



TANTANGAN ETIKA DALAM IMPLEMENTASI KECERDASAN BUATAN DAN PEMBELAJARAN MESIN: SYSTEMATIC LITERATURE REVIEW

Yaulie Deo Y. Rindengan ¹, Febrian Rezki Hemeto ², Ryan Christian Fabian Rattu ³

^{1,2,3} Fakultas Teknik, Program Studi Informatika, Universitas Sam Ratulangi, Manado, Indonesia

E-mail: Rindengan@unsrat.ac.id ¹, febrianhemeto026@student.unsrat.ac.id ²,
ryanrattu026@student.unsrat.ac.id ³

ABSTRACT

The development of artificial intelligence and machine learning technologies has advanced rapidly over the past decade, yet their implementation presents a range of complex ethical challenges. This study aims to identify and analyze the main ethical challenges in the implementation of artificial intelligence and machine learning through a systematic literature review approach. The research methodology uses the PRISMA protocol with literature searches conducted in the IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Link, and Google Scholar databases for the period 2018-2024. From 287 articles identified, 42 high-quality articles were selected for in-depth analysis.

The findings indicate that the primary ethical challenges include algorithmic bias and discrimination, data privacy and security, system transparency and explainability, and accountability in automated decision-making. Significant social impacts include job displacement, the digital divide, and information manipulation through deepfakes. The study also identifies various emerging ethical frameworks and risk-mitigation solutions that can be applied.

The conclusion highlights that responsible implementation of artificial intelligence requires a holistic approach that incorporates technical, legal, and social aspects, supported by collaboration among developers, regulators, and society.

Keywords: *Information Technology Ethics, Artificial Intelligence, Machine Learning, Systematic Literature Review, Algorithmic Bias, Data Privacy*

ABSTRAK

Perkembangan teknologi kecerdasan buatan dan pembelajaran mesin telah mengalami kemajuan pesat dalam dekade terakhir, namun implementasinya menimbulkan berbagai tantangan etika yang kompleks. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis tantangan etika utama dalam implementasi artificial intelligence dan machine learning melalui pendekatan systematic literature review. Metodologi penelitian menggunakan

Article History

Received: Desember 2025

Reviewed: Desember 2025

Published: Desember 2025

Plagiarism Checker No
235

Prefix DOI :

[10.8734/Kohesi.v1i2.365](https://doi.org/10.8734/Kohesi.v1i2.365)

Copyright : Author

Publish by : Kohesi



This work is licensed
under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



protokol PRISMA dengan pencarian literatur pada database IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Link, dan Google Scholar untuk periode 2018-2024. Dari 287 artikel yang ditemukan, 42 artikel berkualitas tinggi dipilih untuk analisis mendalam. Hasil penelitian menunjukkan bahwa tantangan etika utama meliputi bias algoritma dan diskriminasi, privasi dan keamanan data, transparansi dan explainability sistem, serta akuntabilitas dalam pengambilan keputusan otomatis. Dampak sosial yang signifikan mencakup displacement pekerjaan, digital divide, dan manipulasi informasi melalui deepfakes. Penelitian ini juga mengidentifikasi berbagai framework etika yang berkembang dan solusi mitigasi risiko yang dapat diterapkan. Kesimpulan menunjukkan bahwa implementasi kecerdasan buatan yang bertanggung jawab memerlukan pendekatan holistik yang melibatkan aspek teknis, hukum, dan sosial dengan kolaborasi antara pengembang, regulator, dan masyarakat.

Kata Kunci: Etika Teknologi Informasi, Kecerdasan Buatan, Pembelajaran Mesin, Systematic Literature Review, Bias Algoritma, Privasi Data



1. PENDAHULUAN

Era digital saat ini ditandai dengan perkembangan teknologi kecerdasan buatan dan pembelajaran mesin yang sangat pesat (Zhou & Chen, 2025). Teknologi artificial intelligence dan machine learning telah mengubah cara manusia berinteraksi dengan sistem komputer dan mempengaruhi berbagai aspek kehidupan, mulai dari layanan kesehatan, pendidikan, transportasi, hingga sistem keuangan. Implementasi teknologi ini memberikan manfaat yang signifikan dalam meningkatkan efisiensi, akurasi, dan kemampuan prediksi dalam berbagai domain aplikasi (Baeza-Yates, 2024).

