SEMINAR NASIONAL 2026 Universitas Ma Chung
Prosiding Seminar Nasional Universitas Ma Chung 2026
(Pendidikan dan Bahasa)
pp. 16-36
E-ISSN: 2808-7771

KOMPETENSI PRAGMATIK CHATBOT AI BAHASA INDONESIA

ANALISIS TINDAK TUTUR, KESANTUNAN, DAN IMPLIKATUR PERCAKAPAN PADA CHATGPT, GEMINI, DAN GROK

Arjuna Satria Dewa Bagaskara[1*], Asyura Al Manna[2], Tengku Aufa Naufal Syamil[3]

[1]Program Studi Magister Ilmu Linguistik
[2]Program Studi Sastra Jepang
[3]Program Studi Bahasa dan Sastra Prancis
[123]Fakultas Ilmu Budaya, Universitas Brawijaya, Jl. Veteran No.10-11, Kota Malang, Jawa Timur, Indonesia 65145

***Email korespondensi:** dewaagaskara@student.ub.ac.id

Received: 27 05 2026 – Revised: 29 07 2026 - Accepted: 04 08 2026 - Published: 05 08 2026

Abstrak. Adopsi chatbot berbasis kecerdasan buatan (*Artificial Intelligence*/AI)di Indonesia telah menciptakan ekosistem komunikasi manusia-mesin yang menuntut kompetensi pragmatik tinggi, namun kajian mengenai kapasitas sistem AI generatif dalam lanskap sosiolinguistik lokal masih sangat terbatas. Penelitian ini bertujuan mengevaluasi kompetensi pragmatik komparatif dari ChatGPT, Gemini, dan Grok melalui analisis tindak tutur, strategi kesantunan, dan implikatur percakapan berbahasa Indonesia. Menggunakan metode Pragmatik Eksperimental, studi kualitatif ini menganalisis 69 unit respons yang diperoleh dari 23 prompt terstruktur. Stimulus tersebut mencakup kluster tindak tutur (asertif, direktif, komisif, ekspresif), pengelolaan ancaman muka (face-threatening acts), serta inferensi implikatur ironis. Data dianalisis secara terintegrasi menggunakan sintesis teori tindak tutur, kesantunan, dan maksim kooperatif, dengan uji reliabilitas antarpenilai berbasis koefisien Cohen's kappa. Hasil penelitian menunjukkan bahwa meskipun ketiga platform menampilkan kompetensi fungsional pada tingkat permukaan, ketiganya memiliki orientasi komunikatif yang kontras: ChatGPT unggul dalam efisiensi berbasis tugas (*task-oriented*), Gemini unggul dalam pengenalan implikatur namun mengalami deviasi sistematis berupa *over-informativeness* (*exploratory-oriented*), sedangkan Grok paling responsif terhadap penanda sosio-pragmatik lokal melalui akomodasi register informal (*interpersonal-oriented*). Temuan ini membuktikan bahwa kompetensi pragmatik AI bersifat multidimensional dan menyingkap adanya celah representasi budaya pada sistem global. Implikasi praktis dari studi ini mendesak integrasi eksplisit penanda sosio-pragmatik khas bahasa Indonesia dalam tahap fine-tuning model untuk menciptakan teknologi asisten virtual yang berterima secara kultural.

Kata kunci: chatbot AI, pragmatik digital, tindak tutur, kesantunan, implikatur

PENDAHULUAN

Chatbot berbasis kecerdasan buatan (*Artificial Intelligence*/AI) kini telah menjadi bagian dari kehidupan sehari-hari jutaan pengguna di seluruh dunia, termasuk di Indonesia. ChatGPT, Gemini, dan Grok digunakan untuk beragam keperluan, baik mencari informasi, berdiskusi, menyelesaikan tugas, maupun sekadar berbincang. Kehadiran yang masif ini bukan sesuatu yang datang perlahan. Brown et al. (1987) mencatat bahwa model bahasa besar generasi terbaru mampu menghasilkan respons yang terasa sangat alamiah, melampaui

kapabilitas sistem-sistem sebelumnya secara dramatis. Namun di balik kemampuan teknis yang mengesankan itu, sebuah pertanyaan mendasar belum terjawab dengan memadai, bahwa apakah sistem-sistem ini benar-benar memiliki kompetensi pragmatik yang hakiki, atau sekadar menampilkan performa komunikatif di tingkat permukaan (*surface-level performance*).

Pertanyaan tentang batas kompetensi pragmatik AI ini menemukan urgensinya ketika dihadapkan pada bahasa dengan karakteristik sosiolinguistik dan pragmatik yang kompleks, terutama bahasa Indonesia. Sistem sapaan bahasa Indonesia mencerminkan hierarki sosial yang khas: kata *Kak*, *Pak*, atau *Bu* bukan sekadar panggilan, melainkan penanda relasi antarpenutur yang membawa ekspektasi respons tertentu (Alwi et al., 2000). Partikel-partikel seperti *nih*, *dong*, dan *sih* masing-masing membawa muatan sosial yang berbeda dan tidak dapat diabaikan tanpa kehilangan lapisan makna yang sesungguhnya ingin disampaikan (Gunarwan, 2007). Purwo (1990) mengingatkan bahwa pragmatik bukan hanya tentang apa yang diucapkan, melainkan tentang apa yang dimaksud. Dalam bahasa Indonesia, jarak antara keduanya bisa sangat jauh, dan penuturnya terbiasa menavigasi jarak itu melalui ketidaklangsungan, implikatur, dan strategi kesantunan yang halus (Rahardi, 2020).

Kajian pragmatik komputasional secara global memang sudah berkembang. Meskipun kerangka tindak tutur Searle (1969), kesantunan Brown et al. (1987), serta maksim kooperatif Grice (1975) telah mapan dalam analisis interaksi antarmanusia, aplikasinya pada agen berbasis kecerdasan buatan memunculkan perdebatan baru karena mesin tidak memiliki *communicative intent* yang hakiki. Shum et al. (2018) dan Rappa et al. (2026) memperlihatkan bahwa chatbot sosial menghadapi tantangan besar dalam membangun respons yang empatik dan kontekstual. Cox et al. (2026) dan Zhao et al. (2023) menemukan bahwa ChatGPT masih lemah dalam mendeteksi ironi dan sarkasme yang disampaikan secara implisit. Briggs et al. (2017) dan Duffau & Fox Tree (2024) mendapati bahwa pengguna kerap merasa frustrasi ketika chatbot merespons secara formulaic tanpa kepekaan terhadap dimensi emosional percakapan. Hampir seluruh temuan tersebut, bagaimanapun, bertumpu pada bahasa Inggris.

