

**Analisis Etika dalam Sistem Pengambilan Keputusan berbasis AI:
Systematic Literature Review 2020-2025.****Yaulie Deo Y. Rindengan¹, Jacob Alfanicolls Rahayaan², Andre Immanuel Porayou³**^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Sam Ratulangi¹Rindengan@unsrat.ac.id, ²jacobrahayaan026@student.unsrat.ac.id,³andreporayou026@student.unsrat.ac.id**Abstrak**

Penerapan Kecerdasan Buatan (AI) dalam sistem pengambilan keputusan telah berkembang pesat pada periode 2020–2025, namun memunculkan tantangan etika serius seperti bias algoritmik, pelanggaran privasi, dan kurangnya transparansi (black box). Penelitian ini bertujuan untuk memetakan tren, pola, dan isu utama etika spesifik pada sistem pengambilan keputusan berbasis AI dengan menggunakan metode Systematic Literature Review (SLR). Data diperoleh melalui penelusuran sistematis pada pangkalan data internasional bereputasi, di mana 15 artikel terpilih dianalisis untuk menjawab pertanyaan penelitian terkait mitigasi bias, pendekatan desain, dan faktor etika yang berpengaruh. Hasil penelitian menunjukkan tiga temuan utama. Pertama, bias algoritmik teridentifikasi sebagai isu paling kritis yang kini dimitigasi melalui pendekatan hibrida: integrasi Explainable AI (XAI), audit algoritma berkelanjutan, dan jaminan keadilan (fairness). Kedua, terjadi pergeseran paradigma desain dari pedoman normatif menuju kerangka kerja praktis AI by Design yang mengintegrasikan privasi, kontrol manusia bermakna (meaningful human control), dan mekanisme kehati-hatian bawaan (embedded caution). Ketiga, keseimbangan antara transparansi dan akuntabilitas menjadi faktor penentu utama dalam membangun kepercayaan pengguna guna mengatasi fenomena psikologis seperti Outcome Bias dan Moral Distance. Penelitian ini menyimpulkan bahwa adopsi kerangka kerja etika global yang teruji sangat penting untuk mengharmonisasikan standar teknis dengan implementasi UU Pelindungan Data Pribadi (UU PDP) di Indonesia, khususnya dalam menjamin hak penjelasan dan keamanan data.

Kata kunci:

Etika Kecerdasan Buatan, Sistem Pengambilan Keputusan, Bias Algoritmik, Transparansi, Akuntabilitas

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/)

**Abstract**

The application of Artificial Intelligence (AI) in decision-making systems has grown rapidly in the period 2020-2025, but it has raised serious ethical challenges such as algorithmic bias, privacy violations, and lack of transparency (black box). This study aims to map the trends, patterns, and key ethical issues specific to AI-based decision-making systems using the Systematic Literature Review (SLR) method. Data were obtained through a systematic search of reputable international databases, where 15 selected articles were analyzed to answer research questions related to bias mitigation, design approaches, and influential ethical factors. The results of the study show three main findings. First, algorithmic bias was identified as the most critical issue, which is now being mitigated through a hybrid approach: the integration of Explainable AI (XAI), continuous algorithm audits, and fairness guarantees. Second, there has been a paradigm shift in design from normative guidelines to a practical AI by Design framework that integrates privacy, meaningful human control, and embedded caution mechanisms. Third, the balance between transparency and accountability is a key determining factor in building user trust to overcome psychological phenomena such as Outcome Bias and Moral Distance. This study concludes that the adoption of a proven global ethical framework is essential to harmonize technical standards with the implementation of the Personal Data Protection Law (PDP Law) in Indonesia, particularly in guaranteeing the right to explanation and data security.

Keywords: *AI Ethics, Decision-Making System, Algorithmic Bias, Transparency, Accountability*



1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) pada periode 2020-2025 telah membawa perubahan signifikan dalam berbagai aspek kehidupan manusia. Sistem pengambilan keputusan berbasis AI kini digunakan secara luas dalam sektor industri, pemerintahan, pendidikan, kesehatan, keuangan, hingga layanan publik. AI mampu menganalisis data dalam jumlah besar, mengenali pola-pola kompleks, serta menghasilkan rekomendasi yang dianggap lebih cepat dan akurat dibandingkan pengambilan keputusan tradisional. Kemampuan ini membuat AI menjadi bagian penting dalam proses digitalisasi dan otomatisasi organisasi modern (Prasetya & Hanim, 2023).

Meskipun memberikan banyak manfaat, adopsi AI menimbulkan berbagai persoalan etis yang semakin mendapat perhatian. Berbagai penelitian mencatat bahwa sistem AI dapat memperkuat bias, melakukan diskriminasi terhadap kelompok tertentu, dan menghasilkan keputusan yang tidak transparan karena sifat algoritma yang bersifat black box (Harahap & Siregar, 2024). Selain itu, penggunaan data pribadi dalam jumlah besar menimbulkan risiko kebocoran data, pelanggaran privasi, serta isu mengenai siapa yang bertanggung jawab ketika terjadi kesalahan keputusan (Efendi et al., 2022). Tantangan etis ini tidak hanya berdampak pada keakuratan sistem, tetapi juga pada kepercayaan publik dan legitimasi penggunaan AI dalam konteks sosial yang lebih luas.

Urgensi etika AI menjadi semakin krusial seiring dengan pertumbuhan ekosistem digital yang pesat di berbagai belahan dunia. Penerapan AI dalam layanan perbankan, e-government, pendidikan digital, serta industri kesehatan menuntut adanya prinsip-prinsip etis yang universal dan jelas. Namun, beberapa literatur menunjukkan bahwa penerapan pedoman etika AI seringkali masih terbatas dan belum didukung oleh tata kelola implementatif yang menyeluruh (Aulia, 2021). Tantangan ini diperkuat oleh kondisi data global yang beragam dan seringkali tidak terstruktur, yang berpotensi memperkuat bias algoritmik apabila tidak dikendalikan dengan baik (Pabubung, 2024).

Melihat kompleksitas tersebut, penelitian ini melakukan Systematic Literature Review (SLR) untuk menganalisis tren, pola, dan isu utama terkait etika dalam sistem pengambilan keputusan berbasis AI pada periode 2020-2025. Pendekatan SLR memungkinkan peneliti mengidentifikasi relasi antar konsep, pemetaan topik, serta frekuensi kemunculan isu etika seperti keadilan, transparansi, akuntabilitas, keamanan data, dan bias algoritmik secara lebih sistematis. Dengan demikian, penelitian ini dapat memberikan gambaran yang lebih mendalam mengenai perkembangan wacana etika AI dalam literatur ilmiah terkini.

