

Pemetaan Risiko Etika pada Pipeline Embedding Multimodal untuk Representasi Semantik Big Data Smart City: Sintesis Literatur dan Studi Kasus Konseptual

^{*1,2}Endang Sri Sumanti, ¹Didik Dwi Prasetya, ¹Hakkun Elmunsyah, ¹Siti Sendari

¹Departemen Teknik Elektro dan Informatika, Fakultas Teknik, Universitas Negeri Malang, Malang, Indonesia

²Universitas Mbojo Bima. Jl. Pierre Tandean, Kel. Mande, Kec. Mpunda Kota Bima, Nusa Tenggara Barat

*Correspondence e-mail: endangsrsumiyanti@gmail.com

Diterima: Desember 2025; Revisi: Juni 2025; Diterbitkan: Juni 2026

Abstrak

Penelitian ini bertujuan menganalisis manfaat representasi semantik dan memetakan risiko etika pada *pipeline embedding* multimodal dalam pengolahan Big Data Smart City. Penelitian menggunakan desain sintesis literatur yang dipadukan dengan studi kasus konseptual, analisis hermeneutik teknologi, dan pemetaan risiko berbasis *pipeline*. Analisis mencakup lima komponen utama, yaitu *Text Encoder*, *Visual Encoder*, *Cross-Modal Alignment*, *Fusion Layer*, dan *Semantic Output Layer*. Hasil penelitian menunjukkan bahwa integrasi data tekstual dan visual dapat memperkaya konteks semantik, memperkuat hubungan informasi antarmodalitas, serta mempertahankan konsistensi representasi pada dokumen kebijakan yang kompleks. Namun, manfaat tersebut disertai lima risiko utama, yaitu bias representasional, privasi dan *indirect disclosure*, *dual-use*, *surveillance* dan *profiling* otomatis, serta asimetri kekuasaan informasi. *Fusion Layer* teridentifikasi sebagai komponen dengan risiko paling tinggi karena menggabungkan bias dan potensi penyalahgunaan dari beberapa modalitas, sedangkan *Cross-Modal Alignment* menunjukkan mekanisme mitigasi yang masih terbatas. Integrasi prinsip *autonomy*, *beneficence*, *justice*, dan *consent* dengan *fairness*, *accountability*, *transparency*, *explainability*, serta manajemen risiko menghasilkan kerangka evaluasi etika yang dapat diterapkan sepanjang siklus hidup sistem. Penelitian merekomendasikan penerapan *ethical checkpoints*, audit bias, dokumentasi dataset dan model, *logging*, pembatasan tujuan penggunaan, serta *human oversight*. Karena berbasis studi kasus konseptual, kerangka ini masih memerlukan validasi ahli dan pengujian empiris menggunakan dataset Smart City aktual.

Kata kunci: AI governance, Big Data Smart City, embedding multimodal, etika kecerdasan buatan, risiko etika.

Ethical Risk Mapping in Multimodal Embedding Pipelines for Semantic Representation of Smart City Big Data: A Literature Synthesis and Conceptual Case Study

Abstract

This study aims to analyze the benefits of semantic representation and map ethical risks within multimodal embedding pipelines used to process Smart City Big Data. The study employed a literature synthesis design combined with a conceptual case study, technological hermeneutic analysis, and pipeline-based risk mapping. The analysis covered five main components: the Text Encoder, Visual Encoder, Cross-Modal Alignment, Fusion Layer, and Semantic Output Layer. The findings indicate that integrating textual and visual data can enrich semantic context, strengthen cross-modal information relationships, and maintain representational consistency in complex policy documents. However, these benefits are accompanied by five major risks: representational bias, privacy and indirect disclosure, dual use, automated surveillance and profiling, and information-power asymmetry. The Fusion Layer was identified as the component with the highest risk because it combines bias and the potential misuse of information from multiple modalities, whereas Cross-Modal Alignment still has limited mitigation mechanisms. Integrating the principles of autonomy, beneficence, justice, and consent with fairness, accountability, transparency, explainability, and risk management produced an ethical evaluation framework that can be applied throughout the system lifecycle. The study recommends implementing ethical checkpoints, bias audits, dataset and model documentation, logging, purpose limitation, and human oversight. Because the study is based on a conceptual case scenario, the proposed framework still requires expert validation and empirical testing using actual Smart City datasets.

Keywords: AI governance, artificial intelligence ethics, ethical risks, multimodal embedding, Smart City Big Data.

How to Cite: Sumanti, E. S., Prasetya, D. D. ., Elmunsyah, H. ., & Sendari, S. . (2026). Integrasi Embedding Multiformat untuk Representasi Semantik Big Data Smart City: Analisis Etika Penelitian, AI Ethics, dan

Tantangan Publikasi Ilmiah di Era Teknologi Lanjut. *Reflection Journal*, 6(1), 737-750. <https://doi.org/10.36312/rj.v6i2.4193>.



<https://doi.org/10.36312/rj.v6i2.4193>.

Copyright© 2026, Sumanti et al

This is an open-access article under the [CC-BY-SA License](#).



PENDAHULUAN

Perkembangan Big Data dan kecerdasan buatan telah mengubah cara pemerintah kota mengumpulkan, mengolah, menghubungkan, dan menafsirkan informasi untuk mendukung pelayanan publik. Dalam ekosistem Smart City, data tidak lagi tersedia hanya dalam bentuk dokumen tertulis, tetapi juga berupa infografis, gambar, rekaman video, data sensor Internet of Things, peta digital, dan informasi spasial. Keragaman tersebut menyediakan peluang untuk memperoleh pemahaman yang lebih komprehensif mengenai kondisi perkotaan. Namun, karakter data yang heterogen juga menimbulkan persoalan teknis dan metodologis, terutama dalam mempertahankan keterhubungan makna di antara berbagai jenis data. Tantangan penelitian dalam konteks ini bukan sekadar meningkatkan kapasitas komputasi, melainkan membangun representasi semantik yang mampu mengintegrasikan informasi lintas modalitas secara konsisten. Kajian mengenai pembelajaran representasi multimodal menunjukkan bahwa penggunaan informasi komplementer dari beberapa modalitas dapat mendukung proses pencarian informasi, penalaran, klasifikasi, dan pemahaman lintas-modal secara lebih baik dibandingkan penggunaan satu jenis data saja (Manzoor et al., 2023).