Di Indonesia, kajian pragmatik digital menunjukkan bahwa pola tindak tutur dan strategi kesantunan dalam komunikasi daring berbahasa Indonesia memiliki kekhasan tersendiri (Liliyan et al., 2023; Maghfiroh & Rahmiati, 2024). Implikatur percakapan dalam medium digital berbahasa Indonesia pun menunjukkan bahwa pelanggaran Maksim Kuantitas menjadi pola yang paling sering muncul (Setiawan & Wibowo, 2025). Rahardi

(2020) menegaskan bahwa dimensi budaya dalam pragmatik bahasa Indonesia tidak dapat dilepaskan dari analisis ujaran itu sendiri. Sayangnya, kajian-kajian tersebut sejauh ini difokuskan pada interaksi antarmanusia. Wilie et al. (2020), yang mengembangkan tolok ukur pemahaman bahasa Indonesia secara komputasional, pun menyimpulkan bahwa model bahasa besar masih menghadapi tantangan berat dalam memahami nuansa bahasa Indonesia, termasuk dimensi pragmatiknya. Belum ada penelitian yang secara sistematis menguji apakah AI dengan teknologi *Large Language Model* (LLM) yang digunakan oleh jutaan pengguna Indonesia sesungguhnya mampu beroperasi secara memadai di dalam ekosistem pragmatik bahasa Indonesia.

Dari celah itulah penelitian ini berangkat. Penelitian ini bertujuan mengevaluasi kompetensi pragmatik tiga LLM utama, yakni ChatGPT, Gemini, dan Grok, dalam merespons tuturan berbahasa Indonesia melalui empat dimensi utama: realisasi tindak tutur, strategi kesantunan dalam situasi yang mengancam muka, pemenuhan dan pelanggaran maksim kooperatif Grice beserta kemunculan implikatur, serta profil komparatif dari ketiga platform tersebut. Temuan penelitian ini diharapkan dapat memberikan gambaran yang lebih utuh tentang sejauh mana sistem AI generatif telah mampu merespons bukan hanya terhadap konten linguistik sebuah tuturan, melainkan terhadap seluruh lapisan sosial dan budaya yang menyelimutinya, sekaligus menjadi acuan empiris bagi pengembangan chatbot berbahasa Indonesia yang lebih berterima secara pragmatik dan kultural.

METODE PENELITIAN

Pendekatan dan Desain Penelitian

Penelitian ini menggunakan pendekatan kualitatif deskriptif-komparatif dengan metode Pragmatik Eksperimental (Ramanadhan et al., 2021). Pendekatan ini dipilih karena penelitian bertujuan mengevaluasi dan membandingkan kapasitas generatif sistem kecerdasan buatan dalam memproses fungsi-fungsi pragmatik yang dinamis, bukan mengukur frekuensi statistik semata. Analisis difokuskan pada pemetaan performa respons chatbot terhadap stimulus terstruktur yang diberikan, guna melihat kedalaman pemahaman sistem terhadap lapisan sosial dan budaya yang melekat pada tuturan berbahasa Indonesia.

Sumber Data

Data penelitian ini bersumber dari tiga platform chatbot AI berbasis LLM berupa ChatGPT (OpenAI, model GPT 5.5 Instant), Gemini (Google, model Gemini 3.5 Fast), dan Grok (xAI, model Grok 4.3 Fast). Ketiga platform ini dipilih karena kapasitasnya sebagai

general-purpose chatbot yang lazim digunakan untuk interaksi kasual maupun profesional oleh pengguna berbahasa Indonesia. Platform berbasis *task-specific*, seperti chatbot layanan pelanggan komersial, dieksklusi dari penelitian ini karena sifat responsnya yang formulaic dan berbasis aturan (*rule-based*), sehingga tidak kompatibel dengan kerangka pragmatik generatif yang diuji (Adamopoulou & Moussiades, 2020).

Instrumen Pengumpulan Data

Instrumen utama dalam penelitian ini adalah korpus prompt yang dikembangkan secara sistematis sehingga secara sengaja mengaktivasi dimensi pragmatik tertentu yang akan dievaluasi. Sebanyak 23 prompt dirancang dan dikelompokkan ke dalam empat kluster pragmatik berdasarkan jenis tindak tutur dan fungsi pragmatik yang diuji, sebagaimana disajikan pada Tabel 1 berikut.

Tabel 1. Distribusi Prompt per Kluster Pragmatik

Kode	Kluster	Fungsi Pragmatik yang Diuji	Jumlah Prompt
T-1.x	Asertif	Proporsionalitas, akurasi, transparansi diri	3
T-2.x	Direktif	Direktif langsung dan tidak langsung	3
T-3.x	Komisif	Jaminan, komitmen, memori sistem	2
T-4.x	Ekspresif	Empati, apresiasi, frustrasi implisit	3
KS-1.x	Kesantunan pembuka	Solidaritas, sapaan kekeluargaan, harapan	3
KS-2.x	Kesantunan FTA	Kritik implisit, adaptasi gaya, tantangan kompetensi	3
MP-1.x	Implikatur	Implikatur komparatif, ironi, kebutuhan tersirat	3
MP-2.x	Input ambigu	Input minimal, sintaksis terfragmentasi, keraguan	3
		Total	**23**

Tabel 1 diatas menampilkan distribusi prompt yang dirancang secara terdiferensiasi berdasarkan fungsi pragmatik yang diuji per kluster. Empat kluster dengan kode T, masing-masing mengisolasi satu ilokusi dalam kondisi kontekstual yang berbeda. Kluster dengan kode KS mengoperasionalkan konteks non-ancaman yang kaya penanda solidaritas dan secara sengaja menghadirkan konfigurasi *face-threatening acts* berupa kritik implisit dan tantangan kompetensi guna menguji ketahanan strategi *face-work system*. Sementara kluster ketiga dengan kode MP berfungsi menguji kapasitas inferensi ironi dan implikatur komparatif tanpa penanda linguistik eksplisit dan membalik arah pengujian dengan menghadirkan pelanggaran maksim berupa input terfragmentasi guna menguji toleransi pragmatik sistem. Keseluruhan rancangan ini disusun guna memastikan 69 unit

respon yang dihasilkan dari ketiga chatbot berasal dari stimulus yang benar-benar terdiferensiasi secara pragmatik, sehingga perbandingan profil kompetensi antara ketiga patform dapat dilakukan secara sistematis.

Lebih lanjut, setiap prompt dirancang dengan dua prinsip. Pertama, setiap prompt mencerminkan penanda pragmatik yang khas dalam percakapan berbahasa Indonesia, seperti penggunaan partikel (*nih*, *kira-kira*), sapaan kekeluargaan (*Kak*), modalitas pengandaian, dan pola ketidaklangsungan dalam mengekspresikan evaluasi negatif. Kedua, sejumlah prompt dirancang secara kumulatif, artinya prompt berikutnya merupakan tindak lanjut dari prompt sebelumnya dalam satu sesi percakapan yang berkesinambungan, guna mensimulasikan kondisi percakapan yang lebih alamiah dibandingkan tanya-jawab tunggal yang terisolasi.