Namun, di balik kemajuan teknologi yang mengagumkan ini, muncul berbagai tantangan etika yang kompleks dan memerlukan perhatian serius dari para peneliti, praktisi, dan pembuat kebijakan (Modi, 2025). Isu-isu seperti bias algoritma yang dapat menyebabkan diskriminasi sistemik, pelanggaran privasi data pengguna, kurangnya transparansi dalam pengambilan keputusan otomatis, dan ketidakjelasan akuntabilitas ketika sistem artificial intelligence membuat kesalahan, menjadi perhatian utama dalam implementasi teknologi ini (Jaiswal et al., 2023).

Kasus-kasus nyata seperti bias gender dalam sistem rekrutmen otomatis, diskriminasi rasial dalam algoritma pengenalan wajah, dan pelanggaran privasi dalam penggunaan data personal untuk training model machine learning, menunjukkan bahwa tantangan etika dalam implementasi kecerdasan buatan bukan hanya masalah teoretis, tetapi memiliki dampak nyata terhadap kehidupan masyarakat. Hal ini menimbulkan kebutuhan mendesak untuk memahami secara komprehensif berbagai tantangan etika yang ada dan mengembangkan solusi yang tepat.

Penelitian tentang etika dalam teknologi informasi, khususnya dalam konteks artificial intelligence dan machine learning, telah berkembang pesat dalam beberapa tahun terakhir (Cheng et al., 2021). Berbagai framework etika telah diusulkan oleh organisasi internasional, perusahaan teknologi, dan institusi akademik untuk mengatasi tantangan-tantangan ini (Zhou & Chen, 2025). Namun, masih terdapat kesenjangan dalam pemahaman holistik tentang spektrum lengkap tantangan etika dan efektivitas solusi yang telah diusulkan (Jaiswal et al., 2023).

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk mengidentifikasi dan menganalisis tantangan etika utama dalam implementasi artificial intelligence dan machine learning melalui systematic literature review. Penelitian ini



merumuskan empat pertanyaan penelitian utama: (1) tantangan etika utama dalam implementasi teknologi kecerdasan buatan dan pembelajaran mesin, (2) dampak sosial dan ekonomi dari implementasi artificial intelligence dan machine learning, (3) framework etika yang telah dikembangkan untuk mengatasi tantangan etika, dan (4) solusi dan strategi mitigasi yang efektif untuk mengurangi risiko etika. Penelitian ini menggunakan protokol PRISMA dengan pencarian literatur pada database IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Link, dan Google Scholar untuk periode 2018-2024. Hasil penelitian diharapkan dapat memberikan kontribusi bagi pengembangan implementasi teknologi kecerdasan buatan yang bertanggung jawab dan etis.



2. METODE PENELITIAN

2.1 Desain Penelitian (Systematic Literature Review)

Penelitian ini menggunakan pendekatan systematic literature review untuk mengidentifikasi, mengevaluasi, dan mensintesis literatur ilmiah yang relevan tentang tantangan etika dalam implementasi kecerdasan buatan dan pembelajaran mesin. Systematic literature review dipilih sebagai metodologi penelitian karena kemampuannya untuk memberikan analisis yang komprehensif dan objektif terhadap body of knowledge yang ada, serta mengidentifikasi gap penelitian yang dapat menjadi dasar untuk penelitian selanjutnya.

Desain penelitian ini mengikuti protokol PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) yang merupakan standar internasional untuk melakukan dan melaporkan systematic review. Protokol PRISMA memastikan bahwa proses review dilakukan secara sistematis, transparan, dan dapat direproduksi, sehingga meningkatkan validitas dan reliabilitas hasil penelitian.

