

**IMPLEMENTASI TEKNOLOGI *TEXT-TO-SPEECH* PADA
PRODUKSI *VOICE-OVER* ANIMASI “THE TORTOISE
AND THE HARE”**

**SKRIPSI
Program Studi Musik**



**Oleh:
Reza Fauzi Rahman
NIM 21102980131**

**JURUSAN MUSIK
FAKULTAS SENI PERTUNJUKAN
INSTITUT SENI INDONESIA YOGYAKARTA
GENAP 2024/2025**

**IMPLEMENTASI TEKNOLOGI *TEXT-TO-SPEECH* PADA
PRODUKSI *VOICE-OVER* ANIMASI “THE TORTOISE
AND THE HARE”**

**SKRIPSI
Program Studi Musik**



**Oleh:
Reza Fauzi Rahman
NIM 21102980131**

**JURUSAN MUSIK
FAKULTAS SENI PERTUNJUKAN
INSTITUT SENI INDONESIA YOGYAKARTA
GENAP 2024/2025**

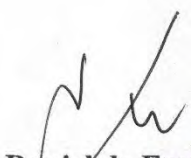
HALAMAN PENGESAHAN

Tugas Akhir berjudul:

IMPLEMENTASI TEKNOLOGI *TEXT-TO-SPEECH* PADA PRODUKSI *VOICE-OVER* ANIMASI “THE TORTOISE AND THE HARE” diajukan oleh Reza Fauzi Rahman, NIM 21102980131, Program Studi S-1 Musik, Jurusan Musik, Fakultas Seni Pertunjukan, Institut Seni Indonesia Yogyakarta (**Kode Prodi: 91221**), telah dipertanggungjawabkan di depan Tim Penguji Tugas Akhir pada tanggal 9 Mei 2025 dan dinyatakan telah memenuhi syarat untuk diterima.

Ketua Tim Penguji

Pembimbing I/Anggota Tim Penguji


Daniel de Fretes, S.Sn., M.Sn.

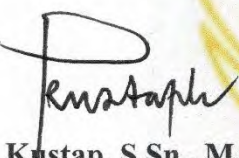
NIP 198401162019031004/
NIDN 2316018401


Daniel de Fretes, S.Sn., M.Sn.

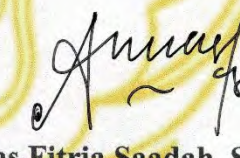
NIP 198401162019031004/
NIDN 2316018401

Penguji Ahli/Anggota Tim Penguji

Pembimbing II/Anggota Tim Penguji


Kustap, S.Sn., M.Sn.

NIP 196707012003121001/
NIDN 0001076707


Annas Fitria Saadah, S.Fil., M.Phil.

NIP 199004022023212050/
NIDN 0502049004

Yogyakarta, **24 - 06 - 25**

Mengetahui,
Dekan Fakultas Seni Pertunjukan
Institut Seni Indonesia Yogyakarta

Koordinator Program Studi
Musik



Dr. I Nyoman Cau Arsana, S.Sn., M.Hum.

NIP 197111071998031002/
NIDN 0007117104



Kustap, S.Sn., M.Sn.

NIP 196707012003121001/
NIDN 0001076707

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya tulis ini bebas dari unsur plagiasi dan belum pernah diajukan untuk memperoleh derajat akademik di suatu perguruan tinggi. Skripsi ini bukan merupakan karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain dan/atau diri saya sendiri sebelumnya, kecuali yang secara tertulis diacu dan dinyatakan dalam naskah ini dan/atau tercantum dalam daftar pustaka. Demikian pernyataan ini saya buat dengan sesungguhnya. Saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku jika di kemudian hari ditemukan bukti bahwa pernyataan ini tidak benar.

Yogyakarta, 9 Mei 2025
Yang menyatakan,



Reza Fauzi Rahman
NIM 21102980131

MOTTO

Fokus ke perkembangan & kemajuan yang aku dapat, bukan kesempurnaan.

Fokus ke hal yang aku kuasai, gak perlu mikirin hal di luar kontrol berlarut-larut.



