Linguistik Indonesia, Agustus 2026, 363-374 *Volume ke-44, No. 2*

ISSN cetak 0215-4846; ISSN online 2580-2429

PEMANFAATAN AI DALAM PENYUSUNAN KAMUS AKRONIM DAN SINGKATAN

Nazarudin[1], R. Muhammad Arie[2]
Universitas Indonesia[1,2]
nazarudin.hum@ui.ac.id[1]; raden.m@ui.ac.id[2]

Abstract

This article presents a reflective case study on the use of artificial intelligence (AI) in compiling the Indonesian Dictionary of Acronyms and Abbreviations (Kamus Akronim dan Singkatan) by the Lexicology and Lexicography Laboratory of the Faculty of Humanities, Universitas Indonesia (LLL FIB UI). Rather than offering a purely conceptual review, this paper draws on the concrete experience of the compiling team in using three AI platforms—ChatGPT, DeepSeek, and Gemini—throughout the process of reviewing and expanding dictionary entries. Our findings suggest that AI does not merely accelerate lexicographic work; it reconfigures the nature of the work itself, shifting its center of gravity from manual information retrieval toward the management of choices and editorial decisions. Every AI output—whether accurate or hallucinated—generates new demands for selection, verification, and lexicographic judgment. Based on these findings, the article proposes a four-phase hybrid lexicographic model that positions AI at the initial exploration stage, while final decision-making authority remains with human lexicographers.

Keywords: *computational lexicography, artificial intelligence, acronyms, NLP*

Abstrak

Artikel ini menyajikan studi kasus reflektif tentang pemanfaatan kecerdasan buatan (AI) dalam penyusunan Kamus Akronim dan Singkatan oleh Laboratorium Leksikologi dan Leksikografi FIB UI (LLL FIB UI). Berbeda dari kajian yang semata-mata bersifat konseptual, tulisan ini bertumpu pada pengalaman konkret tim penyusun dalam menggunakan tiga platform AI—ChatGPT, DeepSeek, dan Gemini—selama proses pemeriksaan dan penambahan entri kamus. Temuan kami menunjukkan bahwa AI tidak sekadar mempercepat kerja leksikografi. Yang terjadi sesungguhnya lebih dari itu: AI merekonfigurasi jenis kerja yang dilakukan leksikografer, menggeser titik beratnya dari pencarian informasi secara manual menuju pengelolaan pilihan dan keputusan editorial. Setiap output AI—baik yang akurat maupun yang mengandung halusinasi—menciptakan kebutuhan seleksi, verifikasi, dan pertimbangan leksikografis baru yang tidak bisa diabaikan. Berdasarkan temuan ini, artikel mengusulkan model leksikografi hibrid empat fase yang menempatkan AI pada tahap eksplorasi awal, sementara otoritas keputusan akhir tetap berada di tangan leksikografer.

Kata Kunci: leksikografi komputasional, kecerdasan buatan, akronim, NLP

PENDAHULUAN

Siapa pun yang sering membaca berita, mengikuti rapat dinas, atau sekadar membuka media sosial, pasti akrab dengan satu fenomena yang kadang terasa seperti bahasa lain: akronim. Di Indonesia, akronim tumbuh begitu subur sehingga seseorang yang tidak mengikuti perkembangan

di satu bidang tertentu bisa merasa tersesat hanya karena tidak tahu kepanjangan dari serangkaian huruf kapital. Ini bukan gejala baru. Lebih dari setengah abad lalu, Anderson (1966) sudah mengidentifikasi bahwa akronim dalam bahasa Indonesia tidak sekadar berfungsi sebagai alat efisiensi komunikasi—ia juga menjadi penanda ideologis dan institusional yang penuh makna. Yang berubah sekarang hanyalah skalanya: di era media sosial, akronim baru lahir setiap hari dari interaksi jutaan pengguna platform digital.

Di sinilah paradoks yang menarik muncul. Akronim diciptakan untuk mempercepat komunikasi, tetapi kemunculannya yang tidak teratur dan tidak konsisten justru menciptakan hambatan baru. Bagi mereka yang tidak familiar dengan domain tertentu, akronim bisa terasa seperti pagar yang memisahkan, bukan jembatan yang menghubungkan (Hartmann & James, 1998). Dari kesadaran inilah Laboratorium Leksikologi dan Leksikografi FIB UI (LLL FIB UI) memulai penyusunan Kamus Akronim dan Singkatan yang merupakan sebuah proyek yang tidak hanya berdimensi leksikografis, tetapi juga merupakan upaya untuk menjaga aksesibilitas bahasa untuk semua kalangan.

Jika kita melihat ke belakang, leksikografi tradisional memiliki keterbatasan yang inheren ketika harus berhadapan dengan bahasa yang bergerak secepat ini. Metode manual yang bergantung sepenuhnya pada keahlian leksikografer sering kali tidak mampu mengimbangi laju kemunculan akronim baru. Terlebih lagi jika leksikografer harus menelusuri semua variasinya lintas domain. Svensén (2009) menyebut kondisi ini sebagai "the lexicographer's dilemma", yaitu gap yang tidak pernah terisi antara kelengkapan dan keterkinian. Di titik inilah kecerdasan buatan (AI) menawarkan kemungkinan yang menarik sekaligus menantang.

Potensi AI dalam leksikografi komputasional memang sudah banyak dibahas secara teoritis (Ide & Pustejovsky, 2017; Dembitz dkk., 2005). Namun, dokumentasi empiris tentang bagaimana AI benar-benar digunakan dalam proyek kamus berbahasa Indonesia masih sangat terbatas. Bagaimana tim leksikografi dengan nyata berinteraksi menggunakan platform AI? Keputusan-keputusan editorial apa yang muncul dari proses itu? Kemudian, apakah AI benar-benar mengubah cara kerja leksikografer, atau hanya mempercepat pekerjaan yang sudah ada? Pertanyaan-pertanyaan inilah yang menjadi fokus artikel ini.

Secara lebih spesifik, penelitian ini bertujuan menjawab tiga pertanyaan:

(1) Bagaimana AI digunakan secara konkret dalam proses penyusunan dan pemeriksaan entri Kamus Akronim dan Singkatan?
(2) Pilihan leksikografis dan keputusan editorial apa yang muncul dari penggunaan AI dalam proses tersebut?
(3) Bagaimana pilihan-pilihan itu mencerminkan pergeseran kerja leksikografer dalam model leksikografi akronim berbantuan AI?

KAJIAN PUSTAKA

Taksonomi Abreviasi, Akronim, dan Singkatan

Sebelum berbicara tentang bagaimana AI membantu pekerjaan leksikografis, ada baiknya kita berhenti sejenak pada pertanyaan yang tampak mudah namun ternyata tidak. Pertanyaan tersebut Adalah apa sebenarnya yang dimaksud dengan "akronim"? Jawabannya, ternyata, jauh dari seragam.

Kridalaksana (2010) membagi gejala pemendekan dalam bahasa Indonesia ke dalam lima kategori: singkatan, penggalan, akronim, kontraksi, dan lambang huruf. Klasifikasi ini berbeda

dari yang dipakai Notosusanto (1979) dan Badudu (1983), yang hanya membedakan dua jenis: singkatan dan akronim. Di ranah internasional, Bauer (1988) membagi abreviasi menjadi paduan dan akronim; Matthews (1997) menambahkan kategori pemenggalan; Haspelmath (2002) memasukkan alfabetisme sebagai kategori tersendiri; sedangkan Beard (1998) mengajukan model yang paling luas dengan empat kategori yang ia kelompokkan di bawah payung "ekspansi stok leksikal".

Keragaman ini bukan sekadar perbedaan istilah di permukaan. Ia mencerminkan pandangan yang berbeda-beda tentang cara kerja pembentukan kata, dan memiliki konsekuensi langsung terhadap bagaimana sistem AI dilatih. Dalam konteks Bahasa Indonesia, Dardjowidjojo (1979) dan Smith-Hefner (2010) mencatat kecenderungan penutur untuk menyamakan akronim dan abreviasi—sebuah kebiasaan yang dikonfirmasi oleh Abidin (2012) ketika ia mengategorikan "SVM" (Support Vector Machine) sebagai akronim, padahal secara formal bentuk itu seharusnya masuk kategori inisialisme karena dilafalkan huruf per huruf. Ketidakkonsistenan kategori seperti ini, jika menjadi dasar data pelatihan AI, akan menghasilkan model yang kurang cermat secara linguistik. Hal ini dalam kerja leksikografi dapat dianggap sebagai ketidakcermatan dan ketidakcermatan itu punya biaya yang nyata.

Leksikografi Preskriptif versus Deskriptif

Setiap tim yang menyusun kamus, sadar atau tidak, akan selalu berhadapan dengan pertanyaan mendasar, seperti haruskah kamus mencatat bahasa sebagaimana ia seharusnya digunakan, atau sebagaimana ia sungguh-sungguh digunakan? Inilah inti dari perdebatan antara pendekatan preskriptif dan deskriptif dalam leksikografi (Hartmann & James, 1998).

Pertanyaan ini sangat nyata dalam konteks Kamus Akronim dan Singkatan, terutama ketika berhadapan dengan akronim yang lahir dari media sosial. Perlu diakui bahwa bentuk-bentuk akronim yang hidup di ranah ini muncul dan berubah dengan cepat, bahkan sering kali melampaui kemampuan norma baku untuk mengejar. Svensén (2009) mengingatkan bahwa dikotomi preskriptif-deskriptif sebenarnya adalah penyederhanaan yang berlebihan. Lebih lanjut, dia menyatakan bahwa setiap keputusan leksikografis, termasuk dalam kamus yang mengaku sepenuhnya deskriptif, selalu mengandung dimensi normatif. Dengan kata lain, yang bisa dipilih bukan apakah suatu kamus bersifat normatif, melainkan dalam hal apa dan sejauh mana ia normatif.

Hal ini memiliki relevansi yang jelas bagi AI. Dengan kapasitasnya menelusuri pola frekuensi dalam korpus berskala besar, AI secara alami cenderung bekerja secara deskriptif dengan merekam apa yang ada. Namun, keputusan tentang apa yang layak masuk ke dalam kamus, definisi mana yang paling tepat, dan label domain apa yang harus disematkan, semua itu tetap membutuhkan pertimbangan manusia yang bernuansa normatif. Apalagi dalam konteks bahasa Indonesia, di mana Badan Pengembangan dan Pembinaan Bahasa (Badan Bahasa) memiliki mandat resmi dalam pembakuan bahasa. Hal ini menjadi sebuah dimensi sosio-politik yang sama sekali tidak bisa dinavigasi oleh sistem AI mana pun.

Paradigma NLP: Dari Aturan ke Statistik, dan Konsekuensinya

Perkembangan Pemrosesan Bahasa Alami (NLP) dapat dibaca sebagai pergeseran bertahap dari kepercayaan pada aturan eksplisit menuju kepercayaan pada pola statistik. Pendekatan berbasis aturan yang mendominasi NLP sejak tahun 1950-an hingga 1980-an memang menawarkan

transparansi. Artinya, setiap keputusan sistem dapat ditelusuri dan dijelaskan dalam istilah linguistik yang jelas. Namun Manning dan Schütze (1999) menunjukkan kelemahan mendasarnya, yaitu pendekatan ini sangat rapuh ketika menghadapi fenomena baru dan ketidakberaturan linguistic. Misalnya saja seperti ketidakberaturan yang berlimpah dalam akronim yang muncul di media sosial Indonesia.

Model-model generatif besar (Large Language Models/LLMs) seperti ChatGPT, Gemini, dan DeepSeek adalah produk dari paradigma statistik berbasis *deep learning* yang mendominasi NLP sejak pertengahan 2010-an. Fleksibilitas dan kemampuan generalisasinya memang mengesankan. Namun, tentu ada harga yang harus dibayar, yaitu model-model ini beroperasi seperti "kotak hitam" yang keputusannya sulit ditelusuri. Model-model ini cenderung menghasilkan output yang terdengar sangat meyakinkan tetapi secara faktual keliru, yang dikenal sebagai fenomena "halusinasi" (Bender dkk., 2021). Pengalaman tim dalam proyek ini mengonfirmasi hal tersebut secara langsung. Beberapa kasusnya akan diuraikan secara rinci dalam bagian Analisis dan Diskusi.

Leksikografi Komputasional: Apa yang Sudah Ada, dan di mana Posisi LLMs

Leksikografi komputasional bukan bidang yang baru. Ia berkembang sejak awal 1980-an, ditopang oleh proyek-proyek besar seperti WordNet (Fellbaum, 1998), FrameNet (Baker dkk., 1998), dan PropBank (Palmer dkk., 2005) yang meletakkan fondasi bagi integrasi antara leksikografi dan NLP. Kemudian, hal yang baru adalah kehadiran LLMs generatif yang bisa digunakan langsung oleh leksikografer biasa tanpa keahlian teknis khusus. Temuan ini dapat dianggap sebagai sebuah lompatan kualitatif yang cukup signifikan.

Studi kasus yang relevan datang dari Dembitz dkk. (2005), yang mendokumentasikan penggunaan alat NLP dalam penyusunan Croatian Encyclopaedic Dictionary. Proyek tersebut menunjukkan bagaimana sistem NLP bisa diadaptasi untuk mengidentifikasi kata-kata dalam definisi kamus yang belum terdaftar sebagai entri utama. Kasus ini dapat menjadi kasus paralel menarik yang dapat dibandingkan dengan pekerjaan tim LLL FIB UI, meskipun tantangan morfologis bahasa Kroasia dan bahasa Indonesia berbeda secara mekanisme.

Satu tantangan struktural yang perlu diakui sejak awal adalah posisi Bahasa Indonesia dalam hierarki sumber daya linguistik komputasional global. Meskipun memiliki lebih dari 270 juta penutur, Bahasa Indonesia masuk kategori bahasa "sumber daya menengah"—jauh di bawah bahasa-bahasa seperti Inggris atau Mandarin dalam hal ketersediaan data pelatihan berkualitas (Wibawa dkk., 2019). Konsekuensinya, LLMs komersial yang dilatih terutama pada data berbahasa Inggris memiliki keterbatasan spesifik dalam menangani akronim birokratis Indonesia, akronim media sosial lokal, dan fenomena campur kode Indonesia-Jawa/Sunda yang belum cukup terwakili dalam data pelatihan mereka.

METODOLOGI

Laboratorium Leksikologi dan Leksikografi FIB UI (LLL FIB UI) mendapatkan hibah dari FIB UI untuk mengembangkan diri. Dana hibah yang diperoleh digunakan LLL FIB UI sebagai laboratorium yang berfokus pada pengembangan studi perkamusan untuk menyusun sebuah kamus akronim dan singkatan yang direncanakan diberi judul Kamus Akronim dan Singkatan Bahasa Indonesia. Kamus ini sudah mulai disusun sejak tiga tahun yang lalu oleh Maman S. Mahayana, M.Hum. dan Nazarudin, M.A. Saat ini, LLL FIB UI membantu akselerasi

penyelesaian hingga penerbitan kamus. Sebelum diterbitkan, lema akronim yang sudah ada kembali ditinjau dan ditambahkan.

Penelitian ini menggunakan pendekatan studi kasus reflektif. Pilihan ini didasari oleh sifat data yang tersedia berdasarkan pengalaman kerja nyata sebuah tim leksikografi selama proses penyusunan kamus. Untuk menganalisis pengalaman itu secara sistematis, kami mengadaptasi model refleksi Rolfe dkk. (2001) yang terdiri dari tiga tahap: *what?* (apa yang terjadi?), *so what?* (apa maknanya?), dan *now what?* (apa implikasinya?). Tahap pertama digunakan untuk mendeskripsikan bagaimana AI sungguh-sungguh digunakan dalam proses penyusunan dan pemeriksaan entri [RQ1]. Tahap kedua menganalisis pilihan-pilihan dan keputusan editorial yang muncul dari output AI [RQ2]. Sementara itu, tahap ketiga merefleksikan apa yang dikatakan semua itu tentang perubahan sifat kerja leksikografer di era AI [RQ3].

Data penelitian bersumber dari proses penyusunan kamus di LLL FIB UI yang berlangsung sejak pertengahan 2024 hingga Maret 2026. Secara konkret, data tersebut mencakup: (1) dokumentasi proses kerja tim; (2) contoh prompt yang digunakan pada platform AI; (3) keluaran AI dari ChatGPT, DeepSeek, dan Gemini; (4) catatan evaluatif tim terhadap keluaran tersebut; (5) contoh entri atau kandidat entri yang diperiksa dengan bantuan AI; serta (6) sumber pembanding yang digunakan untuk memverifikasi kepanjangan, makna, dan konteks penggunaan akronim, mulai dari kamus domain-spesifik, situs resmi lembaga, hingga sumber tekstual primer. Satu hal yang perlu ditegaskan sejak awal mengenai cakupan penelitian ini adalah perbandingan antarplatform yang disajikan dalam artikel ini tidak dimaksudkan sebagai uji *benchmark* yang sistematis. Kami tidak merancang prosedur evaluasi dengan jumlah prompt yang distandarkan, kriteria penilaian yang terukur, atau evaluator independen. Perbandingan ini adalah bagian dari refleksi kasus sebagai dokumentasi pola perbedaan yang teramati selama proses kerja berlangsung. Kesenjangan yang ingin kami isi adalah minimnya catatan empiris tentang bagaimana AI benar-benar bekerja dalam proyek leksikografi akronim Indonesia, dan bagaimana penggunaannya mengubah struktur kerja editorial leksikografer. Artikel ini menjawab kesenjangan itu melalui dokumentasi reflektif berbasis pengalaman proyek yang nyata.

ANALISIS DAN DISKUSI

Bagaimana AI Digunakan dalam Proses Penyusunan Entri (RQ1)

Dalam proses penambahan dan pemeriksaan entri kamus, tim LLL FIB UI mencoba tiga platform AI versi gratis: ChatGPT, DeepSeek, dan Gemini. Dari penggunaan berulang, ChatGPT dan DeepSeek muncul sebagai yang paling bermanfaat dalam kelengkapan penyajian informasi—keduanya bergantian menjadi yang terlengkap tergantung jenis akronim yang ditanyakan.

Secara garis besar, AI dimanfaatkan untuk lima jenis tugas leksikografis, sebagaimana dirangkum dalam Tabel 1 berikut.

Tabel 1. Penggunaan AI dalam Tugas Leksikografis Proyek Kamus

Tugas Leksikografis	Contoh Penggunaan AI	Platform	Fungsi dalam Proyek
Eksplorasi kepanjangan (polisemi)	"ATR is an acronym for what?" → ChatGPT menghasilkan empat kepanjangan lintas domain: Avions de Transport Régional, Average True Range, ATM Transaction Record, Akademi Teknik Rontgen	ChatGPT	Memperluas kemungkinan kepanjangan dengan cepat, menjangkau domain yang tidak dikuasai leksikografer
Pemeriksaan entri lama	"Bisakah Anda mencari akronim dalam file ini?" → DeepSeek mengategorikan kata bui sebagai akronim	DeepSeek	Memunculkan anomali dan false positive yang perlu disaring secara manual
Identifikasi kesalahan data	AI menampilkan kepanjangan "airlines" untuk akronim yang dalam data tim tercatat "airways"	ChatGPT / Gemini	Memicu pengecekan ulang ke sumber resmi; berujung pada koreksi data
Halusinasi (false addition)	Gemini menambahkan entri BTN (Bank Tabungan Negara) meskipun akronim itu tidak muncul dalam teks sumber yang diberikan	Gemini	Menunjukkan perlunya verifikasi independen sebelum entri diterima
Penyusunan draf definisi	AI diminta membuat draf definisi untuk kandidat entri baru berdasarkan konteks penggunaan yang ditemukan	ChatGPT / DeepSeek	Membantu penyeragaman gaya definisi di tahap awal; masih perlu penyuntingan leksikografer

Tabel 1 memperlihatkan bahwa tim tidak memperlakukan AI sebagai satu-satunya sumber kebenaran. Ia digunakan sebagai alat eksplorasi awal yang memperluas kemungkinan pencarian sebagaimana sebuah jaring yang lebar. Setiap output kemudian diverifikasi secara independen sebelum diterima sebagai entri. Pola kerja seperti ini bukan tanpa konsekuensi terhadap distribusi waktu dan jenis kerja yang dilakukan tim.

Pilihan Leksikografis dari Output AI (RQ2)

Bagian inilah yang menjadi inti analisis artikel ini. Pertanyaan yang diajukan bukan sekadar "apakah AI benar atau salah?", melainkan "keputusan editorial apa yang dihasilkan dari output AI?" Perbedaan perspektif ini menjadi salah satu bagian yang penting. Svensén (2009) mengingatkan bahwa setiap keputusan leksikografis mengandung dimensi normative, dalam kasus proyek ini adalah AI tidak mengurangi jumlah keputusan yang harus dibuat, melainkan justru memperbanyaknya dengan menghadirkan beragam kemungkinan yang semuanya perlu ditimbang.

Lima kasus yang kami pilih dari proses kerja tim dapat dilihat secara lebih rinci dalam Tabel 2.

Tabel 2. Analisis Kasus: Output AI dan Keputusan Leksikografis Tim

Kasus	Output AI	Pilihan / Masalah yang Muncul	Keputusan Leksikografis
ATR	Empat kepanjangan: Akademi Teknik Rontgen, Average True Range, ATM Transaction Record, Avions de Transport Régional	Kepanjangan mana yang relevan untuk konteks Indonesia? Mana yang dapat diverifikasi dari sumber terpercaya?	Semua empat kepanjangan diterima setelah verifikasi dengan penambahan label domain (Kes., Eko., Hub.)
BTN	Gemini menambahkan BTN (Bank Tabungan Negara) yang tidak ada dalam teks sumber	Apakah entri yang dikenal secara umum boleh dianggap sebagai temuan dari teks sumber?	BTN ditolak sebagai temuan dari teks; masuk kamus berdasarkan sumber data lain yang terverifikasi
bui	DeepSeek mengategorikan bui sebagai akronim	Apakah bui benar-benar akronim atau sekadar false positive?	Ditolak; bui adalah kata biasa bahasa Indonesia, bukan hasil proses abreviasi apapun
airways / airlines	AI secara konsisten menampilkan kepanjangan "airlines", berbeda dari catatan "airways" dalam data tim	Apakah ada kesalahan dalam data awal tim?	Pengecekan ke sumber resmi mengonfirmasi bahwa "airlines" adalah penulisan yang benar
GIA	AI memberikan kepanjangan Garuda Indonesian Airways; data kamus mencatat nama yang sama	Apakah nama historis atau nama kontemporer yang digunakan sebagai entri?	Kepanjangan historis dipertahankan dengan keterangan konteks; label Hub. ditambahkan

Mari kita telusuri kasus-kasus ini satu per satu karena masing-masing mengungkapkan hal yang berbeda tentang dinamika kerja leksikografi berbantuan AI.

Kasus ATR adalah contoh paling jelas tentang nilai AI sebagai alat eksplorasi polisemi. Ketika tim mengetikkan prompt "ATR is an acronym for what?" ke ChatGPT, yang muncul bukan satu jawaban, melainkan empat kepanjangan dari domain yang berbeda-beda: Akademi Teknik Rontgen (bidang kesehatan), Average True Range (analisis teknikal pasar keuangan), ATM Transaction Record, dan Avions de Transport Régional (produsen pesawat asal Prancis). Dalam praktik leksikografi manual, menelusuri keempat kepanjangan itu dari domain yang berbeda akan membutuhkan konsultasi dengan pakar dan kamus domain-spesifik yang terpisah. Proses semacam ini tentunya bisa memakan waktu berhari-hari. AI menyajikannya dalam hitungan detik. Namun tim tetap harus memverifikasi satu per satu, memilih label domain yang tepat, dan memutuskan urutan kepanjangan berdasarkan frekuensi penggunaan di Indonesia. Pekerjaan

eksplorasi berubah menjadi pekerjaan seleksi yang menjadikannya lebih cepat, tetapi tidak lebih ringan.

Kasus BTN dan bui mengungkapkan dimensi yang sepenuhnya berbeda, yaitu adanya risiko halusinasi. Gemini menambahkan entri BTN (Bank Tabungan Negara) dalam daftar akronim yang ditemukannya, meskipun akronim itu sama sekali tidak muncul dalam teks sumber yang diberikan. Model menghasilkan informasi yang tampak relevan dan masuk akal (BTN memang akronim yang umum dalam konteks perbankan Indonesia), tetapi tidak memiliki basis faktual dalam data yang diberika. Hal inilah yang oleh Bender dkk. (2021) disebut sebagai "halusinasi". BTN akhirnya masuk ke dalam kamus, tetapi bukan karena output Gemini, melainkan karena ada dalam sumber data lain yang terverifikasi. Perbedaan sumber ini penting secara metodologis karena menyangkut keterlacakan entri.

Sementara itu, DeepSeek mengategorikan kata bui sebagai akronim. Padahal, secara linguistik, bui adalah kata biasa dalam bahasa Indonesia yang tidak merupakan hasil proses abreviasi apa pun. Ini adalah kasus *false positive* dalam *named entity recognition*, yaitu model memperlakukan pola tertentu, dalam hal ini penulisan yang pendek dan terasa "asing", sebagai penanda akronim tanpa verifikasi semantik yang memadai. Kasus ini mencerminkan keterbatasan nyata model-model yang terutama dilatih pada data berbahasa Inggris ketika berhadapan dengan kosakata bahasa Indonesia yang kaya.

Kasus airways/airlines, meskipun tampak sepele, justru mengilustrasikan salah satu manfaat yang paling tidak terduga dari penggunaan AI. Ketika AI secara konsisten menampilkan kepanjangan "airlines" untuk akronim yang dalam catatan tim tercatat "airways", perbedaan itu memicu pengecekan ulang. Hasilnya adalah penulisan yang benar memang "airlines". Dalam hal ini, AI tidak memberikan jawaban yang benar secara langsung, tetapi ia membuka pertanyaan yang membawa tim ke jawaban yang benar. Fungsi AI sebagai pemantik pengecekan ulang ini adalah salah satu manfaat yang tidak selalu diantisipasi, namun terbukti berharga dalam praktik.

Pergeseran Kerja Leksikografer (RQ3)

Setelah menelusuri kasus-kasus di atas, kita bisa melihat sebuah pola yang lebih besar. AI tidak datang menggantikan leksikografer, namun ia datang mengubah apa yang leksikografer kerjakan. Untuk memahami perubahan ini dengan lebih tepat, kita bisa menggunakan konsep *affordance* dari Hutchby (2001) yang mengembangkan gagasan Gibson (1979). Dalam kerangka ini, teknologi tidak hanya memiliki "fitur", namun ia membuka dan sekaligus membatasi kemungkinan tindakan tertentu. AI, dalam konteks proyek ini, memiliki *affordance* untuk memperluas pencarian kepanjangan lintas domain, memunculkan kemungkinan polisemi, menawarkan draf definisi, dan menunjukkan anomali dalam data.

Namun setiap *affordance* itu datang berpasangan dengan tuntutan baru. Semakin banyak kemungkinan yang dihasilkan AI, semakin banyak pula keputusan yang harus dibuat leksikografer. Keputusan-keputusan tersebut seperti mana yang sahih, mana yang relevan untuk konteks Indonesia, mana yang perlu diverifikasi lebih lanjut, dan mana yang harus ditolak sama sekali. Zuboff (1988) pernah mencatat bahwa teknologi informasi tidak hanya mengotomasi pekerjaan, namun ia juga mengubah cara kerja diketahui, dipantau, dan dikelola. Pengalaman tim ini mengonfirmasi hal itu secara konkret. Waktu yang dulu dihabiskan untuk mencari kepanjangan kini digunakan untuk mengevaluasi dan menyaring kemungkinan-kemungkinan yang tersaji.

Raisch dan Krakowski (2021) menunjukkan bahwa otomasi dan augmentasi tidak dapat dipisahkan secara sederhana dalam praktik organisasi nyata. Keduanya saling terkait dan saling mengubah. Dalam kasus ini, AI mengotomasi sebagian pekerjaan eksplorasi, tetapi augmentasi yang dihasilkannya, yakni kapasitas leksikografer untuk menjangkau lebih banyak domain dan kemungkinan, pada akhirnya dapat menghasilkan tuntutan kerja editorial yang lebih besar. Inilah yang oleh Ide dan Pustejovsky (2017) disebut sebagai *augmented lexicography*, yaitu ketika AI mengaugmentasi kapasitas leksikografer, bukan menggantikan pertimbangannya.

Tantangan Struktural

Sebagian besar keterbatasan yang ditemui tim dalam penggunaan AI dapat ditelusuri ke satu akar yang sama, yaitu posisi Bahasa Indonesia dalam hierarki sumber daya linguistik komputasional global. Meskipun bahasa ini dituturkan oleh lebih dari 270 juta orang, LLMs komersial dilatih terutama pada data berbahasa Inggris. Jika dibandingkan dengan dengan porsi bahasa Indonesia, ini sangat jauh tidak proporsional dengan jumlah penuturnya (Wibawa dkk., 2019). Hasilnya bisa dirasakan langsung, seperti terbatasnya pengetahuan model tentang akronim birokratis Indonesia, akronim media sosial lokal, dan fenomena campur kode Indonesia-Jawa/Sunda. Kemampuan model memahami konteks sosio-kultural yang dibutuhkan untuk mendisambiguasi akronim berpolisemi dalam situasi Indonesia pun belum memadai.

Di samping itu, ada tantangan teknis yang bersumber dari morfologi Bahasa Indonesia sendiri. Sistem afiksasi yang sangat produktif, misalnya dengan awalan (me-, di-, ke-, ter-), akhiran (-kan, -an, -i), dan konfiks (me-…-kan, ke-…-an), dapat menciptakan variasi bentuk yang tidak mudah ditangani oleh mekanisme lematisasi yang dirancang untuk bahasa-bahasa Indo-Eropa. Hal ini menjelaskan mengapa kasus *false positive* seperti bui bisa terjadi pada model yang kurang terlatih pada data Indonesia. Tanpa disadari, pola yang terlihat "tidak biasa" oleh model cenderung ditandai sebagai akronim, tanpa pemahaman semantis yang cukup.

Model Leksikografi Hibrid: Sebuah Tawaran Berdasarkan Pengalaman

Bertolak dari semua pengalaman yang telah diuraikan, kami mengusulkan model leksikografi hibrid empat fase. Model ini bukan sesuatu yang sepenuhnya baru secara konseptual. Model ini sudah mulai mengoperasionalkan prinsip-prinsip augmented lexicography (Ide & Pustejovsky, 2017) ke dalam konteks spesifik penyusunan kamus akronim bahasa Indonesia dengan LLMs generatif. Hal yang membedakannya adalah penekanannya pada reconfiguration of lexicographic labor. Selain itu, model ini tidak hanya menjelaskan bahwa AI dapat digunakan untuk apa, tetapi bagaimana penggunaannya menggeser distribusi kerja antara eksplorasi, seleksi, dan validasi dalam keseluruhan proses leksikografis.

Fase 1: Eksplorasi Berbasis AI. Pada tahap ini, AI difungsikan sebagai jaring yang lebar, misalnya untuk menelusuri kandidat akronim dari korpus berskala besar, mengeksplorasi polisemi dan kepanjangan alternatif, serta mengumpulkan konteks penggunaan dari berbagai domain. Outputnya adalah basis data kandidat yang kaya, namun belum terverifikasi.

Fase 2: Verifikasi dan Kurasi oleh Leksikografer. Fase ini merupakan fase yang paling membutuhkan keahlian manusia. Setiap kandidat diperiksa terhadap sumber independen yang dapat dipercaya. Keputusan tentang validitas kepanjangan, relevansi konteks, dan tingkat

formalitas entri tidak bisa didelegasikan ke AI. Fase ini juga mencakup identifikasi dan eliminasi halusinasi serta false positive yang dihasilkan AI.

Fase 3: Pengayaan dan Standardisasi Definisi. Setelah entri terverifikasi, AI bisa kembali dilibatkan untuk membantu memformulasikan definisi yang konsisten, mengidentifikasi relasi semantik antarentri, dan menyarankan label domain dan register yang tepat. Keputusan akhir tentang formulasi tetap berada di tangan leksikografer.

Fase 4: Pembaruan Berkelanjutan. Kamus yang baik bukan kamus yang statis. Pada fase ini, AI bisa difungsikan untuk memantau korpus media sosial dan teks baru, mengidentifikasi kemunculan akronim baru atau pergeseran makna yang ada. Hal ini dilakukan untuk kemudian diverifikasi oleh tim leksikografi sebelum masuk sebagai bagian dari pembaruan kamus.

Dalam model ini, posisi AI dan leksikografer masing-masing jelas, yaitu AI unggul dalam kecepatan, cakupan, dan konsistensi pemrosesan data berskala besar. Sementara itu, leksikografer unggul dalam penilaian linguistik yang bernuansa, sensitivitas kontekstual, dan keputusan normatif tentang kualitas entri. Keduanya tidak saling menggantikan, melainkan mereka saling melengkapi sehingga masing-masing ada di tempat yang sesuai.

KESIMPULAN DAN IMPLIKASI KEBIJAKAN

Perjalanan penyusunan Kamus Akronim dan Singkatan oleh tim LLL FIB UI mengajarkan beberapa hal yang, kami kira, relevan melampaui proyek kamus ini sendiri. Pertama, AI memang berguna dan memiliki manfaat yang nyata. Dalam pekerjaan eksplorasi polisemi, pendeteksian anomali data, dan pembuatan draf definisi awal, AI berkontribusi secara substantif dalam mempercepat tahap-tahap yang secara tradisional sangat padat waktu. Leksikografer yang mengabaikan alat ini sama saja dengan menolak menggunakan teropong padahal tugas mereka adalah memandang jauh.

Kedua, AI juga mengandung risiko yang nyata dan tidak boleh diremehkan. Halusinasi dan false positive, seperti yang dialami tim dengan kasus BTN, bui, dan entri yang tidak ada dalam teks sumber, bisa secara diam-diam mencemari database kamus dengan informasi yang tidak terverifikasi. Profil kesalahan setiap platform juga berbeda, misalnya DeepSeek cenderung menghasilkan false positive kategorisasi; Gemini rentan pada halusinasi faktual; dan ChatGPT meskipun lebih akurat namun kadang memberikan informasi lebih dari yang diminta. Memahami profil kesalahan masing-masing platform adalah bagian dari literasi yang harus dimiliki leksikografer di era ini.

Ketiga, yang juga kami pandang sebagai kontribusi utama artikel ini, AI tidak sekadar mempercepat kerja leksikografi. Ia merekonfigurasinya. Kerja leksikografer bergeser dari pencarian informasi secara manual menuju pengelolaan pilihan dan keputusan editorial. Paradoksnya adalah semakin canggih AI, semakin banyak kemungkinan yang ia hasilkan, dan semakin besar pula tuntutan pertimbangan manusia yang diperlukan untuk menyaringnya. AI tidak mengurangi kebutuhan akan leksikografer yang kompeten, namun di sisi lain, ia justru meningkatkan tuntutan terhadap kompetensi editorial mereka.

Dari temuan ini, ada beberapa implikasi kebijakan yang perlu diperhatikan. Pertama, pengembangan korpus bahasa Indonesia yang berkualitas tinggi dan beragam domain adalah investasi yang tidak bisa ditunda. Tanpa korpus yang memadai, model AI, secanggih apapun arsitekturnya, tidak akan pernah bisa bekerja secara andal untuk tugas-tugas berbahasa Indonesia.

Hal ini membutuhkan sinergi antara institusi akademik, Badan Bahasa, dan industri teknologi. Kedua, ada kebutuhan mendesak untuk mengembangkan model NLP yang dioptimasi secara spesifik untuk bahasa Indonesia dan bahasa-bahasa daerah Indonesia, bukan sekadar mengadaptasi model yang dibangun untuk bahasa lain. Terakhir, proyek Kamus Akronim dan Singkatan sendiri berpotensi menjadi dataset berlabel yang sangat berharga untuk melatih dan mengevaluasi model NLP berbahasa Indonesia. Hal ini merupakan sebuah nilai tambah yang mensyaratkan struktur data yang kompatibel dengan standar representasi komputasional sejak tahap awal penyusunan.

Ke depan, evaluasi yang lebih sistematis dengan protokol yang lebih ketat, seperti jumlah *prompt* terkontrol, kriteria penilaian yang terstandar, dan evaluator independen misalnya diperlukan untuk mengukur performa komparatif platform AI secara lebih akurat. Proyek LLL FIB UI, dengan pengalaman reflektif yang didokumentasikan dalam artikel ini, menawarkan titik tolak yang konkret dan empiris untuk penelitian lanjutan tersebut.

CATATAN

Penulis mengucapkan terima kasih kepada mitra bestari yang telah memberikan saran dan masukan berharga untuk perbaikan artikel ini.

DAFTAR PUSTAKA

Abidin, T. F. (2012). Penentuan secara otomatis akronim dan ekspansinya dari data teks berbahasa Indonesia. Prosiding SNATIKA, 1(01).

Anderson, B. (1966). The languages of Indonesian politics. *Indonesia, (1)*, 89–116.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet project. Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics, 1, 86–90.

Badudu, J. S. (1983). *Pelik-pelik Bahasa Indonesia*. Pustaka Prima.

Bauer, L. (1988). *Introducing Linguistic Morphology*. Edinburgh University Press.

Beard, R. (1998). Derivation. Dalam A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (hlm. 44–65). Blackwell.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Dardjowidjojo, S. (1979). Acronymic patterns in Indonesian. Dalam Nguyễn Đ. L. (Ed.), Southeast Asian Linguistic Studies (Vol. 3). Pacific Linguistics, The Australian National University.

Dembitz, Š., Jojić, Lj., & Pavlek, J. (2005). Artificial intelligence in lexicography: A Croatian encyclopaedic dictionary example. Proceedings of the 16th International DAAAM Symposium: Intelligent Manufacturing & Automation, 19–22 Oktober 2005.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Hartmann, R. R. K., & James, G. (1998). *Dictionary of Lexicography*. Routledge.

Haspelmath, M. (2002). *Understanding Morphology*. Arnold.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology, 35*(2), 441–456.

Ide, N., & Pustejovsky, J. (Eds.). (2017). *Handbook of Linguistic Annotation*. Springer.

Kridalaksana, H. (2010). *Pembentukan Kata dalam Bahasa Indonesia*. Gramedia.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Matthews, P. H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford University Press.

Notosusanto, N. (1979). *Tentara Peta pada Jaman Pendudukan Jepang di Indonesia*. Gramedia.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics, 31*(1), 71–106.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210.

Rolfe, G., Freshwater, D., & Jasper, M. (2001). *Critical Reflection in Nursing and the Helping Professions: A User's Guide*. Palgrave Macmillan.

Smith-Hefner, N. J. (2007). Youth language, gaul sociability, and the new Indonesian middle class. *Journal of Linguistic Anthropology, 17*(2), 184–203.

Svensén, B. (2009). *A handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press.

Wibawa, A. P., Fauzi, M. A., Kurniawan, R., & Dwiyanto, F. A. (2019). Natural language processing in bahasa Indonesia: A brief survey. *Jurnal Ilmu Komputer dan Informasi, 12*(2), 119–129.

Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*. Basic Books.