

Submitted: 11/10/2025

Accepted: 8/11/2025

Published: 20/11/2025

Research Article

Inkonsistensi Bahasa Ujaran Navigasi Google Maps: Benarkah AI Belum Sepenuhnya Memahami Bahasa Manusia?

Krismonika Khoirunnisa^{1*}, Sahrul Romadhon², Fakhriyyah Asmay Aidha³,
Siti Zumrotul Maulida⁴, Silfia Qurrotul A'yun⁵

¹ Balai Bahasa Provinsi Jawa Timur, Indonesia

² Universitas Islam Negeri Madura, Indonesia

³⁻⁵ UIN Sayyid Ali Rahmatullah Tulungagung, Indonesia

*Correspondence Author, E-mail: krismonicakhpurnisa@gmail.com

Abstract

Background: Artificial Intelligence (AI) has become essential in digital navigation systems such as Google Maps, whose Indonesian Text-to-Speech (TTS) feature still struggles with mispronouncing abbreviations, acronyms, and proper names, often producing speech that sounds unnatural or confusing. **Purpose:** This study aims to identify the linguistic and technical causes of pronunciation errors in Google Maps' Indonesian TTS system. **Method:** This study employs a qualitative approach combining content analysis and computational linguistics. Audio samples from Google Maps were analyzed to identify recurring pronunciation issues, followed by a technical examination of phonetic modeling, text-to-phoneme conversion, and text-normalization processes within the TTS system. **Results:** The results show that inaccuracies arise from three key factors: insufficient contextual recognition of abbreviations, limited lexical databases, and weaknesses in text-normalization mechanisms. These findings suggest that the challenges are not purely technical but also linguistic and sociocultural, reflecting the complexity of Indonesian with its diverse dialects, orthography, and local naming practices. The study recommends improving lexical coverage, developing more context-aware linguistic models, and training AI with voice and text data that better represent Indonesian language variation. **Conclusion:** Overall, this research contributes to the development of a more accurate, natural, and culturally attuned TTS system while underscoring the importance of integrating AI innovation with linguistic sensitivity.

Kata kunci: navigation code, google maps, artificial intelligence, sistem text-to-speech

PENDAHULUAN

Teknologi kecerdasan buatan (*Artificial Intelligence*) telah membawa revolusi besar dalam kehidupan manusia di era digital. Kecerdasan buatan tidak hanya sekadar sebagai inovasi melainkan telah menjadi bagian dari integral dan rutinitas sehari-hari manusia, seperti halnya kecerdasan buatan sebagai asisten pekerjaan. Hal ini memengaruhi berbagai aspek mulai dari komunikasi, hiburan, pendidikan, hingga mobilitas. Perubahan sistematis dan fundamental menjadi aspek utama yang terlihat dalam bidang navigasi

digital. Pada peradaban lama, manusia bergantung pada peta kertas atau bertanya kepada orang di jalan untuk menemukan arah. Namun, saat ini berkat sistem kecerdasan buatan, seperti aplikasi peta digital dapat diakses dengan mudah melalui ponsel, laptop, ataupun alat digital lainnya. Kemudahan ini tidak hanya meningkatkan efisiensi, tetapi juga secara mendalam mengubah pola perilaku masyarakat dalam berinteraksi dengan ruang geografis di sekitarnya. Di antara sekian banyak layanan navigasi yang tersedia dalam aplikasi Google Maps telah menjadi salah satu platform yang mendominasi dan digunakan secara luas di seluruh dunia. Aplikasi Google Maps menawarkan lebih dari sekadar visualisasi peta tetapi juga dilengkapi dengan berbagai fitur canggih seperti informasi lalu lintas *real-time*, estimasi waktu tempuh, rekomendasi jalur tercepat, hingga panduan suara. Fitur panduan suara menjadi salah satu esensial dari aplikasi Google Maps karena memungkinkan pengguna untuk tetap fokus di jalan tanpa harus terus-menerus memperhatikan layar (gawai, laptop, atau tablet).

Kecerdasan buatan memiliki kelebihan dan kekurangan seperti halnya kecanggihan serta kemudahan yang telah ditawarkan dapat memunculkan sebuah fenomena yang menarik sekaligus problematis, yaitu aspek kebahasaan pada panduan suara aplikasi Google Maps. Banyak pengguna melaporkan bahwa instruksi suara tersebut terdengar kaku, tidak alami, atau bahkan salah dalam pengucapan. Laporan tersebut paling sering muncul pada nama-nama jalan, tempat, atau daerah yang dilafalkan secara tidak sesuai dengan ejaan maupun kaidah pelafalan bahasa Indonesia. Fenomena ini sesungguhnya melampaui masalah teknis semata. Fenomena tersebut juga menyentuh dimensi linguistik komputasional dan sosiokultural. Penelitian ini bertujuan untuk mengkaji secara mendalam fenomena kebahasaan dalam fitur panduan suara Google Maps.

Fokus utamanya adalah memahami mekanisme di balik ketidaksesuaian antara teks dan ujaran suara yang sering kali dialami pengguna. Penelitian ini tidak hanya mengeksplorasi dari perspektif teknis, tetapi juga dari sudut pandang linguistik komputasional untuk mengungkap bagaimana AI memproses, menginterpretasi, dan memproduksi ujaran berbasis teks dalam bahasa Indonesia. Hal ini menimbulkan kebingungan terutama bagi pengguna yang tidak familiar dengan lokasi yang dituju. Hal ini sejalan dengan penelitian yang telah dilakukan oleh Kao dan Stephen (2007) dengan judul *Natural Language Processing and Text Mining* yang menyatakan bahwa pengklasifikasian wacana sering kali memberlakukan sebuah teks sebagai satu kesatuan yang utuh dan homogen. Ketidaksesuaian antara teks tertulis dan ujaran suara menimbulkan pertanyaan mendasar mengenai sejauh mana sistem kecerdasan buatan mampu memahami, mengolah dan menyesuaikan diri dengan kekayaan fonetik, dialek, serta ragam ejaan bahasa lokal yang kompleks.

Penelitian pemodelan linguistik komputasional yang memanfaatkan TTS sudah banyak dilakukan oleh peneliti sebelumnya. Pertama, penelitian dari Adriati, et al (2016) yang berjudul “Pengembangan Aplikasi Text-to-Speech Bahasa Indonesia Menggunakan Metode Finite State Automata Berbasis Android”, yang hasil temuannya berupa pemanfaatan data suara berbahasa Inggris yang dialihbahasakan ke dalam bahasa Indonesia dengan cara menggal kata-kata bahasa Indonesia menjadi suku kata sesuai pola yang telah ditetapkan, menggunakan metode *Finite State Automata* (FSA). Kedua, penelitian dari Ahmad, et al (2024) yang berjudul “Planning the Development of Text-to-Speech Synthesis Models and Datasets with Dynamic Deep Learning”, yang hasil temuannya membahas tentang taksonomi model yang berbasis *deep learning* serta membahas berbagai dataset yang digunakan dalam proses TTS (*Text-to-Speech*). Ketiga, penelitian dari Wang, et al (2025) yang berjudul “Spark-TTS: an Efficient LLM-Based Text-

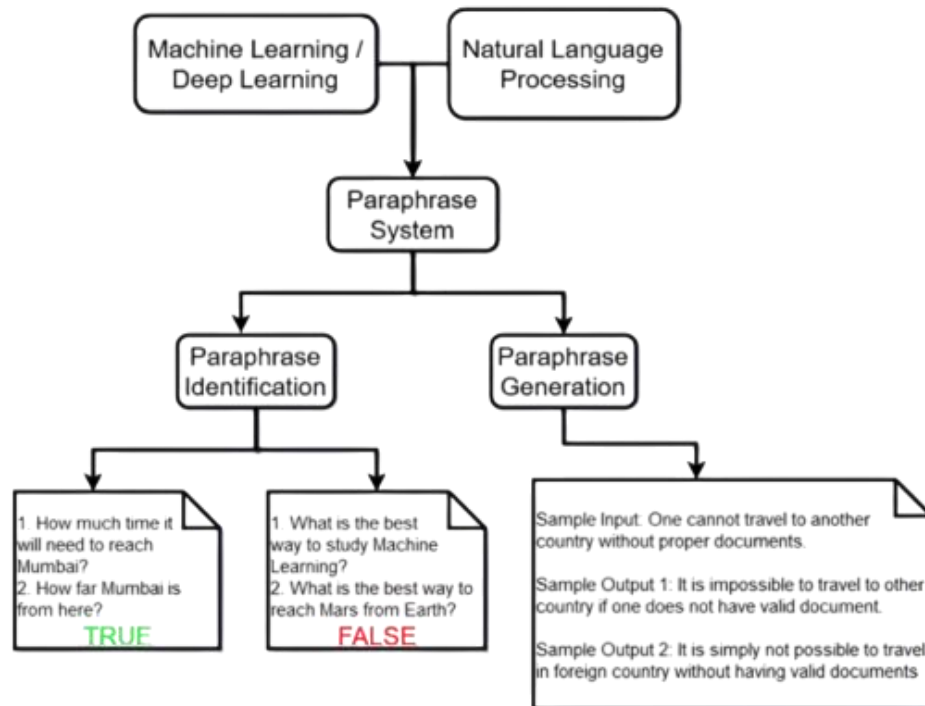
to-Speech Model with Single-Stream Decoupled Speech Tokens”, yang hasil temuannya berupa pembuktian bahwa Spark-TTS tidak hanya mencapai performa terbaik (*state-of-the-art*) dalam *zero-shot voice cloning*, tetapi juga mampu menghasilkan suara yang sangat dapat disesuaikan dan dapat melampaui keterbatasan sintesis berbasis referensi lokal dan global.