HALAMAN PERSEMBAHAN

Skripsi ini aku persembahkan kepada
keluargaku yang tak pernah putus dalam mendoakan,
partner yang tak hentinya menguatkan,
sahabat yang hadir dalam setiap titik perjuangan,
dosen pembimbing yang selalu sabar menghadapi kebandelanku,
serta diriku sendiri yang bertahan dan bertumbuh menghadapi jatuh dan bangun.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kemudahan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Teknologi *Text-to-Speech* pada Produksi *Voice-Over* Animasi ‘The Tortoise and the Hare’”. Skripsi ini diajukan kepada Fakultas Seni Pertunjukan Institut Seni Indonesia Yogyakarta untuk memenuhi salah satu syarat memperoleh gelar Sarjana Seni. Dalam proses penyusunan skripsi ini, penulis menyadari sepenuhnya bahwa tanpa bimbingan, bantuan, dan dukungan dari berbagai pihak, skripsi ini tidak akan terselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Dr. I Nyoman Cau Arsana, S.Sn., M.Hum. selaku Dekan Fakultas Seni Pertunjukan Institut Seni Indonesia Yogyakarta.
2. Kustap S.Sn., M.Sn. selaku Koordinator Program Studi Musik Fakultas Seni Pertunjukan Institut Seni Indonesia Yogyakarta serta dosen penguji ahli yang telah mengoreksi, memberi arahan, serta mendukung penuh dan menyemangati setiap kegiatan yang penulis lakukan.
3. Daniel de Fretes, S.Sn., M.Sn. selaku Sekretaris Program Studi Musik Fakultas Seni Pertunjukan Institut Seni Indonesia Yogyakarta serta dosen pembimbing pertama yang telah membimbing, mengarahkan, dan memberikan masukan berharga pada setiap kegiatan yang penulis jalani. Terutama selama proses penelitian hingga penulisan skripsi ini.

4. Annas Fitria Saadah, S.Fil., M.Phil. selaku dosen pembimbing kedua yang telah membantu memberikan wawasan, kritik, dan saran yang membangun bagi penulis.
5. Seluruh dosen Program Studi Musik Fakultas Seni Pertunjukan Institut Seni Indonesia Yogyakarta yang telah mendidik dan membagikan ilmu selama penulis menempuh studi.
6. Mendiang Bapak, Dudun Badruzaman, serta Mamah, Euis Nurjanah, kemudian Teteh, Resy Fauziah Ratna Nurzaman, dan seluruh kerabat, keluarga, serta guru yang tiada hentinya memberikan doa, kasih sayang, dan dukungan moral maupun material sehingga penulis dapat menjalani kehidupan sebaik-baiknya.
7. Rafa Radithya Swara selaku pasangan yang selalu hadir dan menguatkan satu sama lain dalam setiap jatuh, bangun, serta momen yang membuat penulis bertumbuh.
8. PT Selingkar Literasi Sayang Keluarga sebagai wadah bagi penulis dapat berkembang dan meneliti, khususnya bagi Kak Maya Tauriana selaku CEO perusahaan dan mentor yang telah memberikan izin penelitian, bersedia menjadi narasumber, dan menghadirkan dukungan positif. Penulis juga mengucapkan terima kasih kepada Dimas Irvandi dan Azizah Yuditarsi yang telah bersedia menjadi narasumber dengan setiap *insight* yang berharga.
9. Seluruh rekan Moment Musicaux serta semua sahabat yang selalu memberikan dukungan positif satu sama lain, kebersamai perjuangan,

menjadi teman diskusi umumnya terkait apa pun dan khususnya terkait skripsi penulis, serta berbagi pengalaman yang berharga.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan skripsi ini. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Yogyakarta, 9 Mei 2025

Penulis,



Reza Fauzi Rahman



ABSTRAK

Penelitian ini mengkaji implementasi teknologi *text-to-speech* pada produksi *voice-over* animasi "The Tortoise and the Hare" yang diproduksi oleh PT Selingkar Literasi Sayang Keluarga. Penelitian ini dilatarbelakangi oleh tren pemanfaatan teknologi AI di industri kreatif, namun belum ditemukan penelitian spesifik terkait penggunaan *text-to-speech* untuk produksi *voice-over* animasi edukatif. Penelitian ini bertujuan untuk memaparkan tahapan produksi *voice-over* berbasis *text-to-speech*, mengidentifikasi kendala teknis selama proses produksi, serta menganalisis dampaknya terhadap industri animasi. Penelitian ini menggunakan pendekatan kuasi-etnografi melalui observasi partisipatif, wawancara, serta analisis dokumen berupa catatan produksi, naskah (*script*), *storyboard*, serta luaran *voice-over*. Penelitian dilaksanakan selama 8 bulan dengan durasi pengamatan 7 jam per hari kerja. Analisis proses produksi mengacu pada prosedur umum produksi animasi menurut Laura Moreno dan proses penciptaan musik film menurut Richard Davis. Hasil penelitian menunjukkan bahwa penerapan *text-to-speech* meningkatkan efisiensi waktu, ruang, dan biaya, namun memunculkan tantangan berupa kebutuhan pemahaman *prompting*, keterbatasan ekspresivitas suara, serta proses *trial & error* untuk memperoleh hasil yang sesuai. Dampak implementasi *text-to-speech* tidak serta merta menggantikan peran manusia, tetapi menggeser kompetensi yang dibutuhkan dari keterampilan olah vokal menjadi keterampilan teknis dan teknologi, sejalan dengan teori Substitusi Tenaga Kerja menurut Klaus Schwab dan teori Relasi Manusia dengan Teknologi menurut Don Ihde.