Hasil kajian ini diharapkan dapat memberikan kontribusi bagi peneliti, pengembang teknologi, dan pembuat kebijakan dalam merancang sistem AI yang bertanggung jawab dan selaras dengan nilai-nilai kemanusiaan. Selain itu, temuan penelitian ini dapat menjadi dasar teoretis bagi penyusunan pedoman etis dan kerangka tata kelola AI yang lebih kuat, sehingga pemanfaatan AI dalam pengambilan keputusan dapat berjalan secara adil, transparan, dan akuntabel.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan Systematic Literature Review (SLR), yaitu metode tinjauan literatur yang menggabungkan langkah sistematis dengan analisis semantik untuk mengidentifikasi pola, tren, dan hubungan konsep dalam penelitian terkait etika AI dan sistem pengambilan keputusan. Metode ini dipilih karena mampu memberikan pemetaan topik yang lebih mendalam dibandingkan tinjauan literatur tradisional.



Penelitian ini menggunakan rancangan SLR yang terdiri dari tiga tahap utama:

- (1) Identifikasi (Identification)
Mengumpulkan artikel dari berbagai sumber ilmiah yang relevan.
- (2) Seleksi (Screening & Eligibility)
Memilah artikel berdasarkan kriteria inklusi dan eksklusi.
- (3) Analisis (Semantic Analysis & Synthesis)
Melakukan analisis semantik terhadap kumpulan literatur yang lolos seleksi.

2.1 Research Question (RQ)

Research question adalah pertanyaan mengenai penelitian yang singkat, kompleks, dan terfokus pada topik/permasalahan tertentu. Adapun RQ pada penelitian ini terdapat pada table 1.

RQ	Research Question	Tujuan
RQ1	Bagaimana isu bias algoritmik dibahas dalam sistem pengambilan keputusan berbasis AI?	Mengidentifikasi dan menganalisis cara isu bias algoritmik ditangani dalam pengembangan dan penerapan sistem pengambilan keputusan berbasis AI.
RQ2	Bagaimana pendekatan etika digunakan dalam desain AI?	Menganalisis dan mendeskripsikan berbagai pendekatan etika yang diintegrasikan dalam proses desain, pengembangan, dan implementasi sistem AI.
RQ3	Faktor etika apa yang paling berpengaruh dalam proses pengambilan keputusan AI?	Mengidentifikasi dan mengevaluasi faktor-faktor etika kunci yang memiliki dampak signifikan terhadap proses dan hasil pengambilan keputusan yang dilakukan oleh sistem AI.

Tabel 1. Research Questions

RQ1 membantu dalam menyaring dan mengevaluasi literatur studi utama mengenai deteksi, mitigasi, dan konsekuensi dari bias algoritmik. RQ2 membantu dalam memperoleh informasi mengenai kerangka kerja etika (ethical frameworks) seperti Fairness, Accountability, Transparency (FAT) dan praktik AI by Design. RQ3 memberi pemahaman tentang faktor-faktor etika mana (misalnya: akuntabilitas, transparansi, privasi, keadilan) yang paling kritis dalam memastikan validitas dan penerimaan sistem pengambilan keputusan AI.

2.2 Perencanaan Tinjauan Literatur

Tahap awal SLR dilakukan dengan menentukan kata kunci yang relevan dengan fokus penelitian. Pemilihan kata kunci didasarkan pada penelusuran konsep inti dari tiga aspek utama penelitian, yaitu etika dalam kecerdasan buatan, sistem pengambilan keputusan berbasis AI, dan isu-isu algoritmik seperti bias serta transparansi. Kata kunci kemudian disusun dalam dua bahasa untuk memastikan cakupan literatur yang luas dan representatif.

- “AI ethics”, “ethical AI decision-making”, “algorithmic bias”, “algorithmic accountability”, “fairness in AI”, dan “transparent AI systems”.



Kata kunci tersebut dikombinasikan dengan Boolean Operators seperti AND dan OR untuk memperluas maupun mempersempit hasil pencarian. (“AI Ethics” OR “Ethical Artificial Intelligence”) AND (“Decision-Making System” OR “Decision Support System”) AND (“Algorithmic Bias” OR “Accountability” OR “Transparency”). Proses pencarian literatur dilakukan melalui tools Publish or Perish yaitu *Scopus*, dengan batas waktu publikasi tahun 2020-2025 untuk memastikan bahwa literatur yang dianalisis sesuai dengan perkembangan etika AI dan sistem pengambilan keputusan modern.

2.3 Pelaksanaan Tinjauan

Tahap ini bertujuan untuk menyeleksi, menilai, dan menganalisis literatur yang telah ditemukan agar sesuai dengan fokus penelitian.

a. Kriteria Inklusi dan Eksklusi

Untuk menjamin kualitas dan relevansi literatur yang digunakan, ditetapkan kriteria sebagai berikut:

Kategori	Deskripsi
Kriteria Inklusi	(1) Artikel jurnal ilmiah yang telah peer-reviewed (2) Terbit dalam rentang waktu 2020-2025 (3) Mempunyai fokus pada etika AI, AI governance, atau AI dalam pengambilan keputusan. (4) Ditulis dalam Inggris (5) Tersedia dalam format full-text.
Kriteria Eksklusi	(1) Artikel populer, blog, atau berita, artikel SLR (2) Artikel tanpa sumber sitasi ilmiah yang jelas (3) Hanya membahas aspek teknis pengambilan keputusan berbasis AI tanpa dimensi etika. (4) Publikasi duplikat atau tidak dapat diakses secara penuh.

Tabel 2. Kriteria Inklusi dan Eksklusi

Proses seleksi dilakukan secara sistematis dengan membaca abstrak dan isi artikel untuk memastikan kesesuaian dengan fokus penelitian.

b. Ekstraksi Data Semantik

Pada tahap ini, setiap artikel yang telah memenuhi kriteria inklusi dianalisis dengan mengambil informasi kunci, yaitu abstrak, kata kunci, dan kesimpulan. Data tersebut digunakan sebagai dasar untuk analisis semantik lebih lanjut. Analisis yang dilakukan mencakup:

- 1) Keyword Frequency - menghitung frekuensi kemunculan kata kunci penting untuk mengidentifikasi konsep dan isu yang paling sering dibahas dalam literatur.
- 2) Co-occurrence Network - memetakan hubungan antara kata kunci sehingga terlihat keterkaitan antar-konsep dalam penelitian.
- 3) Semantic Clustering - mengelompokkan artikel atau konsep berdasarkan kemiripan tematik sehingga membentuk klaster yang mewakili subtopik etika AI dalam sistem pengambilan keputusan.

