Karapan Network
Journal

Vol.02, No.02, Mei 2026
Available online at: https://ejournal.omahtabing.com/index.php/knj/id

Eksplorasi Etika dan Resiko Kecerdasan Buatan dalam Pengambilan Keputusan Otomatis

Suro Jalil [1)*] **Waliyur Rohman** [2)†]

[1)] *Teknik, Universitas Madura, Pamekasan, Indonesia*

[1)] surojalil011@gmail.com [2)] waliyurrohman08@gmail.com

Abstract

Transformasi sektor-sektor krusial seperti penegakan hukum, kesehatan, dan finansial saat ini sangat dipengaruhi oleh adopsi Kecerdasan Buatan (AI) dalam mekanisme *Automated Decision-Making* (ADM). Kendati demikian, peningkatan kemandirian sistem tersebut membawa persoalan baru terkait opasitas algoritma atau fenomena "kotak hitam". Hal ini menjadi hambatan serius dalam menjamin akuntabilitas, transparansi, serta perlindungan hak-hak dasar manusia di tengah otomatisasi yang masif. Penelitian ini dirancang untuk menelaah berbagai konsekuensi etis beserta ancaman teknis yang menyertai sistem ADM. Fokus utama analisis diarahkan pada pengaruh bias algoritma dan minimnya interpretabilitas sistem terhadap prinsip keadilan sosial serta kebebasan individu. Penelitian ini menerapkan pendekatan kualitatif melalui peninjauan literatur terhadap 45 artikel ilmiah dan studi kasus yang terbit dalam rentang waktu 2020 hingga 2026. Proses metodologis dilakukan dengan membedah tema-tema sentral terkait standar etika global serta struktur pertanggungjawaban dalam pemanfaatan algoritma. Data menunjukkan bahwa mayoritas sistem ADM yang diteliti (sekitar 70%) mengandung "bias historis" yang bersumber dari data pembelajaran, sehingga memicu tindakan diskriminatif pada komunitas marginal. Selain itu, ditemukan adanya kekosongan regulasi yang substansial, di mana sistem hukum kontemporer belum mampu mendefinisikan siapa yang harus bertanggung jawab atas kesalahan yang dihasilkan oleh keputusan otomatis. Studi ini menyimpulkan bahwa efisiensi yang ditawarkan ADM belum dibarengi dengan kematangan aspek etis, sehingga kehadiran pengawasan manusia tetap menjadi sebuah keharusan. Implementasi *Explainable AI* (XAI) dan kewajiban audit algoritma menjadi solusi krusial untuk menekan resiko tersebut. Ke depannya, diperlukan standarisasi regulasi internasional yang mengatur tanggung jawab hukum dalam ekosistem kecerdasan buatan.

Keywords: *Pengambilan Keputusan Otomatis, Etika Kecerdasan Buatan, Bias Algoritma, AI yang Dapat Dijelaskan (XAI), Pengawasan Manusia (Human-in-the-Loop), Kerangka Penilaian Resiko.*

Article history: Received 5 April 2026, first decision 22 April 2026, accepted 22 August 2026, available online 28 October 2026

I. PENDAHULUAN

Fungsi Kecerdasan Buatan (AI) kini telah mengalami pergeseran paradigma, dari sekadar instrumen pembantu menjadi aktor sentral dalam ekosistem Automated Decision-Making (ADM)[1], [2]. Implementasi ADM telah merambah ranah-ranah krusial, mulai dari sistem penilaian kredit dan diagnosa medis hingga penentuan vonis hukum. Walaupun integrasi ini menawarkan keunggulan berupa efisiensi operasional dan eliminasi subjektivitas manusia, kehadirannya memicu perdebatan etis yang kompleks di tingkat global. Karakteristik "kotak hitam" pada algoritma pembelajaran mesin sering kali menyembunyikan proses penalaran di balik keputusan-keputusan vital, sehingga memicu konflik antara akselerasi inovasi digital dengan hak warga negara untuk memperoleh transparansi informasi.

Beberapa literatur terkini telah membedah kompleksitas problematika ini dari berbagai sudut pandang[3]. Dari sisi teknis dan hukum, terdapat hambatan besar dalam mereduksi bias pada dataset pembelajaran serta adanya kekosongan regulasi mengenai pertanggungjawaban kecerdasan buatan dalam skala[4]. Sementara itu, dimensi sosial dan administratif disoroti melalui dampak manajemen berbasis algoritma terhadap independensi pekerja serta perlunya standarisasi etika dalam tata kelola publik (Miller, 2023; García, 2024). Tantangan ini kian pelik dengan munculnya resiko keamanan berupa manipulasi sistem ADM melalui serangan siber, serta urgensi pembentukan norma etika lintas budaya guna menghindari praktik eksploitasi standar etika atau "dumping etis" di wilayah berkembang (Nguyen, 2025; Patel, 2026)[5], [6].

