Muhammad Musawir

ANATOMI UJARAN KEBENCIAN

KAJIAN TEORETIS DAN APLIKASI LINGUISTIK FORENSIK



Hak cipta dilindungi oleh undang-undang.
Dilarang memperbanyak maupun mengedarkan buku dalam bentuk dan dengan cara apa pun tanpa izin tertulis dari penerbit maupun penulis.

Sanksi Pelanggaran Pasal 72
Undang-undang No. 19 Tahun 2002
Tentang Hak Cipta

(1) Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp 1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp 5.000.000.000,00 (lima miliar rupiah).
(2) Barangsiapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu Ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah).

ANATOMI UJARAN KEBENCIAN

KAJIAN TEORETIS DAN APLIKASI LINGUISTIK FORENSIK

Penyusun:
Muhammad Musawir

Penata Letak:
Iis Nurul Fadhila

Pendesain Sampul:
Tim Ruang Karya

Diterbitkan Oleh:
Ruang Karya

Alamat:
Jl. Martapura lama, Km. 07, RT. 07, Kecamatan Sungai Tabuk, Kelurahan Sungai Lulut, Kabupaten Banjar, Kalimantan Selatan.

Telp. 0897-1169-692

Email:
kirimnaskah@ruangkarya.id

Hak cipta dilindungi oleh undang-undang

Cetakan Pertama September 2026

Copyright 2026
Halaman 342, Ukuran A5

Apabila menemukan kesalahan cetak dan atau kekeliruan informasi pada buku ini, harap menghubungi Penerbit. Terima kasih.

RUANG KARYA

"Berkarya selagi muda, bermanfaat selagi bisa"

ANATOMI UJARAN KEBENCIAN: KAJIAN TEORETIS DAN APLIKASI LINGUISTIK FORENSIK

PENULIS: MUHAMMAD MUSAWIR

EDITOR: ADILAH SABIR

Kata Pengantar

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga buku ini dapat diselesaikan dengan baik. Buku ini hadir sebagai upaya akademik untuk mengkaji fenomena ujaran kebencian dari perspektif linguistik, khususnya dalam kerangka linguistik forensik, yang hingga kini semakin relevan dalam dinamika komunikasi digital dan ruang publik.

Ujaran kebencian bukan sekadar persoalan bahasa yang kasar atau tidak santun, melainkan praktik komunikasi yang sarat dengan ideologi, relasi kuasa, serta potensi dampak sosial yang serius. Dalam konteks ini, bahasa tidak lagi dipahami sebagai alat netral, tetapi sebagai medium yang dapat membentuk realitas sosial, memperkuat stereotip, bahkan memicu konflik. Oleh karena itu, diperlukan pendekatan yang tidak hanya deskriptif, tetapi juga analitis dan kritis dalam memahami bagaimana ujaran kebencian diproduksi, disebarluaskan, dan dimaknai.

Buku ini disusun untuk memberikan landasan konseptual, kerangka teoretis, serta pendekatan metodologis dalam menganalisis ujaran kebencian secara komprehensif. Pembahasan mencakup berbagai aspek, mulai dari definisi dan karakteristik kebahasaan, kerangka analisis wacana dan pragmatik, hingga penerapan linguistik forensik dalam konteks penegakan hukum. Selain itu, buku ini juga mengulas fenomena ujaran kebencian di media sosial, media tradisional, serta ruang publik dengan mempertimbangkan dimensi identitas, regulasi hukum, dan perkembangan teknologi.

Lebih jauh, buku ini diharapkan dapat menjadi rujukan bagi akademisi, peneliti, mahasiswa, praktisi hukum, serta pemangku kebijakan yang memiliki perhatian terhadap isu bahasa dan keadilan sosial. Penulis juga berharap buku ini dapat berkontribusi dalam meningkatkan literasi kritis masyarakat terhadap ujaran kebencian, sekaligus mendorong praktik komunikasi yang lebih etis dan bertanggung jawab.

Penulis menyadari bahwa buku ini masih memiliki keterbatasan, baik dari segi cakupan maupun kedalaman pembahasan. Oleh karena itu, kritik dan saran yang

konstruktif sangat diharapkan untuk pengembangan karya ini di masa yang akan datang.

Akhir kata, penulis berharap buku ini dapat memberikan manfaat yang luas, tidak hanya dalam ranah akademik, tetapi juga dalam kehidupan sosial masyarakat yang semakin kompleks.

Makassar, 2026

Penulis

Daftar Isi

Kata Pengantar iv
Daftar Isi vii
BAB 1 KERANGKA UMUM UJARAN KEBENCIAN DAN LINGUISTIK FORENSIK 1
1.1 Latar Belakang dan Motivasi Studi ..... **Error! Bookmark not defined.**
1.2 Definisi Awal Ujaran Kebencian..... **Error! Bookmark not defined.**
1.3 Ruang Lingkup Kajian dan Batasan Masalah **Error! Bookmark not defined.**
1.4 Tujuan, Manfaat, dan Sistematika Buku..... **Error! Bookmark not defined.**
BAB 2 KONSEP DAN KARAKTERISTIK UJARAN KEBENCIAN 26
2.1. Definisi Ujaran Kebencian dalam Perspektif Internasional..... **Error! Bookmark not defined.**
2.2. Kategori dan Ciri-ciri Kebahasaan Ujaran Kebencian..... **Error! Bookmark not defined.**
2.3. Perbedaan Ujaran Kebencian, Kritik, dan Kebebasan Berekspresi ..... **Error! Bookmark not defined.**

2.4. Konsekuensi Sosial dan Psikologis Ujaran Kebencian..... **Error! Bookmark not defined.**
BAB 3 KERANGKA TEORITIS LINGUISTIK UJARAN KEBENCIAN 55
3.1. Analisis Wacana Kritis (CDA) dan Ideologi Kebencian..... **Error! Bookmark not defined.**
3.2. Teori Tindak Tutur Dan (Im)Politeness Dalam Ujaran Kebencian..... **Error! Bookmark not defined.**
3.3. Perspektif Semantik dan Pragmatik: Implisit vs Eksplisit ..... **Error! Bookmark not defined.**
3.4. Pendekatan Linguistik Korpus dalam Studi Ujaran Kebencian..... **Error! Bookmark not defined.**
BAB 4 LINGUISTIK FORENSIK – PRINSIP DAN PENERAPAN.....83
4.1. Sejarah dan Ruang Lingkup Linguistik Forensik **Error! Bookmark not defined.**
4.2. Bahasa sebagai Bukti (Evidence) dalam Penegakan Hukum Kebencian ... **Error! Bookmark not defined.**
4.3. Aspek Linguistik Forensik: Fonetik, Semantik, Pragmatik..... **Error! Bookmark not defined.**

4.4. Tahapan & Prosedur Analisis Forensik (Pengumpulan Data, Transkripsi, Analisis)..... **Error! Bookmark not defined.**

4.5. Etika dan Validitas dalam Analisis Linguistik Forensik..... **Error! Bookmark not defined.**

BAB 5 METODOLOGI ANALISIS UJARAN KEBENCIAN .118

5.1 Pendekatan Kualitatif dan Kuantitatif dalam Studi Linguistik..... **Error! Bookmark not defined.**

5.2 Teknik Pengumpulan Data (Corpus Digital, Scraping Media Sosial, Wawancara).**Error! Bookmark not defined.**

5.3 Skema Anotasi dan Pengodean (Coding) Ujaran Kebencian..... **Error! Bookmark not defined.**

5.4 Metode Analisis Wacana dan Konten (mis. Analisis Tematik, Framing)..... **Error! Bookmark not defined.**

5.5 Pendekatan Komputasional dan Pemrosesan Bahasa Alami (NLP) .... **Error! Bookmark not defined.**

5.6 Validasi Data dan Keandalan Analisis (Inter-annotator, Triangulasi)..... **Error! Bookmark not defined.**

BAB 6 UJARAN KEBENCIAN DI MEDIA SOSIAL 160

6.1 Karakteristik Ujaran Kebencian pada Platform Media Sosial ..... **Error! Bookmark not defined.**
6.2 Studi Kasus: Ujaran Kebencian di Media Sosial Indonesia..... **Error! Bookmark not defined.**
6.3 Studi Kasus: Ujaran Kebencian di Media Sosial Internasional..... **Error! Bookmark not defined.**
6.4 Strategi Bahasa dan Retorika (Sarkasme, Eufemisme, Meme) dalam Ujaran Kebencian Online **Error! Bookmark not defined.**
BAB 7 UJARAN KEBENCIAN DALAM RUANG PUBLIK DAN MEDIA TRADISIONAL 194
7.1 Media Massa (Surat Kabar, Televisi, Radio) dan Representasi Kebencian ..... **Error! Bookmark not defined.**
7.2 Pidato Politik, Kampanye Publik, dan Propaganda Kebencian ..... **Error! Bookmark not defined.**
7.3 Ujaran Kebencian dalam Media Baru (Blog, Portal Berita, Podcast) **Error! Bookmark not defined.**
7.4 Budaya Populer dan Humor sebagai Wahana Ujaran Kebencian..... **Error! Bookmark not defined.**
BAB 8 UJARAN KEBENCIAN BERDASARKAN IDENTITAS SASARAN 226

8.1 Etnis, Ras, dan Nasionalisme....**Error! Bookmark not defined.**

8.2 Agama dan Kepercayaan... **Error! Bookmark not defined.**

8.3 Gender dan Orientasi Seksual..**Error! Bookmark not defined.**

8.4 Minoritas Sosial dan Kelompok Rentan Lainnya **Error! Bookmark not defined.**

BAB 9 KERANGKA HUKUM DAN PENEGAKAN UJARAN KEBENCIAN 257

9.1 Regulasi dan Kebijakan Internasional tentang Ujaran Kebencian..... **Error! Bookmark not defined.**

9.2 Perbandingan Undang-Undang Anti-Kebencian (Kasus AS, Uni Eropa, Asia)..... **Error! Bookmark not defined.**

9.3 Peran Linguistik Forensik dalam Penyidikan dan Peradilan Kasus Kebencian..... **Error! Bookmark not defined.**

9.4 Tantangan Hukum dan Etis (Anonimitas, Kebebasan Berekspresi, Cross-border).....**Error! Bookmark not defined.**

BAB 10 IMPLIKASI DAN ARAH PENELITIAN LANJUTAN 282

10.1 Rekomendasi bagi Peneliti dan Praktisi ..... **Error! Bookmark not defined.**

10.2 Implikasi Kebijakan: Moderasi Konten dan Pendidikan Literasi Media..... **Error! Bookmark not defined.**

10.3 Peran Teknologi dan Masyarakat Sipil dalam Menangkal Ujaran Kebencian..... **Error! Bookmark not defined.**

10.4 Ringkasan Temuan dan Refleksi Akhir Buku **Error! Bookmark not defined.**

Daftar Pustaka.....304



BAB 1

KETIKA BAHASA MELUKAI: MENGAPA UJARAN KEBENCIAN PENTING DIBICARAKAN

Dalam ruang publik kontemporer—terutama ketika komunikasi berlangsung cepat, ringkas, dan sering kali anonim—ungkapan kebencian tidak selalu hadir dalam bentuk makian gamblang. Ia dapat terselubung melalui pilihan leksikon yang "wajar", sindiran, generalisasi, pelabelan identitas, hingga strategi pragmatik yang menormalisasi diskriminasi. Karena itu, pembacaan ujaran kebencian menuntut ketelitian linguistik: bukan hanya apa yang dikatakan, melainkan bagaimana, kepada siapa, dalam konteks apa, dan dengan konsekuensi apa ujaran itu diproduksi dan ditafsirkan.