Dalam tahap ini, artikel-artikel yang telah dikelompokkan berdasarkan kemiripan



konsep akan diorganisir ke dalam empat tema utama yang menjadi fokus penelitian: bias dan keadilan algoritmik, transparansi dan explainability, akuntabilitas melalui tata kelola yang kuat, dan privasi data dalam pemrosesan skala besar. Tema-tema ini nantinya akan menjadi kerangka untuk menganalisis pola, keterkaitan konsep, dan isu etika yang muncul dalam literatur.

2.4 Pelaporan Hasil secara Deskriptif dan Tematik

Tahap terakhir dalam penelitian ini adalah menyusun hasil sintesis literatur berbasis analisis semantik dalam bentuk laporan yang sistematis dan mudah dipahami. Pelaporan dilakukan secara naratif dan tematik, dengan mengelompokkan temuan ke dalam klaster dan tema semantik yang telah diidentifikasi sebelumnya, seperti bias algoritmik, transparansi, akuntabilitas, dan privasi data. Untuk mendukung proses ini, digunakan beberapa perangkat bantu, antara lain:

- 1) Scopus / Publish or Perish (PoP) untuk pencarian dan ekstraksi artikel. PoP bisa mengambil data dari Scopus, Google Scholar, atau Semantic Scholar, sehingga mempermudah pengumpulan literatur awal.
- 2) Covidence untuk screening artikel (judul, abstrak, full text) dan manajemen proses seleksi literatur secara sistematis.
- 3) Microsoft Excel / Google Sheets untuk pencatatan data dan pengelompokan tema.

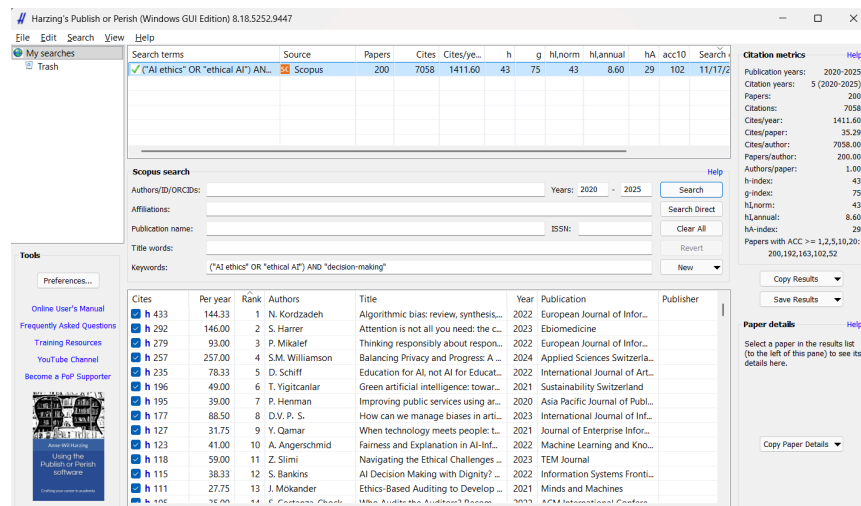
Laporan disusun dengan prinsip transparansi dan replikasi, sehingga dapat menjadi acuan bagi penelitian lanjutan di bidang etika AI dan sistem pengambilan keputusan berbasis AI.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

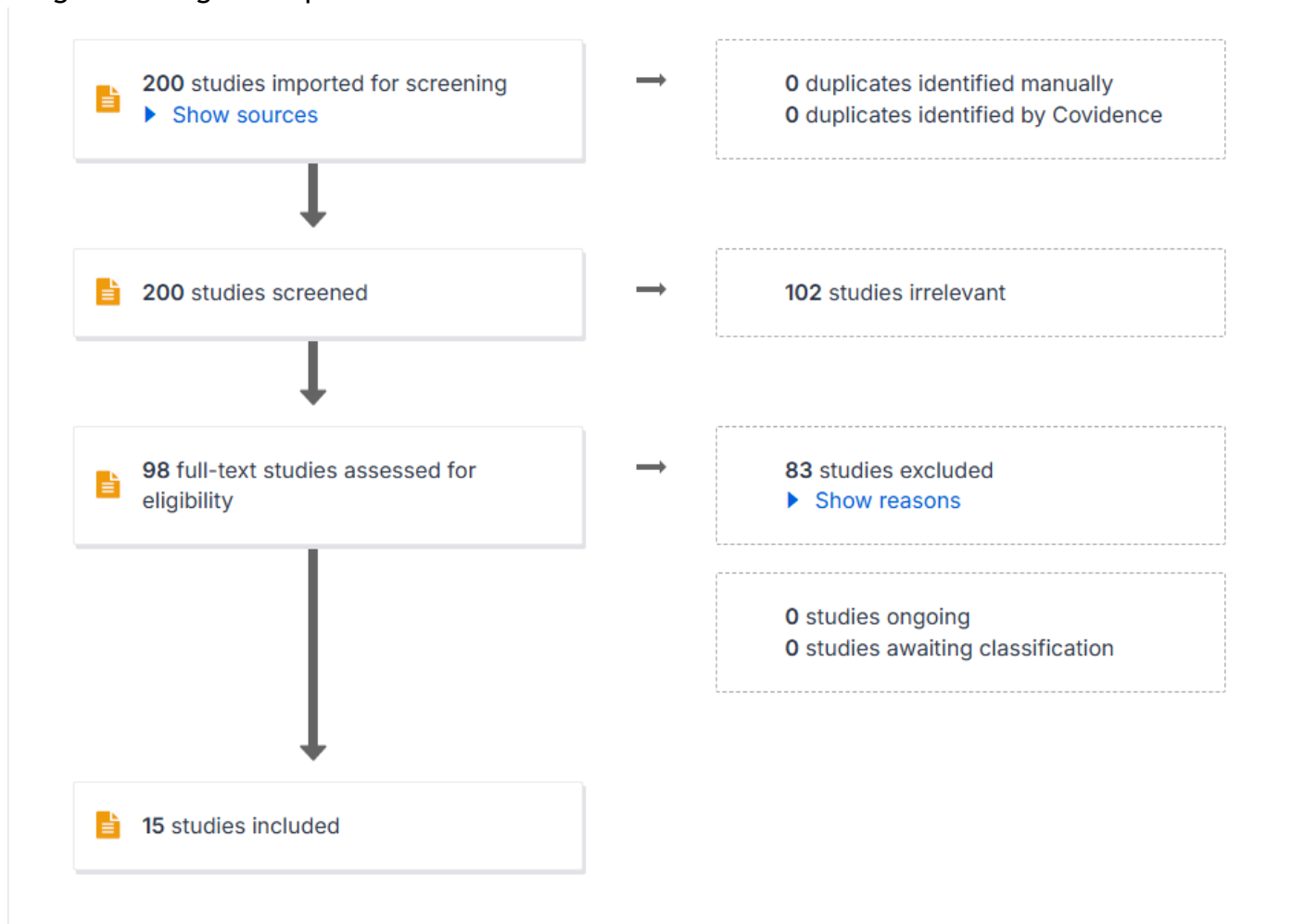
Tahap pertama penelitian dimulai dengan proses pencarian awal menggunakan perangkat Publish or Perish (PoP) untuk mengekstraksi data dari Scopus dan Semantic Scopus. Pencarian dilakukan dengan menggunakan kombinasi kata kunci yang relevan dengan etika kecerdasan buatan dan sistem pengambilan keputusan, sesuai dengan rentang tahun publikasi 2020-2025. Kata kunci tersebut digunakan untuk menghasilkan set artikel awal yang masih bersifat sangat luas dan belum disaring, sehingga seluruh publikasi yang mengandung istilah terkait—baik artikel jurnal, prosiding, maupun laporan ilmiah juga ikut terambil. Proses ini bertujuan memastikan bahwa seluruh literatur potensial dapat terdokumentasi sebelum masuk ke tahap penyaringan berikutnya.