Pendekatan systematic literature review dalam penelitian ini melibatkan beberapa tahap utama, yaitu perumusan pertanyaan penelitian yang spesifik, pengembangan strategi pencarian literatur yang komprehensif, penerapan kriteria inklusi dan eksklusi yang jelas, ekstraksi data yang sistematis, dan analisis serta sintesis temuan dari literatur yang dipilih. Setiap tahap dilakukan dengan dokumentasi yang detail untuk memastikan transparansi dan reproducibility penelitian.

2.2 Strategi Pencarian Literatur

Pencarian literatur dilakukan pada beberapa database akademik utama yang mencakup publikasi berkualitas tinggi dalam bidang computer science, information technology, dan ethics. Database yang digunakan meliputi IEEE Xplore Digital Library yang merupakan database utama untuk publikasi di bidang electrical engineering dan computer science, ACM Digital Library yang mencakup publikasi dari Association for Computing Machinery, ScienceDirect yang menyediakan akses ke jurnal-jurnal Elsevier dalam berbagai bidang ilmu, Springer Link yang mencakup publikasi dari Springer Nature, Google Scholar sebagai database yang komprehensif mencakup berbagai jenis publikasi akademik, dan JSTOR untuk mengakses publikasi dalam bidang philosophy dan ethics yang relevan dengan topik penelitian.

Pemilihan database-database ini didasarkan pada cakupan konten yang relevan dengan topik penelitian, kualitas publikasi yang terindeks, dan aksesibilitas database tersebut. Kombinasi database ini diharapkan dapat memberikan cakupan literatur yang komprehensif



dan representatif terhadap state-of-the-art penelitian dalam bidang etika artificial intelligence dan machine learning.

Strategi pencarian literatur menggunakan kombinasi kata kunci yang dikembangkan berdasarkan konsep-konsep utama dalam penelitian ini. Kata kunci utama yang digunakan meliputi "AI ethics", "artificial intelligence ethics", "machine learning ethics", "algorithmic bias", "AI fairness", "ethical AI", dan "responsible AI". Kata kunci ini dipilih berdasarkan terminologi yang umum digunakan dalam literatur ilmiah tentang etika artificial intelligence.

Kombinasi kata kunci dikembangkan menggunakan operator Boolean untuk meningkatkan precision dan recall pencarian. Kombinasi yang digunakan meliputi ("AI ethics" AND "machine learning") OR ("artificial intelligence" AND "ethical issues"), ("algorithmic bias" AND "fairness") OR ("AI transparency" AND "explainability"), ("data privacy" AND "artificial intelligence") OR ("AI governance" AND "regulation"), dan ("ethical AI" AND "social impact") OR ("responsible AI" AND "accountability").

Strategi pencarian juga menggunakan variasi terminologi dan sinonim untuk memastikan bahwa literatur yang relevan tidak terlewatkan. Pencarian dilakukan dalam bahasa Inggris karena merupakan bahasa utama publikasi ilmiah internasional dalam bidang ini.

Kriteria inklusi untuk pemilihan literatur meliputi publikasi yang membahas aspek etika dalam artificial intelligence atau machine learning, publikasi dalam bentuk journal articles atau conference papers, publikasi dalam bahasa Inggris, publikasi yang dipublikasikan dalam periode 2021-2025 untuk memastikan relevansi dengan perkembangan teknologi terkini, dan publikasi yang tersedia dalam full text.

Kriteria eksklusi meliputi publikasi yang hanya membahas aspek teknis artificial intelligence tanpa dimensi etika, publikasi dalam bentuk editorial, opinion pieces, atau book reviews, publikasi yang tidak tersedia dalam full text, publikasi yang bukan dalam bahasa Inggris, dan publikasi yang dipublikasikan sebelum tahun 2018 kecuali merupakan seminal works yang sangat relevan dengan topik penelitian.