Di sisi lain, ketika ujaran kebencian memasuki ranah pelaporan, penyelidikan, dan pembuktian, bahasa berubah status: dari sekadar medium komunikasi menjadi bukti (evidence). Di titik inilah linguistik forensik menjadi relevan—sebagai disiplin yang memeriksa bahasa secara sistematis untuk kepentingan penegakan hukum dan penilaian profesional yang dapat dipertanggungjawabkan. Kerangka kerja linguistik forensik membantu mengurai relasi antara bentuk kebahasaan (misalnya intensifikasi, dehumanisasi, ajakan/hasutan, implikatur) dan fungsi sosialnya (stigmatisasi, pengucilan, legitimasi kekerasan), sekaligus menilai kekuatan inferensi yang mungkin ditarik dari data kebahasaan.

Namun, kajian ujaran kebencian tidak dapat dipisahkan dari ketegangan normatif antara perlindungan kelompok rentan dan prinsip kebebasan berekspresi. Banyak perdebatan terjadi pada batas-batas: kapan kritik yang keras masih dapat dianggap kritik, kapan ia berubah menjadi kebencian; kapan humor menjadi kendaraan prasangka; kapan ungkapan identitas menjadi serangan terhadap identitas orang lain. Bab ini memosisikan pembaca pada lanskap problematik tersebut—bukan untuk mengabsolutkan satu sisi, melainkan untuk

menyediakan perangkat analitis yang presisi agar pembedaan kategori dapat dilakukan secara argumentatif, konsisten, dan selaras dengan konteks regulasi serta praktik penegakan yang berlaku.

Ketika Kata Berubah Menjadi Bukti

Ujaran kebencian adalah salah satu fenomena kebahasaan yang paling problematik di ruang publik kontemporer karena ia bekerja pada dua level sekaligus. Pertama, sebagai praktik komunikasi yang memproduksi dan mereproduksi relasi kuasa—melalui pelabelan, stereotip, dehumanisasi, serta generalisasi terhadap kelompok tertentu. Kedua, sebagai objek regulasi yang menuntut batas tegas antara ekspresi yang dilindungi dan ekspresi yang patut dibatasi demi mencegah diskriminasi, permusuhan, atau kekerasan. Dalam praktiknya, ujaran kebencian jarang hadir sebagai bentuk "makian telanjang" semata; ia sering muncul sebagai insinuasi, humor yang menormalkan stigma, eufemisme, ironi, hingga ujaran implisit yang bergantung pada konteks bersama dan pengetahuan sosial audiens. Kompleksitas bentuk ini membuat penilaian atas ujaran kebencian tidak dapat berhenti pada identifikasi kata-kata kasar, melainkan

harus menilai fungsi wacana dan daya tindaknya dalam konteks sosial tertentu (Office of the United Nations High Commissioner for Human Rights, 2020).

Di Indonesia, persoalan ini semakin mengemuka karena lanskap komunikasi publik—terutama komunikasi bermediasi digital—memperbesar kecepatan sebar, jangkauan, dan daya replikasi pesan. Pada saat yang sama, perangkat hukum yang sering dijadikan rujukan untuk penanganan ujaran kebencian tersebar pada beberapa rezim normatif (KUHP dan UU sektor tertentu), sehingga memunculkan kebutuhan akan pedoman interpretasi yang lebih konsisten. Salah satu pasal yang kerap menjadi titik rujuk ialah ketentuan mengenai distribusi/transmisi informasi elektronik yang "menghasut, mengajak, atau memengaruhi" sehingga menimbulkan "rasa kebencian atau permusuhan" terhadap individu/kelompok berdasarkan kategori identitas tertentu; ketentuan ini ditegaskan dalam perubahan kedua UU ITE melalui UU No. 1 Tahun 2024 (Republik Indonesia, 2024). Selain itu, rujukan KUHP mengenai pernyataan permusuhan/kebencian/ penghinaan terhadap golongan juga sering hadir dalam perdebatan yuridis dan praktik penegakan, sebagaimana

termuat dalam dokumen putusan pengujian di Mahkamah Konstitusi yang mengutip norma KUHP terkait (Mahkamah Konstitusi Republik Indonesia, 2018b).

Kepolisian pernah menerbitkan pedoman internal melalui Surat Edaran Kapolri No. SE/06/X/2015 yang merangkum bentuk-bentuk tindakan yang sering diasosiasikan sebagai ujaran kebencian (misalnya penghinaan, pencemaran nama baik, penistaan, provokasi, hasutan, dan penyebaran berita bohong), sekaligus menekankan dampak sosialnya bila dibiarkan (Kepolisian Negara Republik Indonesia, 2015). Di ranah antarlembaga, hadir pula Keputusan Bersama (Kominfo/Komdigi, Kejaksaan Agung, dan Polri) tentang pedoman implementasi pasal tertentu dalam UU ITE, yang secara normatif bertujuan menyeragamkan penerapan dan memberi rambu interpretasi (Mahkamah Agung Republik Indonesia, 2021). Keberadaan ragam instrumen ini menunjukkan dua hal: negara menganggap ujaran kebencian sebagai persoalan serius, dan penanganannya menuntut kerja interpretasi yang presisi—karena istilah kunci seperti "menghasut", "mengajak", "memengaruhi", "kebencian", serta "permusuhan" tidak otomatis memiliki batas semantik-pragmatik yang tunggal di semua konteks.

Di titik inilah linguistik forensik menjadi relevan. Ia memandang bahasa bukan sekadar medium, melainkan juga bukti yang dapat dianalisis secara sistematis untuk menjawab pertanyaan-pertanyaan hukum: apa makna yang paling mungkin, bagaimana intensi dan daya ilokusi bekerja, siapa sasaran ujaran, bagaimana konteks membentuk inferensi audiens, serta seberapa kuat hubungan antara bentuk ujaran dan potensi dampak sosialnya. Tradisi kajian language and law menegaskan bahwa banyak problem penegakan hukum berakar pada problem bahasa: ambiguitas, praanggapan, implikatur, serta perbedaan interpretasi antarpartisipan dan institusi (Gibbons, 2003).

Pertanyaan yang tidak dapat dihindari kemudian adalah bagaimana menjembatani ketegangan normatif yang inheren: perlindungan kebebasan berekspresi versus perlindungan martabat dan rasa aman kelompok sasaran. Dalam diskursus internasional, perdebatan tersebut tampak jelas—misalnya dalam argumen tentang harm dan martabat (dignity) yang dirusak oleh ujaran kebencian (Waldron, 2012). Untuk menakar ketegangan ini secara lebih terukur, salah satu rujukan penting adalah Rabat Plan of Action, yang menawarkan uji ambang

(threshold test) enam unsur: konteks sosial-politik, status penutur, intensi, konten dan bentuk, jangkauan diseminasi, serta kemungkinan dampak (termasuk imminence). Kerangka ini penting agar analisis tidak terjebak pada moral panic atau kriminalisasi berlebih, tetapi juga tidak mengabaikan ujaran yang benar-benar meningkatkan risiko kekerasan atau diskriminasi serius (Office of the United Nations High Commissioner for Human Rights, 2020).

Di sisi lain, literatur kontemporer menunjukkan bahwa kategori hate speech sangat luas; karenanya sebagian kalangan mengusulkan fokus yang lebih sempit pada ujaran yang menaikkan risiko kekerasan kelompok (dangerous speech), dengan indikator-indikator yang dapat dioperasionalkan untuk pencegahan (Benesch, 2014). Perspektif ini bukan pengganti pembahasan ujaran kebencian, melainkan pelengkap yang membantu membedakan spektrum ujaran yang ofensif, ujaran diskriminatif yang menormalisasi eksklusi, hingga ujaran yang berpotensi menjadi pemicu tindakan kolektif berbahaya.

Buku ini lahir dari tiga kebutuhan yang saling terkait. Yang pertama adalah kebutuhan akan perangkat analisis

kebahasaan yang mampu memeriksa ujaran kebencian secara berlapis—leksikal, semantik, pragmatik, wacana, hingga multimodal—dan tetap kompatibel dengan kebutuhan pembuktian: transparansi langkah analisis, justifikasi inferensi, serta keterlacakan data (Coulthard dkk., 2017). Yang kedua adalah kebutuhan akan konsistensi penilaian lintas kasus, terutama ketika ujaran bersifat implisit dan sangat bergantung pada konteks lokal—pengetahuan budaya, isu sosial, relasi antarkelompok—sehingga konteks perlu diperlakukan sebagai variabel analitis, bukan sekadar latar cerita (Office of the United Nations High Commissioner for Human Rights, 2020). Yang ketiga adalah kebutuhan akan ruang dialog antara disiplin linguistik, praktik penegakan hukum, dan kerangka kebijakan, agar regulasi tidak berhenti pada norma, melainkan memiliki panduan interpretasi yang dapat diuji dan dievaluasi (Mahkamah Agung Republik Indonesia, 2021).

Dari situ, buku ini mencoba menyusun "anatomi" ujaran kebencian: memetakan komponen kebahasaan yang berulang—strategi pelabelan identitas, predikasi negatif, dehumanisasi, ajakan/hasutan, insinuasi—memeriksa bagaimana komponen itu berinteraksi dengan

konteks, serta menunjukkan konsekuensi interpretatifnya ketika masuk ke ranah forensik. Buku ini ditujukan bagi siapa pun yang membutuhkan kerangka untuk melakukan pembacaan yang lebih presisi terhadap ujaran kebencian—peneliti, pendidik, analis konten, maupun aparat penegak hukum—dengan tetap menjaga kehati-hatian terhadap risiko salah tafsir, bias, dan kriminalisasi yang tidak proporsional.

Ujaran Kebencian: Sebuah Definisi Kerja

Dalam kajian linguistik forensik, "ujaran kebencian" (hate speech) tidak dapat diperlakukan sebagai istilah yang sepenuhnya stabil dan seragam. Di satu sisi, ia adalah konsep sosial-budaya yang hidup dalam praktik komunikasi (offline maupun online). Di sisi lain, ia adalah konsep normatif yang dibatasi oleh rezim hak asasi manusia, kebijakan publik, dan perangkat hukum pidana. Karena itu, definisi awal perlu menempatkan ujaran kebencian sebagai kategori bertingkat: mulai dari ekspresi yang merendahkan (denigrating), ekspresi yang mempromosikan kebencian, hingga ekspresi yang menghasut (incitement) diskriminasi/permusuhan/

kekerasan—yang pada level tertentu dapat memicu konsekuensi hukum.