Dari hasil pencarian awal, diperoleh sejumlah artikel yang kemudian direkap dalam bentuk daftar publikasi lengkap melalui PoP, termasuk informasi mengenai judul, penulis, tahun terbit, jurnal, jumlah sitasi, dan relevansi kata kunci. Data awal ini menjadi dasar untuk memulai proses screening secara bertahap pada tahap berikutnya.



Gambar 1. Pencarian Artikel menggunakan Publish or Perish

Hasil pencarian yang ditampilkan pada gambar menunjukkan bahwa Publish or Perish berhasil mengumpulkan 200 artikel yang sesuai dengan kombinasi kata kunci yang digunakan, dengan rentang tahun publikasi 2020-2025.



Gambar 2. Diagram Alir PRISMA Proses Seleksi Literatur

Setelah memperoleh daftar awal 200 artikel, tahap berikutnya adalah melakukan screening judul dan abstrak untuk menilai relevansi setiap publikasi terhadap fokus penelitian. Proses ini dilakukan secara bersamaan menggunakan platform Covidence, yang memfasilitasi penyaringan literatur secara sistematis dan efisien. Dalam tahap ini, setiap artikel dievaluasi



berdasarkan kriteria inklusi dan eksklusi yang telah ditetapkan sebelumnya. Kriteria inklusi mencakup artikel yang membahas etika kecerdasan buatan, fairness, explainability, bias, dan penerapannya dalam sistem pengambilan keputusan, sedangkan kriteria eksklusi meliputi artikel duplikat, publikasi non-ilmiah, atau artikel yang topiknya tidak relevan dengan ruang lingkup penelitian.

Setiap judul dan abstrak diperiksa secara independen, dan artikel yang memenuhi kriteria langsung ditandai untuk lanjut ke tahap screening full text. Melalui proses ini, jumlah artikel awal berhasil disaring menjadi subset yang lebih fokus dan relevan, sehingga memudahkan langkah analisis mendalam selanjutnya, termasuk identifikasi artikel yang layak dianalisis secara semantik.

Setelah tahap screening judul dan abstrak selesai, artikel yang lolos kemudian masuk ke tahap screening full text. Pada tahap ini, setiap artikel dibaca secara menyeluruh untuk menilai kesesuaian dan kedalaman konten terhadap fokus penelitian. Proses ini dilakukan satu per satu agar setiap publikasi dapat dievaluasi dengan cermat, termasuk metodologi yang digunakan, cakupan topik, relevansi data, dan kontribusi terhadap literatur mengenai etika kecerdasan buatan serta sistem pengambilan keputusan.

Dari proses ini, artikel yang memenuhi kriteria ilmiah dan relevansi tematik dipilih untuk dianalisis lebih lanjut. Hasilnya, sebanyak 15 jurnal/artikel berhasil ditetapkan sebagai literatur yang layak untuk dimasukkan dalam tahap analisis semantik. Pemilihan ini memastikan bahwa dataset penelitian memiliki kualitas tinggi dan mencakup variasi perspektif yang memadai, sehingga hasil analisis berikutnya dapat menggambarkan tren, isu, dan temuan penting dalam literatur terkait.

Setelah tahap screening full text selesai, 15 artikel layak dianalisis lebih lanjut menggunakan analisis semantik. Analisis ini bertujuan untuk mengidentifikasi tema, konsep, dan pola utama dalam literatur yang relevan dengan etika kecerdasan buatan dan sistem pengambilan keputusan. Setiap artikel diperiksa untuk mengidentifikasi:

- 1) Topik utama : fokus penelitian dan isu yang dibahas.
- 2) Metodologi : pendekatan yang digunakan dalam penelitian.
- 3) Temuan utama : hasil penting yang terkait dengan tema penelitian.
- 4) Kontribusi terhadap literatur : bagaimana artikel menambah wawasan atau perspektif baru.
- 5) RQ : pertanyaan penelitian yang dijawab atau relevan dengan artikel tersebut.

No	Judul Artikel	Topik Utama	Metodologi	Temuan Utama	Kontribusi Terhadap Literatur	RQ
1	Thinking responsibly about responsible AI and 'the dark side' of AI	Analisis sisi gelap AI (seperti bias algoritmik) dan pendekatan etika melalui kerangka "AI yang Bertanggung Jawab" (Responsible AI) untuk mengatasinya.	Editorial / Tinjauan Konseptual yang merangkum isu dan mengajukan agenda penelitian.	Bias Algoritmik: Isu utama dalam AI yang dapat menimbulkan diskriminasi dan ketidakadilan (misal: rekrutmen di Amazon & Google).	Menyediakan kerangka terstruktur untuk menganalisis dan memitigasi isu etika seperti bias algoritmik, sekaligus mengajukan pertanyaan penelitian empiris yang spesifik untuk bidang ini.	RQ1
2	Fairness and Explanation in AI-Informed Decision Making	Investigasi empiris tentang efek penjelasan AI (XAI) dan pernyataan	Studi dengan 25 partisipan mensimulasikan pengambilan keputusan AI	Kombinasi antara Penjelasan dan Fairness. Dalam kondisi fairness rendah,	Memberikan bukti empiris tentang hubungan timbal balik dan efek gabungan antara	RQ1, RQ2



