

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Perkembangan *Artificial Intelligence* (AI) membuka peluang pengembangan sistem interaktif berbasis *avatar*. Teknologi ini memungkinkan sistem memahami dan merespons percakapan secara natural, baik vokal maupun visual. Namun, interaksi berbasis teks atau suara saja belum mencukupi untuk komunikasi yang utuh, terutama dalam konteks yang memerlukan ekspresi wajah. Karena itu, dibutuhkan sebuah representasi visual berupa *avatar* yang mampu menyampaikan ekspresi wajah dan emosi pengguna secara *real-time*, guna menciptakan pengalaman komunikasi yang lebih manusiawi dan imersif. Salah satu implementasi AI yang semakin populer adalah *chatbot*, yang awalnya hanya berbasis teks atau suara, namun kini berkembang menjadi sistem yang mengintegrasikan elemen visual seperti *avatar* digital.

Penelitian menunjukkan, *avatar* dengan ekspresi emosi *real-time* meningkatkan kepuasan pengguna [1]. Penggunaan *avatar* dalam *chatbot* tidak hanya memberikan tampilan visual yang menarik, tetapi juga berperan penting dalam komunikasi non-verbal seperti ekspresi wajah, gerakan bibir, dan emosi. Hal ini membantu menciptakan pengalaman interaksi yang lebih natural dan meningkatkan keterlibatan pengguna secara emosional.

Pembangunan *lip-sync* berbasis *visemes* dan volume suara menghadirkan sejumlah tantangan teknis. *Visemes* sebagai representasi visual dari fonem atau unit suara, hanya memberikan visual bentuk mulut, tidak memberikan hasil yang tepat antara tingkat volume suara yang dihasilkan animasi gerakan bibir terbuka dan tertutup. Studi mengenai efek *uncanny valley* mengungkapkan bahwa tampilan yang terlalu realistis namun belum sempurna dapat menimbulkan rasa tidak nyaman pada pengguna, sehingga berdampak negatif terhadap tingkat kepercayaan dan adopsi sistem tersebut [2].

Penelitian dalam penggunaan algoritma *rule-based* untuk *lip-sync* telah menunjukkan potensi dalam menyinkronkan gerakan bibir dengan suara secara *real-time* pada *platform chatbot* digital [3]. Ketika ingin dimasukkan ke dalam lingkungan pengembangan yang realistis seperti *Unreal Engine 5.2*, muncul tantangan baru karena tampilan yang terlalu realistis dan gerakan yang belum

sempurna dapat menimbulkan efek *uncanny valley*.

Selain pendekatan *rule-based* yang sederhana, terdapat berbagai penelitian yang memanfaatkan sinyal audio ke gerakan bibir, sehingga dapat mengatasi keterbatasan sinkronisasi pada sistem berbasis aturan. Penelitian *VASA-Rig* mengembangkan model yang langsung menghasilkan parameter *rig* untuk animasi wajah 3D yang memetakan audio ke parameter ekspresi dalam *Unreal Engine MetaHuman* [4]. Pendekatan *viseme-parametrik* juga dikombinasikan dengan *neural network* untuk memprediksi kurva *viseme* secara dinamis, yang dapat membantu animator dalam menghasilkan animasi bibir yang sesuai dengan volume suara [5].

*ChatGPT* sebagai basis sistem *chatbot* karena kemampuannya menghasilkan respons yang lebih natural, kontekstual, serta mempertahankan alur percakapan secara humanis. Hasil evaluasi performa *ChatGPT*, *GPT-4*, *Claude*, dan *Bard* pada pertanyaan dari dalam *dataset Wordsmiths* menunjukkan bahwa *ChatGPT* secara konsisten menempati posisi kedua terbaik setelah *GPT-4*, mengungguli *Claude* dan *Bard* dalam semua kategori termasuk pemahaman bahasa, logika, dan etika. *ChatGPT* juga menunjukkan performa tinggi dalam kategori “*language understanding*” dan “*coding*”, dua aspek penting dalam pengembangan sistem interaktif berbasis *avatar*. Berdasarkan hasil tersebut, *ChatGPT* merupakan pilihan yang tepat karena menawarkan keseimbangan antara kemampuan teknis, fleksibilitas integrasi, serta kualitas interaksi yang tinggi, menjadikannya sangat sesuai untuk integrasi dalam sistem *chatbot* yang menuntut presisi dan responsif tinggi [6].