Secara internasional, tidak ada satu definisi tunggal yang diterima universal. PBB, misalnya, menggunakan definisi kerja yang menekankan bentuk komunikasi (ucapan, tulisan, perilaku) yang menyerang atau memakai bahasa peyoratif/diskriminatif terhadap individu/kelompok berdasarkan identitas tertentu. Definisi kerja ini penting sebagai payung konseptual, namun tidak otomatis identik dengan ambang pidana (United Nations, t.t.). Dalam kerangka Eropa, Recommendation CM/Rec(2022)16 dari Council of Europe merumuskan hate speech sebagai "semua jenis ekspresi" yang menghasut, mempromosikan, menyebarkan, atau membenarkan kekerasan, kebencian, atau diskriminasi terhadap seseorang/kelompok, atau yang merendahkan mereka, karena karakteristik personal nyata atau yang dilekatkan (misalnya ras, agama, kebangsaan/asal etnis, usia, disabilitas, jenis kelamin, identitas gender, orientasi seksual) (Council of Europe, 2022). Rumusan ini memperlihatkan dua penekanan: fokus pada aksi wacana (incite/promote/spread/justify/

denigrate), dan fokus pada sasaran berbasis identitas (protected characteristics).

Kerangka HAM internasional memberi penanda krusial untuk membedakan ujaran yang sekadar ofensif dari ujaran yang secara normatif wajib dilarang. Pasal 20(2) ICCPR mewajibkan pelarangan "advokasi kebencian" berbasis kebangsaan, ras, atau agama yang merupakan hasutan (incitement) untuk diskriminasi, permusuhan, atau kekerasan (Office of the United Nations High Commissioner for Human Rights, 1966). Selain itu, CERD (khususnya Pasal 4) menekankan pelarangan penyebaran gagasan yang didasarkan pada superioritas/kebencian rasial dan hasutan untuk diskriminasi rasial (Office of the United Nations High Commissioner for Human Rights, t.t.). Untuk mengoperasionalkan ambang "hasutan" secara lebih terukur, Rabat Plan of Action menawarkan uji ambang yang lazim dirujuk dalam praktik: konteks, status/posisi pembicara, niat (intent), isi dan bentuk (content/form), jangkauan penyebaran (extent), serta kemungkinan/ketak-terelakan dampak (likelihood) termasuk imminence (Office of the United Nations High Commissioner for Human Rights, 2020). Bagi linguistik

forensik, uji ambang ini sangat bernilai karena menuntut analisis kebahasaan yang selalu ditambatkan pada konteks produksi–sirkulasi–resepsi wacana, bukan semata kutipan lepas.

Dalam konteks Indonesia, istilah "ujaran kebencian" beroperasi pada beberapa lapis perangkat hukum. KUHP (warisan kolonial yang masih relevan dalam praktik penegakan) memuat rumusan tentang "perasaan permusuhan, kebencian, atau penghinaan terhadap suatu atau beberapa golongan" (Pasal 156) serta tindakan menyiarkan/menempelkan tulisan atau lukisan yang memuat pernyataan permusuhan/kebencian/penghinaan antargolongan (Pasal 157) (Mahkamah Konstitusi Republik Indonesia, 2018a). Dua pasal ini penting karena menunjukkan orientasi klasik hukum pidana Indonesia: ujaran kebencian dipahami terutama sebagai ekspresi publik yang memproduksi permusuhan/kebencian antargolongan, dengan "golongan" dikaitkan dengan diferensiasi identitas (misalnya ras, agama, asal, keturunan, kebangsaan, dan seterusnya) sebagaimana tercermin dalam uraian normatif yang dikutip dalam putusan MK.

Untuk ranah digital, UU ITE (dalam konfigurasi perubahannya) menegaskan aspek "penyebaran/transmisi" informasi elektronik yang bersifat menghasut, mengajak, atau memengaruhi orang lain sehingga menimbulkan kebencian atau permusuhan terhadap individu/kelompok tertentu berdasarkan kategori identitas (ras, kebangsaan, etnis, warna kulit, agama, kepercayaan, jenis kelamin, disabilitas mental/fisik), sebagaimana tampak antara lain dalam ketentuan pidana yang merujuk pada Pasal 28 ayat (2) (Republik Indonesia, 2024). Pada saat yang sama, regulasi transisional juga mengindikasikan adanya peralihan rezim melalui pemberlakuan KUHP Nasional (UU No. 1 Tahun 2023) setelah masa jeda tiga tahun sejak pengundangan (Republik Indonesia, 2024). Konsekuensinya, definisi awal dalam buku ini tidak cukup hanya mengutip norma, melainkan harus siap mengikuti pergeseran locus penanganan—dari pasal-pasal tertentu di UU ITE menuju kodifikasi baru—beserta implikasi interpretasinya.

Dari paparan internasional dan nasional tersebut, ujaran kebencian dapat dilihat selalu memuat tiga komponen pokok, yang akan menjadi fondasi analisis

forensik pada bab-bab berikutnya: komponen target berbasis identitas (protected or attributed characteristics)—ujaran kebencian bukan sekadar serangan personal, ia menautkan serangan pada identitas (nyata atau dilekatkan) dan/atau memposisikan korban sebagai representasi kelompok; komponen tindakan wacana (speech act) yang merendahkan hingga menghasut, dengan spektrum dari denigrating speech (merendahkan/menstigma) sampai inciting speech (menghasut diskriminasi, permusuhan, kekerasan); serta komponen konteks–dampak (context and harm)—dalam perspektif HAM, terutama ketika memasuki ranah pembatasan atau pemidanaan, konteks produksi dan potensi dampak menjadi penentu: siapa yang berbicara, kepada siapa, dalam situasi sosial-politik apa, melalui medium apa, dengan jangkauan seberapa luas, dan dengan probabilitas dampak seperti apa.

Berangkat dari ketiga komponen tersebut, buku ini menggunakan definisi kerja (working definition) berikut untuk kebutuhan pemetaan konseptual dan analisis data:

> *Ujaran kebencian adalah setiap bentuk ekspresi (lisan, tulis, multimodal, atau digital) yang menyerang, merendahkan, menstigma, mempromosikan,*

menyebarkan, membenarkan, atau menghasut kebencian/diskriminasi/permusuhan/kekerasan terhadap individu atau kelompok, terutama ketika serangan tersebut ditautkan pada identitas tertentu (nyata atau dilekatkan) dan diproduksi dalam konteks yang memungkinkan terjadinya kerugian sosial (harm) pada korban maupun kohesi sosial.

Definisi kerja ini sengaja bersifat cukup luas agar mampu menangkap realitas ujaran kebencian sebagai fenomena wacana. Namun, untuk kepentingan penilaian hukum, buku ini membedakan dengan tegas "ujaran kebencian sebagai fenomena wacana" dan "ujaran kebencian yang memenuhi ambang pembatasan/ pemidanaan". Pada tahap pembuktian/ penilaian yuridis, parameter Rabat Plan (konteks–pembicara–niat–isi/bentuk–jangkauan–kemungkinan dampak) digunakan sebagai jembatan metodologis agar analisis kebahasaan tidak terjebak pada penggalan teks, melainkan bergerak sebagai evaluasi forensik yang dapat dipertanggungjawabkan.

Cakupan dan Batas Buku Ini

Agar pembahasan ujaran kebencian tidak melebar menjadi "segala bentuk ujaran ofensif", sekaligus tidak menyempit hanya pada frasa atau kata tertentu, buku ini bekerja pada wilayah yang cukup spesifik. Penetapan cakupan ini diperlukan karena, secara normatif, tidak ada definisi legal internasional tunggal tentang hate speech; klasifikasi "hateful" juga kerap diperdebatkan dan bergantung pada konteks sosial-politik serta kerangka regulatif yang digunakan (United Nations., 2019). Karena itu, buku ini menempatkan ujaran kebencian sebagai fenomena wacana yang dianalisis dengan perangkat linguistik, tetapi tetap menautkannya pada ambang pembatasan/pertanggungjawaban yang menuntut kehati-hatian interpretatif (Office of the United Nations High Commissioner for Human Rights, 2020).

Objek utama buku ini adalah ekspresi kebahasaan yang menargetkan individu/kelompok berbasis identitas (misalnya ras/etnis, agama/kepercayaan, gender, disabilitas, dan kategori serupa) dan bekerja melalui tindakan wacana seperti merendahkan, menstigma, mempromosikan kebencian, atau menghasut permusuhan/kekerasan—sejalan dengan definisi

kebijakan yang menekankan spektrum tindakan wacana (incite/promote/spread/justify/denigrate) dan karakteristik yang "nyata atau dilekatkan" (Council of Europe, 2022). Bahasa diperlakukan sebagai data sekaligus bukti: sebagai data untuk memetakan strategi kebahasaan kebencian (leksikon, metafora, label identitas, generalisasi, dehumanisasi, implikatur, sarkasme), dan sebagai bukti yang dianalisis secara forensik untuk menilai daya tindak ujaran, sasaran, konteks, serta kekuatan inferensi yang wajar dalam pemeriksaan kasus (Coulthard dkk., 2017). Dalam hal ambang "hasutan", buku ini merujuk uji ambang Rabat sebagai perangkat untuk menahan analisis agar tidak jatuh pada over-interpretation atau kriminalisasi berlebih (Office of the United Nations High Commissioner for Human Rights, 2020).

Dari sisi medan data, buku ini bergerak di dua ranah utama: media sosial (misalnya Twitter/X, Instagram, TikTok) berikut strategi retorika khas ruang digital seperti meme, eufemisme, dan sarkasme; serta ruang publik dan media tradisional, termasuk media massa, pidato politik/kampanye, dan media baru seperti blog, portal berita, dan podcast. Pilihan ini selaras dengan

pedoman nasional yang mengakui ujaran kebencian dapat dilakukan melalui berbagai media, dari orasi kampanye sampai jejaring media sosial (Kepolisian Negara Republik Indonesia, 2015). Studi kasus Indonesia menjadi korpus utama, namun pembanding internasional tetap dihadirkan untuk memperjelas variasi rezim konseptual dan kebijakan, terutama pada bab yang membandingkan kasus media sosial internasional. Rujukan normatif yang dipakai mencakup kerangka HAM internasional—khususnya ICCPR Pasal 20(2) mengenai advokasi kebencian yang merupakan hasutan untuk diskriminasi/permusuhan/kekerasan (Office of the United Nations High Commissioner for Human Rights, 1966)—kerangka kebijakan regional seperti Council of Europe Recommendation CM/Rec(2022)16, yang juga membedakan tingkat keparahan ujaran: mana yang masuk ranah pidana, mana yang cukup dengan mekanisme perdata/administratif, dan mana yang sebaiknya direspons dengan langkah non-punitif seperti counterspeech dan edukasi (Council of Europe, 2022)—serta kerangka hukum positif Indonesia, terutama rumusan UU ITE mengenai konten elektronik yang "menghasut, mengajak, atau memengaruhi" sehingga

menimbulkan kebencian/permusuhan berbasis identitas (Republik Indonesia, 2024), berikut pedoman implementasi lintas-lembaga melalui SKB UU ITE (Republik Indonesia, 2024).