Salah satu pendekatan yang dapat digunakan untuk mengintegrasikan data tersebut adalah embedding multimodal. Pendekatan ini memetakan informasi berbasis teks dan visual ke dalam ruang representasi yang dapat dibandingkan, disejajarkan, dan digabungkan. Dalam penerapan Smart City, embedding multimodal dapat digunakan untuk menghubungkan narasi kebijakan dengan grafik indikator, peta wilayah, infografis kependudukan, atau rekaman visual kondisi perkotaan. Integrasi tersebut memungkinkan sistem memahami hubungan antara deskripsi tekstual dan struktur visual yang merepresentasikan fenomena yang sama. Meskipun demikian, peningkatan kemampuan representasi tidak selalu menghasilkan sistem yang lebih adil atau dapat dipertanggungjawabkan. Bias yang terdapat pada satu modalitas dapat berinteraksi dengan bias pada modalitas lain, sehingga proses penggabungan data justru dapat memperkuat distorsi yang telah terdapat dalam sumber data. Penelitian tentang model vision-language memperlihatkan bahwa bias dapat terbentuk secara intra-modal maupun lintas-modal dan memengaruhi cara sistem merepresentasikan kelompok atau kategori tertentu (Srinivasan & Bisk, 2022; Janghorbani & De Melo, 2023).

Persoalan etika menjadi semakin penting ketika embedding multimodal digunakan untuk mendukung keputusan publik. Sistem Smart City sering bergantung pada pengumpulan data yang berkelanjutan melalui kamera pengawas, sensor lingkungan, perangkat bergerak, aplikasi layanan publik, dan sumber informasi digital lainnya. Praktik tersebut dapat memengaruhi privasi, otonomi, kontrol atas data, distribusi manfaat, serta hubungan kekuasaan antara warga, pemerintah, dan penyedia teknologi. Kajian mengenai implikasi etis Smart City menunjukkan bahwa persoalan tersebut mencakup pengelolaan infrastruktur digital, tata kelola data, inklusi sosial, keberlanjutan, dan akuntabilitas institusional (Ziosi et al., 2024). Gstrein (2024) juga menekankan bahwa datafikasi ruang perkotaan tidak cukup dipahami sebagai proses teknis. Penggunaan data publik perlu dianalisis dari perspektif martabat manusia, kontrol individu dan komunitas, serta kemungkinan terjadinya pelanggaran batas antara kepentingan publik dan hak personal.

Kerangka etika penelitian klasik tetap dapat digunakan sebagai dasar untuk menilai persoalan tersebut. Prinsip Nuremberg menekankan pentingnya persetujuan sukarela, sedangkan Deklarasi Helsinki memberi perhatian pada perlindungan subjek, pengurangan risiko, dan prioritas kepentingan manusia. Prinsip Belmont selanjutnya mengembangkan tiga landasan utama, yaitu respect for persons, beneficence, dan justice. Ketiga kerangka tersebut menyediakan dasar normatif untuk menilai proses pengumpulan dan penggunaan data. Namun, karakter sistem AI menimbulkan risiko baru yang tidak sepenuhnya dapat dijelaskan melalui prinsip klasik. Sistem berbasis AI dapat beroperasi secara otomatis, menggunakan data dalam skala besar, menghasilkan inferensi yang tidak langsung terlihat, serta memengaruhi kelompok yang tidak terlibat secara eksplisit dalam proses pengumpulan data. Oleh sebab itu, etika penelitian klasik perlu dilengkapi dengan transparency, accountability, explainability, fairness, dan manajemen risiko sepanjang siklus hidup sistem. Prinsip etika juga perlu diterjemahkan menjadi prosedur dan instrumen yang dapat diterapkan, bukan hanya dicantumkan sebagai pernyataan normatif (Ortega-Bolaños et al., 2024; Tabassi, 2023).

Dalam diskursus AI ethics, kerangka fairness, accountability, transparency, dan explainability atau FATE sering digunakan untuk mengevaluasi sistem kecerdasan buatan. Fairness berkaitan dengan kemampuan sistem menghindari perlakuan yang tidak adil terhadap kelompok tertentu. Accountability menuntut adanya pihak yang bertanggung jawab atas proses pengembangan, penggunaan, dan dampak sistem. Transparency mengharuskan proses pengolahan data serta tujuan penggunaan model dapat diketahui, sedangkan explainability berkaitan dengan kemampuan untuk menjelaskan dasar suatu output atau keputusan. OECD (2023) juga menempatkan fairness, privasi, transparansi, explainability, robustness, dan accountability sebagai unsur penting dalam pengembangan AI yang dapat dipercaya. Berdasarkan perspektif tersebut, embedding multimodal tidak dapat dipahami hanya sebagai persoalan arsitektur komputasi. Teknologi ini merupakan sistem sosio-teknis yang menghubungkan pilihan data, desain model, kepentingan institusi, proses pengambilan keputusan, dan dampak terhadap masyarakat.

Titik kritis dalam pipeline embedding multimodal terutama muncul pada tahap cross-modal alignment dan fusion. Cross-modal alignment berfungsi menyelaraskan representasi teks dan visual agar informasi dari kedua modalitas dapat ditempatkan dalam ruang semantik yang sebanding. Selanjutnya, fusion layer menggabungkan representasi tersebut untuk menghasilkan informasi terpadu. Pada kedua tahap ini, informasi dari satu modalitas tidak hanya ditambahkan pada modalitas lain, tetapi dapat saling memperkuat. Mekanisme yang sama dapat memberikan manfaat teknis sekaligus memperbesar bias, kesalahan pelabelan, atau ketidaksesuaian konteks. Literatur mengenai bias multimodal menunjukkan bahwa kombinasi teks dan visual dapat memperkuat stereotip yang sebelumnya hanya terdapat pada salah satu modalitas (Srinivasan & Bisk, 2022; Janghorbani & De Melo, 2023). Kondisi tersebut menunjukkan bahwa audit etika tidak seharusnya dilakukan hanya setelah output dihasilkan. Evaluasi perlu ditempatkan pada setiap tahapan pipeline, mulai dari pemilihan data, proses encoding, alignment, fusion, hingga penggunaan hasil representasi.

Sebagian besar penelitian terdahulu masih berfokus pada peningkatan performa model, pretraining, retrieval, alignment, dan mekanisme fusion. Sementara itu, pembahasan etika pada implementasi Smart City tersebar dalam kajian privasi, urban computing, kebijakan publik, data governance, dan responsible AI. Pemisahan antara evaluasi teknis dan evaluasi etika menimbulkan kesenjangan karena risiko sosial dapat terbentuk sejak tahap desain data dan model. Kajian Smart City menunjukkan bahwa dampak etis tidak terpisah dari cara infrastruktur digital dibangun, dikelola, dan digunakan dalam proses kebijakan (Ziosi et al., 2024). Oleh karena itu, diperlukan

pendekatan yang mampu menghubungkan komponen teknis embedding multimodal dengan prinsip etika penelitian dan AI governance secara sistematis.