		fairness terhadap kepercayaan pengguna dan persepsi fairness dalam pengambilan keputusan berbasis AI.	dalam skenario asuransi kesehatan dan perawatan medis, dengan variasi jenis penjelasan (fitur, contoh, atau tanpa penjelasan) dan tingkat fairness (tinggi, rendah, atau tanpa informasi).	penjelasan berbasis fitur paling efektif meningkatkan kepercayaan dan persepsi fairness. Skenario aplikasi juga mempengaruhi respons pengguna.	penjelasan AI (XAI) dan fairness dalam membentuk kepercayaan dan persepsi fairness pengguna, dengan implikasi praktis untuk merancang sistem AI yang lebih bertanggung jawab.	
3	AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context	Membandingkan persepsi keadilan interaksional (perlakuan yang hormat dan bermartabat) pekerja terhadap keputusan yang dibuat oleh AI versus manusia dalam konteks Manajemen Sumber Daya Manusia (SDM).	Studi survei eksperimental dengan pertanyaan kualitatif terbuka, memanipulasi pembuat keputusan (AI vs. manusia) dan valensi keputusan (positif vs. negatif) di enam fungsi SDM, dianalisis menggunakan MANOVA, machine learning, dan analisis tematik kualitatif.	Keputusan positif dari AI lebih diterima daripada keputusan negatif dari manusia. Namun, "Bias Hasil" (Outcome Bias) kuat—orang menilai keputusan lebih pada hasilnya (positif/negatif) daripada pada proses atau siapa pembuatnya.	Menantang narasi dominan tentang bias AI dengan menunjukkan bahwa dalam konteks SDM, kekhawatiran tentang bias manusia tetap sangat signifikan.	RQ3
4	Ethics-Based Auditing to Develop Trustworthy AI	Audit etika AI untuk menjembatani prinsip dan praktik dalam tata kelola AI.	Kajian konseptual dengan analisis praktik terbaik dan identifikasi kendala.	Audit etika dapat tingkatkan kualitas keputusan, transparansi, dan akuntabilitas; perlu pendekatan berkelanjutan, holistik, dan terintegrasi dengan kebijakan.	Memberikan kerangka konkret dan roadmap penelitian untuk implementasi audit etika AI, sekaligus mengidentifikasi 16 kendala yang perlu diatasi.	RQ1, RQ2
5	Capable but Amoral? Comparing AI and Human Expert Collaboration in Ethical Decision Making	Persepsi manusia terhadap kolaborasi dengan AI dalam pengambilan keputusan etika, khususnya membandingkan kepercayaan, atribusi tanggung jawab, dan ketergantungan antara ahli manusia dan AI.	Eksperimen simulasi immersive (Unity) dengan desain 2x2, melibatkan 428 partisipan sebagai operator drone yang membuat keputusan etika dengan nasihat ahli (manusia vs. AI) pada dua tingkat otonomi.	AI dipercaya lebih tinggi dalam hal kapasitas, tetapi ahli manusia dipercaya lebih tinggi dalam hal moral.	Memberikan bukti empiris bahwa persepsi manusia terhadap AI dalam keputusan etika bersifat kompleks dan multidimensi (capable but amoral). Menyoroti pentingnya desain kolaborasi manusia-AI yang mempertimbangkan pembagian tanggung jawab dan kalibrasi kepercayaan yang	RQ3



					tepat.	
6	Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions	Faktor-faktor yang memengaruhi persepsi legitimasi keputusan yang dibuat oleh sistem algoritmik dalam konteks bisnis.	Survei faktorial vignette dengan sembilan kondisi yang melibatkan 4.727 responden dari Amazon Mechanical Turk. Vignette memvariasikan jenis keputusan, hasil (outcome), faktor input, dan tata kelola untuk mengukur persepsi legitimasi.	• Kepentingan keputusan: Keputusan penting (rekrutmen) dinilai kurang legitim jika dibuat algoritma. • Hasil (outcome): Hasil baik meningkatkan legitimasi; hasil buruk mengurangi legitimasi. • Faktor input: Input diskriminatif (ras) sangat mengurangi legitimasi, bahkan jika hasil baik atau ada prosedur banding. • Tata kelola: Hanya mekanisme banding ke manusia yang secara signifikan meningkatkan legitimasi, khususnya untuk hasil buruk. Pemberitahuan atau audit kurang efektif.	• Teori Legitimasi: Menguji interaksi faktor keputusan, hasil, input, dan tata kelola dalam konteks algoritma • Etika AI & Bisnis: Menghubungkan etika AI dengan legitimasi organisasi; menekankan bahwa input tidak etis merusak legitimasi secara fundamental. • Kebijakan Publik: Bukti bahwa regulasi prosedural (misal GDPR) mungkin tidak cukup jika algoritma menggunakan faktor diskriminatif.	RQ1, RQ3
7	Ethical Challenges in the Development of Virtual Assistants Powered by Large Language Models	Analisis tantangan etika dan rekomendasi pengembangan Asisten Virtual (VA) yang didukung Large Language Model (LLM), khususnya dalam konteks layanan publik.	Analisis regulasi AI di Eropa dan studi kasus dari proyek VA penulis (NETA, GiDi, XIA).	Integrasi LLM pada VA menimbulkan risiko etika seperti privasi data, bias, misinformasi (hallucinations), dan hilangnya otonomi pengguna. Regulasi Eropa saat ini dinilai belum cukup spesifik.	Menyediakan pedoman etika yang spesifik dan dapat ditindaklanjuti untuk pengembangan VA bertenaga LLM, melengkapi kerangka regulasi yang ada dengan fokus pada auditabilitas, inklusi, dan mitigasi bias.	RQ2