Ada pula beberapa hal yang secara sengaja tidak menjadi fokus utama buku ini, dan penting untuk dinyatakan sejak awal agar ekspektasi pembaca terjaga. Buku ini bukan buku nasihat hukum dan bukan substitusi putusan pengadilan—pembahasan hukum digunakan sebagai kerangka interpretasi dan literasi kebijakan, bukan untuk memberi kesimpulan yuridis yang mengikat; analisis linguistik diposisikan sebagai dukungan argumentasi, sedangkan penetapan pertanggungjawaban pidana/perdata tetap berada pada otoritas peradilan. Fokusnya juga pada ujaran berbasis identitas, bukan seluruh ujaran ofensif—ujaran menghina yang bersifat personal tanpa kaitan identitas kelompok dibahas sejauh relevan untuk membedakan batas konseptual, tetapi tidak menjadi objek utama, konsisten dengan kerangka yang menekankan protected/attributed characteristics (Council of Europe, 2022). Penilaian "hasutan" pun selalu dibatasi oleh uji ambang konteks: buku ini tidak menganggap kata atau ungkapan tertentu otomatis

memenuhi unsur hasutan, karena ambang penilaian mengandalkan variabel konteks, status penutur, niat, isi/bentuk, jangkauan, dan kemungkinan dampak (Office of the United Nations High Commissioner for Human Rights, 2020).

Studi kasus yang dihadirkan bersifat purposif untuk menunjukkan pola kebahasaan dan mekanisme wacana, bukan untuk mengklaim representasi statistik seluruh platform atau seluruh periode—buku ini mengakui adanya dinamika kebijakan platform dan perubahan ekosistem digital yang dapat memengaruhi bentuk ujaran dan praktik moderasi. Pembahasan pendekatan komputasional dan NLP juga dibatasi pada konsep-konsep yang mendukung pembacaan linguistik, seperti ekstraksi fitur, anotasi, dan reliabilitas, bukan pengembangan model prediktif mutakhir atau rekayasa sistem deteksi skala industri; ia diperlakukan sebagai perangkat bantu, bukan fokus tunggal buku. Terakhir, contoh dan korpus yang digunakan diutamakan dari sumber yang dapat diakses secara sah (misalnya materi publik), dengan perhatian pada anonimisasi dan pengurangan risiko reviktimisasi—prinsip kehati-hatian yang sejalan dengan

kebutuhan akuntabilitas dan etika dalam analisis bukti kebahasaan (Coulthard dkk., 2017).

Dengan cakupan dan batas tersebut, pembaca memperoleh ekspektasi yang jelas: buku ini bergerak pada irisan linguistik, konteks sosial, dan kerangka normatif, dengan fokus pada bagaimana ujaran kebencian "bekerja" sebagai praktik wacana dan bagaimana ia dapat dianalisis secara forensik secara terukur dan dapat dipertanggungjawabkan.

Ke Mana Buku Ini akan Membawa Anda

Kajian ujaran kebencian berada pada irisan linguistik, sosial, dan hukum, serta diwarnai perdebatan konseptual karena tidak ada definisi legal internasional tunggal dan karakterisasi "kebencian" sering diperselisihkan (United Nations., 2019). Buku ini disusun untuk memberi kerangka yang operasional bagi analisis kebahasaan sekaligus akuntabel ketika temuan dipakai dalam ruang kebijakan dan penegakan hukum, terutama ketika isu "hasutan kebencian" menuntut ambang penilaian yang tinggi dan berbasis konteks (Office of the United Nations High Commissioner for Human Rights, 2020). Secara umum, buku ini membangun pemahaman komprehensif

mengenai "anatomi" ujaran kebencian—mulai dari konsep, ciri kebahasaan, hingga penerapan analisis linguistik forensik—dengan fokus utama pada studi kasus Indonesia, namun tetap menyediakan pembanding internasional pada bab tertentu.

Perjalanan buku ini tersusun dalam lima bagian besar. Bagian pertama, yang sedang Anda baca, meletakkan landasan konseptual: membedakan ujaran kebencian dari kritik, satire, dan kebebasan berekspresi, sekaligus memetakan ciri-ciri kebahasaan dan konsekuensi sosial-psikologisnya. Bagian kedua membawa pembaca masuk ke kerangka teoretis—Analisis Wacana Kritis, tindak tutur dan (im)politeness, semantik-pragmatik, serta linguistik korpus—sebelum beralih ke prinsip-prinsip linguistik forensik: bagaimana bahasa diperlakukan sebagai bukti (evidence), mencakup tahapan pengumpulan data, transkripsi, analisis, serta pertimbangan etika dan validitas, sejalan dengan tradisi language in evidence yang menekankan antarmuka bahasa dan hukum (Coulthard dkk., 2017).

Bagian ketiga menyajikan perangkat analisis yang dapat dipakai ulang oleh pembaca sendiri—kombinasi cara membaca kualitatif dan kuantitatif, skema anotasi,

teknik analisis wacana dan konten, hingga pengantar pendekatan komputasional (NLP) sebagai pelengkap—dengan penekanan pada bagaimana mengurangi bias interpretasi dan menjaga konsistensi pembacaan lintas kasus. Bagian keempat menguji kerangka ini pada data nyata: media sosial, ruang publik dan media tradisional, serta ragam sasaran berbasis identitas (etnis/ras, agama, gender/orientasi seksual, kelompok rentan lain). Bagian kelima menutup buku dengan mengaitkan analisis linguistik pada kerangka hukum dan penegakan di level internasional dan nasional, termasuk tantangan etis seperti anonimitas dan lintas yurisdiksi (cross-border)—dalam hal ambang "hasutan kebencian", buku ini merujuk uji ambang Rabat (konteks, status penutur, niat, konten/bentuk, jangkauan, dan kemungkinan dampak termasuk imminence) sebagai rambu evaluasi yang menahan over-criminalization sekaligus tetap peka pada risiko bahaya nyata—sebelum ditutup dengan refleksi tentang peran teknologi, kebijakan, dan masyarakat sipil dalam menangkal ujaran kebencian.

Bagi pembaca dari kalangan linguistik dan studi wacana, buku ini memetakan strategi kebahasaan ujaran kebencian—label identitas, dehumanisasi, generalisasi,

insinuasi, eufemisme, sarkasme, dan bentuk multimodal—dan menghubungkannya dengan mekanisme ideologi serta kekuasaan dalam teks dan praktik sosial. Bagi pembaca yang bekerja dengan data dan bukti, buku ini menawarkan prosedur analisis yang transparan—dari pengumpulan data, anotasi, pengodean, hingga validasi—sehingga temuan lebih replikatif dan dapat diaudit, sesuatu yang penting dalam konteks forensik karena analisis bahasa yang diposisikan sebagai bukti menuntut justifikasi yang jelas dan jejak langkah yang dapat ditelusuri (Coulthard dkk., 2017). Bagi pembaca dari kalangan penegakan hukum dan kebijakan, buku ini mengintegrasikan kerangka HAM internasional dan pedoman ambang "hasutan" untuk membantu membedakan ujaran yang sekadar ofensif, ujaran yang diskriminatif/menstigma, hingga ujaran yang berpotensi menghasut kebencian dan membutuhkan respons kebijakan/hukum yang proporsional (Office of the United Nations High Commissioner for Human Rights, 2020); di Indonesia, kebutuhan pedoman interpretasi semacam ini juga relevan mengingat perangkat rujukan penanganan mencakup regulasi dan pedoman implementasi lintas lembaga, seperti perubahan UU ITE dan pedomannya

(Republik Indonesia, 2024). Dan bagi pembaca dari kalangan pendidik serta pegiat literasi media, buku ini menjadi bekal untuk merancang intervensi edukatif—literasi bahasa, literasi digital, counterspeech—berbasis pemahaman mekanisme wacana kebencian, mengingat ujaran kebencian berdampak pada kohesi sosial dan dapat meningkatkan risiko diskriminasi maupun kekerasan (United Nations, t.t.).

Dengan peta perjalanan ini, mari kita mulai dari pertanyaan paling mendasar: apa sebenarnya yang kita maksud dengan ujaran kebencian, dan bagaimana ia dikenali secara konsisten lintas konteks?



BAB 2
ANATOMI KEBAHASAAN UJARAN KEBENCIAN

Bab sebelumnya menetapkan definisi kerja dan cakupan buku ini. Bab ini memperdalamnya dengan memusatkan perhatian pada satu pertanyaan: apa yang dimaksud sebagai ujaran kebencian, dan bagaimana ia dikenali melalui karakteristik yang relatif konsisten lintas konteks?

Ujaran kebencian, dalam pembahasan berikut, dipahami sebagai konsep yang beroperasi pada dua ranah sekaligus. Pada ranah pertama, ia adalah fenomena wacana: tindakan berbahasa yang dapat memproduksi stigma, menormalisasi diskriminasi, dan mengorganisasi "kita–mereka" melalui pelabelan identitas, generalisasi,

dehumanisasi, serta pembenaran moral atas pengucilan. Pada ranah kedua, ia adalah kategori normatif yang beririsan dengan hak asasi manusia, kebijakan publik, dan hukum pidana—sehingga menuntut pembedaan tegas antara ekspresi yang sekadar ofensif dan ekspresi yang mencapai ambang "hasutan" atau menimbulkan risiko bahaya sosial yang nyata.

Karena itu, bab ini tidak hanya memaparkan definisi, tetapi juga menyusun peta kategori dan karakteristik: spektrum bentuk ujaran kebencian (dari yang eksplisit hingga implisit), indikator kebahasaan yang kerap muncul, serta cara ujaran tersebut "bekerja" dalam konteks sosial tertentu. Pemetaan ini penting agar pembacaan pada bab-bab kasus nanti—media sosial, ruang publik, serta sasaran berbasis identitas—tidak terjebak pada pembacaan yang serba intuitif, melainkan bertumpu pada ciri-ciri yang dapat dirumuskan dan diuji. Bab ini juga menyiapkan pembaca untuk perdebatan yang tidak dapat dihindari: keterkaitan ujaran kebencian dengan kebebasan berekspresi, kritik, satire, dan humor; serta isu batas antara kontroversi yang sah dan praktik wacana yang merendahkan martabat kelompok tertentu.

Satu Istilah, Banyak Definisi

Dalam literatur dan instrumen internasional, istilah hate speech tidak bekerja sebagai satu definisi tunggal yang berlaku universal. Ia beroperasi pada dua level yang perlu dibedakan sejak awal: definisi kerja (working definition) untuk kepentingan kebijakan, pendidikan, dan pencegahan; dan rumusan normatif yang terkait kewajiban negara untuk melarang bentuk ekspresi tertentu—terutama ketika ekspresi tersebut mencapai ambang hasutan (incitement) terhadap diskriminasi, permusuhan, atau kekerasan. Perbedaan level ini penting bagi kajian linguistik forensik, sebab analisis kebahasaan harus mampu menjelaskan fenomena wacana secara luas, namun juga mampu menilai kapan sebuah ujaran mendekati atau memenuhi ambang pembatasan yang ketat.