Kontribusi penelitian meliputi. Pertama, penelitian memetakan risiko etika pada lima komponen pipeline, yaitu Text Encoder, Visual Encoder, Cross-Modal Alignment, Fusion Layer, dan Semantic Output Layer. Kedua, analisis teknis dipadukan dengan pendekatan hermeneutik teknologi untuk menelaah bagaimana representasi semantik dapat membentuk cara pemerintah atau sistem memandang kondisi kota, wilayah, dan warganya. Ketiga, hasil analisis diterjemahkan ke dalam matriks risiko dan mitigasi yang dapat digunakan sebagai dasar evaluasi sebelum sistem diterapkan atau dipublikasikan. Pendekatan ini tidak hanya menilai performa model, tetapi juga menghubungkan risiko dengan prinsip autonomy, beneficence, justice, consent, fairness, accountability, transparency, dan explainability.

Tantangan publikasi ilmiah ditempatkan sebagai bagian dari tata kelola penelitian. Dokumentasi dataset, penjelasan konteks penggunaan, pelaporan keterbatasan, transparansi model, serta pengungkapan risiko merupakan unsur penting dalam responsible AI. Model cards, audit trail, dan dokumentasi proses dapat membantu peneliti, reviewer, pemerintah, serta pengguna memahami ruang lingkup dan batas penggunaan sistem (Mitchell et al., 2019). Dengan demikian, kepatuhan etika tidak seharusnya diperlakukan sebagai persyaratan administratif yang terpisah dari kualitas penelitian. Evaluasi etika perlu menjadi bagian dari metode, analisis, dokumentasi, dan komunikasi hasil ilmiah.

Berdasarkan kesenjangan tersebut, penelitian ini menjawab enam pertanyaan: (1) bagaimana embedding multimodal memperkaya representasi semantik Big Data Smart City; (2) apa saja risiko etika utama dalam integrasi teks dan infografis; (3) bagaimana prinsip Nuremberg, Helsinki, dan Belmont dapat diterapkan pada konteks AI; (4) apa tantangan metodologis pada tahap cross-modal alignment dan fusion; (5) bagaimana AI ethics memengaruhi standar pelaporan dan publikasi ilmiah; dan (6) bentuk governance apa yang relevan bagi penelitian embedding multimodal. Keenam pertanyaan tersebut digunakan untuk menyusun sintesis teknis, etis, dan metodologis. Penelitian ini menempatkan embedding multimodal sebagai teknologi sosio-teknis yang manfaatnya harus dinilai bersamaan dengan risiko privasi, bias, profiling, pengawasan, dan akuntabilitas. Posisi tersebut sejalan dengan pendekatan risk-based AI governance yang menekankan identifikasi, pengukuran, mitigasi, pemantauan, dan dokumentasi risiko sepanjang siklus hidup sistem (OECD, 2023; Tabassi, 2023).

METODE

Desain Penelitian

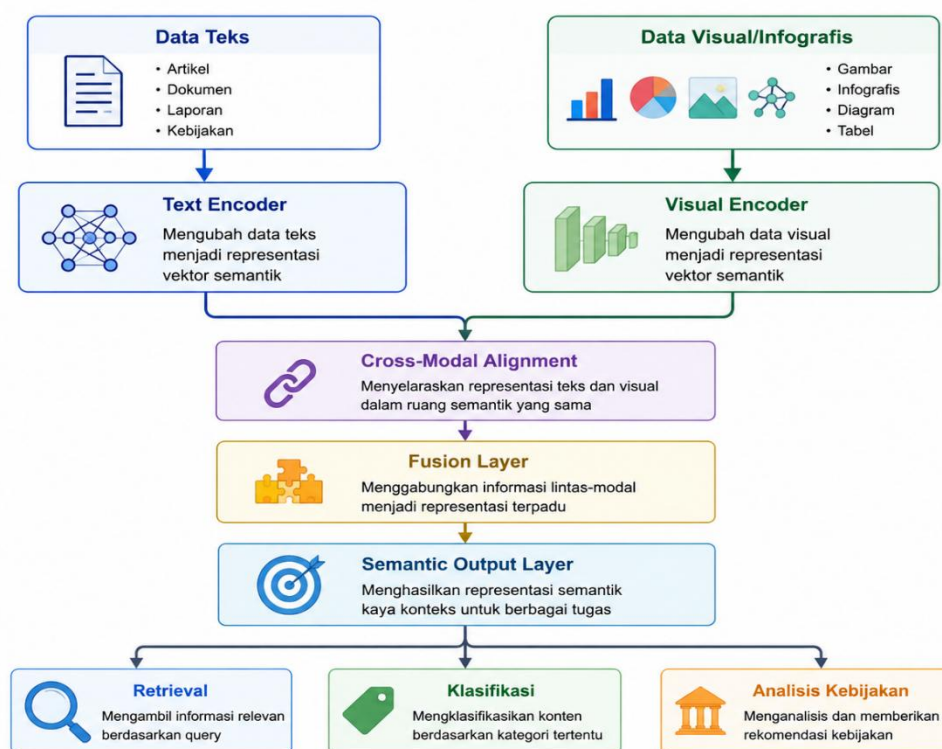
Penelitian ini menggunakan desain sintesis literatur yang dipadukan dengan studi kasus konseptual dan analisis hermeneutik teknologi. Sintesis literatur digunakan untuk mengintegrasikan temuan mengenai representasi multimodal, bias algoritmik, etika Smart City, dan tata kelola kecerdasan buatan. Studi ini tidak dikategorikan sebagai *systematic review* karena proses penelusuran basis data, penyaringan artikel, dan seleksi sumber pada naskah awal tidak didokumentasikan secara lengkap mengikuti protokol PRISMA. Oleh karena itu, sintesis diarahkan untuk membangun pemetaan konseptual, mengidentifikasi hubungan antartema, dan menyusun kerangka evaluasi risiko, bukan untuk menghitung ukuran efek atau menghasilkan generalisasi statistik.

Sumber yang digunakan dikelompokkan ke dalam tiga tema utama, yaitu: (1) representasi multimodal, *embedding*, dan *cross-modal alignment*; (2) bias, keadilan, privasi, serta etika Smart City; dan (3) *AI governance*, akuntabilitas, transparansi, serta manajemen risiko. Sumber yang tidak memiliki informasi bibliografis memadai atau tidak dapat diverifikasi tidak digunakan sebagai dasar utama analisis. Proses sintesis

dilakukan melalui *thematic coding*, *critical synthesis*, dan *methodological mapping*. *Thematic coding* digunakan untuk mengelompokkan isu teknis dan etika, sedangkan *critical synthesis* digunakan untuk membandingkan manfaat teknologi dengan risiko sosial yang mungkin muncul. Selanjutnya, *methodological mapping* menghubungkan setiap komponen pipeline dengan prinsip etika dan bentuk mitigasi yang relevan.