2.3 Ekstraksi Data

Ekstraksi data dilakukan secara sistematis menggunakan form ekstraksi yang telah dikembangkan sebelumnya. Form ekstraksi mencakup informasi bibliografi (penulis, tahun publikasi, judul, journal/conference), metodologi penelitian yang digunakan, fokus utama penelitian, tantangan etika yang diidentifikasi, solusi atau framework yang diusulkan, temuan utama, dan limitasi penelitian.

Proses ekstraksi data dilakukan secara konsisten untuk semua artikel yang dipilih,



dengan fokus pada informasi yang relevan dengan pertanyaan penelitian. Data yang diekstraksi kemudian diorganisir dalam database yang memungkinkan analisis dan sintesis yang efektif.

Kualitas ekstraksi data dijaga melalui penggunaan form yang terstandarisasi dan pilot testing pada subset artikel. Hal ini memastikan bahwa informasi yang diekstraksi konsisten dan komprehensif untuk semua artikel yang dianalisis.

2.4 Analisis dan Sintesis Data

Analisis data dilakukan menggunakan pendekatan thematic analysis untuk mengidentifikasi tema-tema utama yang muncul dari literatur yang direview. Proses analisis dimulai dengan coding artikel-artikel berdasarkan tantangan etika yang diidentifikasi, solusi yang diusulkan, dan temuan utama. Code-code ini kemudian diorganisir menjadi kategori dan tema yang lebih luas.

Sintesis data dilakukan dengan mengintegrasikan temuan dari berbagai studi untuk mengidentifikasi pola, konsistensi, dan inkonsistensi dalam literatur. Pendekatan narrative synthesis digunakan untuk mengorganisir dan menyajikan temuan dalam format yang koheren dan komprehensif.

Analisis juga mencakup identifikasi gap dalam literatur yang ada, area kontroversi atau ketidaksetujuan antar peneliti, dan trend dalam penelitian etika artificial intelligence. Hal ini penting untuk memberikan kontribusi terhadap pengembangan knowledge base dalam bidang ini.



3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis dari studi-studi yang telah diseleksi dan diekstraksi berdasarkan metodologi SLR pada Bab 2. Pembahasan difokuskan pada pola temuan yang muncul, kesesuaian dengan kerangka teori dan regulasi privasi, serta perbandingannya dengan penelitian terdahulu. Struktur utama dalam bab ini mencakup penjabaran hasil analisis tematik, interpretasi temuan, implikasi praktis dan etis, serta penyandingan hasil dengan konteks riset sebelumnya.

3.1 Hasil Pencarian dan Seleksi Literatur

Proses pencarian literatur menghasilkan total 156 artikel dari berbagai database yang digunakan. Distribusi artikel berdasarkan database adalah sebagai berikut: Google Scholar menghasilkan 89 artikel, IEEE Xplore menghasilkan 32 artikel, ACM Digital Library menghasilkan 18 artikel, ScienceDirect menghasilkan 12 artikel, Springer Link menghasilkan 4 artikel, dan JSTOR menghasilkan 1 artikel. Setelah menghilangkan duplikasi, tersisa 134 artikel unik yang kemudian masuk ke tahap screening.

Tahap screening berdasarkan judul dan abstrak menghasilkan 45 artikel yang memenuhi kriteria relevansi awal. Artikel-artikel yang dieliminasi pada tahap ini umumnya karena tidak membahas aspek etika artificial intelligence secara spesifik, fokus pada aspek teknis semata, atau tidak relevan dengan scope penelitian. Screening dilakukan dengan menggunakan kriteria inklusi dan eksklusi yang telah ditetapkan sebelumnya.

Tahap review full text dilakukan terhadap 45 artikel yang lolos screening awal. Pada tahap ini, 37 artikel dieliminasi karena berbagai alasan, termasuk kualitas metodologi yang tidak memadai, tidak tersedia full text, atau setelah dibaca secara mendalam ternyata tidak relevan dengan fokus penelitian. Proses ini menghasilkan 8 artikel berkualitas tinggi yang dimasukkan dalam analisis final.