PBB menggunakan definisi kerja yang banyak dirujuk dalam kebijakan dan literasi publik. Dalam UN Strategy and Plan of Action on Hate Speech, hate speech dipahami sebagai bentuk komunikasi—dalam ujaran, tulisan, atau perilaku—yang menyerang atau menggunakan bahasa peyoratif/diskriminatif terhadap seseorang atau kelompok berdasarkan identitas (misalnya agama, etnis,

kebangsaan, ras, warna kulit, keturunan, gender, atau faktor identitas lain) (United Nations, t.t.). Definisi kerja ini sengaja bersifat inklusif karena dimaksudkan untuk kebutuhan pencegahan dini (early prevention), counter-speech, dan desain intervensi kebijakan lintas sektor. Ia tidak identik dengan "yang harus dipidana", melainkan berfungsi sebagai payung untuk memetakan spektrum praktik wacana yang menyerang martabat berbasis identitas—mulai dari penghinaan yang menstigma hingga ujaran yang berpotensi menghasut kebencian (United Nations, t.t.).

Pada level kewajiban HAM internasional, penanda kuncinya bukan semata label hate speech, melainkan kategori advokasi kebencian yang merupakan hasutan. ICCPR Pasal 20(2) mewajibkan pelarangan oleh hukum terhadap segala advokasi kebencian nasional, rasial, atau agama yang merupakan hasutan untuk diskriminasi, permusuhan, atau kekerasan (Office of the United Nations High Commissioner for Human Rights, 1966). Dengan kata lain, ambang kewajiban internasional bergerak pada hubungan antara kebencian berbasis identitas dan hasutan menuju tindakan atau konsekuensi sosial yang serius. Secara lebih spesifik untuk kebencian rasial, CERD

(Konvensi Penghapusan Diskriminasi Rasial) Pasal 4 meminta negara mengkriminalkan, antara lain, penyebaran gagasan berbasis superioritas/kebencian rasial dan hasutan untuk diskriminasi rasial, termasuk hasutan untuk tindakan kekerasan (Office of the United Nations High Commissioner for Human Rights, t.t.). Kerangka ini memperlihatkan bahwa ujaran kebencian dalam perspektif HAM tidak berhenti pada aspek ofensif, melainkan terkait erat dengan pencegahan diskriminasi struktural dan kekerasan. Konsekuensinya bagi buku ini: ketika "ujaran kebencian" dibahas sebagai fenomena wacana yang spektrumnya luas, "ujaran kebencian" yang menuntut pelarangan dalam kerangka HAM internasional cenderung merujuk pada subset yang memenuhi ambang incitement sebagaimana ditunjukkan ICCPR Pasal 20(2) dan instrumen terkait.

Karena ketegangan antara kebebasan berekspresi dan pelarangan hasutan kebencian selalu rawan disalahgunakan, OHCHR melalui Rabat Plan of Action mendorong ambang yang tinggi dan menawarkan uji enam unsur untuk menilai kapan ekspresi dapat dipandang sebagai incitement to hatred yang layak dikenai pembatasan ketat: konteks sosial-politik,

status/posisi penutur, niat untuk menghasut, konten dan bentuk ujaran, jangkauan diseminasi, serta kemungkinan dampak/kerugian termasuk imminence (Office of the United Nations High Commissioner for Human Rights, 2020). Uji ambang ini memiliki dua fungsi definisional yang penting bagi linguistik forensik. Pertama, ia menegaskan bahwa "ujaran kebencian" tidak dapat dinilai secara adil hanya dari kutipan tekstual; definisi yang bermakna harus memasukkan konteks produksi, sirkulasi, dan resepsi. Kedua, ia menuntut analisis yang membedakan antara ujaran yang menghina (meski problematik) dan ujaran yang menggerakkan audiens menuju diskriminasi/permusuhan/kekerasan.

Di Eropa, Council of Europe melalui Recommendation CM/Rec(2022)16 memberikan definisi yang eksplisit dan relatif luas: hate speech mencakup semua bentuk ekspresi yang menghasut, mempromosikan, menyebarkan, atau membenarkan kekerasan, kebencian, atau diskriminasi terhadap seseorang/kelompok, atau yang merendahkan mereka, karena karakteristik personal nyata atau yang dilekatkan—misalnya ras, warna, bahasa, agama, kebangsaan/asal etnis, usia, disabilitas, jenis kelamin, identitas gender, orientasi seksual (Council of Europe,

2022). Yang relevan bagi buku ini bukan hanya definisinya, melainkan pendekatan bertingkatnya (tiered approach): ada hate speech yang dilarang oleh hukum pidana, ada yang tidak mencapai ambang pidana namun dapat direspons lewat hukum perdata/administratif, dan ada ekspresi ofensif/berbahaya yang tidak cukup berat untuk dibatasi secara sah, tetapi membutuhkan respons alternatif seperti counter-speech dan edukasi (Council of Europe, 2022). Pendekatan bertingkat ini membantu membedakan spektrum fenomena wacana tanpa menyamakan semuanya sebagai objek pemidanaan.

Sebagai pembanding, rezim Uni Eropa cenderung mempersempit definisi untuk ranah pidana. Council Framework Decision 2008/913/JHA tentang rasisme dan xenofobia menuntut negara anggota mengkriminalkan hasutan publik terhadap kekerasan atau kebencian (public incitement to violence or hatred) yang ditujukan kepada kelompok atau anggota kelompok berdasarkan ras, warna, keturunan, agama/kepercayaan, atau asal kebangsaan/etnis (European Union, 2008). Model ini lebih sempit dibanding definisi kerja PBB atau definisi CoE: EU menempatkan inti hate speech dalam ranah pidana pada hasutan publik terhadap kekerasan atau

kebencian—pendekatan yang memudahkan pemetaan unsur delik, tetapi sekaligus dapat meninggalkan wilayah denigrating speech (merendahkan/menstigma) sebagai area yang umumnya ditangani melalui mekanisme non-pidana, kebijakan platform, atau intervensi sosial (European Union, 2008).

Meski bervariasi, definisi-definisi internasional di atas bertemu pada tiga titik yang berguna sebagai fondasi buku ini. Semuanya menekankan sasaran berbasis identitas (protected/attributed characteristics)—baik PBB maupun CoE menyoroti serangan terhadap seseorang/kelompok karena identitas yang nyata atau dilekatkan. Semuanya juga menekankan tindakan wacana (speech act) yang melampaui "sekadar berkata kasar", dengan spektrum dari menyerang/menggunakan bahasa peyoratif (PBB), merendahkan (CoE), hingga menghasut kebencian/kekerasan (ICCPR/EU). Dan semuanya menempatkan konteks serta ambang bahaya sebagai penentu ketika definisi diarahkan untuk pembatasan atau pemidanaan—yang dioperasionalkan melalui uji ambang Rabat. Dengan pijakan ini, "definisi internasional" yang dipakai buku ini tidak dipaksakan menjadi satu kalimat tunggal, melainkan diperlakukan sebagai kerangka

bertingkat: definisi kerja untuk memetakan fenomena wacana, dan definisi berbasis ambang untuk menilai kemungkinan pembatasan/pertanggungjawaban. Model ini menjaga analisis tetap komprehensif, namun tidak terjerumus ke penyamaan otomatis antara "problem wacana" dan "objek pemidanaan".

Mengenali Ujaran Kebencian: Kategori dan Ciri Kebahasaan

Ujaran kebencian adalah fenomena wacana yang bertingkat—dari ekspresi yang merendahkan hingga yang mendekati "hasutan" (incitement). Perspektif bertingkat ini penting untuk dua hal: memahami bagaimana kebencian "dibangun" dalam bahasa, dan memeriksa kapan ujaran berpotensi memenuhi ambang pembatasan yang ketat, misalnya dalam kerangka uji Rabat. Council of Europe sendiri menegaskan bahwa tidak ada metodologi universal untuk menilai ujaran kebencian, sehingga pembacaan perlu dioperasionalkan melalui indikator yang lebih konkret dan tetap sensitif terhadap konteks.

Tiga tingkat keparahan. Dalam praktik internasional, ujaran kebencian sering dipahami dalam spektrum: mulai dari ujaran yang "mengkhawatirkan" (soal

toleransi/sipilitas), hingga yang dapat dikenai sanksi administratif/perdata, dan yang mencapai ambang pidana. Panduan Council of Europe—yang mengacu pada pemikiran Rabat Plan of Action—menekankan pembedaan tiga tingkat tersebut serta pentingnya menilai konteks, penutur, niat, konten/bentuk, jangkauan, dan kemungkinan dampak. Di dalam kerangka bertingkat itu, beberapa pola dapat dibedakan, dengan catatan bahwa dalam praktiknya pola-pola ini dapat tumpang tindih pada satu data. Yang paling ringan adalah denigrasi dan stigmatisasi berbasis identitas—ujaran yang merendahkan martabat kelompok sasaran melalui pelabelan, stereotip negatif, atau penghinaan status sosialnya (misalnya "kelompok X itu bodoh/kotor/tidak pantas"). Berikutnya adalah eksklusi dan diskriminasi—ujaran yang membingkai kelompok sasaran sebagai "tidak berhak" atas sumber daya/ruang publik, atau mendorong pembatasan hak. Lebih jauh lagi ada dehumanisasi—ujaran yang menghilangkan kualitas kemanusiaan sasaran, misalnya lewat metafora hama, penyakit, atau "bukan manusia", yang dalam literatur pencegahan kekerasan dipahami sebagai mekanisme yang mempermudah audiens menerima perlakuan buruk atau

kekerasan terhadap sasaran (Dangerous Speech Project, 2021). Pada titik paling serius berdiri hasutan kebencian/permusuhan/kekerasan (incitement)—ujaran yang tidak hanya mengevaluasi negatif, tetapi mengarahkan audiens pada tindakan (langsung atau tersirat) terhadap kelompok sasaran, yang penilaiannya perlu menimbang variabel Rabat, terutama niat, konten/bentuk, jangkauan, dan kemungkinan dampak. Ada pula pola pembenaran dan normalisasi kebencian—ujaran yang "membenarkan" praktik diskriminasi/kekerasan dengan dalih sejarah, budaya, agama, atau pseudo-sains, misalnya mengutip "fakta" untuk menopang narasi opresif; panduan Council of Europe menempatkan pemeriksaan "fakta" dalam narasi semacam ini sebagai langkah penting dalam analisis.

Eksplisit dan implisit. Selain tingkat keparahan, ujaran kebencian juga dapat dibedakan berdasarkan sejauh mana kebencian "tampak di permukaan". Pada ujaran eksplisit, penamaan kelompok sasaran jelas, evaluasi negatif terang, dan/atau terdapat seruan tindakan yang mudah diidentifikasi. Pada ujaran implisit atau terkode (coded), kebencian disalurkan lewat insinuasi, presuposisi, eufemisme, humor, ironi, atau

Daftar Pustaka

African Union. (1981). *African Charter on Human and Peoples' Rights*.