Berdasarkan latar belakang dan penelitian terdahulu yang telah disebutkan, penelitian ini akan mencoba menjawab empat rumusan masalah, yaitu (1) seberapa akurat Google Maps dalam memberikan informasi yang sesuai dengan pengguna melalui panduan suaranya; (2) bagaimana sistem *Text-to-Speech* (TTS) Google Maps memproses dan memproduksi ujaran dari teks, dan apa yang menyebabkan terjadinya kesalahan pelafalan; (3) bagaimana model fonetik atau konversi teks-ke-fonem (*text-to-phoneme*) bekerja dalam sistem TTS Google Maps, khususnya terhadap akronim; dan (4) bagaimana pengaruh *typing models* yang digunakan terhadap kualitas ujaran yang dihasilkan oleh sistem TTS Google Maps. Melalui rumusan masalah pada penelitian ini, diharapkan dapat memberikan pemahaman yang komprehensif mengenai interaksi antara AI, bahasa, dan teknologi navigasi, serta memberikan kontribusi bagi pengembangan sistem navigasi digital yang lebih akurat.

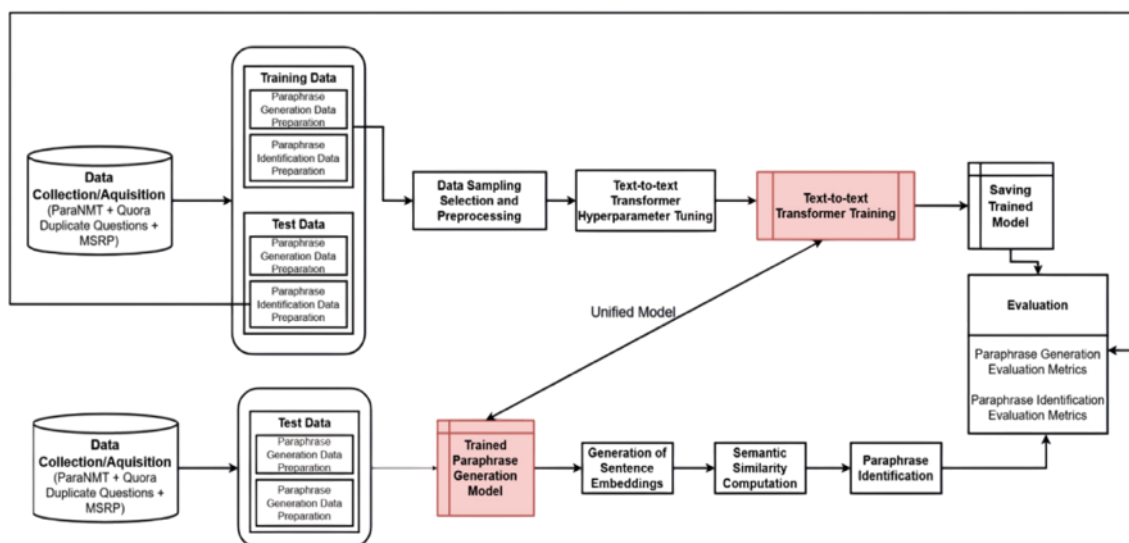
METODE PENELITIAN

Penelitian ini menggunakan metode kualitatif dengan pendekatan analisis konten (*content analysis*) dan analisis linguistik komputasional. Pendekatan ini dipilih untuk memberikan pemahaman yang komprehensif dan mendalam mengenai fenomena kesalahan kebahasaan dalam panduan suara Google Maps. Melalui analisis konten, peneliti akan mengumpulkan dan mengkaji data dari rekaman audio dari panduan suara Google Maps. Data ini akan dikategorikan dan dianalisis untuk mengidentifikasi pola-pola kesalahan yang paling sering terjadi, seperti pelafalan nama jalan yang tidak sesuai dan tantangan pengucapan pada akronim. Selain itu, analisis linguistik komputasional akan digunakan untuk membedah mekanisme teknis di balik kesalahan tersebut. Penelitian ini juga akan membandingkan teks asli yang seharusnya diucapkan dengan ujaran yang dihasilkan oleh sistem *Text-to-Speech* (TTS) dengan menggunakan analisis model ekstraksi data dan pengindeksan.

Analisis ini akan mencakup evaluasi model fonetik, proses konversi teks ke fonem, dan bagaimana algoritma AI berinteraksi dengan struktur fonologi bahasa (Lane et al, 2018; Palivela, 2021). Diagram analisis konten dalam linguistik komputasi memiliki banyak variasi salah satunya adalah generalisasi teks dan normalisasi teks. Pada penelitian ini berfokus pada analisis konten sebagai berikut.



Gambar 1. Analisis konten sistem parafrase AI (Palivela, 2021)



Gambar 2. Proses identifikasi linguistik melalui sampel data AI (Palivela, 2021)

Diagram pertama memperlihatkan dua kemampuan utama sistem parafrase, yaitu mendeteksi kesamaan makna dan menciptakan bentuk kalimat baru dengan makna yang serupa. Diagram kedua menjelaskan tahapan teknis pengembangan model komputasinya. Mulai dari pengumpulan data, pra-pemrosesan, pelatihan dengan model *transformasi linguistiknya*, hingga evaluasi kinerja pemodelan linguistiknya untuk menghasilkan sistem yang cerdas dan akurat sesuai dengan kebutuhan pengguna. Dengan mengintegrasikan kedua pendekatan ini, penelitian ini tidak hanya akan mendeskripsikan masalah yang ada, tetapi juga akan menelusuri penyebab teknis dan linguistiknya serta memberikan wawasan yang lebih kaya tentang interaksi antara AI dan bahasa.

HASIL DAN PEMBAHASAN

Berdasarkan analisis konten dan kajian linguistik komputasional, penelitian ini menemukan bahwa akurasi navigasi Google Maps dapat dibagi menjadi dua aspek, yaitu akurasi data geografis dan akurasi ujaran kebahasaan. Secara umum, akurasi data geografis Google Maps terbukti dengan adanya pengumpulan data dari satelit, informasi pengguna, dan generalisasi konteks.

Hal ini sejalan dengan teori yang dikemukakan oleh Dutoit (1997) yang berjudul *An Introduction to Text-to-Speech Synthesis*, yang menyatakan bahwa sistem akurasi Google Maps memiliki tiga proses pengolahan, yaitu penyaringan sinyal (*low-pass filter*) untuk menghilangkan frekuensi tinggi yang tidak relevan; kedua, sinyal diambil sampelnya (*sampling*) pada interval waktu teratur; dan terakhir, nilai setiap sampel dan kuantisasi (*quantizing*) untuk dibulatkan menjadi nilai digital yang presisi. Proses ini mengubah ucapan analog menjadi data digital yang dapat dianalisis dan disimpan oleh komputer. Hal ini juga sesuai dengan penelitian yang dilakukan oleh Malar, dkk. (2020), dengan judul *Computational Linguistic*, yang membahas tentang cakupan kajian *Natural Language Processing* (NLP) salah satunya adalah meringkas teks dan mengenali ujaran dari pengguna. Meskipun demikian, ketidakakuratan masih bisa terjadi akibat perubahan infrastruktur yang belum diperbarui, kondisi lalu lintas yang sangat dinamis, atau ketersediaan data, dan indeks toponimi yang terbatas di daerah tertentu.