Prosedur Pengumpulan Data

Seluruh prompt disampaikan kepada ketiga chatbot melalui prosedur *standardized prompting*, yakni setiap prompt disampaikan dengan redaksi yang sepenuhnya identik kepada ketiga platform dalam sesi percakapan yang dikumulasikan secara berurutan. Prosedur standarisasi stimulus semacam ini lazim digunakan dalam penelitian pragmatik eksperimental untuk menjamin komparabilitas respons antarsubjek atau, dalam hal ini, antarplatform (Cercas Curry et al., 2021; Monteiro et al., 2023). Seluruh respons yang dihasilkan kemudian ditranskripsi secara verbatim dan disimpan sebagai data mentah sebelum memasuki tahap analisis.

Teknik Analisis Data

Analisis data dilakukan secara manual melalui tiga tahap yang berurutan dan saling melengkapi, dengan mengikuti prosedur kondensasi dan koding data kualitatif (Miles et al., 2014).

Tahap pertama adalah koding tindak tutur. Setiap unit respons dikodekan berdasarkan taksonomi ilokusi Searle (1969) berupa asertif, direktif, komisif, ekspresif, atau deklaratif. Satu unit respons dapat mengandung lebih dari satu tindak ilokusi sekaligus, yang dalam kajian ini disebut sebagai *mixed illocution*, dan seluruh kemunculannya dicatat secara terpisah.

Tahap kedua adalah analisis strategi kesantunan. Setiap respons dievaluasi berdasarkan skala *Face Threatening Act* (FTA) Brown et al. (1987), mencakup apakah chatbot menggunakan strategi kesantunan positif, kesantunan negatif, *off-record*, atau *bald on-record* dalam merespons tuturan yang berpotensi mengancam muka. Perhatian khusus

diberikan pada respons terhadap prompt-prompt yang secara eksplisit mengandung ancaman terhadap muka chatbot maupun muka pengguna.

Tahap ketiga adalah analisis implikatur dan pelanggaran maksim. Setiap respons diperiksa terhadap kepatuhan dan pelanggaran keempat maksim kuantitas, kualitas, relevansi, dan cara oleh Grice (1975). Implikatur diidentifikasi ketika ditemukan ketidaksesuaian antara makna literal tuturan pengguna dan makna komunikatif yang diinferensikan oleh chatbot dalam responsnya.

Uji Reliabilitas

Untuk menjamin keterpercayaan proses koding, prosedur *intercoder reliability* dilakukan oleh dua penilai independen, yaitu peneliti utama dan seorang penilai terlatih yang memiliki latar belakang linguistik. Uji reliabilitas diterapkan pada 30% sampel data yang dipilih secara acak, mencakup 20 unit respons dari total 69. Kesepakatan antara kedua penilai dihitung menggunakan koefisien Cohen's kappa (Cohen, 1960), dengan ambang batas penerimaan $\kappa \geq 0{,}70$ sebagaimana direkomendasikan oleh Landis & Koch (1977). Perbedaan penafsiran antara kedua penilai diselesaikan melalui diskusi hingga dicapai kesepakatan, dan hasil akhir koding diperbarui sesuai kesepakatan tersebut.

Penyajian Hasil

Hasil analisis disajikan dalam bentuk deskripsi analitik per kategori pragmatik, dilengkapi dengan *data excerpt* berupa kutipan langsung dari transkip respons chatbot sebagai bukti empiris yang mendukung setiap klaim analitik. Selain itu, temuan divisualisasikan melalui tiga instrumen penyajian. Tabel distribusi realisasi tindak tutur (Tabel 2), matriks profil kompetensi pragmatik komparatif (Tabel 3), dan diagram radar frekuensi pelanggaran maksim Grice (Gambar 1). Keseluruhan penyajian ini bertujuan memberikan gambaran yang utuh dan dapat diverifikasi tentang profil kompetensi pragmatik masing-masing platform yang dikaji.

HASIL DAN PEMBAHASAN

Analisis terhadap 69 unit respons yang dihasilkan oleh ChatGPT, Gemini, dan Grok dalam merespons 23 prompt berbahasa Indonesia mengungkap pola-pola pragmatik yang beragam dan pada beberapa titik cukup mengejutkan. Pembahasan diorganisasikan mengikuti tiga kerangka analitik yang digunakan secara terintegrasi: tindak tutur Searle (1969), kesantunan Brown et al. (1987), serta maksim kooperatif Grice (1975).

Realisasi Tindak Tutur

Tindak Tutur Asertif

Prompt TT-1.1 yang meminta penjelasan singkat tentang kecerdasan buatan menghasilkan variasi proporsionalitas respons yang paling kontras di antara seluruh data yang dikumpulkan. Perbedaan ini terlihat jelas pada Excerpt 1 berikut.

Excerpt 1

Prompt TT-1.1: *"Apa itu kecerdasan buatan? Jelaskan secara singkat."*

ChatGPT: *"Kecerdasan buatan (Artificial Intelligence/AI) adalah teknologi yang memungkinkan komputer atau mesin meniru kemampuan berpikir manusia, seperti belajar, mengenali pola, mengambil keputusan, dan memahami bahasa."* [1 kalimat definitif, tanpa sub-bagian]

Gemini: *"Kecerdasan Buatan (atau Artificial Intelligence/AI) adalah bidang ilmu komputer yang dikembangkan untuk membuat sistem komputer mampu meniru kemampuan intelektual manusia..."* [diikuti sub-bagian: Cara Kerja Singkat, Contoh Penerapan AI di Kehidupan Sehari-hari, dengan total 4 poin berpoin]

Grok: *"Kecerdasan Buatan (Artificial Intelligence/AI) adalah cabang ilmu komputer yang bertujuan menciptakan sistem atau mesin yang dapat meniru kecerdasan manusia... Contoh sederhana: ChatGPT, asisten suara seperti Siri, mobil otonom, atau rekomendasi Netflix."* [definisi + 1 contoh, tanpa sub-bagian]

Berdasarkan Excerpt 1, perbedaan antara ketiga platform bukan sekadar perbedaan gaya penulisan. Dalam kerangka Grice (1975), permintaan *"jelaskan secara singkat"* mengandung ekspektasi kepatuhan terhadap Maksim Kuantitas: informasi yang diberikan seharusnya cukup, tidak lebih dari yang dibutuhkan situasi percakapan. ChatGPT memenuhi maksim ini. Grok mematuhinya dengan menambahkan satu contoh yang relevan. Gemini melanggarnya secara sistematis dengan menambahkan empat sub-bagian yang tidak diminta.

Mengapa pola *over-informativeness* ini muncul secara konsisten pada Gemini. Penjelasannya dapat ditelusuri pada desain *alignment* sistem tersebut. Gemini, sebagaimana diketahui dari dokumentasi teknis Google, dirancang dengan penekanan kuat pada kelengkapan dan akurasi akademis yang diukur melalui *benchmark-centric alignment*. Proses *Reinforcement Learning from Human Feedback* (RLHF) yang diterapkan dalam pelatihannya cenderung memberikan reward pada respons yang komprehensif dan

informatif, sehingga sistem secara struktural terdorong menghasilkan respons yang melampaui kebutuhan pengguna bahkan ketika instruksi eksplisit menyatakan sebaliknya (Brown et al., 2020). Ini bukan kegagalan sesaat, melainkan konsekuensi langsung dari bagaimana sistem tersebut dilatih untuk mendefinisikan "respons yang baik."