Allchorn, W. (2024, 22 October). *A digital megaphone: The far right's use of podcasts for radicalisation*. Global Network on Extremism & Technology (GNET).

Amnesty International. (2022). *Myanmar: Facebook's systems promoted violence against Rohingya* (Press release).

Amnesty International. (2023). *A death sentence for my father: Meta's contribution to human rights abuses in northern Ethiopia* (Report & press materials).

Anderson, B. (1983/2006). *Imagined communities: Reflections on the origin and spread of nationalism* (Rev. ed.). Verso.

Anderson, L., & Barnes, M. (2022). *Hate Speech*. https://plato.stanford.edu/entries/hate-speech/?utm

Aranda Serna, F. J. (2024). *Memes as symbols of hate speech: The influence of graphic humour on freedom of expression and politics. VISUAL REVIEW:*

International Visual Culture Review. https://doi.org/10.62161/revvisual.v16.5222

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*(4), 555–596. doi:10.1162/coli.07-034-R2

Assagaf, A. S. A. (2023). Legal analysis of freedom of expression and online humour in Indonesia. *The European Journal of Humour Research, 11*(3), 105–122. https://doi.org/10.7592/EJHR.2023.11.3.807

Associated Press. (2025, January 29). *Over 100 Rohingya refugees fleeing persecution in Myanmar arrive in Indonesia's Aceh*. (R. Mirza, Author)

Austin, J. L. (1962). *How to do things with words*. Clarendon Press.

Bach, K. (1994). *Conversational Impliciture*. *9*(2), 124–162.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. A&C Black.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum

seekers in the UK press. *Discourse & Society*, *19*(3), 273–306. https://doi.org/10.1177/0957926508088962

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge University Press. https://doi.org/10.1017/CBO9780511920103

Beaver, D. I. (2001). *Presupposition and Assertion in Dynamic Semantics: A Critical Review of Linguistic Theories of Presupposition and a Dynamic Alternative*. Center for the Study of Language and Information. https://press.uchicago.edu/ucp/books/book/distributed/P/bo3639443.html

Benesch, S. (2014). *Countering Dangerous Speech: New Ideas for Genocide Prevention* (SSRN Scholarly Paper No. 3686876). Social Science Research Network. https://doi.org/10.2139/ssrn.3686876

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. *Cambridge University Press*.

Billig, M. (1995). *Banal nationalism*. SAGE.

Billig, M. (2005). *Laughter and ridicule: Towards a social critique of humour*. SAGE. https://doi.org/10.4135/9781446211779

Bousfield, D. (2008). Impoliteness in Interaction. Dalam *Pbns.167*. John Benjamins Publishing Company. https://benjamins.com/catalog/pbns.167

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

BSR. (2018). *Human rights impact assessment: Facebook in Myanmar* (Report).

Butler, J. (1997). *Excitable Speech: A Politics of the Performative*. Routledge.

Cabrera, A., et al. (2025). *A sarcasm and hate speech case study*. arXiv.

CAHSS Research Ethics Committee, & Edinburgh Centre for Data, Culture and Society. (n.d.). *Social media research: Ethical guidance for researchers*.

Calhoun, K., & Fawcett, A. (2023). “They Edited Out her Nip Nops”: Linguistic innovation as textual censorship avoidance on TikTok. *Language@Internet, 21*, 1–30. https://doi.org/10.14434/li.v21.37371

California Department of Justice. (2024). *Meta California AB 587 Terms of Service Report (Q3/Q4 2024)*.

Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy*, *54*(3), 330–349. https://doi.org/10.1111/phib.12022

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545–547.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Cohen, S., et al. (2023). Enhancing social network hate detection using back-translation and adversarial training (implicit hate includes sarcasm, humor, euphemism). *Information Fusion* (ScienceDirect record).

Committee on the Elimination of Racial Discrimination. (2013). *General recommendation No. 35 (2013) on combating racist hate speech (CERD/C/GC/35*).

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, *297*, 35–46. https://doi.org/10.1016/j.forsciint.2019.01.016

Coulthard, M., Alison Johnson, & David Wright. (2017). *An Introduction to Forensic Linguistics: Language in Evidence*. Routledge & CRC Press. https://www.routledge.com/An-Introduction-to-Forensic-Linguistics-Language-in-Evidence/Coulthard-Johnson-Wright/p/book/9781138641716

Coulthard, M., & Johnson, A. (2010). *The Routledge Handbook of Forensic Linguistics*. Routledge.

Court of Justice of the European Union. (2008). *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law*. EUR-Lex.

Council of Europe. (2001). Convention on Cybercrime (ETS No. 185).

Council of Europe. (2021). *European Convention on Human Rights* (as amended by Protocol No. 15, in force 1 August 2021).

Council of Europe. (2022). Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence.

Council of Europe. (2022). *Recommendation CM/Rec(2022)16 of the Committee of Ministers to member States on combating hate speech*. Council of Europe.

Council of Europe, Committee of Ministers. (2022). *Recommendation CM/Rec(2022)16 to member States on combating hate speech*.

Council of the European Union. (2008). *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law*.

Council of Europe. (t.t.). *Toolkit for analysing a case of hate speech: Advanced guide*.

Creswell, J. W. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124–130.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.)*. SAGE.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research (3rd ed.)*. SAGE.

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics*, *25*(3), 349–367. https://doi.org/10.1016/0378-2166(95)00014-3

Culpeper, J. (2011). *Impoliteness: Using Language to Cause Offence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975752

Dangerous Speech Project. (2021). *Dangerous Speech: A Practical Guide*. https://www.dangerousspeech.org/libraries/guide?utm

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). *Automated Hate Speech Detection and the Problem of Offensive Language* (arXiv:1703.04009). arXiv. https://doi.org/10.48550/arXiv.1703.04009

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186). Association for Computational Linguistics.

Dewan Pers. (2006). *Kode Etik Jurnalistik* (ditetapkan melalui SK Dewan Pers No. 03/SK-DP/III/2006; disahkan sebagai Peraturan Dewan Pers No. 6/Peraturan-DP/V/2008).

Dewan Pers. (2012). *Pedoman Pemberitaan Media Siber (Peraturan Dewan Pers Nomor 1/Peraturan-DP/III/2012)*. Dewan Pers.

Dijk, T. A. van. (2011). *Discourse and Ideology*. https://discourses.org/wp-content/uploads/2022/07/Teun-A.-van-Dijk-2011-Discourse-and-Ideology.pdf?utm.com

Dijk, T. A. van. (2015). *Critical discourse analysis. Dalam D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.)*.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. OUP Oxford.

Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., & Niemi, T. (2015). *Methodological Guidelines for*

Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition.

ENFSI. (2016). *ENFSI guideline for evaluative reporting in forensic science*. European Network of Forensic Science Institutes.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*.

European Commission. (2016). *EU Code of conduct on countering illegal hate speech online. European Commission*.

European Commission. (2022). *EU Code of Conduct on countering illegal hate speech online: Seventh evaluation* (Press release).

European Commission. (2023, October). *Request for information to Meta under the Digital Services Act* (Press corner item).

European Commission. (2025). *Code of conduct on countering illegal hate speech online+* (library entry).

European Commission against Racism and Intolerance (ECRI). (2015). *ECRI General Policy Recommendation No. 15 on combating hate speech*.

European Court of Human Rights. (1976). *Handyside v. the United Kingdom* (Application no. 5493/72).

European Union. (2008). *Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law. EUR-Lex*. https://eur-lex.europa.eu/EN/legal-content/summary/framework-decision-on-combating-certain-forms-and-expressions-of-racism-and-xenophobia-by-means-of-criminal-law.html

Fadhil, H. M., & Imtyas, R. (2024). *Hate Speech in Cyberspace: Fueling the Rejection of Rohingya Refugees in Indonesia*.

Fairclough, N. (2013). Critical discourse analysis: The critical study of language (2nd ed.). Routledge. https://doi.org/10.4324/9781315834368

Fairclough, N. (1989). *Language and Power*. Longman.

Fairclough, N. (1992). *Discourse and Social Change* (Reprinted). Polity Press.

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language* (2. ed., [Nachdr.]). Routledge.

Firdaus, M., & Rizki, P. N. M. (2024). *BIJAKAWEB: Kumpulan data komentar ofensif di portal berita daring Bahasa Indonesia. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*.

Ford, T. E., & Ferguson, M. A. (2004). Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review, 8*(1), 79–94.

Ford, T. E., Wentzel, E. R., & Lorion, J. (2008). Effects of exposure to sexist humor on perceptions of normative tolerance of sexism. *European Journal of Social Psychology, 38*(7), 1083–1095.

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys, 51*(4), Article 85.

Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge.

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0.*

Freedom House. (2024). *Indonesia: Freedom on the Net 2024 Country Report*.

Frenda, S., et al. (2018). The role of sarcasm in hate speech: A multilingual perspective. (IronITA/Evalita-related paper).

Gardiner, B., Mansfield, M., Anderson, I., Holder, J., Louter, D., & Ulmanu, M. (2016, April 12). *The dark side of Guardian comments. The Guardian*.

Germany. (n.d.). *German Criminal Code (StGB), Section 130: Incitement to hatred* (English translation). UNODC SHERLOC / Gesetze im Internet.

Gibbons, J. (2003). *Forensic Linguistics: An Introduction to Language in the Justice System*. Wiley.Com. https://www.wiley.com/en-us/Forensic+Linguistics%3A+An+Introduction+to+Language+in+the+Justice+System-p-9780631212478

Godioli, A., Young, J., & Fiori, B. M. (2022). Laughing matters: Humor, free speech and hate speech at the European Court of Human Rights. *International Journal for the Semiotics of Law, 35*(6), 2241–2265. https://doi.org/10.1007/s11196-022-09949-8

Greenpeace Indonesia. (2019, August 20). *Hentikan rasisme, diskriminasi dan kekerasan pada rakyat Papua*.

Grice, H. P. (Ed.). (1975). *Logic and conversation* (5. print). Academic Press.

Gries, S. Th. (2013). Statistics for linguistics with R: A practical introduction (2nd rev. ed.). De Gruyter Mouton.

Hakoköngäs, E., Halmesvaara, O., Sakki, I., & Pirttilä-Backman, A.-M. (2020). Persuasion through bitter humor: Multimodal discourse analysis of rhetoric in internet memes of two far-right groups in Finland. *Social Media + Society*.

Hakoköngäs, E., Halmesvaara, O., Sakki, I., & Suominen, A. H. (2020). *Persuasion through bitter humor: Multimodal discourse analysis of hateful internet memes*.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34.

Hangartner, D., Gennaro, G., Marie, A., et al. (2021). Empathy-based counterspeech can reduce racist hate speech. *Proceedings of the National Academy of Sciences*.

Hickey, D., Fessler, D. M. T., Lerman, K., & Burghardt, K. (2025). *X under Musk's leadership: Substantial hate and no reduction in inauthentic activity*. PLOS ONE, 20(2), e0313293. doi:10.1371/journal.pone.0313293

Hom, C. (2008). The Semantics of Racial Epithets. *The Journal of Philosophy*, *105*(8), 416–440.