Arsitektur Konseptual Embedding Multimodal

Unit analisis penelitian terdiri atas lima komponen pipeline, yaitu *Text Encoder*, *Visual Encoder*, *Cross-Modal Alignment*, *Fusion Layer*, dan *Semantic Output Layer*. *Text Encoder* berfungsi mengubah narasi kebijakan atau laporan tekstual menjadi representasi semantik. *Visual Encoder* mengekstraksi elemen visual, simbol, grafik, warna, label, dan struktur informasi dari infografis. Selanjutnya, *Cross-Modal Alignment* menyelaraskan representasi teks dan visual melalui mekanisme proyeksi atau *contrastive learning*. Representasi yang telah disejajarkan kemudian digabungkan pada *Fusion Layer* melalui *concatenation*, *attention*, atau *adaptive gating*. Hasil penggabungan diproses pada *Semantic Output Layer* untuk mendukung pencarian informasi, klasifikasi, rekomendasi, dan analisis kebijakan.



Gambar 1. Arsitektur konseptual integrasi embedding multimodal untuk pengolahan Big Data Smart City

Evaluasi etika tidak hanya dilakukan pada hasil akhir, tetapi ditempatkan pada setiap tahap pipeline. Pada *Text Encoder*, evaluasi diarahkan pada bias naratif dan ketimpangan representasi bahasa. Pada *Visual Encoder*, perhatian diberikan pada pilihan label, warna, simbol, dan distribusi data. Tahap *Cross-Modal Alignment* diperiksa terhadap kemungkinan penguatan bias antarmodalitas dan ketidaksesuaian konteks. *Fusion Layer* dianalisis karena berpotensi menggabungkan bias dari beberapa sumber sekaligus, sedangkan *Semantic Output Layer* dievaluasi dari aspek transparansi, *explainability*, dan keterlacakan keputusan.

Kerangka Evaluasi Etika

Evaluasi etika dilakukan melalui dua lapisan. Lapisan pertama menggunakan prinsip etika klasik yang mencakup *autonomy*, *beneficence*, *justice*, dan *voluntary consent*. Prinsip tersebut digunakan untuk menilai kontrol atas data, perlindungan terhadap risiko, keadilan distribusi manfaat, serta kejelasan persetujuan penggunaan data. Penerapan prinsip tersebut dirangkum dalam Tabel 1.

Tabel 1. Pemetaan prinsip etika dalam konteks Big Data Smart City

Prinsip etika	Relevansi terhadap embedding multimodal	Contoh implementasi
<i>Autonomy</i>	Menjamin kontrol pengguna, transparansi tujuan, dan hak menolak penggunaan data	<i>Consent notice</i> , pembatasan tujuan, serta mekanisme <i>opt-in</i> atau <i>opt-out</i>
<i>Beneficence</i>	Menilai manfaat dan risiko sebelum model diterapkan	<i>Impact assessment</i> dan audit bias secara berkala
<i>Justice</i>	Menjamin keadilan representasi dan distribusi manfaat maupun risiko	Evaluasi subkelompok dan pemeriksaan ketidakseimbangan data
<i>Voluntary consent</i>	Memastikan persetujuan penggunaan data diberikan secara jelas dan dapat dipahami	Pemberitahuan penggunaan data dan mekanisme persetujuan yang sesuai dengan konteks

Lapisan kedua menggunakan *data ethics*, *algorithmic justice*, *AI governance*, dan FATE yang mencakup *fairness*, *accountability*, *transparency*, dan *explainability*. Kerangka ini digunakan untuk menilai privasi, minimalisasi data, diskriminasi algoritmik, keterlacakan proses, reproduktibilitas, serta tanggung jawab institusional.

Studi Kasus Konseptual dan Validasi

Studi kasus berbentuk skenario sistem prediksi kepadatan lalu lintas yang menggabungkan laporan teks, visual CCTV, dan infografis kepadatan. Skenario tersebut tidak merepresentasikan audit terhadap kota tertentu, tetapi digunakan untuk menguji konsistensi kerangka dalam mengidentifikasi pelanggaran etika, dampak sosial, aktor yang terlibat, dan bentuk mitigasi yang diperlukan.

Validasi kerangka dilakukan melalui *Logical Consistency Check* dan *Scenario-Based Ethical Testing*. Pemeriksaan konsistensi digunakan untuk memastikan bahwa setiap risiko memiliki hubungan logis dengan prinsip etika dan tindakan mitigasinya. Pengujian berbasis skenario digunakan untuk menilai apakah kerangka tetap relevan ketika beberapa jenis data digunakan secara bersamaan. Karena penelitian tidak menggunakan data primer, hasilnya dibatasi pada pemetaan konseptual dan masih memerlukan validasi ahli serta pengujian empiris pada konteks Smart City yang nyata.

HASIL DAN PEMBAHASAN

Representasi Semantik melalui Embedding Multimodal

Hasil sintesis menunjukkan bahwa penggabungan informasi tekstual dan visual berpotensi menghasilkan representasi semantik yang lebih kaya dibandingkan pemrosesan data dengan satu modalitas. Dalam konteks Big Data Smart City, teks umumnya memuat uraian kebijakan, laporan kejadian, tujuan program, serta penjelasan indikator, sedangkan infografis menyajikan distribusi spasial, kecenderungan waktu, perbandingan data, dan hubungan antarkomponen. Ketika kedua jenis informasi tersebut

dipetakan ke dalam ruang representasi yang saling terhubung, sistem dapat menafsirkan suatu istilah dengan mempertimbangkan konteks visual yang menyertainya. Temuan ini sejalan dengan kajian pembelajaran representasi multimodal yang menempatkan informasi komplementer sebagai dasar untuk meningkatkan kemampuan *retrieval*, klasifikasi, penalaran, dan pemahaman lintas modalitas (Manzoor et al., 2023).

Meskipun demikian, hasil tersebut tidak dapat ditafsirkan sebagai bukti bahwa embedding multimodal selalu lebih akurat daripada embedding unimodal. Penelitian ini tidak menggunakan dataset primer dan tidak melakukan eksperimen komputasional dengan metrik seperti *precision*, *recall*, *F1-score*, *cosine similarity*, atau *retrieval accuracy*. Oleh karena itu, istilah “lebih kaya” merujuk pada kelengkapan konteks dan keterhubungan informasi, bukan pada peningkatan kinerja numerik. Pembatasan ini penting karena literatur mengenai AI yang dapat dipercaya menekankan bahwa kualitas sistem tidak cukup ditentukan oleh kemampuan teknis, tetapi juga oleh keandalan, keamanan, transparansi, dan kesesuaian penggunaannya (Thiebes et al., 2021).