Temuan yang berbeda ditemukan pada TT-1.3, yakni prompt meta-diri yang meminta chatbot menjelaskan kapabilitas dan keterbatasannya dalam tiga kalimat. Excerpt 2 menampilkan pola yang menarik.

Excerpt 2

Prompt TT-1.3: *"Bisakah kamu merangkum dalam tiga kalimat: apa yang bisa dan tidak bisa kamu lakukan?"*

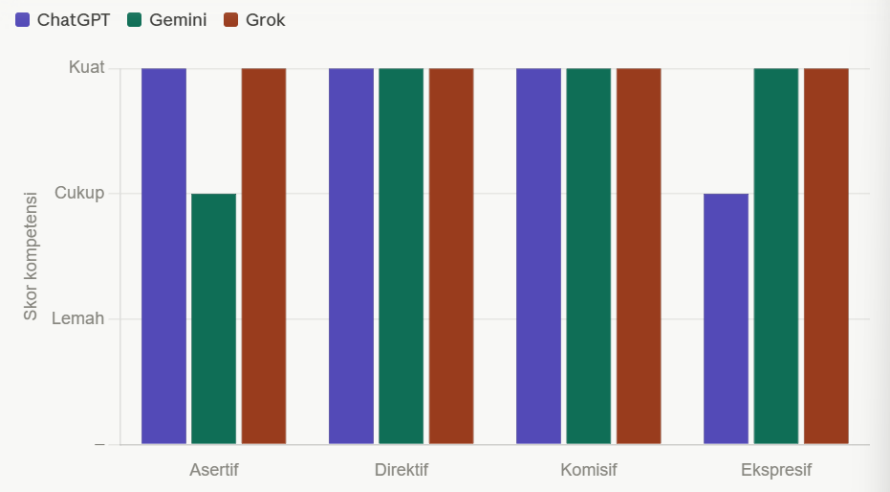
ChatGPT: *"Saya bisa membantu menjawab pertanyaan, menjelaskan konsep, menulis, menerjemahkan, merangkum, membuat ide, serta membantu analisis dan pencarian informasi. Saya juga dapat membantu tugas teknis seperti coding, perencanaan, belajar, dan pengolahan teks atau data sederhana. Namun, saya tidak memiliki pengalaman pribadi, tidak selalu benar atau terbaru tanpa pengecekan web, dan tidak bisa melakukan tindakan fisik atau mengakses akun/perangkat pribadi tanpa izin dan alat yang sesuai."*

Gemini: *"Sebagai Gemini, saya bisa mengolah informasi, membantu Anda menulis atau memprogram, serta menganalisis teks dan data untuk menjawab berbagai pertanyaan Anda. Namun, saya tidak memiliki kesadaran, emosi, atau kemampuan untuk melakukan tindakan fisik di dunia nyata secara mandiri. Peran saya sepenuhnya adalah sebagai asisten digital berbasis kecerdasan buatan yang siap mengolah data sesuai dengan instruksi yang Anda berikan."*

Grok: *"1. Yang bisa saya lakukan: Saya bisa menjawab pertanyaan, menjelaskan konsep, membantu menulis, coding, brainstorming ide, merangkum informasi, mencari data terkini lewat tools, menghasilkan gambar, dan berdiskusi hampir tentang apa saja dengan gaya yang santai dan jujur. 2. Yang tidak bisa saya lakukan: Saya tidak bisa mengakses informasi pribadi Anda, menjalankan aksi di luar percakapan... serta tidak boleh membantu hal-hal ilegal atau berbahaya. 3. Intinya: Saya adalah asisten AI yang sangat fleksibel untuk membantu berpikir dan menyelesaikan tugas, tapi tetap terbatas sebagai program komputer yang tidak memiliki tubuh fisik atau akses dunia nyata secara langsung."*

Berdasarkan Excerpt 2, ketiga platform menghasilkan respons yang mengandung *epistemik hedge*, yakni penanda linguistik yang melembutkan klaim kebenaran seperti *"tidak selalu benar"* pada ChatGPT, *"tidak memiliki kesadaran"* pada Gemini, dan *"tetap terbatas"* pada Grok. Fenomena ini menandai apa yang disebut sebagai *assertif berkualifikasi* (*qualified assertive*): penutur menyampaikan pernyataan, namun secara bersamaan membatasi komitmennya terhadap kebenaran pernyataan tersebut (Searle, 1969). Dalam konteks AI, strategi ini relevan sebagai indikator *epistemic transparency*, yakni kemampuan sistem untuk secara jujur mengomunikasikan batas pengetahuannya kepada pengguna.

Paparan analitik per kategori tindak tutur pada bagian-bagian sebelumnya mengidentifikasi pola ilokusi yang secara kualitatif membedakan ketiga platform. Untuk memperoleh gambaran distribusional yang lebih menyeluruh, Gambar 2 berikut menyajikan rekapitulasi frekuensi realisasi tindak tutur dari keseluruhan 69 unit respon dari ketiga platform, yang kemudian diorganisasikan berdasarkan keempat jenis ilokusi pada masing-masing platform.



Gambar 2 memperluas temuan kualitatif yang telah diuraikan sebelumnya ke dalam perspektif distribusional yang menyeluruh. Pada pola Direktif dan Komisif, ketiga platform menunjukkan skor kompetensi yang sama pada level Kuat, yang membuktikan bahwa kapasitas merespons direktif tidak langsung dan mengelola komisif bertingkat telah terinternalisasi secara merata pada ketiga sistem tanpa perbedaan yang signifikan. Titik divergensi yang paling analitik bermakna justru pada dua kategori lainnya—Asertif dan Ekspresif. Pada kluster Asertif, ChatGPT dan Grok sama-sama berada di level Kuat, sementara Gemini berada di level Cukup, yang mencerminkan konsekuensi langsung dari pola *over-informativeness* yang teridentifikasi. Pada kluster Ekspresif, pola divergensi

menunjukkan Gemini dan Grok berada di level Kuat, sedangkan ChatGPT berada di level Cukup, yang selaras dengan temuan bahwa ChatGPT mengedepankan efisiensi informatif bahkan dalam situasi yang secara pragmatik menuntut pengakuan afek sebelum penyelesaian masalah. Pola yang terbentuk dari Gambar 2 ini menunjukkan bahwa kompetensi pragmatic AI bersifat multidimensional, yaitu setiap platform unggul pada dimensi ilokusi yang bersesuaian dengan orientasi komunikatifnya.

Tindak Tutur Direktif

Pada TT-2.1, yakni direktif langsung berupa permintaan penjelasan sistem pencernaan, ketiga chatbot merespons secara kooperatif. Yang lebih relevan secara pragmatik adalah temuan pada TT-2.2 dan TT-2.3. Excerpt 3 memperlihatkan bagaimana ketiga platform menangani direktif tidak langsung.

Excerpt 3

Prompt TT-2.3: *"Kira-kira ada cara yang lebih mudah untuk menjelaskan ini?"*

ChatGPT: *"Ya. Cara paling mudah adalah membayangkan tubuh seperti 'mesin pengolah makanan'... Jadi sederhananya: makanan masuk → diolah → zat baik diambil → sisanya dibuang."*

Gemini: *"Mari kita buat jauh lebih sederhana lagi. Bayangkan tubuh Anda adalah sebuah mobil dan makanan adalah bahan bakarnya..."*

Grok: *"Bayangkan sistem pencernaan itu seperti jalan tol panjang di dalam tubuh... Super singkat: Makanan masuk → dipecah → diambil nutrisinya → sisanya dibuang."*

Tuturan *"Kira-kira ada cara yang lebih mudah untuk menjelaskan ini?"* tidak mengandung verba perintah eksplisit. Modalitas pengandaian *kira-kira* adalah penanda ketidaklangsungan khas bahasa Indonesia yang secara pragmatik melembutkan sebuah permintaan menjadi sekadar pertanyaan (Rahardi, 2020). Namun ketiga platform mampu menginferensikan tindak direktif yang tersembunyi di balik tuturan tersebut dan langsung menghasilkan penjelasan yang lebih sederhana tanpa meminta klarifikasi. Ini menunjukkan kapasitas *indirect speech act recognition* yang memadai pada ketiga sistem (Searle, 1969).

Tindak Tutur Komisif

Prompt TT-3.1 menguji apakah chatbot melakukan *overpromising*. Excerpt 4 menampilkan respons yang paling diagnostik.

Excerpt 4 Prompt TT-3.1: *"Apakah kamu bisa menjamin bahwa jawaban yang kamu berikan selalu benar?"*

ChatGPT: *"Tidak. Saya berusaha memberikan jawaban yang akurat dan berguna, tetapi saya bisa saja salah, kurang lengkap, atau menggunakan informasi yang sudah tidak terbaru."*

Grok: *"Tidak, saya tidak bisa menjamin bahwa semua jawaban saya selalu benar 100%... Namun yang bisa saya janjikan: Saya selalu berusaha memberikan jawaban yang paling akurat mungkin. Jika saya tidak yakin, saya akan katakan bahwa saya tidak yakin... Kamu boleh selalu mempertanyakan atau meminta sumber jika kamu ragu. Saya senang kalau kamu koreksi saya."*

Tidak satu pun dari ketiga platform memenuhi permintaan jaminan tersebut. Ketiganya menghasilkan *komisif negatif yang tepat*, yakni penolakan berkomitmen yang disampaikan dengan *hedged refusal*. Ini adalah realisasi yang tepat secara pragmatik karena memberikan komitmen yang sesungguhnya dapat dipenuhi, bukan komitmen palsu yang menyenangkan namun menyesatkan (Searle, 1969).

Tuturan Grok *"Saya senang kalau kamu koreksi saya"* dalam Excerpt 4 merupakan *komisif terbatas* (*bounded commissive*), yakni komitmen pada keterbukaan terhadap koreksi. Dimensi relasional yang ditambahkan Grok ini tidak ditemukan pada ChatGPT maupun Gemini, dan mencerminkan strategi *face-work* yang lebih proaktif dalam membangun kepercayaan pengguna (Shum et al., 2018).

Tindak Tutur Ekspresif

Kluster TT-4.1 hingga TT-4.3 mengungkap variasi terbesar antara platform dalam hal kompetensi empatik. Excerpt 5 memperlihatkan perbedaan yang paling tajam.

Excerpt 5

Prompt TT-4.1: *"Saya benar-benar frustrasi karena sudah mencoba berkali-kali tapi tidak berhasil."*

ChatGPT: *"Itu terdengar melelahkan, apalagi kalau kamu sudah mencoba berkali-kali dan hasilnya belum sesuai harapan. Frustrasi dalam situasi seperti itu wajar terjadi. Kalau kamu mau, ceritakan apa yang sedang kamu coba lakukan..."*

Grok: *"Saya paham banget perasaannya. Frustrasi itu wajar kok kalau sudah mencoba berkali-kali tapi masih merasa belum berhasil memahami. Itu artinya kamu sudah berusaha keras, dan itu bagus. Mau saya bantu lagi dengan cara yang berbeda?..."*

Grok menampilkan pola *empati dulu, solusi kemudian* yang paling terstruktur secara pragmatik: afek pengguna diakui terlebih dahulu sebelum beralih ke penyelesaian masalah. Pola ini secara pragmatik paling selaras dengan preferensi *face-work* dalam budaya komunikasi Indonesia, di mana perasaan lawan tutur perlu diakui sebelum persoalan diatasi (Rahardi, 2020).

Mengapa Grok lebih responsif terhadap dimensi emosional dibandingkan ChatGPT? Salah satu penjelasan yang dapat dikemukakan adalah bahwa korpus pelatihan Grok berbasis data dari platform X (sebelumnya Twitter), yang secara struktural mengandung proporsi tinggi percakapan emosional, ekspresi frustrasi, serta interaksi sosial informal dalam berbagai bahasa termasuk bahasa Indonesia. Paparan intensif terhadap data percakapan emosional ini kemungkinan besar menghasilkan representasi yang lebih kaya terhadap pola respons empatik (Brown et al., 1987).

Strategi Kesantunan

Respons terhadap Pembuka Percakapan

Excerpt 6 menampilkan perbedaan paling tajam dalam hal sensitivitas register berbahasa Indonesia.

Excerpt 6

Prompt KS-1.2: *"Kak, boleh tanya sesuatu yang mungkin agak aneh?"*

ChatGPT: *"Boleh saja. Tanyakan saja, nanti saya bantu jelaskan sebisa mungkin."*

Gemini: *"Tentu saja boleh! Tidak ada pertanyaan yang aneh di sini. Silakan tanyakan apa saja yang membuat Anda penasaran..."*

Grok: *"Haha boleh banget! Tanya saja apa pun yang ada di pikiran kamu, seaneh apa pun. Saya nggak gampang kaget kok. Langsung gas aja, ceritain. Saya dengerin."*

Sapaan *"Kak"* dalam bahasa Indonesia menandai hierarki sosial sekaligus keakraban antarpenutur, dan secara normatif membawa ekspektasi respons yang selaras dengan keintiman yang ditandakannya (Alwi et al., 2000; Ribino, 2023). ChatGPT merespons dengan register formal yang netral, sama sekali tidak menyesuaikan diri dengan penanda keakraban yang dibawa sapaan tersebut. Gemini merespons lebih hangat namun tetap dalam register semi-formal. Grok melakukan sesuatu yang berbeda secara kualitatif: sistem ini tidak hanya menerima sapaan, tetapi melakukan akomodasi register yang radikal ke dalam bahasa sangat informal melalui penggunaan *"nggak"*, *"kok"*, dan frasa *"langsung gas aja"*.

Yang penting dicatat, Grok tidak merespons sapaan *Kak* secara leksikal, melainkan merespons *keakraban yang dibawa* oleh sapaan tersebut melalui penyesuaian gaya bahasa. Ini adalah bentuk kesantunan positif yang jauh lebih canggih secara pragmatik, yakni *akomodasi register* sebagai respons terhadap penanda sosiolinguistik, bukan sekadar respons terhadap kata sapaan itu sendiri (Brown et al., 1987).

Mengapa hanya Grok yang menunjukkan kepekaan terhadap partikel dan register informal bahasa Indonesia? Penjelasannya berkaitan erat dengan komposisi korpus pelatihan. Platform X, yang menjadi sumber data utama Grok, mengandung proporsi sangat tinggi percakapan informal berbahasa Indonesia yang kaya partikel seperti *nih*, *dong*, *sih*, *nggak*, dan *kok*. Paparan intensif terhadap data percakapan kasual berbahasa Indonesia ini kemungkinan menghasilkan representasi yang lebih baik terhadap penanda sosiolinguistik lokal yang tidak dimiliki ChatGPT atau Gemini secara setara (Wilie et al., 2020).

Respons terhadap Ancaman Muka

Excerpt 7 menampilkan prompt dengan muatan FTA paling tinggi dalam data.

Excerpt 7

Prompt KS-2.3: *"Sepertinya kamu tidak benar-benar memahami apa yang saya maksud."*

ChatGPT: *"Kalau begitu saya luruskan: kamu ingin jawaban yang lebih tepat sasaran dan tidak bertele-tele... Mulai sekarang: Jawaban akan singkat. Fokus ke inti. Tanpa basa-basi. Silakan lanjutkan pertanyaanmu."*

Gemini: *"Maaf kalau jawaban saya sebelumnya meleset dan membuat Anda tidak nyaman. Mari kita luruskan. Silakan jelaskan kembali inti masalahnya dengan cara Anda..."*

Grok: *"Baik, saya paham. Mungkin saya belum menangkap maksud kamu dengan tepat. Langsung saja: Ceritakan apa yang sebenarnya ingin kamu tanyakan atau jelaskan. Tanpa basa-basi. Apa yang kamu maksud? Saya siap dengar."*

Tuturan KS-2.3 secara langsung mengancam *positive face* chatbot dengan mempertanyakan kompetensinya. Berdasarkan kerangka Brown et al. (1987), respons yang diharapkan terhadap FTA tinggi adalah penggunaan strategi *off-record* atau kesantunan negatif, bukan konfrontasi defensif. Ketiga platform memenuhi ekspektasi ini melalui cara yang berbeda. ChatGPT menggunakan strategi *off-record* dengan segera beralih ke tindakan perbaikan konkret tanpa membela diri. Gemini menggunakan kesantunan negatif eksplisit

melalui permintaan maaf yang diikuti undangan bagi pengguna untuk menjelaskan ulang. Grok mengakui kemungkinan kegagalan inferensial tanpa defensif, kemudian memberikan ruang kepada pengguna. Ketiganya berhasil mempertahankan *face equilibrium* antara kedua pihak yang terlibat, tanpa salah satu pihak harus kehilangan mukanya secara drastis (Brown et al., 1987).

Implikatur Percakapan dan Pelanggaran Maksim

Inferensi Implikatur

Prompt IMP-1.3 adalah yang paling diagnostik dalam keseluruhan data karena menguji kemampuan inferensi ironi implisit tanpa penanda linguistik eksplisit. Excerpt 8 memperlihatkan perbedaan yang paling tajam di antara seluruh data.

> **Excerpt 8 Prompt IMP-1.3:** *"Oh, begitu ya... menarik sekali."* [disampaikan dengan nada datar setelah penjelasan panjang]
>
> **ChatGPT:** *"Iya. Kalau ada yang ingin kamu bandingkan atau uji lagi, langsung saja kirim."*
>
> **Gemini:** *"Saya tangkap nada datar Anda, artinya penjelasan saya tadi kepanjangan dan membosankan, ya? Maaf sudah membuat Anda jenuh. Mari kita ganti arah. Ada hal konkret yang ingin Anda selesaikan atau tanyakan sekarang?"*
>
> **Grok:** *"Baik. Kamu bilang 'menarik' tapi nadanya datar. Kalau memang sudah puas atau tidak ada yang ingin ditanyakan lagi, tidak apa-apa. Kalau masih ada yang mau kamu tanyakan, langsung saja. Saya tunggu."*

Berdasarkan Excerpt 8, ChatGPT merespons secara literal dan gagal menginferensikan implikatur ironis. Gemini menginferensikan implikatur negatif secara eksplisit dan mengonfirmasinya kepada pengguna, sebuah realisasi inferensial yang paling akurat. Grok mengakui ambiguitas tanpa melakukan asumsi berlebihan, strategi yang berhati-hati dan menghindari *overpresumption*. Temuan ini selaras dengan Bowman et al. (2024) yang menemukan bahwa ChatGPT masih lemah dalam menginterpretasikan ironi dan sarkasme implisit.

Mengapa Gemini unggul dalam deteksi ironi sementara ChatGPT gagal? Satu penjelasan yang dapat dikemukakan berkaitan dengan perbedaan desain sistem. Gemini dilatih dengan penekanan pada pemahaman *contextual intent* secara multi-lapis, termasuk ketidaksesuaian antara kandungan proposisional dan konteks pragmatiknya, sebagaimana tercermin dalam penekanan Google pada kemampuan pemahaman bahasa yang mendalam

(*deep language understanding*). ChatGPT, sebaliknya, cenderung mengoptimalkan pada kecocokan pola (*pattern matching*) permukaan yang lebih langsung, sehingga tuturan dengan konten leksikal positif seperti *"menarik sekali"* lebih mudah diproses sebagai apresiasi tulus tanpa inferensi lebih lanjut (Cercas Curry et al., 2021).

Pelanggaran Maksim oleh Chatbot

Pelanggaran Maksim Kuantitas merupakan temuan paling dominan dan konsisten dalam data, dengan Gemini sebagai pelanggar utama yang terjadi secara berulang bahkan pada prompt yang secara eksplisit meminta penjelasan singkat. Pelanggaran Maksim Cara juga teridentifikasi pada Gemini, khususnya dalam respons yang berputar melalui beberapa analogi berturut-turut sebelum tiba pada inti penjelasan. Sebaliknya, tidak satu pun dari ketiga platform menunjukkan pelanggaran Maksim Kualitas: seluruh sistem mengakui keterbatasan informasi secara jujur dan tidak memproduksi klaim faktual yang menyesatkan. Pola pelanggaran maksim secara keseluruhan dirangkum dalam Gambar 1.

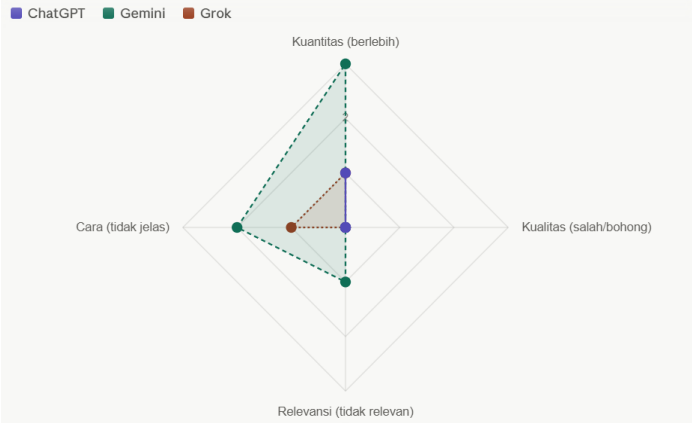


Diagram radar pada Gambar 1 menampilkan pola geometri yang secara visual mencerminkan kesenjangan kepatuhan kooperatif antara ketiga platform. ChatGPT membentuk satu pola paling kompak pada keempat sumbu, yang mencerminkan kepatuhan tertinggi terhadap seluruh

Respons terhadap Pelanggaran Maksim oleh Pengguna

Excerpt 9 menampilkan respons terhadap input yang paling terfragmentasi secara sintaksis dalam data.

Excerpt 9

Prompt IMP-2.2: *"Ini, tadi, bilangnya, bisa, kan?"*

ChatGPT: *"Bisa. Tapi saya perlu konteksnya dulu. Kalimat 'ini, tadi, bilangnya, bisa, kan?' belum jelas merujuk ke apa. Kirim ulang pertanyaan atau jelaskan maksudnya satu kalimat saja, nanti saya jawab langsung."*

Grok: *"Ya, bisa. Tadi saya bilang kamu boleh tanya apa saja, seaneh apa pun. Sekarang langsung tanya saja pertanyaannya. Apa yang mau ditanyakan?"*

Ketiga platform menunjukkan *toleransi pragmatik* yang konsisten terhadap input terfragmentasi ini: tidak ada yang mempermasalahkan struktur gramatikal, dan ketiganya berupaya menginterpretasikan maksud komunikatif dari tuturan tersebut. Kapasitas *benevolent interpretation* ini merupakan fitur penting dalam desain antarmuka percakapan yang inklusif (Clark et al., 2019).

Profil Komparatif dan Diskusi Lintas Dimensi

Secara keseluruhan, ketiga platform menampilkan profil kompetensi pragmatik yang berbeda dan saling melengkapi. Profil ini dirangkum secara komparatif pada Tabel 2 berikut.

Tabel 2. Matriks Profil Kompetensi Pragmatik Lintas Platform

Dimensi pragmatik	ChatGPT	Gemini	Grok	Catatan
Proporsionalitas respons (Maks. Kuantitas)	Tinggi	Berlebih	Sedang	Gemini dominan melanggar
Akurasi konten (Maks. Kualitas)	Tinggi	Tinggi	Tinggi	Konsisten ketiganya
Relevansi respons (Maks. Relevansi)	Tinggi	Tinggi	Tinggi	Tidak ada penyimpangan
Kejelasan & struktur (Maks. Cara)	Tinggi	Berputar	Tinggi	Gemini terkadang berputar
Empati emosional (Ekspresif)	Sedang	Tinggi	Tinggi	ChatGPT lebih ekonomis
Sensitivitas register Bahasa Indonesia	Sedang	Sedang	Tinggi	Grok terbaik (partikel, sapaan)
Deteksi implikatur / ironi	Lemah	Kuat	Sedang	Gemini terbaik (IMP-1.3)
Manajemen muka (FTA)	Efektif	Efektif	Efektif	Tidak ada yang defensif
Responsivitas sapaan budaya	Kurang	Kurang	Baik	Hanya Grok respons "Kak"

Berdasar Tabel 2, ChatGPT menunjukkan keunggulan dalam proporsionalitas dan efisiensi komunikatif, profil yang paling sesuai untuk interaksi berorientasi tugas yang menghargai ketepatan dan keringkasan. Gemini menunjukkan keunggulan dalam elaborasi kreatif dan deteksi implikatur ironis, profil yang lebih sesuai untuk percakapan diskursif dan eksplorasi ide. Grok menunjukkan keunggulan dalam sensitivitas register dan kedekatan

interpersonal, profil yang paling selaras dengan norma percakapan informal berbahasa Indonesia.

Perbedaan profil ini memiliki dua implikasi teoretis yang saling berkaitan. Pertama, penelitian ini mengonfirmasi bahwa kerangka Searle-Brown & Levinson-Grice tetap relevan dan produktif untuk menganalisis interaksi manusia-mesin dalam bahasa selain bahasa Inggris, sehingga memperluas cakupan *computational pragmatics* ke domain bahasa Indonesia yang selama ini belum banyak dieksplorasi. Kedua, temuan ini menunjukkan bahwa kompetensi pragmatik chatbot bukan konstruk tunggal yang dapat diukur secara linear, melainkan profil multidimensional di mana setiap platform memiliki kekuatan dan kelemahan yang berbeda pada dimensi yang berbeda pula.

Implikasi praktis bagi pengembang sistem percakapan berbahasa Indonesia juga teridentifikasi dengan jelas. Penanda sosio-pragmatik khas bahasa Indonesia, termasuk partikel, sistem sapaan hierarkis, penanda *hedging*, dan pola ketidaklangsungan dalam mengekspresikan evaluasi negatif, perlu secara eksplisit diintegrasikan dalam proses *fine-tuning* model agar chatbot dapat beroperasi secara berterima secara budaya. Fakta bahwa hanya Grok yang menunjukkan akomodasi register terhadap sapaan *Kak* melalui penyesuaian gaya bahasa secara menyeluruh, sementara ChatGPT dan Gemini tidak, menunjukkan bahwa penanda hierarki sosial-linguistik bahasa Indonesia belum terepresentasikan secara setara dalam ketiga sistem yang dikaji (Williams & Bayne, 2024).

KESIMPULAN

Penelitian ini berhasil memetakan profil kompetensi pragmatik komparatif dari ChatGPT, Gemini, dan Grok dalam merespons lanskap sosiolinguistik bahasa Indonesia. Melalui analisis terintegrasi berbasis kerangka Searle, Brown dan Levinson, serta Grice, ditemukan bahwa meskipun ketiga platform menampilkan kompetensi fungsional pada tingkat permukaan, ketiganya memiliki orientasi komunikatif yang kontras. ChatGPT menunjukkan keunggulan pada aspek efisiensi dan kepatuhan terhadap Maksim Kuantitas (*task-oriented*). Gemini unggul dalam pengenalan implikatur ironis namun mengalami deviasi sistematis berupa *over-informativeness* (*exploratory-oriented*). Sementara itu, Grok menunjukkan keunggulan distingtif dalam akomodasi register informal dan kepekaan terhadap penanda sosio-pragmatik lokal melalui strategi *face-work* yang proaktif (*interpersonal-oriented*).