8	Beyond Discrimination : Generative AI Applications and Ethical Challenges in Forensic Psychiatry	Aplikasi dan tantangan etika Kecerdasan Buatan Generatif (GenAI) dalam psikiatri forensik dan peradilan pidana.	Analisis konseptual dengan membandingkan AI Generatif dan Diskriminatif, serta tinjauan literatur tentang tantangan etika yang muncul.	GenAI dapat mentransformasi praktik forensik (seperti asesmen risiko, pelatihan simulasi) tetapi memperburuk bias, menimbulkan tantangan baru seperti "hallucinations", deepfake, dan ketergantungan berlebihan.	Memberikan analisis komprehensif pertama yang memetakan aplikasi transformatif GenAI dan spektrum lengkap tantangan etiknya dalam konteks psikiatri forensik, melampaui isu bias yang sudah dikenal.	RQ1
9	Moral distance, AI, and the ethics of care	Analisis bagaimana AI meningkatkan "jarak moral" dalam pengambilan keputusan dan usulan etika kepedulian (ethics of care) sebagai kerangka solusi.	Analisis filosofis konseptual dengan mengidentifikasi dua jenis jarak moral (proximity dan bureaucratic) yang diperparah oleh AI.	AI meningkatkan jarak moral melalui penghapusan interaksi tatap muka (jarak fisik, temporal, budaya) dan proses birokrasi yang kompleks, yang mengurangi empati dan tanggung jawab.	Memperkenalkan etika kepedulian sebagai kerangka kerja etika yang komplementer untuk AI, yang menekankan hubungan interdependen, konteks, kerentanan, dan suara pihak terdampak.	RQ3
10	Moral Decision Making in Human-Agent Teams: Human Control & Explanations	Bagaimana manusia bisa tetap memegang kendali bermakna (Meaningful Human Control) saat bekerja sama dengan AI dalam tugas yang melibatkan keputusan moral, seperti triase medis.	Uji tiga pola kerja tim manusia-AI (Team Design Patterns): • AI cuma kasih saran, manusia putuskan. • AI pilih tugas mana yang bermoral untuk manusia, sisanya AI kerjakan. • AI putuskan semua, manusia hanya awasi. Simulasi triase pasien saat pandemi, dengan dokter/ahli medis sebagai partisipan. Data dikumpulkan lewat think-aloud, wawancara, dan kuesioner.	Pola 1 paling disukai manusia merasa paling berkuasa. Pola 3 paling tidak disukai manusia merasa tidak bertanggung jawab. Penjelasan dari AI dianggap berguna, tapi jarang dibaca karena waktu terbatas. Kontrol terasa kuat jika efeknya langsung terlihat. Kalau terlambat (misal lewat setting awal), manusia tidak merasa mengendalikan.	Menawarkan pola desain tim manusia-AI untuk tugas moral. Menekankan bahwa explainable AI harus mempertimbangkan kapan dan bagaimana penjelasan diberikan. Simulasi triase bisa dipakai lagi untuk penelitian sejenis.	RQ2
11	Incompleteness of moral choice and	Tantangan filosofis dan metodologis	Analisis filosofis dan konseptual terhadap	Pemikiran etika AI saat ini terjebak dalam pendekatan	Mengkritik landasan filosofis etika AI yang ada dan	RQ2



	evolution towards fully autonomous AI	dalam menciptakan Agen Moral Buatan (Artificial Moral Agents/AMAs) yang sepenuhnya otonom, dengan fokus pada sistem senjata otonom.	pendekatan normatif dalam etika AI, dengan mengkritik fondasi teori moral tradisional (seperti deontologi Kantian).	normatif yang tidak praktis ("ambulans moral Kant"). Pilihan moral adalah fenomena psikologis dan sosial yang membutuhkan solusi empiris, bukan hanya prinsip abstrak.	menganjurkan pergeseran dari pendekatan normatif murni ke arah penelitian empiris dan naturalistik tentang pengambilan keputusan moral untuk AI.	
12	Ethical machine decisions and the input-selection problem	Masalah "pemilihan input" dalam pengambilan keputusan mesin yang etis, yang melibatkan ketidakpastian faktual dan trade-off yang diperlukan untuk menguranginya.	Analisis filosofis dengan menggunakan contoh-contoh pemikiran (thought experiments), terutama dari konteks kendaraan otonom, untuk menguraikan masalah konseptual.	Etika keputusan mesin adalah masalah tiga sisi: apa yang harus dilakukan mesin, ketidakpastian keputusan apa yang dapat diterima, dan trade-off apa yang dapat diterima untuk mengurangi ketidakpastian tersebut. Mengurangi ketidakpastian dengan menambah input data menimbulkan trade-off terhadap privasi, transparansi, dan waktu respons.	Memperkenalkan dan memformalkan "masalah pemilihan input" sebagai tantangan sentral dalam etika AI, yang mengharuskan analisis normatif terhadap trade-off antara akurasi keputusan dan nilai-nilai seperti privasi dan transparansi.	RQ3
13	Embedding Fear in Medical AI: A Risk-Averse Framework for Safety and Ethics	Kerangka kerja untuk menanamkan mekanisme "takut" atau kehati-hatian dalam Kecerdasan Buatan (AI) medis guna meningkatkan keselamatan pasien dan mengurangi risiko.	Konseptual dan interdisipliner, menggabungkan teknik komputasi seperti Bayesian risk assessment, reinforcement learning dengan penalti, dan pemodelan ketidakpastian, serta analogi dengan sistem biologis.	Diusulkan sebuah arsitektur modul "takut" yang dapat diintegrasikan ke dalam AI medis untuk mendeteksi risiko, menghentikan tindakan otonom, dan mengalihkannya ke pengawasan manusia ketika ketidakpastian atau potensi bahaya tinggi.	Memperkenalkan konsep "kehati-hatian bawaan" (embedded caution) sebagai lapisan keamanan baru untuk AI medis berisiko tinggi, menjembatani literatur tentang etika AI, keselamatan pasien, dan sistem otonom.	RQ2
14	The AI ethics of digital COVID-19 diagnosis and their legal,	Kerangka kerja etika AI untuk diagnosis COVID-19 digital yang	Konseptual, dengan menganalisis pedoman kebijakan etika AI	Diidentifikasi 9 persyaratan etika inti (seperti Akuntabilitas, Otonomi,	Menyediakan kerangka kerja etika yang dapat ditindaklanjuti dan spesifik untuk	RQ2



	medical, technological, and operational managerial implications	menjembatani kesenjangan antara penelitian dan penerapan di rumah sakit, dengan mempertimbangkan perspektif interdisipliner.	yang ada (WHO, UE, Jerman) dan memetakannya ke dalam proses pengembangan sistem (pengumpulan data, pelatihan algoritma, pengembangan aplikasi, penggunaan klinis).	Transparansi) dan dampaknya bagi masing-masing pemangku kepentingan (Teknologi, Hukum, Kedokteran, Manajemen Operasional) dalam pengembangan alat diagnosis.	penggunaan kasus diagnosis COVID-19, menekankan pendekatan interdisipliner untuk mengatasi skeptisisme dan tantangan implementasi AI di bidang kesehatan.	
15	Supporting Ethical Decision-Making for Lethal Autonomous Weapons	Pengembangan model untuk mendukung dan mengkalibrasi kepercayaan manusia dalam pengambilan keputusan etis oleh Senjata Otonom Mematikan (LAWS).	Penelitian ini menggunakan model keputusan etis Bayesian yang mengintegrasikan Hukum Perang dan faktor taktis untuk menghasilkan skor etis "strike/no-strike". Metode ini diuji melalui studi kalibrasi dengan operator dan evaluasi antarmuka pengguna bersama para ahli (SME).	Model Bayesian dapat menghasilkan rekomendasi yang selaras dengan keputusan operator dalam skenario yang jelas, tetapi terdapat ketidakselarasan dalam skenario berisiko rendah di mana operator cenderung lebih permisif. Antarmuka model dinilai sangat berguna untuk meningkatkan kesadaran situasi etis.	Menawarkan metodologi baru yang komputasional dan transparan untuk mengoperasikan prinsip-prinsip etis dalam LAWS, menjembatani kesenjangan antara teori etika dan penerapan praktis dalam operasi militer.	RQ3

Tabel 3. Ekstraksi Data

3.2 Pembahasan

Berdasarkan hasil ekstraksi data pada Tabel 3, bagian ini akan mendiskusikan sintesis temuan untuk menjawab ketiga pertanyaan penelitian (RQ) dan relevansinya dengan konteks regulasi di Indonesia.