Temuan kedua memperlihatkan adanya *cross-modal reinforcement*, yaitu kondisi ketika informasi pada satu modalitas memperjelas informasi pada modalitas lain. Sebagai contoh, laporan tekstual mengenai peningkatan kepadatan lalu lintas dapat dipahami secara lebih spesifik ketika dipadukan dengan peta wilayah, grafik waktu, dan visual kondisi jalan. Sebaliknya, infografis kepadatan dapat kehilangan konteks apabila tidak dilengkapi penjelasan mengenai penyebab, durasi, dan dampak kejadian. Interaksi tersebut memungkinkan sistem membentuk hubungan semantik antara istilah, lokasi, indikator, dan struktur visual.

Namun, penguatan lintas modalitas juga membawa risiko. Jika teks mengandung kategori yang bias dan visual menggunakan label yang tidak netral, proses alignment dan fusion dapat menghasilkan representasi yang konsisten secara matematis, tetapi tetap bermasalah secara sosial. Penelitian terdahulu menunjukkan bahwa bias pada model vision-language dapat terbentuk dalam satu modalitas maupun melalui interaksi antarmodalitas (Srinivasan & Bisk, 2022). Penilaian bias yang melampaui kategori gender dan ras juga memperlihatkan bahwa stereotip dapat muncul pada berbagai kelompok dan konteks populasi (Janghorbani & De Melo, 2023). Dengan demikian, hubungan semantik yang kuat tidak selalu berarti hubungan tersebut adil, benar, atau sesuai dengan konteks sosial.

Temuan ketiga berkaitan dengan kestabilan representasi pada dokumen kebijakan yang kompleks. Integrasi narasi, grafik, peta, dan indikator membantu mempertahankan keterhubungan antarinformasi ketika dokumen memuat banyak kategori dan lapisan data. Dalam penelitian ini, stabilitas dipahami secara kualitatif sebagai konsistensi hubungan makna, bukan sebagai ketahanan statistik terhadap gangguan data. Penafsiran tersebut membatasi klaim agar tidak melampaui desain sintesis literatur dan studi kasus konseptual yang digunakan.

Risiko Etika Berdasarkan Prinsip Etika Klasik

Evaluasi berdasarkan prinsip etika klasik menghasilkan empat kategori utama, yaitu *autonomy*, *beneficence*, *justice*, dan *consent*. Hasil pemetaan menunjukkan bahwa persoalan etika tidak hanya muncul ketika sistem menghasilkan output, tetapi telah terbentuk sejak data dikumpulkan, dikategorikan, disimpan, dan dihubungkan dengan sumber lain. Prinsip-prinsip tersebut tetap relevan, tetapi penerapannya dalam sistem AI memerlukan penjabaran yang lebih operasional.

Dari perspektif *autonomy*, risiko utama berkaitan dengan keterbatasan kontrol masyarakat atas data yang dikumpulkan dalam lingkungan Smart City. Kamera pengawas, sensor, aplikasi layanan publik, perangkat bergerak, dan data lokasi dapat merekam aktivitas warga secara terus-menerus. Walaupun sebagian data berasal dari

ruang publik, penggunaannya tetap dapat menghasilkan inferensi mengenai identitas, kebiasaan, mobilitas, dan hubungan sosial. Konsep *data autonomy* menekankan bahwa persoalan tersebut tidak terbatas pada penyalahgunaan data pribadi, tetapi juga mencakup pelanggaran batas sosial dan berkurangnya kemampuan individu atau komunitas untuk mengendalikan proses datafikasi (Gstrein, 2024).

Prinsip *beneficence* mengharuskan manfaat sistem dibandingkan secara kritis dengan risiko yang ditimbulkan. Sistem prediksi kepadatan, misalnya, dapat membantu pemerintah mengatur lalu lintas dan meningkatkan efisiensi layanan. Namun, sistem yang sama dapat menghasilkan salah klasifikasi, memperluas pengawasan, atau digunakan untuk tujuan yang berbeda dari tujuan awalnya. Klaim manfaat publik karena itu tidak dapat diterima tanpa *impact assessment*, pengujian risiko, dan mekanisme koreksi. Kerangka etika AI UNESCO juga menempatkan proporsionalitas, keselamatan, perlindungan hak asasi manusia, dan pengawasan manusia sebagai prinsip penting dalam penggunaan AI (UNESCO, 2021).

Prinsip *justice* berkaitan dengan keadilan representasi dan distribusi manfaat serta risiko. Data Smart City dapat lebih banyak berasal dari wilayah yang memiliki sensor, konektivitas, dan partisipasi digital yang baik. Sebaliknya, wilayah dengan keterbatasan infrastruktur dapat kurang terwakili. Ketimpangan tersebut berpotensi menghasilkan sistem yang lebih akurat bagi kelompok tertentu, tetapi kurang tepat bagi kelompok lainnya. Studi multidimensional mengenai bias vision-language memperlihatkan bahwa bias tidak berdiri sebagai satu kategori tunggal, melainkan dapat melibatkan identitas, atribut sosial, konteks bahasa, dan hubungan antara objek visual dengan deskripsi tekstual (Ruggeri & Nozza, 2023).

Prinsip *consent* menghadapi persoalan khusus karena data perkotaan sering dikumpulkan tanpa interaksi langsung antara pengelola sistem dan individu yang terekam. Mekanisme persetujuan individual tidak selalu dapat diterapkan pada CCTV atau sensor ruang publik. Meskipun demikian, keadaan tersebut tidak menghapus kewajiban untuk menyediakan pemberitahuan, menjelaskan tujuan penggunaan, membatasi retensi, dan menyediakan mekanisme pengaduan. Persetujuan perlu dipahami sebagai bagian dari tata kelola yang lebih luas, bukan sekadar formulir administratif.

Tabel 2. Temuan risiko berdasarkan prinsip etika klasik

Prinsip	Temuan risiko	Rekomendasi
<i>Autonomy</i>	Tujuan, cakupan, dan penggunaan data belum selalu dipahami oleh masyarakat	Dokumentasi tujuan, pemberitahuan publik, pembatasan penggunaan, dan mekanisme kontrol
<i>Beneficence</i>	Manfaat sering diasumsikan tanpa evaluasi memadai terhadap risiko sosial	<i>Impact assessment</i> , evaluasi manfaat-risiko, dan audit algoritmik berkala
<i>Justice</i>	Representasi data dapat tidak seimbang antarkelompok dan antarwilayah	Evaluasi subkelompok, pemeriksaan distribusi data, dan perbaikan representasi
<i>Consent</i>	Persetujuan sulit diterapkan pada data yang dikumpulkan di ruang publik	Pemberitahuan yang mudah dipahami, pembatasan tujuan, dan mekanisme keberatan

Prinsip klasik tetap berfungsi sebagai dasar normatif, tetapi belum cukup untuk menangani risiko algoritmik yang dinamis. *Autonomy* perlu diterjemahkan menjadi kontrol data, *beneficence* menjadi penilaian dampak, *justice* menjadi evaluasi kelompok, dan *consent* menjadi mekanisme pemberitahuan serta pembatasan penggunaan.