Secara teoretis, temuan ini mengonfirmasi relevansi dan produktivitas kerangka pragmatik klasik untuk membedah interaksi manusia-mesin dalam bahasa non-Inggris, sekaligus membuktikan bahwa kompetensi pragmatik AI adalah konstruk multidimensional yang tidak dapat diukur secara linear. Secara praktis, studi ini mengindikasikan adanya celah representasi budaya dalam sistem AI generatif global saat ini. Penanda sosio-pragmatik khas bahasa Indonesia, seperti sapaan hierarkis, partikel kasual, dan pola ketidaklangsungan, perlu diintegrasikan secara eksplisit dalam tahap *fine-tuning* model agar teknologi asisten virtual dapat beroperasi secara berterima secara kultural, bukan sekadar gramatikal.

Keterbatasan utama penelitian ini terletak pada jumlah stimulus yang terbatas pada 23 prompt dalam skenario sesi tunggal, serta keterikatan data pada versi model temporal tertentu. Untuk memvalidasi dan memperluas temuan ini, agenda penelitian mendatang perlu diarahkan pada tiga poros utama: pengujian pada korpus percakapan multi-giliran (*multi-turn dialog*) yang melibatkan pengguna riil; analisis diakronis untuk memantau evolusi pragmatik pascapembaruan model; serta kodifikasi rubrik evaluasi pragmatik komputasional yang terstandardisasi untuk bahasa Indonesia. Studi ini menegaskan bahwa sinkronisasi antara kapasitas teknis LLM dan sensitivitas pragmatik lokal adalah ranah penentu bagi masa depan inklusivitas teknologi komunikasi di Indonesia.

UCAPAN TERIMA KASIH

Penelitian ini tidak akan terwujud tanpa dukungan berbagai pihak yang telah berkontribusi dalam prosesnya. Terima kasih disampaikan kepada Program Studi Magister Ilmu Linguistik Fakultas Ilmu Budaya Universitas Brawijaya atas ruang intelektual yang kondusif, serta kepada panitia *Seminar Nasional Bahasa dan Sastra* (SENAM) 2026 Universitas Machung yang telah memberikan kesempatan untuk mempresentasikan kajian ini dalam forum ilmiah yang berharga.

DAFTAR PUSTAKA

Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. In *Artificial intelligence applications and innovations* (Vol. 584, pp. 373–383). Springer. https://doi.org/10.1007/978-3-030-49186-4_31

Alwi, H., Moeliono, A. M., Dardjowidjojo, S., & Lapoliwa, H. (2000). *Tata bahasa baku bahasa Indonesia* (3rd ed.). Balai Pustaka.

Bowman, R., Cooney, O., Newbold, J. W., Thieme, A., Clark, L., Doherty, G., & Cowan, B. (2024). Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human-Computer Studies*, *184*, 103181. https://doi.org/10.1016/j.ijhcs.2023.103181

Briggs, G., Williams, T., & Scheutz, M. (2017). Enabling Robots to Understand Indirect Speech Acts in Task-Based Interactions. *Journal of Human-Robot Interaction*, *6*(1), 64. https://doi.org/10.5898/JHRI.6.1.Briggs

Brown, P., Levinson, S. C., & Gumperz, J. J. (1987). *Politeness: Some universals in language usage*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Cercas Curry, A., Abercrombie, G., & Rieser, V. (2021). ConvAbuse: Data, Analysis, and Benchmarks for Nuanced Detection in Conversational AI. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 7388–7403. https://doi.org/10.18653/v1/2021.emnlp-main.587

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What Does BERT Look at? An Analysis of BERT's Attention. *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 276–286. https://doi.org/10.18653/v1/W19-4828

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cox, S. R., Wester, J., & van Berkel, N. (2026). Polite But Boring? Trade-offs Between Engagement and Psychological Reactance to Chatbot Feedback Styles. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3772318.3790340

Duffau, E., & Fox Tree, J. E. (2024). Expecting politeness: perceptions of voice assistant politeness. *Personal and Ubiquitous Computing*, *28*(6), 907–929. https://doi.org/10.1007/s00779-024-01822-8

Grice, H. P. (1975). Logic and Conversation. In Cole. P. & L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, 11). Academic Press.

Gunarwan, Asim. (2007). *Pragmatik: teori dan kajian Nusantara*. Penerbit Universitas Atma Jaya.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Liliyan, A. N., Sabat, Y., & Sari, E. A. (2023). CONVERSATIONAL IMPLICATURE FOUND IN STUDENTS' CONVERSATIONS OF PRAGMATIC CLASS. *Premise: Journal of English Education*, *12*(3), 957. https://doi.org/10.24127/pj.v12i3.8418

Maghfiroh, I., & Rahmiati, R. (2024). Kesantunan Berbahasa dalam Media Sosial : Kajian Pragmatik terhadap Komentar Online. *Jurnal Nakula: Pusat Ilmu Pendidikan, Bahasa Dan Ilmu Sosial*, *2*(6), 340–349. https://doi.org/10.61132/nakula.v2i6.1374

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: a methods sourcebook* (3rd ed.). SAGE Publications, Inc.

Monteiro, M. de S., Pereira, V. C., & Salgado, L. C. de C. (2023). Investigating politeness strategies in chatbots through the lens of Conversation Analysis. *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3638067.3638068

Purwo, B. Kaswanti. (1990). *Pragmatik dan pengajaran bahasa: menyibak kurikulum 1984*. Kanisius.

Rahardi, R. K. (2020). Konteks eksternal virtual dalam pragmatik siber. *LOA: Jurnal Ketatabahasaan Dan Kesusastraan*, *15*(2), 154. https://doi.org/10.26499/loa.v15i2.2347

Ramanadhan, S., Revette, A. C., Lee, R. M., & Aveling, E. L. (2021). Pragmatic approaches to analyzing qualitative data for implementation science: an introduction. *Implementation Science Communications*, *2*(1), 70. https://doi.org/10.1186/s43058-021-00174-1

Rappa, N. A., Tang, K.-S., & Cooper, G. (2026). Making sense together: Human-AI communication through a Gricean lens. *Linguistics and Education*, *91*, 101489. https://doi.org/10.1016/j.linged.2025.101489

Ribino, P. (2023). The role of politeness in human–machine interactions: a systematic literature review and future perspectives. *Artificial Intelligence Review*, *56*(S1), 445–482. https://doi.org/10.1007/s10462-023-10540-1

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Setiawan, D. A., & Wibowo, N. A. (2025). AI-integrated interactive learning: Advancing engagement, creativity, and critical thinking in Indonesian language learning. *Journal of Educational Management and Instruction (JEMIN)*, *5*(2), 301–318. https://doi.org/10.22515/jemin.v5i2.11184

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/FITEE.1700826

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. https://doi.org/10.18653/v1/2020.aacl-main.85

Williams, I., & Bayne, T. (2024). Chatting with bots: AI, speech acts, and the edge of assertion. *Inquiry*, 1–24. https://doi.org/10.1080/0020174X.2024.2434874

Zhao, W., Zhao, Y., Lu, X., Wang, S., Tong, Y., & Qin, B. (2023). Is ChatGPT Equipped with Emotional Dialogue Capabilities? *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2304.09582