Kata kunci: Kecerdasan buatan (AI), *voice-over*, *text-to-speech*, animasi, teknologi

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Rumusan Masalah.....	3
C. Tujuan Penelitian.....	4
D. Manfaat Penelitian.....	5
1. Manfaat Teoretis	5
2. Manfaat Praktis	5
E. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	7
A. Tinjauan Pustaka.....	7
B. Landasan Teori.....	12
1. Produksi Animasi.....	13
2. <i>Film Spotting</i>	17
3. Substitusi Tenaga Kerja	20
4. Relasi Manusia dengan Teknologi.....	21
BAB III METODE PENELITIAN	23
A. Jenis Penelitian	23
B. Waktu dan Lokasi Penelitian	23
C. Teknik Pengumpulan Data.....	24
D. Sumber Data	25
E. Teknik Analisis Data.....	25

BAB IV HASIL PENELITIAN DAN PEMBAHASAN	27
A. Hasil Penelitian.....	27
1. Proses Produksi <i>Voice-Over</i> Animasi dengan AI	28
2. Kendala pada Proses Produksi <i>Voice-Over</i> Animasi dengan AI	53
3. Dampak Penggunaan <i>Text-to-Speech</i> terhadap Industri Animasi.....	56
4. Analisis Proses Produksi <i>Voice-Over</i> Animasi dengan Teknologi <i>Text-to-Speech</i>	61
5. Analisis Kendala dalam Proses Produksi <i>Voice-Over</i>	65
6. Analisis Dampak Penggunaan Teknologi <i>Text-to-Speech</i> terhadap Industri Animasi	67
B. Pembahasan	70
BAB V PENUTUP.....	75
A. Kesimpulan.....	75
B. Saran	76
DAFTAR PUSTAKA	77
LAMPIRAN.....	79
Lampiran 1. Letter of Acceptance Program MSIB <i>Batch 7</i>	80
Lampiran 2. Surat Izin Penelitian	81
Lampiran 3. Denah Lokasi Penelitian	82
Lampiran 4. Dokumen Observasi	83
1. Naskah Cerita (<i>Script</i>).....	83
2. <i>Storyboard</i>	86
3. Catatan Produksi pada Aplikasi Notion	90
4. Catatan Produksi pada Aplikasi Taskade.....	92
Lampiran 5. Dokumentasi Observasi & <i>Link</i> Video Animasi	95
Lampiran 6. Pedoman Wawancara	96
Lampiran 7. Transkrip Wawancara	99
Narasumber 1	99
Narasumber 2	106
Narasumber 3	114
Lampiran 8. Dokumentasi Wawancara	120

DAFTAR GAMBAR

Gambar 1. Bagan alir penelitian	26
Gambar 2 . Bagan alir produksi <i>text-to-speech</i>	29
Gambar 3. Catatan analisis referensi	31
Gambar 4. Penggalan latar cerita dan tokoh dalam naskah	32
Gambar 5. <i>Voice library</i> ElevenLabs.....	33
Gambar 6. Perangkat keras yang digunakan.....	35
Gambar 7. Penggalan <i>storyboard</i> dengan deskripsi musik latar.....	35
Gambar 8. Tampilan proyek DAW pada awal proses produksi	37
Gambar 9. Naskah sebagai acuan pengembangan <i>storyboard</i>	38
Gambar 10. Antarmuka <i>text-to-speech</i> ElevenLabs	40
Gambar 11. Penggeseran klip audio untuk jeda dan <i>time warp</i> untuk mengubah ritme penceritaan (<i>pacing</i>)	41
Gambar 12. Riwayat konversi <i>text-to-speech</i>	42
Gambar 13. Antarmuka <i>voice changer</i> ElevenLabs	43
Gambar 14. Penggunaan <i>automation</i> untuk mensimulasikan perubahan suara saat bicara setelah menguap	44
Gambar 15. Perubahan tempo untuk keperluan transisi	45
Gambar 16. <i>Pitch</i> sebelum diubah VariAudio	48
Gambar 17. <i>Pitch</i> setelah diubah VariAudio	48
Gambar 18. Notasi tema utama.....	48
Gambar 19. Pemetaan tempo <i>scene</i> 1	49
Gambar 20. Tampilan DAW tahap <i>mixing</i>	51
Gambar 21. Tampilan DAW tahap <i>mastering</i>	52