Human Rights Committee. (1983). *General Comment No. 11: Prohibition of propaganda for war and inciting national, racial or religious hatred (Article 20*).

Human Rights Committee. (2011). *General Comment No. 34: Article 19—Freedoms of opinion and expression (CCPR/C/GC/34)*.

Husin, T. (2022, Mei 25). *Alat bukti dan barang bukti dalam KUHAP*. Pengadilan Tinggi Banda Aceh. https://www.pt-nad.go.id/new/content/artikel/20220525091259894497854628d90abe4c2e.html

Ibrahim, N., Qura, U., & Rahman, F. (2020). Speech act of Indonesian stand up comedian that potentially implicated to racist problem (Linguistic forensic analysis). *Humanus, 19*(2), 192–205.

https://doi.org/10.24036/humanus.v19i2.109638

Ibrohim, M. O., & Budi, I. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. Dalam S. T. Roberts, J. Tetreault, V. Prabhakaran, & Z. Waseem (Ed.), *Proceedings of the Third Workshop on Abusive Language Online* (hlm. 46–57). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3506

Ibrohim, M. O., & Budi, I. (2023). Hate speech and abusive language detection in Indonesian social media: Progress and challenges. *Heliyon*, *9*(8), e18647. https://doi.org/10.1016/j.heliyon.2023.e18647

Idris, I., Wijaya, D., Wijanarko, M. I., & Tim AJI Indonesia. (2024). *Monitoring report on hate speech against vulnerable groups in the 2024 election*. AJI Indonesia – Monash Data & Democracy Research Hub.

India. (1860). *The Indian Penal Code* (as published on India Code). India Code.

India, Ministry of Home Affairs. (2023). *The Bharatiya Nyaya Sanhita, 2023 (No. 45 of 2023)*.

International Association of Forensic Linguists. (2013). *IAFL Code of Practice*.

International IDEA. (2025). *When words become weapons: How hate speech threatens democracy*. https://www.idea.int/news/when-words-become-weapons-how-hate-speech-threatens-democracy

ISO/IEC 17025:2017. (t.t.). ISO. Diambil 2 Maret 2026, dari https://www.iso.org/standard/66912.html

Janks, H. (1997). Critical Discourse Analysis as a Research Tool. *Discourse: Studies in the Cultural Politics of Education*, *18*(3), 329–342. https://doi.org/10.1080/0159630970180302

Japan. (2016). *Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior against Persons Originating from Outside Japan* (English translation). Japanese Law Translation.

Jia, Y. (2025). Tackling hate speech online: The effect of counter-speech. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational Researcher, 33(7), 14–

26. https://doi.org/10.3102/0013189X033007014

Johnstone, B. (2018). *Discourse Analysis*. John Wiley & Sons.

Jowett, G. S., & O'Donnell, V. (2019). *Propaganda & persuasion* (7th ed.). SAGE.

Kapil, P., & Ekbal, A. (2024). A survey on combating hate speech through detection and prevention in English. In *Proceedings of ICON 2024*. Association for Computational Linguistics.

Karttunen, L. (1974). *Presupposition and Linguistic Context*. *1*(1–3), 181–194. https://doi.org/10.1515/thli.1974.1.1-3.181

Katharina, R. (2019, August). *Insiden asrama mahasiswa Papua di Surabaya* (Info Singkat, Vol. XI, No. 16/II/Puslit/Agustus/2019). Pusat Penelitian Badan Keahlian DPR RI.

Kepolisian Negara Republik Indonesia. (2015). Surat Edaran Kepala Kepolisian Negara Republik Indonesia Nomor SE/06/X/2015 tentang Penanganan Ujaran Kebencian (Hate Speech). *Tribrata News Polda Sulawesi Utara*.

https://tribratanews.sulut.polri.go.id/surat-edaran-kapolri-ujaran-kebencian-hate-speech/

KhosraviNik, M., & Esposito, E. (2018). Online hate, digital discourse and critique: Exploring digitally-mediated discursive practices of gender-based hostility. *Lodz Papers in Pragmatics*. https://doi.org/10.1515/lpp-2018-0003

Kiela, D., Firooz, H., Mohan, A., Goswami, V., Singh, A., Ringshia, P., & Testuggine, D. (2020). The hateful memes challenge: Detecting hate speech in multimodal memes. *NeurIPS 2020*.

Kiela, D., et al. (2021). The hateful memes challenge: Competition report. *Proceedings of Machine Learning Research*.

Knustad, T. (2021). *Anonymity, online commenting, and self-disclosure: The role of inhibition. Social Sciences, 10*(6).

Komisi Penyiaran Indonesia. (2012). *Pedoman Perilaku Penyiaran dan Standar Program Siaran (P3SPS)*.

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. European Journal of General

Practice, 24(1), 120–124. doi:10.1080/13814788.2017.1375092

Koto, F., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. In *Proceedings of COLING 2020* (pp. 757–770).

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research, 30*(3), 411–433.

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability* (Technical report).

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology (4th ed.)*. SAGE.

Kruk, J., et al. (2024). Silent signals, loud impact: LLMs for word-sense disambiguation of coded dog whistles. *Proceedings of ACL 2024*.

Lamb, K., & Doherty, B. (2019, August 22). *West Papua protests: Indonesia deploys 1,000 soldiers to quell unrest, cuts internet. The Guardian*.

Langton, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy and Public Affairs*, *22*(4), 293–330.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. https://direct.mit.edu/books/monograph/2629/Presumptive-MeaningsThe-Theory-of-Generalized

Litosseliti, L. (2018). *Research Methods in Linguistics*. Bloomsbury Academic. Bloomsbury. https://www.bloomsbury.com/uk/research-methods-in-linguistics-9781350043459/?utm

Locher, M. A., & Watts, R. J. (2005). *Politeness Theory and Relational Work*. *1*(1), 9–33. https://doi.org/10.1515/jplr.2005.1.1.9

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Communication Research, 28(4), 587–604.

Magu, R., & Luo, J. (2018). Determining Code Words in Euphemistic Hate Speech Using Word Embedding Networks. Dalam D. Fišer, R. Huang, V. Prabhakaran, R. Voigt, Z. Waseem, & J. Wernimont (Ed.), *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)* (hlm. 93–100). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-5112

Mahkamah Agung Republik Indonesia. (2021). *Keputusan bersama tentang pedoman implementasi atas pasal tertentu dalam UU tentang Informasi dan Transaksi Elektronik (UU ITE).* https://www.mahkamahagung.go.id/id/keputusan/4783/keputusan-bersama-tentang-pedoman-implementasi-atas-pasal-tertentu-dalam-uu-tentang-informasi-dan-transaksi-elektronik-uu-ite

Mahkamah Konstitusi Republik Indonesia. (2018a). *Putusan Mahkamah Konstitusi Republik Indonesia (memuat kutipan norma Pasal 156 dan Pasal 157 KUHP serta uraian relevan)*. Mahkamah Konstitusi RI. https://www.mkri.id/public/content/persidangan/putusan/76_PUU-XVI_2018.pdf

Mahkamah Konstitusi Republik Indonesia. (2018b). *Putusan Persidangan*. https://www.mkri.id/public/content/persidangan/putusan/76_PUU-XVI_2018.pdf

Mahkamah Konstitusi Republik Indonesia. (2025). *Putusan Nomor 105/PUU-XXII/2024*.

Mahsun. (2018). *Linguistik forensik: Memahami forensik berbasis teks dengan analogi DNA*. Rajagrafindo.

Malaysia. (1998). *Communications and Multimedia Act 1998, Section 233 (Improper use of network facilities or services)*. CommonLII / UNODC repository.

Malaysian Bar. (2015). *Section 233(1)(a) of the Communications and Multimedia Act 1998…* (commentary).

Marwick, A. E., & boyd, d. (2011). Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114–133.

Matamoros-Fernández, A., Bartolo, L., & Troynar, L. (2023). Humour as an online safety issue: Exploring solutions to help platforms better address this form of expression. *Policy & Internet / Internet Policy Review*. https://doi.org/10.14763/2023.1.1677

Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication, 58*(2), 258–279.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*.

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University

Press. https://doi.org/10.1017/CBO9780511981395

McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science, 21*(8), 1141–1149.

Mehta, H., et al. (2022). Social media hate speech detection with explainable AI. *Algorithms, 15*(8), 291.

Meta. (n.d.). *Hateful conduct* (Community Standards).

Meta. (2025a). *Journey to 1000 models: Scaling Instagram's recommendation system* (Meta Engineering).

Meta. (2025b). *Instagram Feed Recommendations: Explaining ranking* (Meta Transparency).

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). Qualitative Data Analysis: A Methods Sourcebook. Third Edition. Dalam *SAGE Publications Ltd (CA)*. SAGE Publications.

Musawir, M. (2025a). Honesty in Language in Public Figures' Narratives in Tempo Magazine: A Forensic Linguistic Study. *TOFEDU: The Future of Education Journal*, *4*(5), 1452–1463. https://doi.org/10.61445/tofedu.v4i5.639

Musawir, M. (2025b). *Kejujuran berbahasa dalam perspektif linguistik forensik*. Penerbit Adab.

Musawir, M., & Alifah, R. F. (2025). Semiotic Analysis in Forensic Linguistics: The Meaning of Symbols in The Tempo News Office Terrorism Case. *Lingue : Jurnal Bahasa, Budaya, Dan Sastra*, *7*(1), 129–143. https://doi.org/10.33477/lingue.v7i1.9473

Musawir, M., Amir, J., & Saleh, M. (2022). Honesty in Language in Special Fugitive Scandal: Forensic Linguistic Studies in Broadcasting Talks Indonesia Lawyers Club. *Journal of Asian Multicultural Research for Social Sciences Study*, *3*(1), 40–49. https://doi.org/10.47616/jamrsss.v3i1.256

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1).

Office for Democratic Institutions and Human Rights (Ed.). (2020). *Understanding the needs of hate crime victims*. OSCE Office for Democratic Institutions and Human Rights (ODIHR).

Office of the High Commissioner for Human Rights. (2023, November 4). *UN Human Rights Chief condemns rise in hatred* (Press release).

Office of the United Nations High Commissioner for Human Rights. (1966). *International Covenant on Civil and Political Rights (ICCPR) (ketentuan terkait advokasi kebencian dan hasutan)*. OHCHR. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

Office of the United Nations High Commissioner for Human Rights. (1965). International Convention on the Elimination of All Forms of Racial Discrimination.

Office of the United Nations High Commissioner for Human Rights. (2013). Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (A/HRC/22/17/Add.4).

Office of the United Nations High Commissioner for Human Rights. (2020). *One-pager on "incitement to hatred": The Rabat threshold test*. OHCHR. https://www.ohchr.org/en/documents/tools-and-resources/one-pager-incitement-hatred-rabat-threshold-test

Office of the United Nations High Commissioner for Human Rights. (t.t.). International Convention on the Elimination of All Forms of Racial Discrimination (CERD) (ketentuan mengenai pelarangan propaganda kebencian rasial). OHCHR. Diambil 2 Januari 2026, dari https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial

Olsson, J. (2008). *Forensic linguistics* (2nd ed.). Continuum.