3.2 Pembahasan

Implementasi artificial intelligence (AI) dan machine learning (ML) menghadapi berbagai tantangan etika yang kompleks, terutama terkait bias algoritma yang dapat muncul pada seluruh siklus pengembangan sistem. Bias ini dapat berasal dari data historis, representasi data yang tidak seimbang, penggunaan model yang tidak sesuai untuk populasi tertentu, atau metrik evaluasi yang tidak adil. Dampaknya terlihat di berbagai sektor seperti rekrutmen, layanan keuangan, serta sistem peradilan, dan menimbulkan konsekuensi sosial maupun hukum yang serius. Selain itu, isu privasi data semakin penting seiring meningkatnya kebutuhan data dalam jumlah besar untuk melatih model AI, yang memunculkan tantangan seperti consent management, risiko re-identification, dan ketidakselarasan dengan prinsip data minimization.

Transparansi dan interpretabilitas menjadi tantangan lain dalam pengembangan sistem AI modern, khususnya dengan semakin kompleksnya model deep learning. Pengguna dan stakeholder sering kali kesulitan memahami bagaimana suatu keputusan dihasilkan, sehingga diperlukan pendekatan seperti LIME, SHAP, dan mekanisme perhatian untuk memberikan penjelasan yang lebih jelas. Namun, trade-off antara akurasi dan interpretabilitas tetap



menjadi isu utama. Tantangan-tantangan ini juga berdampak pada kepercayaan, auditabilitas, dan akuntabilitas sistem AI.

Dari perspektif sosial dan ekonomi, perkembangan AI membawa risiko displacement pekerjaan, terutama pada jenis pekerjaan rutin yang mudah diotomatisasi. Meskipun beberapa pekerjaan baru muncul, dampaknya tidak merata dan cenderung memperburuk ketimpangan bagi kelompok dengan pendidikan atau keterampilan rendah. Selain itu, digital divide semakin terlihat ketika masyarakat yang tidak memiliki akses teknologi atau literasi digital yang memadai tertinggal dalam memanfaatkan sistem berbasis AI, sehingga memperlebar kesenjangan dalam akses layanan kesehatan, pendidikan, pekerjaan, dan layanan publik.

Ancaman manipulasi informasi melalui deepfakes menjadi tantangan etika yang semakin kritis. Dengan kemampuan menghasilkan konten sintetis yang sangat realistis, teknologi ini dapat digunakan untuk menyebarkan disinformasi politik, penipuan, dan bentuk penyalahgunaan lainnya. Upaya mitigasi seperti deteksi deepfake, watermarking, dan regulasi konten diperlukan, meskipun perkembangan teknologi pembuatan konten palsu terus melampaui kemampuan deteksi.

Dalam aspek regulasi, berbagai negara dan organisasi internasional telah mengembangkan kerangka hukum untuk memastikan penggunaan AI yang bertanggung jawab. Uni Eropa memelopori pendekatan risk-based regulation melalui AI Act, sementara IEEE, ISO, dan perusahaan teknologi besar mengembangkan standar dan best practices terkait privasi, fairness, dan manajemen risiko. Audit dan monitoring sistem AI menjadi komponen penting untuk memastikan kepatuhan terhadap prinsip etika, meskipun masih menghadapi hambatan seperti kurangnya transparansi, akses terhadap model, dan keterbatasan metodologi audit.

Solusi mitigasi risiko etika mencakup pendekatan teknis seperti fairness-aware machine learning, differential privacy, federated learning, serta peningkatan interpretabilitas model; pendekatan organisasional berupa governance yang kuat, ethics review, dan program pelatihan; serta pendekatan regulatori melalui standar industri dan kerangka hukum yang jelas. Tren terbaru menunjukkan pergeseran ke arah pendekatan etika yang lebih praktis dan kontekstual, peningkatan kolaborasi lintas disiplin, serta kebutuhan koordinasi global. Implikasi bagi pengembang dan organisasi mencakup pentingnya menerapkan ethics by design, memperkuat governance, berinvestasi dalam edukasi, serta melakukan audit dan monitoring berkelanjutan untuk memastikan pengembangan AI yang aman, adil, dan bertanggung jawab..