Bias, Fairness, dan Interaksi Antarmodalitas

Hasil analisis memperlihatkan bahwa bias dapat muncul pada setiap komponen pipeline. Pada *Text Encoder*, bias dapat berasal dari pilihan istilah, narasi kebijakan, kategori administratif, atau representasi bahasa. Pada *Visual Encoder*, bias dapat muncul dari warna, simbol, anotasi, sudut pengambilan gambar, kualitas visual, dan distribusi objek. Ketika kedua representasi disejajarkan, hubungan yang bias dapat diperkuat melalui *Cross-Modal Alignment*.

Penelitian mengenai model vision-language menunjukkan bahwa model dapat membangun asosiasi sosial yang tidak seimbang meskipun data tampak netral (Cabello et al., 2023). Temuan tersebut memperlihatkan bahwa penghilangan kata atau kategori yang secara eksplisit menunjukkan identitas belum tentu menghapus bias. Model masih dapat menggunakan petunjuk visual, konteks, lokasi, pakaian, atau objek lain sebagai proksi. Karena itu, evaluasi fairness tidak cukup dilakukan dengan memeriksa satu atribut sensitif.

Strategi mitigasi bias juga tidak dapat dibatasi pada pembersihan dataset. Pendekatan *adversarial learning* menunjukkan bahwa intervensi pada proses pembelajaran model dapat membantu mengurangi asosiasi bias pada vision-language models (Berg et al., 2022). Namun, teknik tersebut tidak menjamin bahwa seluruh bias telah dihapus. Debiasing dapat memperbaiki satu kategori, tetapi meninggalkan atau memindahkan bias ke kategori lain. Oleh karena itu, mitigasi perlu melibatkan pengujian berulang, analisis subkelompok, pemeriksaan konteks, dan evaluasi oleh manusia.

Fairness pada embedding multimodal juga perlu dinilai pada tiga tingkat. Pertama, fairness pada setiap modalitas sebelum integrasi. Kedua, fairness pada hubungan hasil alignment. Ketiga, fairness pada output setelah fusion. Model dapat menunjukkan performa yang seimbang pada teks dan visual secara terpisah, tetapi menghasilkan hubungan lintas modalitas yang bias. Dengan demikian, evaluasi fairness harus mencakup keseluruhan pipeline dan tidak hanya berfokus pada prediksi akhir.

Accountability, Transparency, dan Explainability

Risiko berikutnya adalah terbatasnya akuntabilitas dalam pipeline Smart City. Ketika sistem tidak memiliki dokumentasi yang memadai, pihak pengembang, pemerintah, dan masyarakat akan kesulitan menelusuri hubungan antara data, versi model, proses alignment, fusion, dan output. Persoalan menjadi lebih serius ketika output digunakan dalam keputusan publik karena kesalahan sistem dapat memengaruhi distribusi layanan, pengawasan, mobilitas, atau kebijakan wilayah.

Penggunaan *model cards* dapat memperkuat transparansi dengan menyediakan informasi mengenai tujuan model, konteks penggunaan, performa, keterbatasan, kelompok yang dievaluasi, serta risiko penyalahgunaan (Mitchell et al., 2019). Namun, model card tidak seharusnya diperlakukan sebagai pengganti audit. Dokumentasi hanya berguna apabila informasinya lengkap, dapat diverifikasi, dan diperbarui ketika model atau dataset berubah.

Explainability juga menjadi persoalan penting karena representasi multimodal terbentuk dari interaksi beberapa sumber. Sistem tidak cukup hanya menjelaskan bahwa suatu output dihasilkan oleh model. Pengguna perlu mengetahui elemen teks dan visual yang memengaruhi output, cara kedua modalitas disejajarkan, serta alasan informasi tertentu memperoleh bobot lebih besar. Dalam keputusan publik, penjelasan merupakan bagian dari legitimasi karena warga berhak memahami dan mempertanyakan dasar keputusan yang memengaruhi mereka.

Kerangka *trustworthy AI* menempatkan explainability, fairness, robustness, dan accountability sebagai dimensi yang saling berhubungan (Thiebes et al., 2021). Sistem yang mudah dijelaskan tetapi menghasilkan ketidakadilan tetap bermasalah. Sebaliknya,

sistem yang memiliki performa relatif adil tetapi tidak dapat diaudit juga belum memenuhi prinsip tata kelola yang bertanggung jawab. Oleh sebab itu, transparansi tidak dapat dinilai secara terpisah dari dampak sosial dan distribusi tanggung jawab.

Kajian sistematis mengenai perangkat etika AI menunjukkan bahwa banyak organisasi telah mengembangkan pedoman, daftar pemeriksaan, kerangka penilaian, dan alat audit, tetapi penerapannya masih tidak merata (Ortega-Bolaños et al., 2024). Temuan tersebut mendukung perlunya *ethical checkpoints* pada setiap komponen pipeline, bukan hanya pernyataan etika pada bagian akhir penelitian.

Hasil Studi Kasus Konseptual Smart City

Studi kasus konseptual menggambarkan sistem prediksi kepadatan lalu lintas yang mengintegrasikan laporan insiden berbasis teks, rekaman CCTV, dan infografis peta. Data diproses melalui encoder yang berbeda, disejajarkan pada *Cross-Modal Alignment*, kemudian digabungkan melalui *Fusion Layer*. Skenario ini tidak mewakili audit terhadap kota tertentu, tetapi digunakan untuk menguji kemampuan kerangka dalam mengidentifikasi risiko.

Risiko pertama adalah privasi dan *indirect disclosure*. Rekaman CCTV dapat menangkap wajah, nomor kendaraan, lokasi, dan pola perjalanan. Walaupun identitas langsung dihapus, penggabungan waktu, lokasi, rute, dan laporan tekstual dapat memungkinkan identifikasi tidak langsung. Risiko meningkat apabila data disimpan terlalu lama atau digunakan kembali untuk kepentingan lain.