BAB I

PENDAHULUAN

A. Latar Belakang

Penggunaan teknologi kecerdasan buatan (AI) telah mengubah dinamika industri kreatif secara signifikan dalam beberapa tahun terakhir. Tren ini telah menyebar ke berbagai aspek produksi, mulai dari penulisan naskah, penciptaan karakter, hingga produksi suara. Penelitian terkini oleh Hanifa et al. (2023) dan Smink (2024) menunjukkan besarnya dampak transformasi ini, terutama dalam aspek produktivitas kinerja. Penelitian tersebut menghasilkan temuan bahwa AI dapat membantu pekerja kreatif dalam mengembangkan ide dan inovasi serta meningkatkan efisiensi produksi, sehingga menghadirkan peluang dan tantangan baru bagi para pelaku industri.

PT Selingkar Literasi Sayang Keluarga telah mengadopsi tren implementasi AI dalam produksi animasi fabel klasik “The Tortoise and the Hare”. Perusahaan ini bergerak di bidang pendidikan literasi dengan mengemban misi untuk meningkatkan taraf literasi anak-anak Indonesia. Di antaranya melalui lokakarya, kelas kreatif, serta konten *storytelling* yang mengintegrasikan berbagai bidang ilmu, seperti bahasa, matematika, sains, budaya, dan teknologi (Tauriana, 2019). Pada tahun 2025, perusahaan ini melakukan perubahan strategi dengan meluncurkan platform pembelajaran melalui kanal YouTube “The Edventures of Mia & Juno” yang menyajikan materi literasi melalui media video lagu anak dan adaptasi cerita klasik berbahasa Inggris. Dengan menargetkan pasar internasional, perusahaan ini

berkomitmen untuk menghadirkan konten berkualitas tinggi, termasuk pada aspek *voice-over* dengan aksen dan logat bahasa Inggris *native*.

Salah satu tantangan dalam pengembangan kanal “The Edventures of Mia & Juno” adalah kebutuhan akan *voice-over* dengan aksen dan logat bahasa Inggris *native* yang sulit dipenuhi karena keterbatasan sumber daya dan waktu produksi. Oleh karena itu, dalam produksi animasi fabel klasik “The Tortoise and the Hare”, tim produksi yang dipimpin oleh CEO PT Selingkar Literasi Sayang Keluarga sebagai *supervisor*, bersama tiga orang peserta magang memutuskan untuk mengimplementasikan teknologi *text-to-speech* berbasis AI dari ElevenLabs sebagai solusi untuk memenuhi kebutuhan *voice-over*.

Text-to-speech atau *speech synthesis* merupakan sistem komputasi yang mengonversi teks tertulis menjadi suara buatan secara otomatis (Khanam et al., 2022). *Text-to-speech* kerap dimanfaatkan untuk pembuatan *voice-over* sebagai narasi atau dialog tanpa perlu merekam langsung suara manusia (Smink, 2024). *Voice-over* sendiri didefinisikan sebagai rekaman suara yang digunakan sebagai narasi, dialog, atau pengisi suara karakter dalam media audiovisual seperti berita, konten promosi, film, dan animasi (Alburger, 2014). Dalam proyek animasi “The Tortoise and the Hare”, ElevenLabs dipilih sebagai platform *text-to-speech* karena kemampuannya untuk mereplikasi emosi manusia melalui algoritma *deep learning* (Barrington et al., 2023). Selain itu, ElevenLabs dinilai memiliki antarmuka yang mudah dipahami, menyediakan pilihan *voice model* dengan berbagai aksen dan karakteristik suara, serta dapat digunakan secara gratis (Aaron S., 2025; White, 2024).

Berbagai penelitian telah membahas implementasi teknologi AI dalam industri kreatif, terutama dalam produksi film dan audio. Penelitian Anantrasirichai & Bull (2022) menunjukkan bahwa AI telah berhasil menciptakan film dengan teknologi *face-swapping* dan *voice-generation*. Sementara itu, Bird & Lotfi (2023) membahas etika penggunaan teknologi AI generatif dalam mengkloning suara dan memanipulasi ucapan secara *real-time*. Lebih lanjut, Bharathi & Vardhan (2024) mengungkapkan bahwa peran-peran dalam produksi film, termasuk *sound designer*, sangat terdampak oleh perkembangan AI. Bahkan, saat ini AI mampu merekonstruksi vokal manusia tanpa kesan robotik melalui proses *lipreading* dalam video *multi-angle* (Kumar et al., 2018).