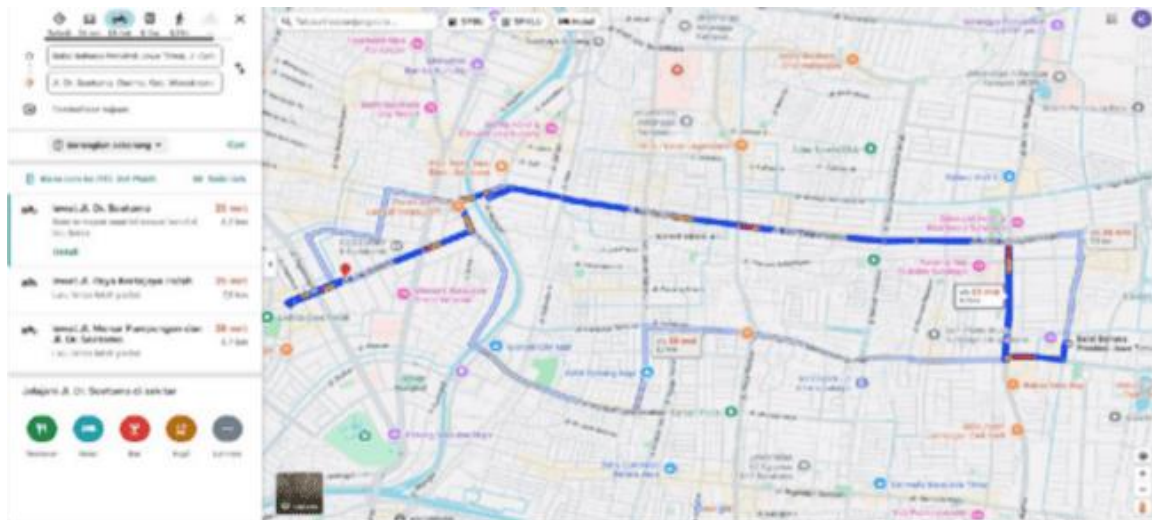
Namun, fokus utama penelitian ini adalah akurasi ujaran kebahasaan yang ditemukan jauh lebih kompleks dan problematis. Hasil analisis konten ujaran navigasi menunjukkan bahwa sistem *Text-to-Speech* (TTS) Google Maps sering kali gagal dalam pelafalan singkatan, akronim, dan nama-nama khusus. Contoh-contoh yang teridentifikasi meliputi pelafalan "Dr." dan "dr." menjadi "de-er" alih-alih "Doktor" dan "dokter", serta "Tb" menjadi "tebe" yang seharusnya "Tambak". Kesalahan ini tidak hanya mengurangi kenyamanan pengguna, tetapi juga dapat menimbulkan kebingungan, meskipun rute yang diberikan sudah benar. Analisis juga menemukan adanya inkonsistensi dalam pelafalan nama seperti "Soekarno," yang diucapkan "Sukarno," dan "Manyar Kerta Adi," yang dinormalisasi menjadi "Kerto Adi" sebelum diucapkan.

Analisis Konten Ujaran Navigasi

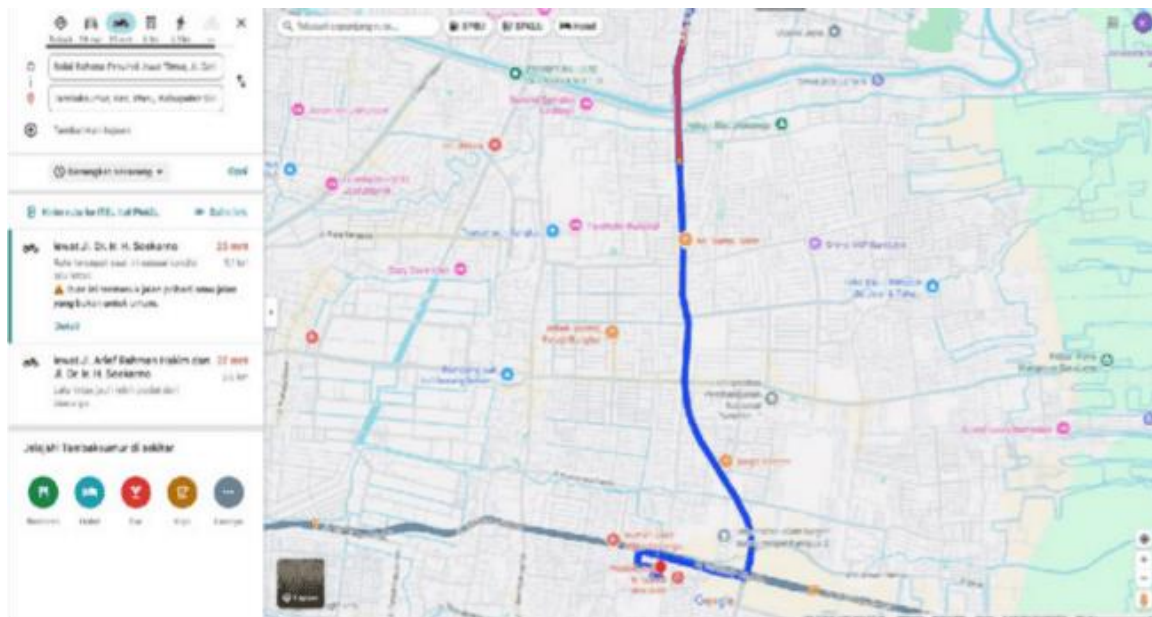
1. Kesalahan Pelafalan pada Singkatan dan Akronim

Data yang ditemukan menunjukkan bahwa sistem Google Maps masih menghadapi tantangan serius dalam aspek kebahasaan, terutama dalam melafalkan singkatan dan akronim. Kecenderungan yang terlihat adalah sistem membaca singkatan tersebut secara harfiah sesuai huruf yang tertulis, alih-alih menyesuaikannya dengan kaidah pelafalan yang benar dalam bahasa Indonesia.

Fenomena ini menegaskan bahwa kecerdasan buatan yang digunakan dalam fitur *text-to-speech* (TTS) pada Google Maps belum mampu mengenali konteks linguistik secara mendalam, sehingga gagal membedakan apakah sebuah singkatan merujuk pada nama orang, gelar akademik, nama jalan, atau bentuk akronim lain yang biasa digunakan dalam kehidupan sehari-hari. Fenomena ini sesuai dengan pendapat Bolshakov dan Alexander (2004); Alexander (2011) yang menyatakan bahwa proses TTS pada komputer memproses bahasa manusia berdasarkan dengan asumsi pengguna. Jika pengguna tidak menuliskan atau memberikan ejaan yang sesuai, maka informasi yang disampaikan akan mengandalkan konteks general. Sebagai contoh, "dr. Soetomo" yang seharusnya dibaca sebagai "dokter Soetomo" justru dilafalkan menjadi "de-er Soetomo", sehingga menghilangkan makna semantis yang sebenarnya.

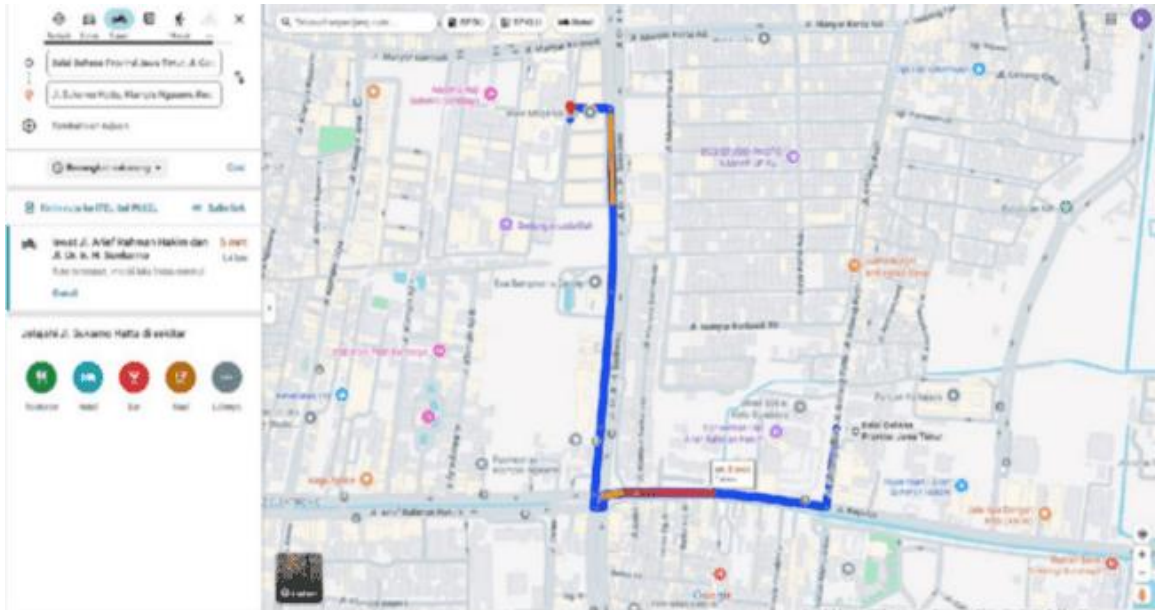


Hal serupa juga terjadi pada akronim "Tb" yang biasa dipahami sebagai singkatan dari "Tambak", tetapi justru diucapkan menjadi "tebe", yang tidak hanya menyalahi kaidah bahasa Indonesia, melainkan juga berpotensi membingungkan pengguna yang sedang mencari lokasi tertentu. Temuan ini membuktikan adanya permasalahan pada tahap *text normalization*, yaitu proses penting dalam sistem TTS yang berfungsi untuk menguraikan singkatan, simbol, atau angka ke dalam bentuk kata yang lengkap dan sesuai dengan aturan bahasa sebelum kemudian diubah menjadi bentuk ujaran.



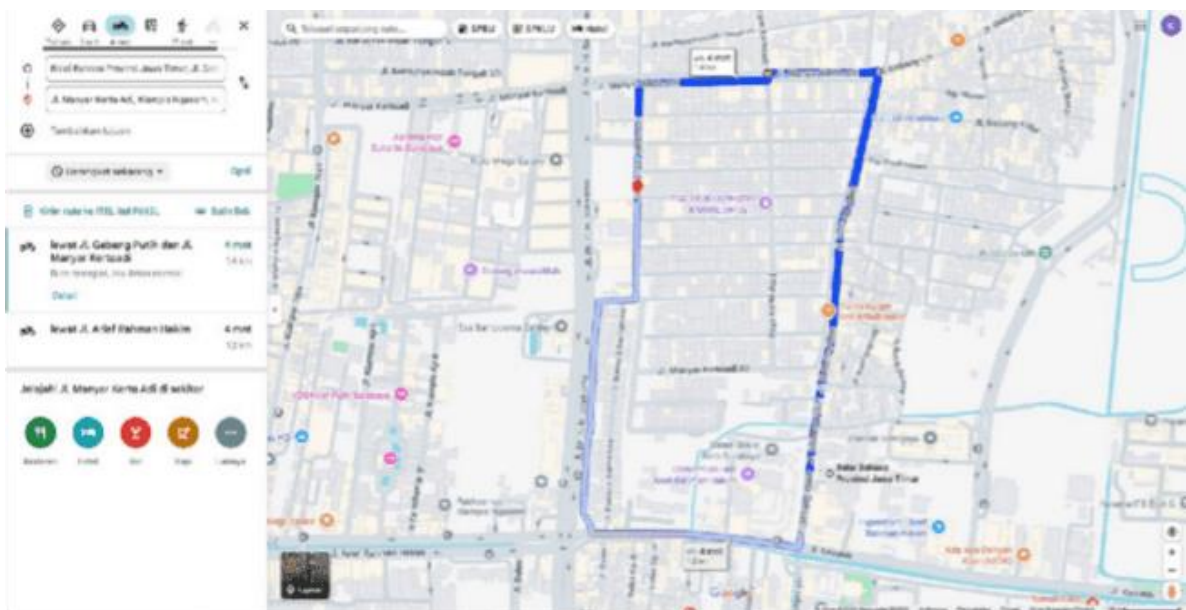
2. Inkonsistensi Pelafalan Nama Khusus dan Ejaan Lama

Penelitian ini menemukan bahwa kecerdasan buatan dalam sistem navigasi digital, khususnya Google Maps, masih sering menghadapi kesulitan ketika harus melafalkan nama- nama khusus atau nama diri yang memiliki karakteristik linguistik tertentu, terutama yang menggunakan ejaan lama. Salah satu contoh paling menonjol adalah pada nama "Soekarno", yang oleh sistem diucapkan menjadi "Sukarno". Kesalahan pelafalan ini bukan hanya sekadar perbedaan fonetis, melainkan juga mencerminkan adanya bias dalam model linguistik yang digunakan oleh AI.



Model tersebut condong pada bentuk ejaan yang lebih umum digunakan secara global atau lebih dominan muncul dalam data digital, sehingga mengesampingkan kaidah ejaan yang berlaku dalam konteks sejarah. Selain itu, penelitian ini juga menemukan kasus inkonsistensi dalam pelafalan nama tempat, misalnya pada "Manyar Kerta Adi" yang oleh sistem dinormalisasi menjadi "Manyar Kerto Adi" atau pemangkasan menjadi "Kerto Adi."

Hal ini sejalan dengan penelitian yang dikemukakan oleh Li Deng (2006) dalam bukunya *Dynamic Speech Models*. Deng menyatakan bahwa ketidakkonsistenan pelafalan disebabkan oleh adanya *modeling speech dynamics* yang memiliki dua aspek relatif. Pertama, perubahan kode bahasa manusia memiliki sifat non-linier dan kontekstual, sehingga suatu bunyi ujaran dapat mengalami variasi bentuk bergantung pada lingkungan fonetis maupun situasi tutur. Kedua, proses dinamis ini berkaitan dengan adaptasi motorik dan kognitif, di mana sistem produksi ujaran manusia secara terus-menerus menyesuaikan pola pelafalan untuk mencapai kelancaran komunikasi.



Perubahan fonem ini mungkin tampak sederhana, tetapi sesungguhnya menunjukkan adanya misinterpretasi pada tahap pra-pemrosesan teks, di mana sistem secara keliru melakukan substitusi atau penyederhanaan berdasarkan pola data yang dianggap lebih umum. Kesalahan semacam ini berpotensi menimbulkan kebingungan, terutama bagi pengguna yang tidak familiar dengan wilayah tersebut, karena perubahan sekecil apa pun dalam nama toponimi dapat memengaruhi ketepatan navigasi. Dengan demikian, fenomena salah pelafalan nama khusus ini tidak hanya mengungkap keterbatasan teknis dalam proses normalisasi teks, tetapi juga memperlihatkan tantangan yang lebih luas dalam mengintegrasikan sensitivitas linguistik dan budaya lokal ke dalam sistem kecerdasan buatan global.

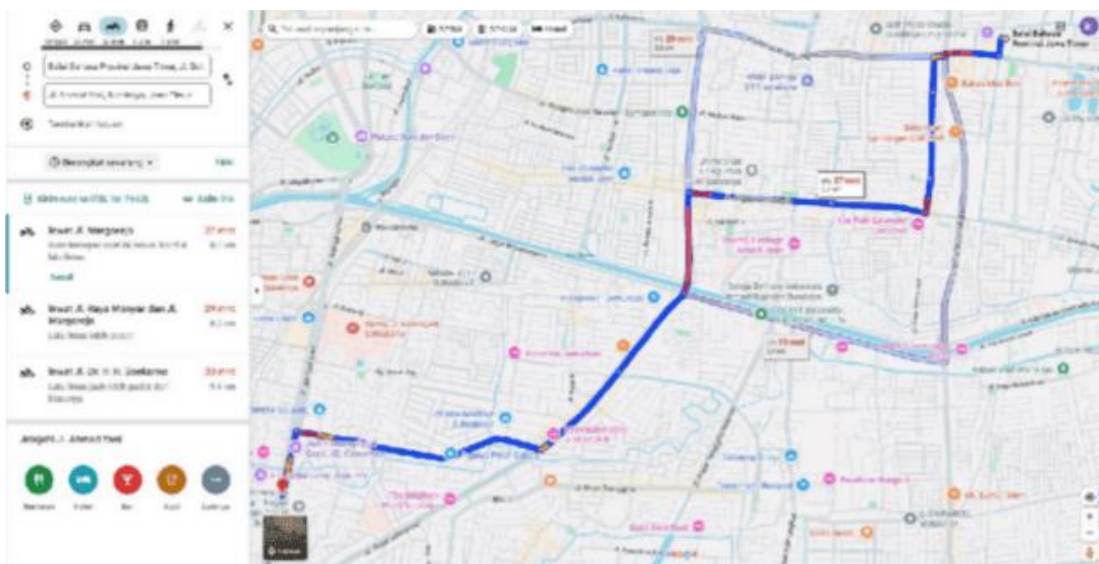
Analisis Linguistik Komputasional

1. Mekanisme Kesalahan pada Proses *Text-to-Speech* (TTS)

Kesalahan dalam pengucapan navigasi tidak terjadi begitu saja, melainkan merupakan hasil dari beberapa tahapan pemrosesan yang tidak sempurna dalam model *Text-to-Speech* (TTS).

a. Analisis Teks

Analisis Teks adalah langkah pertama dan paling mendasar dalam proses *Text-to-Speech* (TTS). Pada tahap ini, sistem kecerdasan buatan menganalisis teks masukan yang berisi informasi navigasi, seperti nama jalan, tanda baca, angka, dan singkatan. Hal ini sesuai dengan pendapat Desikan (2018), yang menyatakan bahwa analisis teks akan lebih mudah terkumpul secara konteks melalui kumpulan data besar yang didapat dari internet, sehingga google maps akan lebih mudah melakukan meta- analisis. Tujuannya adalah untuk memecah teks menjadi unit-unit yang dapat diproses lebih lanjut. Misalnya, teks "Belok kanan di Jl. A. Yani No. 25" akan diuraikan menjadi komponen-komponen terpisah, yaitu "Belok", "kanan", "di", "Jl.", "A.", "Yani", "No.", dan "25".



Proses ini tidak hanya mengenali kata-kata biasa, tetapi juga mendeteksi tanda baca seperti titik (.) yang menandakan singkatan, serta angka. Kualitas analisis teks sangat menentukan akurasi seluruh proses TTS. Jika pada tahap awal ini AI gagal mengenali singkatan dengan benar seperti menganggap "Jl." sebagai dua huruf "j"

dan "I" daripada sebagai singkatan dari "Jalan", maka seluruh proses selanjutnya akan menghasilkan ujaran yang sama. Inilah alasan mengapa kesalahan yang terjadi pada tahap awal ini dapat berdampak pada pengalaman pengguna.

b. Normalisasi Teks

Pada tahap ini, sistem AI dituntut untuk dapat mengubah berbagai bentuk tulisan yang tidak standar, seperti singkatan, akronim, angka, simbol, ataupun tanda baca ke dalam bentuk kata yang lengkap dan sesuai dengan kaidah kebahasaan. Dengan demikian, ketika teks sudah masuk ke tahap berikutnya, yaitu konversi huruf menjadi fonem, sistem dapat melafalkannya secara alami dan mudah dipahami pengguna. Namun, dalam praktiknya, data menunjukkan bahwa tahap normalisasi teks masih menyisakan banyak permasalahan, terutama jika leksikal pada rekam digitalnya yang digunakan tidak lengkap atau aturan normalisasi yang diterapkan terlalu sederhana. Akibat dari keterbatasan tersebut, sistem sering kali gagal mengenali konteks linguistik dari suatu singkatan, sehingga memilih untuk melafalkannya secara general, yaitu per huruf. Contoh yang paling jelas adalah pada singkatan "dr." yang seharusnya diinterpretasikan sebagai "dokter", tetapi justru dilafalkan "de-er" (tidak diurai dengan benar dan dilafalkan per huruf) dan "lr." yang seharusnya diujarkan "Insinyur" tetapi justru diuraikan "l-r" dan "lr".

Ketidakmampuan sistem dalam melakukan normalisasi teks secara tepat menunjukkan keterbatasan kecerdasan buatan dalam memahami keragaman dan kekayaan bahasa Indonesia. Hal ini sejalan dengan pendapat Mitkov (2003) yang menyatakan bahwa normalisasi teks pada Google Maps telah melalui proses formalisme sintaksis, yang memungkinkan sistem AI dan komputer tetap dapat memproses dan memahami. Akan tetapi, pendapat dari Clark (2010) menyatakan bahwa normalisasi teks pada sistem AI dan komputer bergantung dengan algoritma dan analogi yang dimanfaatkan sistem untuk menyusun generatif transformasional dengan memberikan hasil seakurat mungkin.

Selain itu, normalisasi teks ini tidak sekadar bersifat teknis, tetapi juga menyangkut aspek praktis dan fungsional, karena kesalahan pelafalan dapat mengurangi keakuratan navigasi, menurunkan kenyamanan pengguna, dan bahkan berpotensi menimbulkan kesalahpahaman dalam situasi perjalanan yang membutuhkan kejelasan instruksi. Dengan demikian, fenomena ini penting untuk dikaji lebih lanjut guna mengidentifikasi akar permasalahan linguistik yang dihadapi oleh sistem navigasi digital, serta merumuskan strategi perbaikan agar teknologi global seperti Google Maps dapat lebih adaptif terhadap konteks kebahasaan lokal.

Kegagalan normalisasi teks tidak hanya terbatas pada singkatan gelar, tetapi juga dapat terjadi pada bentuk lain seperti alamat, angka rumah, nama jalan yang disingkat, hingga simbol tertentu yang lazim dipakai dalam penamaan tempat. Hal ini menunjukkan bahwa kemampuan normalisasi teks dalam model TTS masih bersifat terbatas dan belum sepenuhnya teradaptasi dengan keragaman ekspresi bahasa Indonesia. Dengan demikian, perbaikan pada tahap normalisasi teks menjadi krusial, baik melalui pengayaan kamus leksikal, pembaruan aturan linguistik, maupun integrasi konteks semantik, agar sistem navigasi berbasis suara dapat memberikan pengalaman yang lebih akurat, alami, dan sesuai dengan kebutuhan pengguna lokal.

c. Konversi *Grapheme-to-Phoneme* (GTP)

Setelah teks melalui tahap normalisasi, sistem *Text-to-Speech* (TTS) memasuki proses penting berikutnya, yaitu konversi huruf menjadi fonem atau yang dikenal dengan konversi *grapheme-to-phoneme* (GTP). Pada tahap ini, setiap kata yang sudah diproses sebelumnya diuraikan menjadi rangkaian fonem yang kemudian digunakan sebagai dasar pengucapan. Idealnya, proses GTP mampu menangkap variasi fonetis dari sebuah kata secara tepat, termasuk perbedaan ejaan, pelafalan, maupun pengaruh konteks bahasa lokal.

Namun, hasil penelitian menunjukkan bahwa tahap ini masih menyisakan kelemahan signifikan, terutama ketika sistem harus melafalkan nama-nama diri atau kata-kata dengan ejaan lama. Inkonsistensi dalam pelafalan nama seperti "Soekarno" yang diucapkan menjadi "Sukarno" merupakan bukti keterbatasan model GTP yang digunakan.

Pemodelan GTP yang terdapat pada sistem AI dan komputer pada umumnya memiliki delapan figur sinkronisasi, yaitu dari aspek fonologi, morfologi, leksikal, sintaksis, semantik, analisis wacana, pragmatik, dan level hasil (Liddy, 2001). Kedelapan figur sinkronisasi yang dijelaskan Liddy (2001) tersebut dimanfaatkan sistem AI dan komputer untuk mengimplementasikan pemrosesan bahasa sesuai dengan kebutuhan interpretasi pengguna. Pemrosesan tingkat rendah, seperti analisis kata dan kalimat, akan diimplementasikan secara lebih menyeluruh, karena unit-unit analisisnya lebih kecil dan diatur oleh aturan bahasa yang jelas. Sebaliknya, pemrosesan tingkat tinggi yang berhubungan dengan pemahaman teks hanya diatur oleh pola-pola yang lebih kompleks.

Kesalahan tersebut menunjukkan bahwa pemodelan pengenalan AI tidak secara khusus dilatih untuk mengenali variasi ejaan bahasa Indonesia yang memiliki kekhasan historis maupun kultural. Alih-alih mengacu pada aturan fonetis bahasa Indonesia, sistem cenderung melakukan generalisasi berdasarkan bentuk ejaan yang lebih sering muncul dalam data yang tersistem global, sehingga variasi lokal terabaikan. Fenomena ini juga berimplikasi pada pelafalan nama-nama tempat, jalan, atau tokoh yang menggunakan kombinasi huruf menurut standar fonetis bahasa Inggris atau bahasa lain yang dominan dalam korpus dan sistem AI.

Akibatnya, pengucapan menjadi tidak konsisten, bahkan menimbulkan distorsi makna. Dengan demikian, kelemahan pada tahap GTP tidak hanya menunjukkan keterbatasan teknis dari sisi algoritma, tetapi juga mencerminkan masalah representasi data, di mana bahasa Indonesia belum mendapatkan porsi yang memadai dalam pemodelan linguistik secara global. Hal ini menegaskan perlunya pengembangan model GTP yang lebih peka terhadap variasi ejaan lokal serta didukung dengan data fonetis bahasa Indonesia agar sistem TTS mampu menghasilkan pelafalan yang akurat dengan konteks kebahasaan.

d. Prediksi Prosodi

Pada tahap ini, AI menambahkan informasi tentang intonasi, ritme, dan penekanan pada urutan fonem yang telah dihasilkan sebelumnya. Prosodi adalah "melodi" dari sebuah bahasa; tanpanya, ujaran akan terdengar datar dan monoton seperti suara robot. Dalam konteks navigasi, prosodi membantu membedakan antara pertanyaan, perintah, atau pernyataan. Misalnya, nada suara akan dinaikkan pada akhir kalimat seperti "Belok kanan?" jika itu adalah pertanyaan, tetapi akan datar dan tegas pada kalimat "Belok kanan" yang merupakan sebuah instruksi. Hal ini sejalan

dengan penelitian yang telah dilakukan oleh Bekmurodovna, dkk (2024), dalam penelitiannya yang berjudul *Linguistics in the 21st Century: Artificial Intelligence and Natural Language Processing*, yang menyatakan bahwa analisis berbasis komputasi prosodi ini memungkinkan peneliti untuk menemukan keteraturan maupun variasi bahasa yang sebelumnya sulit diidentifikasi dengan metode tradisional karena keterbatasan skala dan kompleksitas data.

Sayangnya, dalam praktiknya, sistem TTS pada Google Maps masih sering menghasilkan prosodi yang tidak konsisten, sehingga suara terdengar kaku, kurang alami, atau bahkan menimbulkan ambiguitas makna. Hal ini mengindikasikan bahwa meskipun prosodi sudah diintegrasikan dalam tahap sintesis suara, model AI belum sepenuhnya mampu menangkap prosodi bahasa Indonesia, yang memiliki pola intonasi dan penekanan berbeda dengan bahasa lain yang lebih dominan dalam data referen. Dengan demikian, pengembangan prosodi dalam TTS menjadi aspek penting untuk dikembangkan sesuai dengan pengalaman navigasi yang tidak hanya akurat secara teknis, tetapi juga alami, mudah dipahami, dan selaras dengan kebiasaan tutur penutur bahasa Indonesia.

e. Sintesis Ujaran

Setelah semua data linguistik, fonem, serta prosodi berhasil diproses, tahap terakhir dalam sistem *Text-to-Speech* (TTS) adalah sintesis ujaran. Pada tahap inilah seluruh informasi abstrak yang sebelumnya hanya berupa representasi teks dan simbol fonetis diubah menjadi gelombang suara yang dapat didengar oleh pengguna. Dengan kata lain, sintesis ujaran merupakan titik krusial di mana kecerdasan buatan (AI) benar-benar menciptakan suara yang menyerupai suara manusia.

Proses ini tidaklah sederhana, karena AI harus menyatukan ribuan fonem yang telah dipetakan, kemudian mengombinasikannya dengan intonasi, ritme, dan penekanan yang tepat agar menghasilkan tuturan yang tidak hanya akurat secara linguistik, tetapi juga terdengar alami dan mudah dipahami. Hal ini sejalan dengan penelitian yang telah dilakukan oleh Badawi (2021), yang menyatakan bahwa keterbatasan dalam merepresentasikan pengetahuan menggunakan basis data relasional dapat diatasi dengan pemodelan data ontologi. Ontologi mampu menggambarkan bentuk struktur hierarkis yang memuat definisi, hubungan antarentitas, karakteristik atau properti. Dengan demikian, melalui model ontologi, data yang disimpan dapat dipandang sebagai wujud konkret dari pengetahuan karena memiliki makna, saling berhubungan, dan terkait dengan data lainnya.

Kompleksitasnya semakin meningkat ketika sistem harus memastikan adanya kelancaran transisi antar-fonem, kejelasan artikulasi, serta keutuhan makna kalimat yang disampaikan. Kualitas sintesis ujaran sangat bergantung pada dua aspek utama, yaitu arsitektur model yang digunakan dan ketersediaan data pelatihan. Semakin banyak data suara manusia yang digunakan, semakin besar kemungkinan model untuk menghasilkan suara yang natural, ekspresif, dan mirip dengan pola tutur manusia. Dengan demikian, tahap sintesis ujaran bukan hanya masalah teknis semata, melainkan juga menyangkut pengalaman pengguna secara langsung, sehingga pengembangan model TTS yang lebih adaptif terhadap kekayaan bahasa dan prosodi lokal menjadi kebutuhan yang mendesak agar teknologi dapat berfungsi secara optimal di berbagai konteks budaya dan linguistik.

2. Pengaruh Pemodelan Linguistik (*Typing Models*)

Model linguistik, atau yang sering disebut sebagai *typing models*, memegang peran penting dalam menentukan kualitas ujaran yang dihasilkan oleh sistem *Text-to-Speech* (TTS) pada Google Maps. Model ini dapat diibaratkan sebagai “otak” yang pertama kali menerima masukan berupa teks, kemudian bertugas untuk mengidentifikasi struktur bahasa, menafsirkan makna, serta normalisasi data sebelum akhirnya diubah menjadi suara.

Dengan fungsi strategis tersebut, kualitas output TTS secara langsung dipengaruhi oleh kecanggihan dan kelengkapan model linguistik yang digunakan. Jika model tidak memiliki kamus leksikal yang kaya atau aturan linguistik yang akurat, maka berbagai bentuk singkatan, akronim, dan variasi ejaan akan diproses secara keliru. Contoh nyata dari kelemahan ini terlihat pada kasus singkatan “dr.” yang seharusnya dikenali sebagai “dokter” atau dalam beberapa konteks sebagai “jalan”, tetapi justru diperlakukan sebagai rangkaian huruf “de-er”.

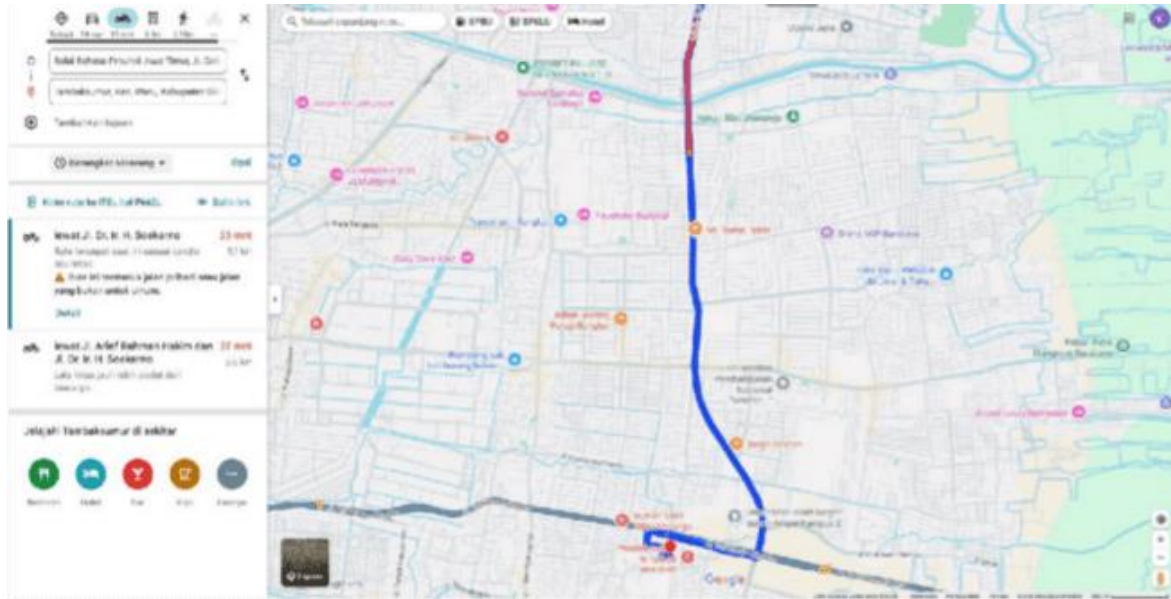
Kesalahan semacam ini menunjukkan keterbatasan model dalam memahami konteks kebahasaan, karena model hanya berfokus pada pola grafem tanpa memperhitungkan kemungkinan variasi semantik di baliknya. Akibatnya, output suara yang dihasilkan terdengar janggal, tidak alami, bahkan berpotensi membingungkan pengguna, terutama ketika digunakan dalam situasi navigasi yang membutuhkan kejelasan instruksi. Dengan kata lain, kualitas ujaran yang akurat, alami, dan sesuai dengan norma bahasa lokal sangat bergantung pada seberapa baik model linguistik dilatih untuk mengenali kekhasan suatu bahasa.

Jika *typing models* hanya bertumpu pada data global yang tidak merepresentasikan keragaman singkatan, ejaan lama, maupun nama-nama lokal di Indonesia, maka sistem TTS akan cenderung melakukan generalisasi konteks. Fenomena ini menegaskan bahwa pengembangan model linguistik yang lebih adaptif terhadap konteks bahasa Indonesia sesuai dengan kebutuhan. Tanpa adanya representasi data dan referen yang sesuai, sistem navigasi berbasis suara seperti Google Maps akan terus menghadapi masalah akurasi.

3. Tantangan terhadap Akronim dan Keterbatasan Data

Tantangan terbesar dalam implementasi sistem *Text-to-Speech* (TTS) pada layanan navigasi digital salah satunya terletak pada pengucapan akronim. Berbeda dengan kata biasa yang memiliki aturan pelafalan lebih stabil, akronim sering kali memiliki variasi pelafalan yang bergantung pada konteks sosial, kebiasaan penutur, serta konvensi linguistik yang berlaku secara lokal.

Dalam konteks bahasa Indonesia, permasalahan semakin kompleks karena banyak singkatan atau akronim yang tidak mengikuti kaidah fonetis baku, melainkan terbentuk dari praktik sehari-hari. Permasalahan utama muncul ketika sebuah akronim memiliki arti atau konteks yang khas, namun sistem AI tidak mampu mengidentifikasinya secara otomatis. Hal ini terutama disebabkan oleh keterbatasan data pelatihan yang digunakan dalam membangun model TTS. Jika referen tidak mencakup variasi pelafalan dari nama-nama lokal, singkatan populer, atau ejaan tidak standar yang digunakan di Indonesia, maka model akan cenderung melakukan generalisasi teks dengan cara melafalkan per huruf.



Fenomena "Tb." yang seharusnya dibaca sebagai "Tambak" tetapi justru dilafalkan "tebe" menjadi contoh nyata dari kegagalan ini. AI dalam kasus tersebut hanya mengenali urutan grafem tanpa memiliki pemahaman semantik maupun kontekstual yang lebih dalam. Kekeliruan semacam ini bukan sekadar masalah teknis, tetapi juga menyangkut aspek kepraktisan, karena pelafalan yang tidak sesuai berpotensi menimbulkan kebingungan dalam navigasi, terutama ketika pengguna berada di lokasi yang belum familiar. Untuk itu, peningkatan akurasi TTS memerlukan strategi pengembangan yang lebih serius, seperti penyusunan kamus akronim khusus untuk bahasa Indonesia, integrasi aturan linguistik berbasis konteks, serta pelatihan model dengan data yang lebih kaya, bervariasi, dan representatif terhadap realitas kebahasaan lokal.

Dengan demikian, penelitian ini memberikan gambaran komprehensif mengenai interaksi antara AI, bahasa, dan teknologi navigasi. Temuan yang diperoleh tidak hanya menjelaskan "apa" yang salah dalam panduan suara Google Maps, tetapi juga menyingkap "mengapa" kesalahan tersebut terjadi, baik dari sisi teknis pemrosesan bahasa maupun dari sisi keterbatasan representasi linguistik dalam data pelatihan yang digunakan oleh AI.

4. Peran *Machine Learning* dan Data Referen untuk Identifikasi Konteks

Proses pembuatan ujaran navigasi oleh kecerdasan buatan (AI) sangat erat kaitannya dengan ketersediaan data referensi serta metode mesin (*machine learning*) yang digunakan. Dalam praktiknya, miliaran data teks dari beragam sumber daring dimanfaatkan untuk melatih model bahasa agar mampu mengenali pola sintaksis, semantik, dan struktur kebahasaan. Di sisi lain, ribuan rekaman suara manusia dijadikan bahan untuk melatih model akustik, sehingga sistem dapat menghasilkan suara yang menyerupai intonasi, ritme, dan penekanan dalam ujaran manusia yang alami. Akan tetapi, penelitian ini mengungkapkan bahwa kualitas hasil TTS sangat dipengaruhi oleh keterbatasan data referen yang digunakan. Apabila data tidak cukup merepresentasikan keragaman bahasa Indonesia, terutama variasi pelafalan nama lokal, singkatan, atau ejaan yang tidak standar, maka sistem AI cenderung melakukan generalisasi.

Contoh konkret dari fenomena ini dapat dilihat pada pelafalan "dr." yang justru dibaca sebagai "de-er" alih-alih "dokter", serta singkatan "Tb" yang salah diucapkan sebagai "tebe" padahal dalam konteks toponimi wilayah Sukolilo, Surabaya, "Tb" pada umumnya berarti "Tambak". Kesalahan ini memperlihatkan bahwa model AI hanya memandang teks sebagai rangkaian huruf tanpa memahami makna kontekstual lokalnya. Implikasi dari temuan ini penting, karena menunjukkan bahwa keberhasilan sebuah sistem TTS tidak hanya ditentukan oleh kecanggihan algoritma, tetapi juga oleh kelengkapan, variasi, dan relevansi data. Oleh karena itu, untuk menghasilkan *output* suara yang lebih akurat dan alami, pengembangan model TTS di Indonesia perlu didukung dengan korpus bahasa yang lebih representatif, yang mencakup kekayaan singkatan, variasi ejaan, nama tempat, serta istilah khas lokal yang sering digunakan dalam kehidupan sehari-hari.

Fenomena kesalahan interpretasi konteks linguistik tidak hanya terjadi pada sistem *text-to-speech* (TTS), tetapi terjadi pada model generative AI berbasis *Large Language Models* (LLMs) yang digunakan dalam produksi teks. Sistem generatif seperti GPT dan LLaMa melakukan pemrosesan bahasa dengan pendekatan prediksi probabilistik berbasis token sehingga akurasi dalam memahami makna akronim sangat bergantung pada konteks data latihannya.

Dalam konteks bahasa Indonesia, akronim memiliki fungsi sosial dan pragmatik yang khas, seperti mengandung makna budaya atau toponimi lokal yang tidak secara eksplisit dapat disimpulkan dari bentuk leksikal. Misalnya, akronim "RS" dalam wacana kesehatan berarti Rumah Sakit sedangkan dalam konteks akademik dapat bermakna Riset Sosial. Perbedaan ini menunjukkan bahwa model generatif cenderung mengalami abiguitas semantik ketika berhadapan dengan bentuk-bentuk singkatan yang bersifat polisemantis dan kontekstual (Bender & Koller, 2020).

Secara teoretis kemampuan model dalam mengenali akronim dapat dijelaskan melalui pendekatan *Distributional Semantics* (Harris, 1954) dan teori *Contextual Embedding* (Devlin et al., 2019) menjelaskan bahwa makna kata diturunkan dari distribusional dalam korpus. Namun, dilihat dari perspektif sosiolinguistik dan pragmatik bahwa makna akronim tidak hanya ditentukan oleh distribusi leksikal, tetapi juga dari kesepakatan sosial dan fungsi komunikatif dalam masyarakat tutur (Leech, 1983; Mey, 2001). Penelitian perbandingan oleh Nurhadi dan Sari (2023) menunjukkan bahwa model GPT-3 menghasilkan kesalahan interpretasi terhadap akronim lokal bahasa Indonesia sebesar 31% lebih tinggi dibandingkan model dwibahasa yang dilatih langsung menggunakan korpus Indonesia-Inggris. Hal ini mengindikasikan bahwa sistem generative AI masih menghadapi keterbatasan dalam menangkap fungsi sosial bahasa dan inferensi pragmatik yang muncul dari interaksi manusia terutama dalam konteks budaya lokal. Dapat disimpulkan, pengembangan model generatif bahasa Indonesia sebaiknya mempertimbangkan integrasi antara pendekatan statistik dengan analisis pragmatik dan sosiolinguistik agar dapat menghasilkan interpretasi yang lebih kontekstual, relevan, dan komunikatif dalam representasi akronim.