Berdasarkan hasil evaluasi komparatif terhadap sejumlah API *text-to-speech* (TTS), *Microsoft Azure* dipilih sebagai solusi utama dalam proyek ini karena menunjukkan kinerja yang unggul secara konsisten pada berbagai parameter. *Azure* tidak hanya menawarkan kualitas suara berbasis *neural network* yang tinggi dan natural, tetapi juga menyediakan fitur *viseme timestamp* secara eksplisit yang sangat krusial dalam proses sinkronisasi animasi bibir. *Azure* mendukung cakupan bahasa yang sangat luas serta dokumentasi teknis yang komprehensif, memudahkan pengembang dalam proses integrasi sistem. Dibandingkan dengan layanan seperti *Google Cloud TTS* dan *ElevenLabs*, *Microsoft Azure* menunjukkan keunggulan dalam kemampuan untuk menghasilkan *viseme* yang terstruktur untuk animasi *lip-sync* [7].

*Unreal Engine 5.2* merupakan salah satu *platform* pengembangan *game* dengan visual yang realistis, seperti *MetaHuman*, yang mendukung pembuatan

karakter digital dan sistem animasi dengan tingkat realisme tinggi [8]. Fitur-fitur tersebut memungkinkan pembuatan *avatar* dengan ekspresi wajah yang dinamis. Namun, meskipun *Unreal Engine 5.2* memberikan fondasi teknis yang kuat, integrasi modul *lip-sync* yang mengandalkan *visemes* dan volume suara yang *real-time* membutuhkan penelitian mendalam untuk memastikan bahwa *output* dari TTS dapat diterjemahkan secara akurat ke dalam animasi yang konsisten dan natural.

Penelitian ini menggunakan metode *System Usability Scale* (SUS) yang bertujuan untuk mendapatkan pemahaman awal yang objektif dan terstandarisasi mengenai persepsi pengguna terhadap kemudahan penggunaan dan kenyamanan interaksi dengan *avatar chatbot* yang dikembangkan. SUS adalah sebuah kuesioner yang terdiri dari 10 pertanyaan dengan 5 poin skala likert yang dirancang untuk memberikan penilaian cepat terhadap kegunaan suatu sistem. Metode ini digunakan secara luas dalam bidang interaksi manusia dan komputer karena kesederhanaannya, kecepatan dalam pelaksanaan, serta kemampuannya dalam memberikan gambaran umum mengenai kegunaan suatu sistem. Keunggulan SUS terletak pada kemampuannya untuk memberikan hasil evaluasi yang valid dengan jumlah responden yang relatif sedikit, serta dapat digunakan dalam berbagai konteks sistem, termasuk prototipe awal dan produk yang sudah matang [9].

Berdasarkan tantangan teknis dan psikologis yang telah diuraikan, penelitian ini bertujuan untuk merancang dan membangun sebuah sistem *avatar chatbot* yang dilengkapi dengan animasi *lip-sync* berbasis *volume-visemes* di dalam lingkungan *Unreal Engine 5.2*. Sistem ini dirancang agar mampu menyinkronkan gerakan bibir secara *real-time* berdasarkan informasi volume suara dan *visemes* dari *text-to-speech*, guna menghasilkan animasi wajah yang lebih realistis dan ekspresif. Dengan memanfaatkan teknologi seperti *MetaHuman* dan *Control Rig*, diharapkan hasil animasi yang dihasilkan dapat mengurangi efek *uncanny valley*, meningkatkan kepercayaan pengguna, serta memberikan pengalaman interaksi yang lebih natural. Penelitian ini diharapkan menjadi langkah awal dalam pengembangan *avatar chatbot* yang lebih canggih dan adaptif untuk berbagai aplikasi digital, seperti layanan pelanggan, edukasi interaktif, dan simulasi sosial di dunia virtual.

## 1.2 Rumusan Masalah

Dalam penelitian ini, beberapa permasalahan utama yang perlu dipecahkan adalah sebagai berikut:

1. Bagaimana cara merancang dan membangun *virtual avatar chatbot* dengan

*lip-sync* berbasis *visemes* dan volume untuk *platform* pengembangan *game* dan aplikasi *Unreal Engine 5.2* secara *real-time*?

2. Bagaimana cara mengukur tingkat penerimaan pengguna terhadap *virtual avatar chatbot* yang dikembangkan, menggunakan metode *System Usability Scale (SUS)*?

### 1.3 Batasan Permasalahan

Agar penelitian ini tetap fokus dan dapat diselesaikan dalam ruang lingkup yang jelas, maka batasan-batasan berikut diterapkan:

1. Bahasa percakapan yang digunakan dalam sistem ini dibatasi hanya pada bahasa Inggris.
2. *Avatar* yang digunakan dalam sistem ini dibatasi hanya satu jenis, yaitu *template MetaHuman* yang tersedia di *Unreal Engine 5.2*.
3. *Chatbot* yang digunakan dalam penelitian ini dibatasi dengan menggunakan layanan API *ChatGPT*.
4. Sistem TTS yang digunakan dalam penelitian ini dibatasi dari layanan *Microsoft Azure TTS* yang dapat menghasilkan data *visemes*.
5. Ekspresi wajah yang digunakan dibatasi hanya pada tiga jenis yaitu natural, senang, dan marah.
6. Bahasa non-verbal selain *lip-sync*, seperti animasi ekspresi wajah, gestur tangan, dan pergerakan kepala, tidak termasuk dalam lingkup penelitian.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan sebelumnya, berikut merupakan tujuan penelitian:

1. Merancang dan membangun *virtual avatar chatbot* dengan *lip-sync* berbasis *visemes* dan volume untuk *platform* pengembangan *game* dan aplikasi *Unreal Engine 5.2* secara *real-time*.
2. Mengetahui tingkat penerimaan pengguna terhadap *virtual avatar chatbot* menggunakan instrumen *System Usability Scale (SUS)*.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam berbagai aspek, baik dari segi akademik, teknis, maupun praktis. Berikut adalah beberapa manfaat yang dapat diperoleh dari penelitian ini:

1. Penelitian ini dapat dijadikan sebagai referensi teknis mengenai penerapan sistem *lip-sync* berbasis *volume-visemes* pada *platform Unreal Engine 5.2*.
2. Penelitian ini memberikan dasar bagi pengembangan sistem yang lebih kompleks di masa depan, termasuk penggunaan ekspresi emosi yang lebih beragam dan dukungan terhadap bahasa selain Inggris.

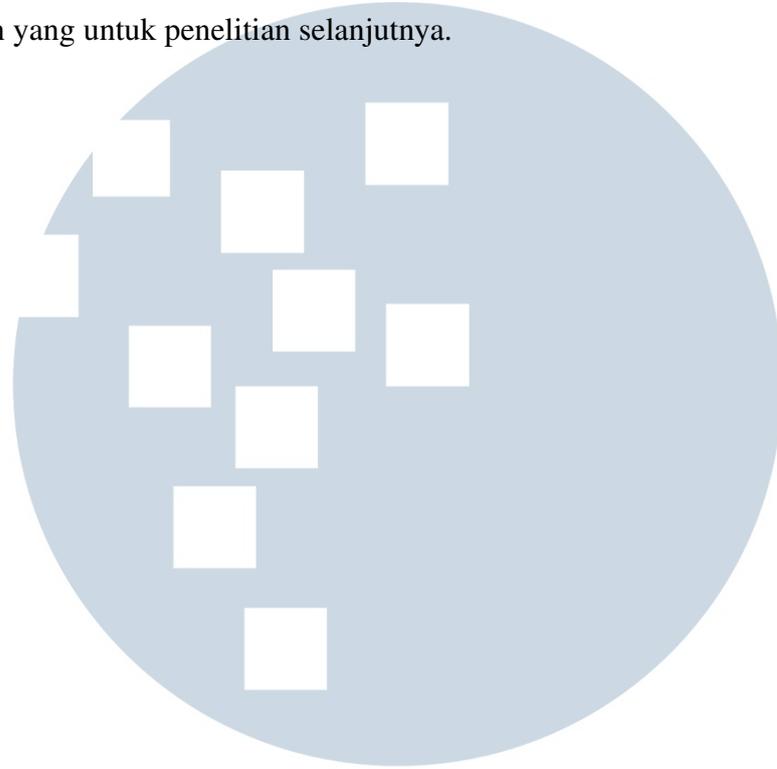
## 1.6 Sistematika Penulisan

Laporan tugas akhir ini disusun secara sistematis agar pembahasan lebih terstruktur dan mudah dipahami. Adapun sistematika penulisan laporan ini terdiri dari beberapa bab sebagai berikut:

- Bab 1 PENDAHULUAN  
Bab 1 berisi penjelasan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- Bab 2 LANDASAN TEORI  
Bab 2 berisi penjelasan mengenai teori - teori yang digunakan dalam penelitian ini. Mulai dari dari konsep *uncanny valley*, teknologi *ChatGPT*, *Text-to-Speech*, *viseme*, *MetaHuman*, *lip-sync*, fonem, dan *viseme*. Bab ini juga mencakup teknik pengolahan data seperti rata-rata, desibel, normalisasi, dan *smoothing*, serta metode evaluasi sistem menggunakan *System Usability Scale (SUS)* dan skala likert.
- Bab 3 METODOLOGI PENELITIAN  
Bab 3 berisi penjelasan mengenai objek penelitian, metodologi pengembangan sistem, *flowchart*, perancangan desain antarmuka, dan pembuatan aplikasi.
- Bab 4 HASIL DAN DISKUSI  
Bab 4 berisi penjelasan mengenai identifikasi masalah, analisa kebutuhan, desain sistem, desain antarmuka, dan pembuatan aplikasi.

- Bab 5 SIMPULAN DAN SARAN

Bab 5 berisi penjelasan mengenai kesimpulan akhir dari penelitian beserta saran yang untuk penelitian selanjutnya.



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA