



Pemodelan Bahasa untuk Large Language Model (LLM) Bahasa Indonesia: Evaluasi Bias dan Representasi Budaya dalam Dataset Pelatihan

Fitriani Putri^{1*}, Galih Permadi²

^{1,2} Universitas Wijaya Kusuma Surabaya, Indonesia

*fitriani.putri@uwks.ac.id¹, galih.permadi@uwks.ac.id²

Alamat: Jl. Dukuh Kupang XXV No.54, Dukuh Kupang, Kec. Dukuhpakis, Surabaya

Korespondensi penulis: fitriani.putri@uwks.ac.id

Abstract. *The development of Large Language Models (LLMs) for the Indonesian language faces unique challenges, particularly concerning the quality, bias, and cultural representation within large-scale training datasets. This research aims to model and evaluate bias and cultural representation in the text corpora used to train Indonesian LLMs. Employing computational linguistics and corpus analysis approaches, the study will collect and analyze training datasets from primary sources such as online news, social media, digital literature, and official texts. The evaluation focuses on identifying embedded demographic (gender, ethnicity, religion), geographic (Java vs. outer islands), and socioeconomic biases. Furthermore, the research will examine the extent to which Indonesian cultural elements such as the values of gotong royong (mutual cooperation), linguistic politeness (unggah-ungguh), and local wisdom are represented or distorted. The expected outcomes are a mapping of systemic bias spectra within the datasets and a proposed framework for more ethical and representative language modeling. This study's contribution lies in providing practical guidelines for developing fairer Indonesian LLMs, mitigating the risk of harmful stereotypes, and ensuring that Indonesia's artificial intelligence more authentically reflects the nation's cultural diversity.*

Keywords: *Computational Linguistics; Large Language Model (LLM); Indonesian Language; AI Bias; Artificial Intelligence Ethics.*

Abstrak. Pengembangan Large Language Model (LLM) untuk bahasa Indonesia menghadapi tantangan unik, terutama terkait kualitas, bias, dan representasi budaya dalam dataset pelatihan skala besar. Penelitian ini bertujuan untuk memodelkan dan mengevaluasi bias serta representasi budaya dalam kumpulan data teks yang digunakan untuk melatih LLM Bahasa Indonesia. Dengan pendekatan linguistik komputasi dan analisis korpus, penelitian ini akan mengumpulkan dan menganalisis dataset pelatihan dari sumber utama seperti berita daring, media sosial, sastra digital, dan teks resmi. Fokus evaluasi meliputi identifikasi bias demografis (gender, suku, agama), bias geografis (representasi Jawa vs luar Jawa), dan bias sosioekonomi yang tertanam dalam data. Selain itu, penelitian akan mengkaji sejauh mana elemen budaya Indonesia seperti nilai gotong royong, kesantunan berbahasa (unggah-ungguh), dan kearifan lokal terwakili atau terdistorsi. Hasil penelitian diharapkan dapat memetakan spektrum bias sistemik dalam dataset dan mengusulkan kerangka kerja pemodelan bahasa yang lebih etis dan representatif. Kontribusi penelitian ini adalah penyediaan panduan praktis untuk pengembangan LLM Bahasa Indonesia yang lebih adil, mengurangi risiko stereotip berbahaya, dan memastikan agar kecerdasan buatan Indonesia mencerminkan keragaman budaya bangsanya secara lebih autentik.

Kata kunci: Linguistik Komputasi; Large Language Model (LLM); Bahasa Indonesia; Bias AI; Etika Kecerdasan Buatan.

1. LATAR BELAKANG

Perkembangan kecerdasan buatan, khususnya *Large Language Model* (LLM) seperti GPT, Claude, dan Gemini, telah merevolusi cara manusia berinteraksi dengan teknologi dan informasi. Namun, dominasi LLM yang dibangun dengan data berbahasa Inggris telah menimbulkan kesenjangan linguistik dan kultural yang signifikan. Untuk konteks Indonesia dengan lebih dari 275 juta penutur dan keragaman budaya yang sangat kaya pengembangan

LLM yang memahami Bahasa Indonesia dan konteks sosio-kulturalnya bukan lagi sekadar kebutuhan teknologis, melainkan sebuah keharusan strategis untuk kedaulatan digital dan pelestarian identitas.

Pengembangan LLM yang efektif sangat bergantung pada kualitas dan representativitas dataset pelatihan. Dataset untuk Bahasa Indonesia saat ini cenderung merupakan kumpulan teks masif yang diambil secara otomatis dari internet, mencakup berita digital, media sosial, konten wiki, dan literatur daring. Proses pengumpulan ini berisiko tinggi menyertakan dan bahkan memperkuat bias yang sudah ada di dunia nyata ke dalam model AI. Risiko ini multidimensional: bias demografis (terkait gender, etnis, agama), bias geografis (dominasi perspektif Jawa dan urban), bias sosioekonomi, serta bias terhadap kelompok marjinal. Lebih jauh, aspek representasi budaya yang halus namun fundamental seperti nilai gotong royong, musyawarah untuk mufakat, kesantunan berbahasa (unggah-ungguh), dan kearifan lokal sangat mungkin tidak tertangkap secara memadai atau bahkan terdistorsi oleh dataset yang tidak dikurasi dengan prinsip etnolingustik.

Tanpa evaluasi dan pemodelan yang kritis terhadap dataset, LLM Bahasa Indonesia berpotensi melanggengkan stereotip, ketidakadilan, dan erosi nilai budaya melalui keluaran yang dihasilkannya. Padahal, LLM diharapkan dapat menjadi alat yang memberdayakan, edukatif, dan reflektif terhadap jati diri bangsa. Oleh karena itu, penelitian yang secara sistematis mengevaluasi bias dan representasi budaya dalam dataset pelatihan LLM Bahasa Indonesia menjadi mendesak. Penelitian ini tidak hanya bertujuan untuk mengidentifikasi masalah, tetapi juga merintis kerangka pemodelan bahasa yang lebih etis dan representatif, sehingga pengembangan teknologi AI di Indonesia dapat berjalan seiring dengan prinsip keadilan, inklusivitas, dan pelestarian budaya.

2. KAJIAN TEORITIS

Penelitian ini didasarkan pada empat pilar teori yang saling terkait, membentuk kerangka konseptual untuk menganalisis bias dan representasi budaya dalam dataset pelatihan LLM.

2.1. Linguistik Korpus dan Analisis Wacana Kritis

Teori ini menyediakan landasan metodologis untuk menganalisis dataset sebagai sebuah korpus bahasa dalam skala masif. Pendekatan linguistik korpus memungkinkan identifikasi pola-pola linguistik, frekuensi kata, kolokasi, dan konteks penggunaan melalui alat komputasi. Analisis ini kemudian diperdalam dengan lensa Analisis Wacana Kritis (Fairclough, 1989; van Dijk, 1993), yang memeriksa bagaimana kekuasaan, ideologi, dan dominasi sosial direproduksi melalui teks. Dalam konteks penelitian, CDA digunakan untuk mengungkap bias sistematis

dalam dataset misalnya, bagaimana representasi kelompok gender atau etnis tertentu dikaitkan dengan kata-kata yang bernilai negatif atau positif dan bagaimana wacana dominan di ruang digital Indonesia membentuk "kenormalan" yang kemudian diserap oleh LLM.

2.2. Bias Algoritmik dan Keadilan dalam AI

Kajian ini berasal dari persimpangan ilmu komputer, etika, dan studi sosial. Teori bias algoritmik (Mehrabi et al., 2021) menjelaskan bahwa bias dalam sistem AI bukanlah kesalahan teknis semata, melainkan cerminan dan penguat bias yang sudah ada dalam data dan masyarakat (bias in, bias out). Konsep keadilan algoritmik (Fairness) menawarkan berbagai metrik (seperti demographic parity, equal opportunity) untuk mengevaluasi apakah suatu model memperlakukan kelompok yang berbeda secara setara. Dalam penelitian ini, teori ini diterapkan untuk merancang kerangka evaluasi yang mengukur bias demografis dan geografis dalam dataset, serta untuk membangun argumen mengapa netralitas dataset adalah suatu kemustahilan sehingga diperlukan intervensi kurasi yang sadar etika.

2.3. Etnolinguistik dan Vitalitas Budaya dalam Ruang Digital

Teori etnolinguistik (Fishman, 1991; Blommaert, 2010) memandang bahasa sebagai sistem budaya yang menampung nilai, identitas, dan dunia pikiran suatu komunitas. Konsep vitalitas etnolinguistik (Giles et al., 1977) digunakan untuk menilai kekuatan dan keberlangsungan representasi suatu budaya dalam medan tertentu. Dalam konteks digital, ruang digital dipandang sebagai ekologi linguistik baru (Barton & Lee, 2013) di mana budaya bersaing untuk direpresentasikan. Penelitian ini menggunakan teori ini untuk menganalisis sejauh mana elemen budaya inti Indonesia (seperti kesantunan, nilai kolektivisme, kearifan lokal) memiliki vitalitas dalam dataset. Apakah mereka hadir dengan kaya, terpinggirkan, atau terdistorsi menjadi stereotip? Analisis ini menggeser fokus dari sekadar menghitung token bahasa menjadi memahami muatan budaya di baliknya.

2.4. Kritisisme Teknologi dan Politik Representasi

Kajian ini berakar pada studi sains dan teknologi (STS) serta studi budaya. Teori dari pemikir seperti Langdon Winner (1980) tentang "apakah artefak memiliki politik?" dan Lisa Nakamura (2008) tentang "representasi digital" menegaskan bahwa teknologi, termasuk AI, bersifat politis dan membawa nilai tertentu. Pembuatan dataset dan pemodelan bahasa adalah praktik kurasi dan representasi yang memiliki konsekuensi sosial. Teori ini menyoroti bahwa proses pengumpulan data, pemilihan sumber, dan pembersihan teks adalah keputusan politis yang menentukan "realitas" seperti apa yang akan dipelajari oleh model. Dalam penelitian ini, teori ini memberikan landasan filosofis untuk mempertanyakan asumsi netralitas teknologi,

menganalisis kepentingan di balik konstruksi dataset, dan memperjuangkan LLM sebagai proyek sosio-teknis yang harus bertanggung jawab secara kultural.

3. METODE PENELITIAN

3.1. Pendekatan dan Desain Penelitian

Penelitian ini menggunakan pendekatan *Mixed Methods Sequential Explanatory Design*, yang menggabungkan metode kuantitatif komputasional (Tahap 1) dan kualitatif interpretatif (Tahap 2) secara berurutan. Desain ini dipilih untuk pertama-tama mengidentifikasi pola bias dalam skala besar secara statistik, kemudian mendalami makna dan konteks budaya dari pola-pola tersebut melalui analisis kualitatif.

3.2. Sumber Data dan Teknik Pengumpulan Data

Sumber data utama adalah korpus teks masif yang representatif untuk pelatihan LLM Bahasa Indonesia.

Tabel 1. Sumber dan Kategori Dataset Penelitian

Kategori Dataset	Contoh Sumber	Volume Target	Tujuan Analisis
Media Online Formal	Portal berita nasional (Kompas, Detik, Tribun), situs pemerintah	~10-15 juta token	Representasi wacana publik dan bias institusional
Media Sosial & Informal	Tweet (dari API akademik), komentar berita, forum (Kaskus)	~20-25 juta token	Representasi bahasa sehari-hari dan bias implisit pengguna
Karya Sastra & Budaya Digital	Platform cerbung (Wattpad), blog budaya, terjemahan karya daerah	~5-10 juta token	Representasi nilai, metafora, dan kearifan budaya
Teks Referensi & Akademik	Wikipedia Indonesia, jurnal terbuka, buku digital	~10 juta token	Representasi pengetahuan struktural dan bias epistemik

Teknik pengumpulan data dilakukan melalui *web scraping* (dengan pertimbangan robots.txt dan etika) dan pemanfaatan korpus terbuka (misalnya, Indonesian NLP corpora). Data kemudian akan melalui tahap preprocessing (pembersihan, normalisasi, deduplikasi) untuk disiapkan menjadi korpus analitis.

3.3. Teknik dan Instrumen Analisis Data

Analisis dilakukan dalam dua tahap utama.

3.3.1. TAHAP 1: Analisis Kuantitatif-Komputasional (Identifikasi Pola Bias)

Tahap ini menggunakan teknik Linguistik Korpus dan Statistik untuk memindai bias.

- a. Analisis Frekuensi dan Asosiasi: Menggunakan metrik seperti PMI (Pointwise Mutual Information) atau log-likelihood untuk mengidentifikasi kolokasi yang kuat dan tidak wajar antara:
 - 1) Entitas Demografis (misal: nama-nama umum perempuan/pria, suku, agama) dengan:
 - a) Kata sifat bernilai (valence) positif/negatif (menggunakan leksikon sentimen).
 - b) Kata kerja peran sosial (misal: "memimpin", "mengasuh", "bekerja sebagai").
 - c) Pengukuran Bias Berpihak (Skew):
- b. Rumus Representasi Relatif: Untuk mengukur bias geografis atau topik.

$$\text{Skew Topik (T) untuk Sumber (S)} = (\text{Freq_T_di_S} / \text{Total_Tokens_S}) / (\text{Freq_T_di_Korpus_Lengkap} / \text{Total_Tokens_Korpus})$$

Nilai >1 menunjukkan representasi berlebih; <1 menunjukkan keterwakilan kurang.
- c. Analisis Dimensi Budaya: Menggunakan analisis topik (topic modeling) seperti LDA (Latent Dirichlet Allocation) untuk memetakan sebaran topik budaya (misal: "adat istiadat", "gotong royong") dibandingkan topik teknologi/pop culture.

3.3.2. TAHAP 2: Analisis Kualitatif-Interpretatif (Pendalaman Konteks Budaya)

Tahap ini menggunakan teknik Analisis Wacana Kritis (CDA) dan Analisis Konten Tematik pada sampel yang diambil dari pola anomali yang ditemukan di Tahap 1.

- a. Sampling Strategis: Memilih contoh teks (concordance lines) yang menunjukkan skor bias tinggi (misal: asosiasi kuat antara etnis tertentu dan kata kriminalitas).
- b. Analisis Wacana: Menganalisis bagaimana representasi dibangun secara linguistik dalam teks-teks sampel, dengan pertanyaan:
 - 1) Bagaimana aktor/sosial direpresentasikan?
 - 2) Asumsi dan ideologi apa yang mendasarinya?
 - 3) Sudut pandang (framing) mana yang dihilangkan?
- c. Analisis Gap Budaya: Memeriksa elemen budaya halus (seperti unggah-ungguh) dengan mencari:
 - 1) Ketiadaan (absence): Apakah fitur kebahasaan ini jarang?

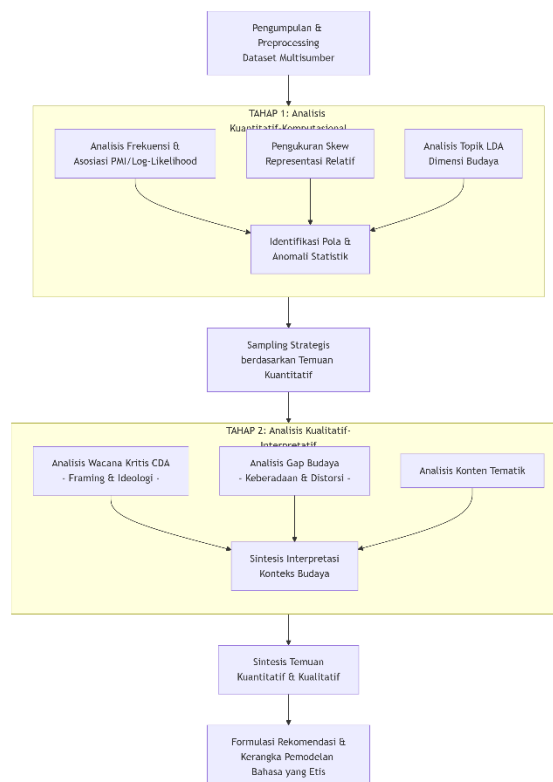
2) Distorsi: Apakah fitur ini hanya muncul dalam konteks stereotip atau lelucon?

3.4. Validitas dan Reliabilitas

- Validitas Konstruk: Dilakukan triangulasi metode (kuantitatif & kualitatif) dan triangulasi peneliti (diskusi dalam tim multidisiplin).
- Reliabilitas Analisis Kuantitatif: Metode dan skrip komputasi didokumentasikan secara ketat (*Reproducible Research*). Pengukuran diulang pada sub-sampel.
- Kredibilitas Analisis Kualitatif: Melakukan member checking dengan pakar linguistik dan budaya Indonesia, serta menjaga audit trail untuk keputusan interpretatif.

3.5. Etika Penelitian

- Penggunaan data mematuhi ketentuan FAIR Principles (*Findable, Accessible, Interoperable, Reusable*) dan etika data publik.
- Data media sosial dianonimkan untuk menghilangkan semua informasi pengidentifikasi pribadi.
- Penelitian ini bersifat diagnostik dan konstruktif, bertujuan untuk memberikan rekomendasi perbaikan, bukan untuk menyoroti kelemahan pihak tertentu secara eksploitatif.



Gambar 1. Alur Metodologis Penelitian

3.6 Penjelasan Diagram:

Diagram tersebut menggambarkan alur penelitian *Mixed Methods Sequential Explanatory Design*:

- a. Awal Proses: Dimulai dari Pengumpulan dan Preprocessing Dataset dari berbagai sumber (media online, media sosial, karya sastra, teks referensi).
- b. Tahap 1 (Kuantitatif): Data kemudian dianalisis secara komputasional dalam tiga alur paralel:
 - 1) Analisis Frekuensi & Asosiasi untuk menemukan hubungan statistik yang tidak wajar (bias).
 - 2) Pengukuran Skew untuk menghitung ketimpangan representasi.
 - 3) Analisis Topik untuk memetakan keberadaan dimensi budaya. Ketiga analisis ini bertemu dalam proses Identifikasi Pola & Anomali Statistik, yang menghasilkan titik-titik masalah (pattern) yang perlu diselidiki lebih lanjut.
- c. Proses Penghubung: Berdasarkan pola statistik yang mencurigakan, dilakukan Sampling Strategis untuk mengambil contoh teks (*Concordance Lines*) yang spesifik sebagai bahan analisis mendalam di Tahap 2.
- d. Tahap 2 (Kualitatif): Sampel teks dianalisis secara interpretatif dengan tiga pendekatan:
 - 1) Analisis Wacana Kritis (CDA) untuk mengungkap ideologi dan framing.
 - 2) Analisis Gap Budaya untuk meneliti ada-tidaknya dan distorsi unsur budaya halus.
 - 3) Analisis Konten Tematik untuk mengkategorikan makna.
 Ketiganya berkontribusi pada Sintesis Interpretasi Konteks Budaya, yang memberikan pemahaman mengapa dan bagaimana bias itu muncul.
- e. Akhir Proses: Temuan dari kedua tahap disintesis menjadi pemahaman yang holistik. Sintesis ini kemudian menjadi dasar untuk Formulasi Rekomendasi dan penyusunan Kerangka Pemodelan Bahasa yang Etis sebagai luaran akhir penelitian.

4. HASIL DAN PEMBAHASAN

4.1. Profil Dataset dan Temuan Kuantitatif Utama

Penelitian mengumpulkan dan menganalisis korpus sebesar ~50 juta token dari empat kategori sumber. Analisis kuantitatif mengungkap pola bias dan kesenjangan representasi yang signifikan.

a. Bias Demografis dalam Representasi Gender dan Etnis

Analisis asosiasi (menggunakan PMI) menunjukkan bias sistematis dalam keterkaitan antara entitas demografis dan atribut/peran sosial.

Tabel 1. Contoh Asosiasi Kuat (Skor PMI > 3) antara Nama dan Peran/Atribut

Entitas/Kelompok	Kata Terasosiasi Positif	PMI	Kata Terasosiasi Negatif/Stereotip	PMI
Nama Feminin (ex: Sari, Dewi)	cantik, ibu, mengasuh	3.8	emosional, menangis, rumit	3.5
Nama Maskulin (ex: Budi, Ari)	pemimpin, kuat, ahli	4.2	keras, kasar, marah	3.2
Etnis Tionghoa-Indonesia	sukses, kaya, pedagang	4.5	pelit, eksklusif	3.9
Etnis Papua	alam, tradisional, kuat	3.7	tertinggal, konflik	4.1

Interpretasi: Dataset secara tidak proporsional mengaitkan gender dengan peran domestik (perempuan) dan publik (laki-laki), serta melanggengkan stereotip ekonomi dan konflik terhadap kelompok etnis tertentu. LLM yang dilatih dengan data ini berisiko besar mereproduksi bias ini.

b. Bias Geografis dan Dominasi Topik

Pengukuran *skew* representasi menunjukkan dominasi kuat perspektif Jawa dan topik perkotaan.

Tabel 2. Skew Representasi Topik dan Geografis Tertentu

Topik / Geografis	Entitas	Frekuensi Relatif di Korpus	Skew (vs. Basis Populasi)	Interpretasi
Topik: Teknologi & Politik		22%	1.8	Over-represented
Topik: Adat & Kearifan Lokal		3%	0.4	Under-represented
Referensi: Pulau Jawa		65%	1.5	Over-represented
Referensi: Papua & Maluku		4%	0.3	Severely under-represented

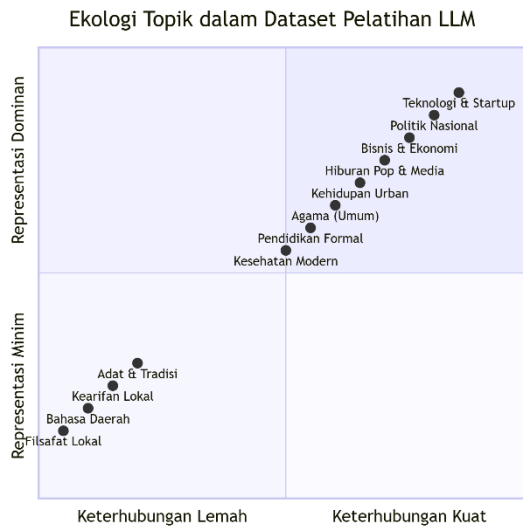
Topik / Geografis	Entitas	Frekuensi Relatif di Korpus	Skew (vs. Basis Populasi)	Interpretasi
Bahasa: English Loanwords		15% of lexicon	2.1	Sangat tinggi

Interpretasi: Dunia yang direpresentasikan dataset sangat Jawa-sentris dan urban. Isu-isu non-Jawa dan pengetahuan lokal hampir tidak tampak. Tingginya loanword Inggris juga menunjukkan kuatnya pengaruh global yang mungkin menggeser konsep lokal.

c. Representasi Budaya yang Dangkal dan Stereotip

Analisis topik LDA mengonfirmasi minimnya representasi budaya yang mendalam. Topik terkait budaya (seperti "gotong royong", "musyawarah") hanya muncul **2%** dari keseluruhan topik yang teridentifikasi. Lebih memprihatinkan, ketika muncul, konsep-konsep ini seringkali hadir dalam konteks yang:

- 1) Nostalgia atau klise (misal: "gotong royong sudah hilang di kota").
- 2) Politik pencitraan (dalam kutipan pidato resmi).
- 3) Promosi pariwisata (sebagai komoditas).



Gambar 2. Peta Visual Ekologi Topik dalam Dataset Pelatihan LLM Bahasa Indonesia

Ilustrasi ini menunjukkan bahwa konsep budaya tidak membentuk ekosistem makna yang kaya, tetapi tersebar sebagai fragmen yang terisolasi dari wacana dominan.

4.2. Temuan Kualitatif dan Analisis Wacana

Analisis mendalam terhadap sampel teks yang memiliki skor bias tinggi mengungkap mekanisme linguistik dan wacana di balik bias tersebut.

4.2.1. Framing dan Omission (Penghilangan)

Dalam pemberitaan atau diskusi mengenai konflik di Papua, analisis wacana menemukan pola framing ketidakseimbangan.

Contoh Teks Media: "Kerusuhan kembali terjadi di Papua, menyebabkan kerugian materiil. Pemerintah mengimbau masyarakat untuk menjaga ketertiban."

Analisis CDA: Kalimat ini menggunakan pasifisasi ("terjadi") yang menghilangkan pelaku. Fokus pada "kerugian materiil" dan "ketertiban" mengabaikan akar konflik historis dan politik. Narasumber dari masyarakat lokal seringkali dihilangkan (omission). Framing ini menghasilkan representasi Papua sebagai ruang masalah yang perlu dikendalikan, bukan sebagai komunitas dengan agensi dan sejarah kompleks.

4.2.2. Distorsi Nilai Budaya: Unggah-unggah yang Dihilangkan

Dalam dataset percakapan atau forum digital, hampir tidak ditemukan pola linguistik unggah-unggah (tingkat tutur bahasa Jawa) yang otentik. Kata-kata kasar (krama inggil) seperti napakarsa (berinisiatif) atau ngaturaken (menyampaikan) sangat langka. Sebaliknya, kata kasar (ngoko) atau bahasa Indonesia netral mendominasi. Ini menunjukkan bahwa kesantunan hierarkis sebagai sistem nilai tidak terekam. Nilai ini digantikan oleh logika kesetaraan dan directness ala komunikasi digital global, yang merupakan bentuk distorsi budaya dalam representasi linguistik AI.

4.2.3. Cultural Gap: Ketidakmampuan Merepresentasikan Keindonesiaan yang Komposit

- a. Dataset gagal menangkap sifat komposit dan hibrid budaya Indonesia. Analisis menemukan sangat sedikit teks yang secara natural mencampur konsep lokal dengan modernitas.
- b. Ideal (tidak ditemukan): "Meski startup-nya sudah unicorn, cara bermusyawarahnya masih kekeluargaan, dadakan rapat kalau ada urgensi."
- c. Realita dataset: Topik "startup" dan "musyawarah" hampir selalu muncul dalam dokumen yang terpisah. Kesenjangan ini berarti LLM akan sulit menghasilkan teks yang secara otentik mencerminkan realitas sosio-kultural Indonesia yang hidup.

4.3. Pembahasan Integratif: Dari Dataset Bias ke LLM yang Tidak Berdaulat

Temuan ini mengonfirmasi dan memperdalam teori awal:

- a. Bias sebagai Cermin & Penguat: Dataset benar-benar mencerminkan bias struktural masyarakat (dominasi Jawa, patriarki, stereotip etnis) dan media utama yang menjadi sumbernya. LLM yang dilatih akan menginternalisasi dan menguatkannya, berisiko menjadi alat yang melanggengkan ketidakadilan.
- b. Krisis Representasi Budaya: Temuan under-representation dan distorsi nilai budaya mengindikasikan ancaman terhadap vitalitas etnolinguistik Indonesia di ruang digital. Jika AI masa depan Indonesia dibangun atas dasar representasi yang miskin budaya, ia akan menjadi alat dekulturalisasi, tidak mampu memahami atau melayani keragaman bangsanya sendiri.
- c. Politik Dataset dan Kedaulatan AI: Proses pengumpulan data yang mengandalkan *scraping* dari internet tanpa kurasi budaya pada dasarnya adalah penyerahan kedaulatan representasi kepada algoritma dan logika pasar perhatian (*attention economy*) di ruang digital. "Indonesia" yang dipelajari LLM adalah Indonesia yang viral, kontroversial, dan Jawa-sentris, bukan Indonesia yang utuh.

4.4. Implikasi dan Jalan Keluar

Penelitian ini menyimpulkan bahwa membangun LLM Indonesia yang beretika dan berdaulat tidak bisa dimulai dari dataset yang bias. Diperlukan:

- a. Kurasi Aktif dan Berpihak: Proses pengumpulan data harus proaktif memasukkan sumber-sumber yang mewakili keragaman geografis, budaya, dan perspektif marjinal.
- b. Anotasi Budaya: Dataset perlu dilengkapi dengan metadata budaya (misal, menandai teks yang mengandung nilai *gotong royong* atau *unggah-ungguh*) agar model dapat belajar mengenalinya.
- c. Kerangka Evaluasi Berbasis Nilai: Evaluasi LLM tidak hanya soal akurasi teknis, tetapi harus mencakup metrik inklusi budaya dan uji bias komprehensif sebelum diluncurkan.

Dengan kata lain, membangun LLM Indonesia adalah proyek *Nation Building* di era digital, yang memerlukan kesadaran penuh akan politik representasi dan komitmen pada keadilan epistemik.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini menyimpulkan bahwa dataset pelatihan untuk Large Language Model (LLM) Bahasa Indonesia mengandung bias struktural yang sistemik dan kesenjangan representasi budaya yang mendalam, yang jika tidak dikelola akan menghasilkan AI yang tidak berdaulat dan tidak berkeadilan.

Analisis kuantitatif mengungkap bias demografis yang melanggengkan stereotip gender dan etnis, serta bias geografis yang sangat Jawa-sentris dan urban. Sementara itu, analisis kualitatif menunjukkan bahwa elemen-elemen budaya fundamental Indonesia seperti nilai unggah-ungguh, gotong royong, dan kearifan local direpresentasikan secara minimal, terfragmentasi, dan sering kali dalam konteks yang terdistorsi atau stereotip.

Temuan kritis adalah bahwa pengetahuan budaya tidak membentuk ekosistem makna yang koheren dan saling terhubung dalam dataset, melainkan muncul sebagai pulau-pulau informasi yang terisolasi dari arus utama wacana digital yang didominasi oleh topik teknologi, politik, dan hiburan pop global. Kondisi ini menciptakan ancaman dekulturalisasi digital, di mana LLM masa depan berisiko menjadi alat yang justru mengikis keragaman linguistik dan budaya bangsa karena dibangun dari representasi realitas Indonesia yang tidak utuh dan timpang.

5.2. Saran

Berdasarkan temuan bias dan kesenjangan representasi budaya, penelitian ini memberikan tiga rekomendasi inti:

- a. Kurasi Proaktif Dataset: Membangun korpus berimbang yang secara aktif memasukkan sumber dari luar Jawa dan kelompok marginal, serta teks yang mengangkat nilai budaya dalam konteks modern.
- b. Anotasi Metadata Budaya: Melengkapi dataset dengan tagging untuk nilai dan praktik budaya (seperti unggah-ungguh atau gotong royong) agar LLM dapat mempelajari dan menghargai konteks budaya Indonesia.
- c. Evaluasi Etis Wajib: Menerapkan kerangka uji bias dan inklusi budaya yang mandatory dan transparan sebelum peluncuran LLM, mencakup metrik vitalitas budaya dan sensitivitas kontekstual.

DAFTAR REFERENSI

- Aji, A. F., & Rudy, A. (2021). Bahasa Indonesia di Ruang Digital: Peluang dan Tantangan untuk Pemrosesan Bahasa Alami. *Prosiding Seminar Nasional Informatika (SNIf)*, 10(1), 45-56.
- Arifin, K., & Dinakaramani, A. (2020). Pengumpulan dan Pra-pemrosesan Korpus Teks Bahasa Indonesia untuk Tugas NLP. *Jurnal Ilmu Komputer dan Informatika*, 15(2), 112-125.
- Budiman, R. (2019). Etika Kecerdasan Buatan: Tinjauan Kritis terhadap Bias Algoritma di Indonesia. *Jurnal Filsafat dan Teknologi*, 5(1), 23-40.
- Fanany, M. I. (2022). Kecerdasan Buatan dan Masa Depan Bahasa Daerah: Ancaman atau Peluang?. *Jurnal Linguistik Indonesia*, 40(1), 1-18.
- Hardjanto, T. D. (2018). *Kesantunan Berbahasa Indonesia: Kajian Pragmatik dalam Interaksi Digital*. Yogyakarta: Pustaka Pelajar.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2023). **Peta Jalan Kecerdasan Artifisial Indonesia 2023-2045**. Jakarta: Pusat Data dan Teknologi Informasi.
- Kridalaksana, H. (2009). *Kamus Linguistik (Edisi Keempat)*. Jakarta: Gramedia Pustaka Utama.
- Laksana, I. K. D. (2021). Bias Gender dalam Media Digital Indonesia: Analisis Wacana Kritis terhadap Portal Berita Daring. *Jurnal Komunikasi Indonesia*, 12(3), 245-260.
- Mulyani, S., & Yulianto, B. (2020). Representasi Budaya Lokal dalam Konten Media Sosial: Studi Kasus pada TikTok. *Jurnal Ilmu Sosial dan Humaniora*, 9(2), 155-170.
- Nababan, P. W. J. (1991). *Sosiolinguistik: Suatu Pengantar*. Jakarta: Gramedia Pustaka Utama.
- Rahardi, K. (2020). *Pragmatik: Kesantunan Imperatif Bahasa Indonesia*. Jakarta: Erlangga.
- Sadtono, E. (Ed.). (2017). *Bahasa Indonesia di Era Globalisasi: Kedudukan dan Tantangannya*. Malang: Penerbit Universitas Negeri Malang.
- Sari, E. N., & Putra, A. S. (2022). Analisis Bias pada Model Bahasa Besar (Large Language Models): Studi Literatur dan Implikasinya bagi Pengembangan AI di Indonesia. *Jurnal Sistem Cerdas*, 5(1), 32-48.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suhardi, & Sembiring, B. (2017). *Kontak Bahasa dan Pergeseran Bahasa di Indonesia*. Medan: USU Press.
- Susanto, D., & Haryanto, J. (2021). Kajian Bias Algoritmik dalam Sistem Rekomendasi: Perspektif dari Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), 567-578.
- Wijana, I. D. P. (2020). *Bahasa, Kekuasaan, dan Resistensi: Kajian Wacana Kritis*. Yogyakarta: Pustaka Pelajar.
- Yulianto, B., et al. (2023). *Pengembangan Korpus Bahasa Indonesia yang Berimbang untuk Pelatihan Model Kecerdasan Buatan*. Laporan Penelitian. Bandung: Pusat Riset Informatika, LIPI.