Olsson, J., & Luchjenbroers, J. (2014). *Forensic Linguistics*. Bloomsbury Academic.

Organization of American States. (1969). *American Convention on Human Rights* (Article 13).

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.55

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). SAGE.

Petra, B., & Judit, B. (2020). *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*.

Polli, C. (2024). Multimodal computation or interpretation? Automatic vs. sociosemiotic challenges in detecting hateful racist memes. *Journal of Pragmatics* (ScienceDirect record).

Potts, C., & Potts, C. (2004). *The Logic of Conventional Implicatures*. Oxford University Press.

Rasinger, S. M. (2013). *Quantitative Research in Linguistics*. Bloomsbury Academic. Bloomsbury. https://www.bloomsbury.com/uk/quantitative-research-in-linguistics-9781441180100/

Reisigl, M., & Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. Routledge & CRC Press. https://www.routledge.com/Discourse-and-Discrimination-Rhetorics-of-Racism-and-Antisemitism/Reisigl-Wodak/p/book/9780415231503

Reisigl, M., & Wodak, R. (2009). *The Discourse-Historical Approach (DHA)*.

Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*.

Republik Indonesia. (2024). *Undang-Undang Republik Indonesia Nomor 1 Tahun 2024 tentang Perubahan*

Kedua atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik. https://jdih.komdigi.go.id/produk_hukum/view/id/884/t/undangundang+nomor+1+tahun+2024

Reuters. (2023, October 25). *US antisemitic incidents up about 400% since Israel-Hamas war began, report says*.

Reuters. (2025, January 20). *Facebook, X, YouTube to do more against online hate speech, EU says*.

Reuters. (2025, August 27). *Indonesia urges TikTok, Meta to act against harmful online content*.

Reyes, A., et al. (2022). Linguistic-based approach for recognizing implicit hate speech: Irony and sarcasm. *Computación y Sistemas* (SciELO record).

Rolando. (2019, August 20). *Komnas HAM didesak selidiki dugaan rasisme di asrama Papua Surabaya. detikNews*.

Rychlý, P. (2008). *A Lexicographer-Friendly Association Score*. 6–9.

Röttger, P., Vidgen, B., Nguyen, D., Waseem, Z., Margetts, H., & Pierrehumbert, J. B. (2021). HateCheck: Functional tests for hate speech detection models. *Proceedings of the 59th Annual Meeting of the*

Association for Computational Linguistics (ACL 2021).

Saha, K., Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects of Hateful Speech in Online College Communities. *Proceedings of the ... ACM Web Science Conference. ACM Web Science Conference*, *2019*, 255–264. https://doi.org/10.1145/3292522.3326032

Saul, J. (2018). Dogwhistles, Political Manipulation, and Philosophy of Language. Dalam D. Fogal, D. W. Harris, & M. Moss (Ed.), *New Work on Speech Acts* (hlm. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0013

Schmid, U. K. (2025). Humorous hate speech on social media: A mixed-methods investigation of users' perceptions and processing of hateful memes. New Media & Society, 27(3), 1588–1606. https://doi.org/10.1177/14614448231198169

Schmid, U. K., Schulze, H., & Drexel, A. (2024). Memes, humor, and the far right's strategic mainstreaming. Information, Communication & Society, 28(3), 1–20.

https://doi.org/10.1080/1369118X.2024.2329610

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Shmulewitz, D., Levitin, M. D., Skvirsky, V., Vider, M., Lev-Ran, S., & Mikulincer, M. (2025). Exposure to online hate speech is positively associated with post-traumatic stress disorder symptom severity. *Scientific Reports*, *15*(1), 29869. https://doi.org/10.1038/s41598-025-16168-1

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Shuy, R. (1996). *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom*. John Wiley & Sons.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Singapore. (1990). *Maintenance of Religious Harmony Act 1990*. Singapore Statutes Online.

Singapore. (1871/2020 ed.). *Penal Code 1871 (provisions including sections 298/298A)*. Singapore Statutes Online.

Singapore. (2014). *Protection from Harassment Act 2014*. Singapore Statutes Online.

Spencer-Oatey, H. (2008). *Politeness and Rapport Management*. 11–47.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition, 2nd ed* (hlm. viii, 326). Blackwell Publishing.

Stalnaker, R. (1974). *Pragmatic presuppositions*.

Stalnaker, R. C. (1999). Pragmatic Presuppositions. Dalam R. C. Stalnaker (Ed.), *Context and Content: Essays on Intentionality in Speech and Thought* (hlm. 0). Oxford University Press. https://doi.org/10.1093/0198237073.003.0003

Steen, E., Yurechko, K., & Klug, D. (2023). You Can (Not) Say What You Want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media + Society*.

Suler, J. (2004). *The online disinhibition effect. CyberPsychology & Behavior, 7*(3), 321–326.

Swiss Confederation (Federal Department of Foreign Affairs). (2025). *The Rabat Plan of Action in the Digital Age* (Side event brief, 19 September 2025).

Tannen, D., Hamilton, H. E., & Schiffrin, D. (2015). *The handbook of discourse analysis* (2nd edition). Wiley Blackwell.

The Christchurch Call. (2019–2021). *Christchurch Call to Action / community consultation materials* (Public documents).

Tiersma, P. M. (1999). *Legal Language*. University of Chicago Press.

Treem, J. W., & Leonardi, P. M. (2012). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association.

Törnberg, A., & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media*, *13*, 132–142. https://doi.org/10.1016/j.dcm.2016.04.003

Undang-Undang Nomor 1 Tahun 2024. (2024). https://jdih.komdigi.go.id/index.php/produk_hukum/view/id/884/t/undangundang+nomor+1+tahun+2024

UNESCO. (2025). *Silenced voices: Why freedom of expression is receding worldwide*. https://www.unesco.org/en/articles/silenced-

voices-why-freedom-expression-receding-worldwide

United Nations. (t.t.). *What is hate speech?* United Nations. United Nations. Diambil 2 Januari 2026, dari https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech?utm

United Nations. (1948). *Convention on the Prevention and Punishment of the Crime of Genocide*.

United Nations. (2019). *United Nations Strategy and Plan of Action on Hate Speech*.

United Nations. (2019). *UN Strategy and Plan of Action on Hate Speech (termasuk penjelasan bahwa tidak ada definisi legal internasional tunggal tentang hate speech).* https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf?utm.com

United Nations Educational, Scientific and Cultural Organization. (2023). *Addressing Hate Speech Through Education: A Guide for Policy-makers*. United Nations.

United Nations Office at Geneva. (2024, Juni 14). *Genocide prevention adviser warns of global threat posed by hate speech | The United Nations Office at Geneva*. https://www.ungeneva.org/en/news-media/news/2024/06/94472/genocide-prevention-adviser-warns-global-threat-posed-hate-speech

United States. (n.d.). *18 U.S.C. § 249: Hate crime acts*. Legal Information Institute, Cornell Law School.

United States Department of Justice. (2025). *Hate crimes: Laws and policies*.

U.S. Supreme Court. (1969). *Brandenburg v. Ohio*, 395 U.S. 444. Oyez / Legal Information Institute.

U.S. Supreme Court. (1992). *R.A.V. v. City of St. Paul*, 505 U.S. 377. Legal Information Institute.

U.S. Supreme Court. (1993). *Wisconsin v. Mitchell*, 508 U.S. 47. Legal Information Institute.

U.S. Supreme Court. (2003). *Virginia v. Black*, 538 U.S. 343. Legal Information Institute.

Van de Kerkhof, J. (2025). Article 22 Digital Services Act: Building trust with trusted flaggers. *Internet Policy Review*.

Van Dijk, T. A. (1988). *News as discourse*. Lawrence Erlbaum.

Van Dijk, T. A. (1991). *Racism and the press*. Routledge.

Vidgen, B., & Derczynski, L. (2020). Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLOS ONE, 15(12), e0243300. https://doi.org/10.1371/journal.pone.0243300

Wachs, S., Gámez-Guadix, M., & Wright, M. F. (2022). Online Hate Speech Victimization and Depressive Symptoms Among Adolescents: The Protective Role of Resilience. *Cyberpsychology, Behavior, and Social Networking*, *25*(7), 416–423. https://doi.org/10.1089/cyber.2022.0009

Wachs, S., et al. (2024). From hate speech to HateLess: The effectiveness of a prevention program. *Computers in Human Behavior*.

Waldron, J. (2012). *The Harm in Hate Speech*. Harvard University Press. https://www.jstor.org/stable/j.ctt2jbrjd

WALHI Jawa Timur. (2019, August 20). *Aliansi Masyarakat Sipil Jatim, Mengecam Tindak Kekerasan Kepada Mahasiswa Papua*.

Waseem, Z. (2016). Are you a racist or am I seeing things? Annotator influence on hate speech detection on Twitter. *Proceedings of the First Workshop on NLP and Computational Social Science*, 138–142.

Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 88–93.

West Point CTC Sentinel. (2019). The Christchurch attacks: Livestream terror in the viral video age.

Wiberg, A., & Adar, E. (2025). *Speaking in code: Unmasking implicit hate speech with ALBERT* (Bachelor's thesis).

Wijanarko, M. I., Susanto, L., Pratama, P. A., Idris, I., Hong, T., & Wijaya, D. (2024). Monitoring hate speech in Indonesia: An NLP-based classification of social media texts. In *Proceedings of EMNLP 2024: System Demonstrations* (pp. 142–152). Association for Computational Linguistics.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., et al. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv*.

Wodak, R., & Meyer, M. (2001). *Methods of Critical Discourse Analysis*. SAGE Publications, Ltd. https://doi.org/10.4135/9780857028020

Wodak, R., & Meyer, M. (Eds.). (2009). *Methods for critical discourse analysis (2nd ed.)*. SAGE.

Wodak, R. (2015). *The politics of fear: What right-wing populist discourses mean*. SAGE.

X. (n.d.a). *X Rules* (Safety and policies).

X. (n.d.b). *Hateful conduct policy* (Rules and policies).

X. (n.d.c). *About different types of Posts* (Help Center).

Yoanita, D., Hadi, I. P., Setiawan, A., Aritonang, A. I., & Maureen, V. (2019). *Spirit of writing and sharing as motives of Kompasiana users. Mattingley Publishing Co., Inc*.

Yogyakarta Principles. (2017). *The Yogyakarta Principles plus 10*.

Zhou, L., et al. (2025). Stylometric characteristics of code-switched offensive language on social media. *Language Sciences* (ScienceDirect record).

ANATOMI
UJARAN KEBENCIAN

RUANG KARYA
Jl. Martapura Lama km. 07 Kec. Sungai Tabuk. Kel. Sungai Lulut. Kab. Banjar. Kalimantan Selatan. Komplek Karya Budi Utama Raya 2. Blok A No. 17.
Instagram: @ruangkar_ya
Whatsapp: 08971169692