KESIMPULAN

Penelitian systematic literature review ini telah mengidentifikasi dan menganalisis berbagai tantangan etika utama dalam implementasi teknologi kecerdasan buatan dan pembelajaran mesin berdasarkan 42 artikel berkualitas tinggi yang dipublikasikan dalam periode 2018-2024. Hasil penelitian menunjukkan bahwa tantangan etika dalam artificial intelligence dan machine learning bersifat multidimensional dan kompleks, mencakup aspek teknis, sosial, hukum, dan ekonomi yang saling berinteraksi.

Tantangan etika utama yang diidentifikasi meliputi bias algoritma dan diskriminasi yang dapat menyebabkan ketidakadilan sistemik, isu privasi dan keamanan data yang berkaitan dengan penggunaan data personal dalam jumlah besar, kurangnya transparansi dan explainability dalam sistem artificial intelligence yang kompleks, serta tantangan



akuntabilitas dan tanggung jawab ketika sistem membuat keputusan otomatis. Setiap tantangan ini memiliki karakteristik dan implikasi yang unik, namun seringkali saling terkait dan memerlukan pendekatan holistik untuk mengatasinya.

Dampak sosial dan ekonomi dari implementasi artificial intelligence mencakup displacement pekerjaan dan perubahan struktur pasar tenaga kerja, digital divide yang dapat memperburuk ketidaksetaraan yang sudah ada, serta risiko manipulasi informasi melalui teknologi seperti deepfakes. Dampak-dampak ini menunjukkan bahwa implementasi artificial intelligence tidak hanya memiliki implikasi teknis tetapi juga implikasi sosial yang luas yang perlu dipertimbangkan dalam pengembangan dan deployment sistem.

Berbagai framework etika telah dikembangkan oleh organisasi internasional, perusahaan teknologi, dan institusi akademik untuk mengatasi tantangan-tantangan ini. Framework-framework ini umumnya menekankan prinsip-prinsip seperti fairness, transparency, accountability, privacy, dan human oversight. Namun, masih terdapat kesenjangan antara prinsip-prinsip yang diusulkan dan implementasi praktisnya dalam pengembangan sistem artificial intelligence.

DAFTAR PUSTAKA

- [1] J. Zhou and F. Chen, *AI ethics: From principles to practice*. **AI & Society**, Data Science Institute, University of Technology Sydney, Australia, 2025. Available at: <https://orcid.org/0000-0001-6034-644X>
- [2] B. Modi, *Ethics and artificial intelligence*. **International Journal of Medical Research**, vol. 13, no. 3, pp. 61-65, 2025. <https://doi.org/10.55489/ijmr.1303202582>
- [3] J. Savulescu, A. Giubilini, R. Vandersluis, and A. Mishra, *Ethics of artificial intelligence in medicine*. **Singapore Medical Journal**, vol. 65, pp. 150-158, 2024.
- [4] W. Holmes and K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, challenges, and debates*. New York & London: Routledge, 2022.
- [5] L. Judijanto, A. Hardiansyah, and O. Arifudin, *Ethics and security in artificial intelligence and machine learning: Current perspectives in computing*. **International Journal of Society Reviews (INJOSER)**, vol. 3, no. 2, pp. 374-380, 2025.
- [6] R. K. Jaiswal, S. S. Sharma, and R. Kaushik, *Ethics in AI and machine learning*. **Journal of Nonlinear Analysis and Optimization**, vol. 14, no. 1, pp. 8-12, 2023. <https://doi.org/10.36893/JNAO.2023.V14I1.0008-0012>
- [7] R. Baeza-Yates, *Introduction to responsible AI*. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24)*, pp. 1-4, 2024.



<https://doi.org/10.1145/3616855.3636455>

[8] L. Cheng, K. R. Varshney, and H. Liu, *Socially responsible AI algorithms: Issues, purposes, and challenges*. **Journal of Artificial Intelligence Research**, vol. 71, pp. 1137-1181, 2021.