Risiko kedua adalah bias dalam pelabelan wilayah. Infografis dapat menggunakan istilah seperti “zona rawan”, “wilayah padat”, atau “area berisiko”. Kategori tersebut mungkin didasarkan pada data historis yang tidak seimbang. Jika label digunakan sebagai *ground truth*, model dapat mereproduksi asumsi lama dan memperkuat prioritas pengawasan pada wilayah tertentu.

Risiko ketiga berkaitan dengan distribusi dampak kebijakan. Prediksi kepadatan dapat digunakan untuk menentukan pengalihan arus, lokasi patroli, atau distribusi layanan. Kesalahan representasi dapat menyebabkan suatu wilayah menerima perhatian berlebihan, sedangkan wilayah lain diabaikan. Karena itu, output model tidak seharusnya menjadi satu-satunya dasar keputusan.

Penerapan AI pada sektor publik memang berkembang dalam berbagai layanan pemerintahan, tetapi manfaatnya sangat bergantung pada kapasitas institusi, kualitas data, dan tata kelola (van Noordt & Misuraca, 2022). Pemerintah tidak cukup hanya membeli atau menerapkan sistem. Institusi perlu memastikan adanya pembagian tanggung jawab, kompetensi teknis, prosedur audit, dan mekanisme pengaduan.

Risiko keempat adalah *dual-use*. Sistem lalu lintas dapat dikembangkan untuk meningkatkan mobilitas, tetapi representasi yang sama dapat digunakan untuk pemantauan individu atau profiling kelompok. Prinsip OECD menekankan pentingnya transparansi, akuntabilitas, keamanan, dan pengelolaan risiko sepanjang siklus hidup AI (OECD, 2023). NIST AI Risk Management Framework juga mengarahkan organisasi untuk mengidentifikasi, mengukur, mengelola, dan memantau risiko sebelum serta selama sistem digunakan (Tabassi, 2023).

Pemetaan Risiko pada Pipeline Embedding Multimodal

Hasil pemetaan menunjukkan bahwa setiap komponen memiliki bentuk risiko yang berbeda. Text Encoder berkaitan dengan bias naratif, Visual Encoder dengan bias visual dan pelabelan, Cross-Modal Alignment dengan penguatan bias, Fusion Layer dengan penggabungan risiko dan *dual-use*, sedangkan Semantic Output Layer berkaitan dengan explainability dan keterlacakan hasil.

Tabel 3. Matriks risiko etika pada pipeline embedding multimodal

Komponen pipeline	Risiko utama	Dampak potensial	Status mitigasi
<i>Text Encoder</i>	Bias naratif dan distorsi konteks	Diskriminasi dan hilangnya konteks	Belum optimal
<i>Visual Encoder</i>	Bias visual, simbol, dan pelabelan	Kesalahan representasi	Sebagian
<i>Cross-Modal Alignment</i>	Penguatan bias antarmodalitas	Salah interpretasi	Minim
<i>Fusion Layer</i>	Penggabungan bias dan <i>dual-use</i>	Surveillance dan dampak gabungan	Risiko tinggi
<i>Semantic Output Layer</i>	Rendahnya explainability dan transparency	Output sulit ditelusuri	Rendah

Fusion Layer menjadi komponen paling kritis karena informasi dan risiko dari beberapa modalitas digabungkan pada tahap ini. Bias yang relatif kecil pada teks dan visual dapat menghasilkan dampak lebih besar setelah fusion. Namun, kategori “risiko tinggi” dalam penelitian ini merupakan klasifikasi konseptual, bukan hasil pengukuran probabilitas. Penilaian tersebut didasarkan pada potensi dampak, luas penggunaan, dan keterbatasan mitigasi.

Cross-Modal Alignment juga memerlukan perhatian karena pengujiannya sering berfokus pada kesamaan representasi, bukan pada keadilan hubungan antarmodalitas. Model dapat menganggap teks dan visual selaras secara matematis, meskipun hubungan tersebut mengandung stereotip atau kesalahan konteks. Evaluasi alignment perlu memasukkan pemeriksaan kelompok, wilayah, kategori sosial, dan kondisi data yang berbeda.

Implikasi Filosofis, Metodologis, dan Praktis

Secara filosofis, embedding bukan sekadar transformasi data menjadi vektor. Proses tersebut melibatkan pilihan mengenai data yang dimasukkan, kategori yang digunakan, hubungan yang dipertahankan, dan informasi yang diabaikan. Dalam lingkungan Smart City, representasi semantik dapat membentuk cara pemerintah melihat warga, wilayah, mobilitas, dan masalah sosial. Kajian etika Smart City menegaskan bahwa teknologi perkotaan harus dinilai melalui relasinya dengan governance, inklusi sosial, privasi, kekuasaan, dan keberlanjutan (Ziosi et al., 2024).

Secara metodologis, evaluasi etika perlu dilakukan sepanjang pipeline. Text Encoder memerlukan audit narasi dan bahasa. Visual Encoder membutuhkan pemeriksaan label, simbol, dan distribusi visual. Cross-Modal Alignment memerlukan evaluasi fairness lintas modalitas. Fusion Layer membutuhkan analisis dampak gabungan dan kemungkinan penyalahgunaan. Semantic Output Layer memerlukan explainability, logging, dan human review.

Secara praktis, pemerintah kota perlu menetapkan tujuan penggunaan data sebelum pengumpulan dilakukan. Perubahan tujuan harus dievaluasi dan didokumentasikan. Prinsip *privacy by design*, minimisasi data, pembatasan retensi, pengendalian akses, dan pengawasan manusia perlu diterapkan sejak tahap desain. Rekomendasi UNESCO menekankan bahwa AI harus tetap berada di bawah tanggung jawab manusia dan tidak boleh mengurangi martabat, hak, atau kebebasan individu (UNESCO, 2021).

Sintesis hasil menunjukkan bahwa prinsip etika klasik dan kontemporer perlu digunakan secara bersamaan. *Autonomy* diterjemahkan menjadi kontrol dan transparansi data. *Beneficence* diwujudkan melalui penilaian manfaat-risiko. *Justice* diterapkan melalui evaluasi representasi dan subgroup fairness. *Accountability* diwujudkan melalui dokumentasi, audit trail, oversight, dan pembagian tanggung jawab institusional. Pendekatan tersebut tidak menghilangkan prinsip klasik, tetapi mengoperasionalkannya dalam konteks AI yang otomatis, kompleks, dan berdampak luas.