KESIMPULAN

Penelitian ini menegaskan bahwa keberadaan fitur panduan suara Google Maps berbasis AI telah memberikan kemudahan besar bagi mobilitas masyarakat, namun sekaligus memperlihatkan keterbatasan signifikan dalam konteks kebahasaan, khususnya bahasa Indonesia. Akurasi pengucapan yang dihasilkan sistem *Text-to-Speech* (TTS) sering kali tidak memadai dan dipengaruhi oleh tiga faktor utama. Pertama, kegagalan

sistem dalam mengenali konteks singkatan menyebabkan pelafalan dilakukan secara harfiah, misalnya “Tb.” yang seharusnya dilafalkan “Tambak” justru dibaca “tebe.” Kedua, keterbatasan data leksikal membuat banyak nama diri atau nama tempat dengan ejaan tertentu mengalami distorsi, seperti “Soekarno” yang diucapkan “Sukarno.” Ketiga, kelemahan dalam tahap normalisasi teks menimbulkan penghilangan atau perubahan bagian nama, misalnya “Manyar Kerta Adi” yang dipotong menjadi “Kerto Adi.” Kesalahan semacam ini bukan hanya mengurangi kenyamanan, tetapi juga dapat menimbulkan kebingungan, terutama bagi pengguna di wilayah yang belum familiar.

Berdasarkan temuan tersebut, penelitian ini merekomendasikan tiga langkah strategis untuk meningkatkan akurasi sistem TTS. Pertama, penyempurnaan basis data leksikal yang lebih komprehensif, yang mencakup singkatan, akronim, nama tempat, dan nama tokoh dalam bahasa Indonesia. Kedua, integrasi model kontekstual yang lebih canggih agar AI mampu membedakan makna singkatan sesuai lingkungan teksnya, misalnya membedakan “dr” sebagai “dokter”, Dr. sebagai “Doktor”, atau Jl. sebagai “Jalan.” Ketiga, pemodelan AI dengan data suara dan teks yang lebih bervariasi akan menghasilkan pengucapan yang lebih natural, kontekstual, dan sesuai dengan kebiasaan masyarakat lokal.

Temuan ini sejalan dengan tren global dalam AI *localization*, yaitu upaya untuk menyesuaikan sistem kecerdasan buatan agar mampu memahami dan beradaptasi dengan konteks linguistik, sosial, dan budaya lokal dari pengguna di berbagai wilayah. Pendekatan ini menandai pergeseran paradigma dalam pengembangan AI. Mulai dari sistem yang bersifat universal dan berorientasi pada bahasa dominan (seperti bahasa Inggris), menuju sistem yang lebih kontekstual dan peka terhadap keragaman bahasa manusia.

Dalam konteks linguistik, arah pengembangan AI *localization* semakin nyata melalui hadirnya pemodelan bahasa seperti IndoBERT, IndoGPT, dan GPT-ID, yang secara khusus dirancang untuk memahami struktur bahasa Indonesia dalam segala kompleksitasnya. Model-model ini tidak hanya dilatih pada data berbahasa Indonesia, tetapi juga memperhitungkan morfologis (bentuk kata dan afiksasi), sintaksis (susunan kalimat), serta fonologis (pola bunyi) yang membedakan bahasa Indonesia dengan bahasa lain.

Dengan demikian, dapat disimpulkan bahwa persoalan akurasi ujaran navigasi bukan sekadar tantangan teknis, melainkan juga linguistik dan kultural. Pengembangan sistem TTS di masa depan harus mengintegrasikan pendekatan teknologi dengan sensitivitas linguistik komputasi, agar lebih inklusif terhadap keragaman bahasa lokal sesuai dengan toponimi. Implikasi dari penelitian ini tidak hanya penting bagi pengembang aplikasi navigasi digital, tetapi juga bagi kajian linguistik komputasional yang berupaya memahami interaksi antara bahasa dan teknologi, sekaligus membuka ruang penelitian lebih lanjut mengenai adaptasi AI terhadap dinamika bahasa lokal.

DAFTAR PUSTAKA

- Adriati, Rieke, et al. (2016). "Pengembangan Aplikasi Text-to-Speech Bahasa Indonesia Menggunakan Metode Finite State Automata Berbasis Android". *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 15(1), 1—8. https://www.researchgate.net/publication/309625428_Pengembangan_Aplikasi_Text-to-Speech_Bahasa_Indonesia_Menggunakan_Metode_Finite_State_Automata_Berbasis_Android.
- Ahmad, Hawraz, et al. (2024). "Planning the Development of Text-To-Speech Synthesis Models and Datasets with Dynamic Deep Learning". *Journal of King Saud University - Computer and Information Sciences*, ELSEIVER, 1—18. <https://www.sciencedirect.com/science/article/pii/S1319157824002209>.
- Badawi, A. (2021). "The Effectiveness of Natural Language Processing (NLP) as a Processing Solution and Semantic Improvement". *International Journal of Economic, Technology and Social Sciences*, 2(1), 36—44. <https://jurnal.ceredindonesia.or.id/index.php/injects/article/view/194>
- Bekmurodovna, Y. D., dkk. (2024). "Linguistics in the 21st Century: Artificial Intelligence and Natural Language Processing". *Web of Teachers: Inderscience Research*, 2(12), 366—370. <https://webofjournals.com/index.php/1/article/view/2616>
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185—5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bolshakov, I. A. dan Alexander, G. (2004). *Computational Linguistics Models, Resources, Applications*. Mexico: Universidad Nacional Autónoma. <https://www.gelbukh.com/clbook/Computational-Linguistics.pdf>
- Clark, A., Fox., Lappin. (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. United Kingdom: Blackwell Publishing. <http://ndl.ethernet.edu.et/bitstream/123456789/14228/1/28%20pdf.pdf>
- Deng, Li. (2006). *Dynamic Speech Models: Theory, Algorithms, and Applications*. United States of America: Morgan & Claypool Publisher. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Deng-Book2006.pdf>
- Desikan, B. S. (2018). *Natural Language Processing and Computational Linguistics: a Practical Guide to Text Analysis with Python, Gensim, spaCy and Keras*. Birmigham: Packt Publishing Ltd
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171—4186. <https://doi.org/10.48550/arXiv.1810.04805>
- Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. London: Springer.
- Gelbukh, Alexander. (2011). *Computational Linguistics and Intelligent Text Processing*. Springer-Verlag Berlin Heidelberg. <http://ndl.ethernet.edu.et/bitstream/123456789/11965/1/12pdef.pdf>
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2—3), 146—162. <https://doi.org/10.1080/00437956.1954.11659520>
- Kao dan Stephen. (2007). *Natural Language Processing and Text Mining*. London: Springer.
- Lane, H., Cole, dan Hannes. (2018). *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*. New York: Manning

- Publications. https://manning-content.s3.amazonaws.com/download/a/c9fc557-b088-4f1f-87a8-7ea2e488d262/Lane_NLPiA_MEAP_V10_ch1.pdf
- Leech, G. N. (1983). *Principles of pragmatics*. London: Longman.
- Liddy, E. (2001). "Natural Language Processing". In *Encyclopedia of Library and Information Science, 2nd Ed.* NY. Marcel Decker, Inc: School of Information Studies (iSchool).
https://surface.syr.edu/cgi/viewcontent.cgi?params=/context/istpub/article/1043/&path_info=Natural_Language_Processing.pdf
- Malar, N., Swetha, dan Yamini. (2020). "Computational Linguistic". *International Journal of Advance Research and Innovative Ideas in Education*, 2544—2549.
- Mitkov, R. (2003). *The Oxford Handbook of Computational Linguistics Second Edition*. Britania Raya: Oxford University Press.
https://api.pageplace.de/preview/DT0400.9780191625534_A43561407/preview-9780191625534_A43561407.pdf
- Mey, J. L. (2001). *Pragmatics: An introduction* (2nd ed.). Oxford: Blackwell.
- Nurhadi, A., & Sari, D. R. (2023). Comparative Evaluation of Acronym Interpretation in Multilingual Language Models. *Journal of Computational Linguistics and AI Studies*, 5(2), 112—128.
- Palivela, Hemant. (2021). "Optimization of Paraphrase Generation and Identification Using Language Models in Natural Language Processing". *International Journal of Information Management Data Insights, ELSEIVER*, 1—9.
<https://www.sciencedirect.com/science/article/pii/S2667096821000185>
- Wang, Xinsheng, et al. (2025). "Spark-TTS: an Efficient LLM-Based Text-to-Speech Model with Single-Stream Decoupled Speech Tokens". *ARXIV: Cornell University*.
<https://arxiv.org/html/2503.01710v1>