Terlepas dari banyaknya penelitian atas kemajuan dan implementasi AI di industri kreatif, belum ditemukan penelitian yang secara spesifik membahas implementasi teknologi AI dalam produksi animasi edukatif, khususnya pada aspek *voice-over* dengan *text-to-speech*. Padahal, teknologi AI terus berkembang secara pesat dan telah mencapai tingkat kecanggihan yang tinggi dalam menghasilkan konten audio maupun visual yang menyerupai manusia. Oleh karena itu, penelitian ini menjadi relevan untuk mengidentifikasi potensi implementasi AI untuk menghasilkan konten animasi edukatif yang berkualitas tinggi, khususnya pada aspek *voice-over*.

B. Rumusan Masalah

Dari latar belakang yang telah dijabarkan, dapat dirumuskan pertanyaan penelitian sebagai berikut:

1. Bagaimana proses produksi *voice-over* animasi “The Tortoise and the Hare” dengan implementasi teknologi *text-to-speech*?
2. Apa saja kendala yang dihadapi pada proses produksi *voice-over* animasi “The Tortoise and the Hare” dengan implementasi teknologi *text-to-speech* untuk menciptakan karakter suara yang sesuai dengan kebutuhan?
3. Bagaimana dampak penggunaan teknologi *text-to-speech* terhadap industri animasi?

C. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, maka tujuan dari penelitian ini adalah:

1. Mendalami dan menjabarkan proses produksi *voice-over* animasi “The Tortoise and the Hare” yang mengimplementasikan teknologi *text-to-speech*.
2. Mengidentifikasi kendala yang dihadapi pada proses produksi *voice-over* animasi “The Tortoise and the Hare” yang mengimplementasikan teknologi *text-to-speech* untuk menciptakan karakter suara yang sesuai dengan kebutuhan.
3. Menganalisis dampak penggunaan teknologi *text-to-speech* terhadap industri animasi.

D. Manfaat Penelitian

1. Manfaat Teoretis

Secara teoretis, hasil penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan selanjutnya, khususnya dalam bidang produksi di industri kreatif dengan memanfaatkan teknologi AI.

2. Manfaat Praktis

- a. Menyediakan wawasan dan panduan praktis bagi studio animasi kecil atau independen yang ingin mengimplementasikan teknologi *text-to-speech* dalam proses produksinya.
- b. Membantu pengembang teknologi AI dalam memahami kebutuhan spesifik industri animasi untuk pengembangan teknologi *text-to-speech* yang lebih efektif.

E. Sistematika Penulisan

Hasil penelitian disusun dengan sistematika penulisan yang diuraikan ke dalam lima bagian. Bagian pertama ialah bab pendahuluan dengan sub-bab pertama yakni latar belakang, yang menguraikan tentang perkembangan teknologi AI dalam industri kreatif, deskripsi proyek animasi yang menjadi objek penelitian, alasan penggunaan teknologi *text-to-speech* dalam proyek animasi “The Tortoise and the Hare”, serta urgensinya. Kemudian sub-bab rumusan masalah berisi butir-butir pertanyaan penelitian berdasarkan permasalahan yang diuraikan pada latar belakang. Dalam hal ini, permasalahan yang dirumuskan perihal penggunaan *text-to-speech* pada produksi *voice-over* animasi. Kemudian sub-bab tujuan penelitian

berisi pernyataan tentang capaian penelitian sebagai target yang hendak dicapai serta ukuran keberhasilan penelitian. Sub-bab manfaat penelitian menguraikan manfaat teoretis dan manfaat praktis. Kemudian sebagai penutup bab pertama, sub-bab sistematika penulisan menguraikan keseluruhan penelitian secara struktural.

Bab kedua berisikan sub-bab tinjauan pustaka sebagai acuan untuk menemukan kesamaan dan kebaruan topik yang diteliti. Kemudian sub-bab berikutnya menjabarkan teori serta prosedur baku yang dapat menjadi pisau analisis utama dalam melaksanakan penelitian ini.

Pada bab ketiga, diuraikan langkah-langkah dalam menjalankan penelitian kuasi-etnografi dengan urutan sub-bab jenis penelitian, waktu dan lokasi penelitian, teknik pengumpulan data, sumber data, teknik analisis data, serta diagram alir penelitian.

Bab keempat memaparkan hasil yang didapatkan setelah melakukan langkah-langkah penelitian yang dijabarkan dalam sub-bab teknik analisis data.

Kemudian penelitian ditutup dengan bab kelima yang berisi kesimpulan, saran, serta lampiran penelitian.