Temuan penelitian pada akhirnya menunjukkan bahwa manfaat embedding multimodal bersifat bersyarat. Integrasi teks dan visual dapat memperkaya representasi semantik, tetapi juga dapat memperkuat bias, meningkatkan risiko privasi, dan memperluas kapasitas pengawasan. Oleh karena itu, kualitas sistem tidak dapat dinilai hanya berdasarkan performa teknis. Fairness, transparency, explainability, accountability, privasi, dan pengawasan manusia harus diperlakukan sebagai bagian dari desain dan evaluasi ilmiah. Kerangka konseptual ini masih memerlukan validasi ahli dan pengujian empiris menggunakan dataset Smart City aktual sebelum dapat digunakan sebagai dasar generalisasi atau kebijakan

KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi *embedding* multimodal pada Big Data Smart City mampu memperkaya representasi semantik melalui penggabungan informasi tekstual dan visual. Hubungan antarmodalitas membantu mempertahankan konteks kebijakan, memperjelas keterkaitan antara narasi, indikator, peta, dan infografis, serta mendukung interpretasi data yang lebih komprehensif. Namun, manfaat tersebut tidak dapat dipisahkan dari risiko etika yang muncul sepanjang *pipeline*, meliputi bias representasional, privasi dan *indirect disclosure*, *dual-use*, *surveillance* dan *profiling* otomatis, serta asimetri kekuasaan informasi. Pemetaan terhadap lima komponen *pipeline* menunjukkan bahwa *Fusion Layer* memiliki risiko paling tinggi karena menggabungkan bias dan potensi penyalahgunaan dari beberapa modalitas, sedangkan *Cross-Modal Alignment* masih memiliki mekanisme mitigasi yang terbatas. Prinsip *autonomy*, *beneficence*, *justice*, dan *consent* tetap relevan, tetapi perlu dioperasionalkan melalui *fairness*, *accountability*, *transparency*, *explainability*, tata kelola data, dan manajemen risiko. Karena penelitian ini menggunakan sintesis literatur dan studi kasus konseptual, hasilnya belum dapat diperlakukan sebagai bukti empiris terhadap sistem atau kota tertentu. Implementasi kerangka memerlukan validasi ahli, pengujian menggunakan dataset Smart City aktual, audit bias, dokumentasi model dan dataset, *logging*, pembatasan tujuan penggunaan, serta *human oversight*. Dengan demikian, kualitas *embedding* multimodal perlu dinilai berdasarkan performa teknis sekaligus tanggung jawab sosial, keterlacakan, keadilan, dan perlindungan otonomi warga sepanjang seluruh siklus hidup pengembangan dan penerapan sistem AI perkotaan.

REKOMENDASI

Pengembangan dan penerapan *embedding* multimodal dalam sistem Big Data Smart City perlu menempatkan evaluasi etika sebagai bagian dari setiap tahap *pipeline*, bukan hanya sebagai pemeriksaan setelah model menghasilkan output. Peneliti dan pengembang disarankan melakukan audit bias secara terpisah pada *Text Encoder* dan *Visual Encoder*, kemudian melanjutkannya dengan pengujian *fairness* pada tahap *Cross-Modal Alignment* dan *Fusion Layer*. Perhatian khusus perlu diberikan kepada *Fusion Layer* karena komponen ini memiliki potensi tertinggi untuk menggabungkan bias, memperluas fungsi penggunaan, dan meningkatkan risiko *surveillance*.

Pemerintah kota dan penyedia teknologi perlu menerapkan prinsip *privacy by design*, minimalisasi data, pembatasan masa penyimpanan, pengendalian akses, serta pembatasan tujuan penggunaan. Dokumentasi melalui *dataset cards*, *model cards*, *audit trail*, dan *logging* juga perlu diwajibkan agar proses pengambilan keputusan dapat ditelusuri dan dipertanggungjawabkan. Output sistem yang berdampak pada pelayanan publik, pengawasan, atau distribusi sumber daya tidak seharusnya digunakan tanpa verifikasi manusia.

Penelitian selanjutnya disarankan menguji kerangka ini menggunakan dataset Smart City aktual, melibatkan ahli AI, etika, hukum, pemerintah, dan perwakilan masyarakat. Pengujian empiris juga perlu menggunakan metrik performa, *fairness*, privasi, *explainability*, dan evaluasi subkelompok agar efektivitas mitigasi dapat diukur secara lebih objektif.

DAFTAR PUSTAKA

- Berg, H., Hall, S., Bhalgat, Y., Kirk, H., Shtedritski, A., & Bain, M. (2022). A prompt array keeps the bias away: Debiasing vision-language models with adversarial learning. *AAACL-IJCNLP 2022*. <https://aclanthology.org/2022.aacl-main.61/>
- Cabello, L., Bugliarello, E., Brandl, S., & Elliott, D. (2023). Evaluating bias and fairness in gender-neutral pretrained vision-and-language models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. <https://aclanthology.org/2023.emnlp-main.525/>
- Gstrein, O. J. (2024). Data autonomy: Beyond personal data abuse, sphere transgression, and datafied gentrification in smart cities. *Ethics and Information Technology*, 26, Article 61. <https://doi.org/10.1007/s10676-024-09799-x>
- Janghorbani, S., & De Melo, G. (2023). Multi-modal bias: Introducing a framework for stereotypical bias assessment beyond gender and race in vision-language models. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 1725–1735). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.eacl-main.126>
- Manzoor, M. A., Albarri, S., Xian, Z., Meng, Z., Nakov, P., & Liang, S. (2023). Multimodality representation learning: A survey on evolution, pretraining, and its applications. *ACM Transactions on Multimedia Computing, Communications, and Applications*. <https://arxiv.org/abs/2302.00389>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
- Organisation for Economic Co-operation and Development. (2023). *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI* (OECD Digital Economy Papers No. 349). OECD Publishing. <https://doi.org/10.1787/2448f04b-en>
- Ortega-Bolaños, R., Bernal-Salcedo, J., Ortiz, M. G., Galeano Sarmiento, J., Ruz, G. A., & Tabares-Soto, R. (2024). Applying the ethics of AI: A systematic review of tools for developing and assessing AI-based systems. *Artificial Intelligence Review*, 57, Article 110. <https://doi.org/10.1007/s10462-024-10740-3>
- Ruggeri, F., & Nozza, D. (2023). A multi-dimensional study on bias in vision-language models. *Findings of the Association for Computational Linguistics: ACL 2023*. <https://aclanthology.org/2023.findings-acl.403/>

- Srinivasan, T., & Bisk, Y. (2022). Worst of both worlds: Biases compound in pre-trained vision-and-language models. In *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing* (pp. 77–85). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.gebnlp-1.10>
- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
- United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Government Information Quarterly*, 39(3), Article 101714. <https://doi.org/10.1016/j.giq.2022.101714>
- Ziosi, M., Hewitt, B., Juneja, P., Taddeo, M., & Floridi, L. (2024). Smart cities: Reviewing the debate about their ethical implications. *AI & Society*, 39, 1185–1200. <https://doi.org/10.1007/s00146-022-01558-0>