* membuat navigasi di lingkungan yang dinamis dan asing menjadi sulit, seringkali menimbulkan resiko keselamatan dan mengurangi partisipasi sosial [1].

Selama ini, alat bantu konvensional seperti tongkat putih (*white cane*) dan anjing pemandu (*guide dogs*) menjadi andalan utama. Tongkat putih efektif untuk mendeteksi rintangan di permukaan tanah pada jarak dekat, namun memiliki keterbatasan signifikan dalam mendeteksi objek yang lebih tinggi (seperti dahan pohon atau papan nama), rintangan yang bergerak, atau bahaya seperti genangan air. Sementara itu, anjing pemandu, meskipun sangat canggih, memerlukan biaya perawatan dan pelatihan yang sangat tinggi, serta ketersediaannya sangat terbatas dan tidak dapat diakses oleh semua individu [2].

Perkembangan pesat dalam teknologi Kecerdasan Buatan (AI), khususnya di bidang *Computer Vision*, telah membuka jalan bagi pengembangan teknologi asistif yang lebih canggih. Berbagai penelitian telah berfokus pada sistem pengenalan objek (*object recognition*) untuk membantu tunanetra. Namun, kebanyakan sistem generasi awal seringkali terbatas pada pelabelan objek secara individual (misalnya, "kursi", "meja", "orang") tanpa memberikan pemahaman kontekstual yang mendalam. Sistem ini mungkin gagal menginterpretasikan situasi yang kompleks, seperti "sebuah kursi menghalangi jalan" atau "seseorang sedang berjalan cepat ke arah Anda" [3].

Keterbatasan inilah yang coba diatasi oleh kemunculan *AI multimodal*. Model *AI* seperti *Gemini* dari *Google* memiliki kemampuan revolusioner untuk memproses dan memahami informasi dari berbagai modalitas seperti gambar, video, dan teks secara bersamaan. Kemampuan ini memungkinkan *AI* untuk tidak hanya "melihat" objek, tetapi juga "memahami" skenario, hubungan antar objek, dan konteks dari situasi yang sedang berlangsung. *AI multimodal* dapat memberikan deskripsi yang jauh lebih kaya dan lebih manusiawi terhadap suatu pemandangan [4].

Oleh karena itu, penelitian ini mengusulkan implementasi *AI multimodal Gemini* untuk membangun sebuah sistem pengenalan objek *real-time* yang dirancang khusus untuk tunanetra. Sistem ini bertujuan untuk menangkap data visual dari lingkungan sekitar (misalnya, melalui kamera smartphone), memprosesnya secara cepat menggunakan model *Gemini* untuk mendapatkan pemahaman situasional, dan menyampaikannya kepada pengguna melalui umpan balik audio yang intuitif dan deskriptif. Diharapkan, sistem ini dapat menjadi "mata" artifisial yang cerdas, memberikan deskripsi lingkungan yang komprehensif, dan secara signifikan meningkatkan kemandirian, keamanan, serta kualitas hidup penyandang tunanetra dalam bernavigasi.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang tersebut, maka peneliti dapat merumuskan masalah dalam penelitian ini yaitu:

1. Bagaimana mengimplementasikan *AI Multimodal Gemini* kedalam sistem pendeteksian objek *realtime*?
2. Bagaimana sistem dapat memberikan informasi kepada pengguna dalam bentuk perintah suara secara *real-time*?
3. Bagaimana performa sistem dalam mendeteksi objek dan kecepatan respon suara?

1.3 BATASAN MASALAH

Agar penelitian ini lebih terfokus dan terarah, maka batasan masalah dalam perancangan sistem ini ditentukan sebagai berikut:

1. Penelitian ini berfokus pada implementasi *AI Multimodal Gemini* dan tidak membahas perbandingan performa model *Gemini* dengan model deteksi objek lainnya seperti *YOLO*, *SSD*, atau model *multimodal* lain.
2. *Input* yang digunakan oleh sistem terbatas pada aliran video (*video stream*) secara *real-time* yang ditangkap oleh satu perangkat kamera.
3. Sesuai dengan rumusan masalah (RM 2), *output* informasi kepada pengguna hanya terbatas pada umpan balik suara (*audio*).
4. Pengujian performa sistem akan difokuskan pada dua parameter utama yaitu akurasi deskripsi objek atau situasi yang dihasilkan, dan latensi atau kecepatan respon.
5. Penelitian ini berfokus pada pengembangan dan pengujian fungsionalitas purwarupa (*prototype*).

1.4 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah untuk:

1. Mengimplementasikan *AI Multimodal Gemini* untuk membangun sebuah purwarupa sistem yang mampu mendeteksi objek dan menganalisis lingkungan sekitar secara *real-time*.
2. Membangun mekanisme integrasi pada sistem agar dapat menyajikan informasi hasil deteksi secara *real-time* kepada pengguna dalam bentuk umpan balik suara (*audio*).
3. Menguji dan menganalisis performa sistem untuk mengetahui tingkat keberhasilan (akurasi) pendeteksian objek serta kecepatan respon (latensi) dalam memberikan umpan balik suara.

1.5 MANFAAT PENELITIAN

Manfaat yang didapat dari penelitian ini antara lain:

1. Memberikan solusi teknologi untuk membantu penyandang tunanetra dalam mengenali lingkungan sekitar.
2. Memberikan solusi alat bantu baru berbasis suara yang dapat meningkatkan keamanan dan kemandirian pengguna.
3. Menjadi referensi dalam pengembangan perangkat berbasis *AI* untuk aplikasi asistif.

1.6 SISTEMATIKA PENULISAN

Laporan tugas akhir ini disusun dengan sistematika sebagai berikut :

BAB 1 : PENDAHULUAN

Bagian ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 : LANDASAN TEORI

Bagian ini berisi uraian teori-teori dan konsep fundamental yang mendukung penelitian meliputi Kecerdasan Buatan Multimodal (*Multimodal AI*), arsitektur dan kapabilitas model *Gemini*, pemrosesan video *real-time*, teknologi *TTS*.

BAB 3 : METODOLOGI PENELITIAN

Bagian ini menjelaskan metode penelitian yang digunakan, alur penelitian, alat dan bahan, serta prosedur pelaksanaan penelitian.

BAB 4 : ANALISIS DAN PERANCANGAN

Bagian ini berfokus pada tahap perancangan teknis sistem mencakup analisis kebutuhan fungsional dan non-fungsional sistem, serta perancangan arsitektur sistem yang menunjukkan alur data dari input kamera ke *API Gemini* dan *output audio*.

BAB 5 : IMPLEMENTASI DAN PENGUJIAN

Bagian ini menjelaskan hasil pengujian sistem secara rinci, terutama analisis performa terkait kecepatan respon (latensi) dan akurasi/relevansi deskripsi audio yang dihasilkan sistem.

BAB 6 : PENUTUP

Bagian ini berisi kesimpulan dan saran dari hasil penelitian.