1.) Mitigasi Bias dalam Pengambilan Keputusan Otomatis (RQ1)

Analisis terhadap literatur 2020-2025 menunjukkan bahwa penanganan bias algoritmik telah bergeser dari sekadar identifikasi masalah menuju solusi mitigasi yang bersifat teknis-sosiologis. Temuan dari artikel ke-1 dan ke-3 mengonfirmasi bahwa bias sering kali muncul dari data historis yang diskriminatif, namun solusinya kini lebih kompleks.

Strategi mitigasi utama yang ditemukan meliputi:

1. Integrasi Explainability dan Fairness: Artikel ke-2 menunjukkan temuan empiris bahwa transparansi melalui Explainable AI (XAI) saja tidak cukup. Dalam kondisi di mana keadilan (fairness) sistem rendah, penjelasan fitur justru dapat menurunkan kepercayaan pengguna jika tidak disertai perbaikan pada akurasi keadilan itu sendiri.
2. Audit Berkelanjutan: Artikel ke-4 menekankan urgensi Ethics-Based Auditing (Audit Berbasis Etika). Audit tidak boleh hanya dilakukan di akhir (post-hoc), melainkan harus



terintegrasi sepanjang siklus hidup pengembangan AI untuk menjembatani kesenjangan antara prinsip etika abstrak dan praktik tata kelola.

3. Masalah Seleksi Input: Artikel ke-12 menyoroiti tantangan teknis input-selection problem. Upaya mengurangi ketidakpastian keputusan dengan menambah data input sering kali memunculkan trade-off etika baru, seperti pelanggaran privasi dan penurunan transparansi.

2.) Evolusi Pendekatan Desain: Dari Normatif ke AI by Design (RQ2)

Terdapat pergeseran paradigma yang signifikan dalam cara sistem etis didesain. Artikel ke-11 mengkritik pendekatan tradisional yang bersifat normatif ("ambulans moral") karena dianggap tidak praktis untuk menangani otonomi sistem modern. Sebaliknya, literatur merekomendasikan pendekatan AI by Design yang lebih empiris dan tertanam dalam arsitektur sistem.

Dua pendekatan desain inovatif yang teridentifikasi adalah:

1. Mekanisme Kehati-hatian Bawaan (Embedded Caution): Artikel ke-13 memperkenalkan konsep menanamkan "rasa takut" (fear) atau kehati-hatian ke dalam algoritma medis. Sistem didesain untuk mendeteksi ambang batas risiko dan secara otomatis menghentikan tindakan otonom atau mengalihkan kendali ke manusia saat ketidakpastian tinggi.
2. Kontrol Manusia yang Bermakna (Meaningful Human Control): Dalam desain tim manusia-agen (artikel ke-10), pola desain yang paling etis adalah yang menempatkan manusia sebagai otoritas moral terakhir, bukan sekadar pengawas pasif. Hal ini penting untuk menjaga rasa tanggung jawab (sense of agency) pada pengguna manusia.

3.) Faktor Penentu Kepercayaan: Transparansi dan Jarak Moral (RQ3)

Faktor etika yang paling krusial dalam penerimaan sistem pengambilan keputusan bukan semata-mata akurasi, melainkan bagaimana sistem mengatasi bias psikologis pengguna.

- Fenomena Outcome Bias: Temuan dari artikel ke-3 dan ke-6 mengungkapkan bahwa pengguna cenderung menilai legitimasi keputusan AI berdasarkan hasil akhirnya (positif/negatif), bukan berdasarkan keadilan prosesnya. Hal ini menunjukkan bahwa transparansi proses sangat vital untuk mencegah kepercayaan buta terhadap hasil yang menguntungkan namun bias.
- Jarak Moral (Moral Distance): Artikel ke-9 mengidentifikasi bahwa mediasi teknologi dalam birokrasi menciptakan "jarak moral", di mana pengambil keputusan kehilangan empati karena tidak berhadapan langsung dengan subjek yang terdampak. Oleh karena itu, faktor etika yang penting adalah kemampuan sistem untuk memfasilitasi "etika kepedulian" (ethics of care), bukan hanya efisiensi.
- Kalibrasi Kepercayaan: Pada sistem berisiko ekstrem seperti senjata otonom (Artikel 15), model probabilistik (Bayesian) terbukti efektif membantu operator mengkalibrasi kepercayaan mereka agar keputusan tetap selaras dengan hukum humaniter.

4.) Implikasi bagi Indonesia:

Harmonisasi dengan UU PDP Temuan-temuan global ini memiliki relevansi strategis bagi Indonesia, khususnya dalam mendukung implementasi Undang-Undang No. 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP).

Konsep Embedded Caution dan Auditability yang dibahas dalam artikel ke-4 dan ke-13



sangat relevan untuk diadopsi sebagai standar teknis bagi Pengendali Data Pribadi dalam memenuhi kewajiban keamanan pemrosesan data. Selain itu, temuan mengenai pentingnya penjelasan (explanation) dalam artikel ke-2 mendukung hak subjek data di Indonesia untuk mendapatkan informasi terkait mekanisme pemprofilan otomatis (automated decision making). Dengan mengadopsi kerangka kerja AI by Design yang teruji secara global, pengembangan sistem AI di Indonesia dapat lebih siap menghadapi tantangan regulasi dan etika lokal.

4. SIMPULAN

Penelitian ini telah melaksanakan Systematic Literature Review (SLR) terhadap literatur ilmiah internasional bereputasi untuk memetakan perkembangan etika dalam sistem pengambilan keputusan berbasis Kecerdasan Buatan (AI) periode 2020-2025. Berdasarkan sintesis terhadap studi-studi global tersebut, penelitian ini menghasilkan tiga simpulan utama yang menjawab tujuan penelitian.

1. Literatur global menyoroti bias algoritmik sebagai tantangan etika universal yang paling mendesak. Studi menunjukkan bahwa bias merupakan risiko sosial yang memperburuk diskriminasi, yang kini dimitigasi secara teknis melalui pendekatan hibrida: integrasi Explainable AI (XAI), audit algoritma berkelanjutan, dan jaminan keadilan (fairness) untuk memulihkan kepercayaan publik.
2. Terdapat pergeseran paradigma desain di tingkat internasional, dari sekadar pedoman normatif menuju kerangka kerja praktis. Penerapan AI by Design yang mengintegrasikan privasi, kontrol manusia bermakna (meaningful human control), dan mekanisme "kehati-hatian bawaan" (embedded caution) telah menjadi standar baru dalam pengembangan sistem cerdas yang bertanggung jawab.
3. Faktor penentu utama dalam keberhasilan implementasi keputusan AI adalah keseimbangan antara transparansi dan akuntabilitas, mengingat adanya fenomena psikologis seperti Outcome Bias dan Moral Distance yang mempengaruhi legitimasi interaksi manusia-mesin.

Meskipun berbasis pada data global, temuan ini memberikan implikasi strategis bagi pengembangan ekosistem digital di Indonesia. Hasil studi ini berfungsi sebagai rujukan bagi pembuat kebijakan untuk mengharmonisasikan standar teknis global dengan implementasi UU Perlindungan Data Pribadi (UU PDP), khususnya terkait hak penjelasan (explainability) dan keamanan data. Bagi pengembang teknologi dan akademisi di Indonesia, temuan ini menekankan perlunya mengadopsi kerangka kerja etika praktis yang telah teruji secara internasional untuk mencegah risiko kegagalan sistem pada sektor krusial.

Penelitian ini memiliki keterbatasan karena berfokus pada literatur internasional berbahasa Inggris, sehingga mungkin belum menangkap nuansa tantangan spesifik lokal secara mendalam. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan studi komparasi antara kerangka etika global ini dengan realitas implementasi di Indonesia, guna menguji efektivitas prinsip-prinsip tersebut dalam konteks budaya dan regulasi lokal yang beragam.

DAFTAR PUSTAKA

- Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and Explanation in AI-Informed Decision Making. *Machine Learning and Knowledge Extraction*, 4(2), 556-579. <https://doi.org/10.3390/make402026>
- Aulia, S. R.,. (2024). PERTIMBANGAN ETIKA DALAM KECERDASAN BUATAN MEMASTIKAN PENGEMBANGAN YANG BERTANGGUNG JAWAB. *GuardKom: Jurnal Ilmiah Teknologi*



- Informasi Dan Komputer, 1(1), 11-20. Retrieved from <https://jurnal.stmikiba.ac.id/index.php/GuardKom/article/view/98>
- Bankins, S., Formosa, P., Griep, Y., & others. (2022). AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, 24, 857-875. <https://doi.org/10.1007/s10796-021-10223-8>
- Benanti, P., Kõiv, H., Benen, A., da Silveira, M. B., & Chaniotis, N. (2024). Is moral AI necessary or impossible? Comments on moral domain and competence from multiple academic perspectives. *AI & Society*. Advance online publication. <https://doi.org/10.1016/j.aisoc.2024.100115>
- Efendi, Z., Hanim, M. A. F., & Santoso, A. (2025). KECERDASAN BUATAN (AI) DALAM PENDIDIKAN: TINJAUAN LITERATUR SISTEMATIS TENTANG PELUANG, MASALAH ETIKA, DAN IMPLIKASI PEDAGOGIS. *Jurnal Pendidikan, Kebudayaan Dan Keislaman*, 4(3), 134-152. <https://doi.org/10.24260/jpkk.v4i3.5052>
- Harahap, I. P., Busyro, M., & Faisal, M. (2025). Analisis Etika dan Legalitas Penggunaan Artificial Intelligence (AI) dalam Pengambilan Keputusan Hukum. *Jurnal Pendidikan Tambusai*, 9(1), 6196-6204. <https://doi.org/10.31004/jptam.v9i1.25455>
- Hauer, T. (2022). Incompleteness of moral choice and evolution towards fully autonomous AI. *Humanities and Social Sciences Communications*, 9, 38. <https://doi.org/10.1057/s41599-022-01060-4>
- Kitchenham, B., Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report EBSE-2007-01, Version 2.3). Keele University and University of Durham.
- Kohn, S., Cohen, M., Johnson, A., Terman, M., Weltman, G., & Lyons, J. (2024). Supporting Ethical Decision-Making for Lethal Autonomous Weapons. *Journal of Military Ethics*, 23(1), 12-31. <https://doi.org/10.1080/15027570.2024.2366094>
- Lundgren, B. (2021). Ethical machine decisions and the input-selection problem. *Synthese*, 199, 11423-11443. <https://doi.org/10.1007/s11229-021-03296-0>
- Martin, K., & Waldman, A. (2023). Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. *Journal of Business Ethics*, 183, 653-670. <https://doi.org/10.1007/s10551-021-05032-7>
- Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, 31(3), 257-268. <https://doi.org/10.1080/0960085X.2022.2026621>
- Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds & Machines*, 31, 323-327. <https://doi.org/10.1007/s11023-021-09557-8>
- Pabubung, M. R. (2024). Era kecerdasan buatan dan tantangan etika terhadap martabat manusia. *Jurnal Filsafat Indonesia*. Link: <https://ejournal.undiksha.ac.id/index.php/JFI/article/view/49293>
- Piñeiro-Martín, A., García-Mateo, C., Docío-Fernández, L., & López-Pérez, M. d. C. (2023). Ethical Challenges in the Development of Virtual Assistants Powered by Large Language Models. *Electronics*, 12(14), 3170. <https://doi.org/10.3390/electronics12143170>
- Prasetya, R. N., Kusdiyanto, A., Radiana, U., & Wicaksono, L. (2024). Etika dalam Pengembangan Artificial Intelligence: Tinjauan Pedoman dan Penerapannya. *Juwara: Jurnal Wawasan Dan Aksara*, 4(1), 240-253. <https://doi.org/10.58740/juwara.v4i1.271>
- Thurzo, A., & Thurzo, V. (2025). Embedding Fear in Medical AI: A Risk-Averse Framework for Safety and Ethics. *AI*, 6(5), 101. <https://doi.org/10.3390/ai6050101>
- Tolmeijer, S., Christen, M., Kandul, S., Kneer, M., & Bernstein, A. (2022). Capable but amoral? Comparing AI and human expert collaboration in ethical decision making. *CHI Conference on Human Factors in Computing Systems (CHI '22)*, 1-17. <https://doi.org/10.1145/3491102.3517732>



- Tortora, L. (2024). Beyond discrimination: Generative AI applications and ethical challenges in forensic psychiatry. *Frontiers in Psychiatry*, 15, 1346059.
<https://doi.org/10.3389/fpsy.2024.1346059>
- Van der Waa, J., Verdult, S., van den Bosch, K., van Diggelen, J., Haije, T., van der Stigchel, B., & Cocu, I. (2021). Moral decision making in human-agent teams: Human control and the role of explanations. *Frontiers in Robotics and AI*, 8, 640647.
<https://doi.org/10.3389/frobt.2021.640647>
- Villegas-Galaviz, C., & Martin, K. (2024). Moral distance, AI, and the ethics of care. *AI & Society*, 39, 1695-1706. <https://doi.org/10.1007/s00146-023-01642-z>