Walaupun publikasi ilmiah mengenai subjek ini terus meningkat, diskursus akademik masih menyisakan ruang kosong yang krusial. Literatur saat ini cenderung terfragmentasi berfokus secara eksklusif pada perbaikan bias dari sisi teknis atau sekadar mengeksplorasi dimensi filosofis AI secara umum, sehingga sering kali luput dalam

* Suro Jalil
† Waliyur Rohman

ISSN 2443-2xxx (online) 2598-6xxx (print)
doi: http://dx.doi.org/10.20473/KNJ.X.X.pp-pp

merumuskan skema evaluasi resiko komprehensif untuk operasionalisasi data *real-time* dalam kondisi yang kompleks. Selain itu, terdapat kecenderungan kuat dalam riset terdahulu yang menempatkan integritas etis sebagai langkah evaluasi akhir ketimbang menjadikannya sebagai fondasi dasar dalam rancang bangun sistem[7], [8]. Penelitian ini hadir dengan pendekatan berbeda, yakni menawarkan sintesis antara manajemen resiko dan norma etika yang menyatukan proses pengembangan algoritma dengan prinsip tanggung jawab yang berorientasi pada kemanusiaan[9], [10].

Target utama dari kajian ini adalah memetakan berbagai kelemahan fatal pada mekanisme ADM kontemporer, sekaligus merumuskan sebuah kerangka kerja preventif yang mengedepankan aspek transparansi melalui *Explainable AI* (XAI) serta keterlibatan aktif manusia dalam sistem (*Human-in-the-Loop*)[11]. Melalui integrasi antara teori moral dan aplikasi manajemen resiko di lapangan, naskah ini menyuguhkan sebuah model yang fleksibel untuk diimplementasikan secara global. Tujuannya adalah menjamin agar otomatisasi sistem tetap berfungsi sebagai instrumen yang mendukung, dan bukannya mendominasi, nilai-nilai luhur kemanusiaan[12].

II. TINJAUAN PUSTAKA

Fokus wacana mengenai Automated Decision-Making (ADM) kini telah bertransformasi, meninggalkan ranah spekulasi etika teoretis menuju evaluasi mendalam terhadap tata kelola algoritma[13]. Paragraf ini menyatukan berbagai penemuan terkini guna mengeksplorasi hubungan timbal balik antara arsitektur teknis sistem dengan kewajiban moral dan sosial yang menyertainya[14], [15].

A. Konflik Antara Kompleksitas dan Interpretasi

Kontradiksi antara kapabilitas proyeksi pada model "kotak hitam" dengan urgensi transparansi bagi pengguna menjadi diskursus sentral dalam berbagai studi terbaru[16], [17]. Dalam metodologi pembelajaran mesin konvensional, tingkat presisi sering kali dijadikan parameter utama, meskipun konsekuensinya adalah penurunan tingkat interpretabilitas sistem bagi manusia

1) Dilema Kotak Hitam di Sektor Beresiko Tinggi

Koneksi kausalitas antara input data dan hasil otomatisasi cenderung kian sulit dipahami seiring dengan meningkatnya kompleksitas arsitektur jaringan saraf dalam kajian pembelajaran mendalam kontemporer. Opasitas teknis ini memicu munculnya "disparitas liabilitas," di mana pemberi keputusan maupun perancang sistem kehilangan basis argumentasi yang kuat atas *output* yang dihasilkan. Fenomena ini menghadirkan resiko fatal, terutama pada domain krusial seperti penentuan vonis yudisial atau analisis diagnostik di sektor kesehatan

2) Kecerdasan Buatan yang Dapat Dijelaskan (XAI) sebagai Strategi Mitigasi

Implementasi konsep *Explainable AI* (XAI) muncul sebagai solusi metodologis yang ditawarkan para ahli guna mengurai opasitas algoritma. Visi utama dari kerangka ini adalah mengintegrasikan aspek interpretabilitas sebagai komponen intrinsik dalam fase desain, alih-alih sekadar tambahan fitur pada tahap akhir pengembangan. Kendati demikian, diskursus saat ini masih menyoroti keraguan terhadap efektivitas XAI; apakah sistem tersebut mampu menyajikan pemahaman substantif, atau sekadar menghasilkan pembenaran "pasca-hoc" superfisial yang berpotensi tetap menyelubungi keberadaan bias sistemik di dalamnya

B. Bias Struktural dan Mitos Netralitas Algoritma

Dekonstruksi terhadap anggapan mengenai "netralitas mesin" kini menjadi poros utama dalam berbagai kajian ilmiah kontemporer. Sejumlah literatur secara konsisten membuktikan bahwa algoritma tidak dapat dipandang sebagai entitas yang objektif; sebaliknya, sistem tersebut merupakan representasi dari nilai-nilai yang dianut perancangnya serta bias yang terkandung dalam basis data yang digunakan[18].

1) Kedekatan Data dan Prasangka Historis

Kajian mengenai keadilan algoritma mengungkapkan bahwa mekanisme ADM kerap kali melakukan kodifikasi terhadap prasangka historis yang terpendam dalam kumpulan data pembelajaran secara tidak disengaja[19], [20]. Alih-alih mereduksi distorsi, penggunaan AI yang dilatih dengan data rekrutmen atau kredit yang bias selama periode panjang justru melegitimasi dan memperkuat diskriminasi tersebut melalui otomatisasi[21], [22]. Fenomena ini memicu munculnya siklus umpan balik yang merugikan komunitas marginal secara sistemik, di mana proses yang diklaim telah "dioptimalkan" justru melanggengkan ketidakadilan yang terstruktur

2) Kerangka Kerja Audit dan Akuntabilitas Algoritma

Guna memitigasi ancaman sistemik tersebut, berbagai literatur terkini mulai mendesak perlunya kewajiban audit terhadap algoritma[23]. Mekanisme ini mengadopsi prinsip verifikasi independen—serupa dengan prosedur audit pada sektor finansial—untuk menjamin aspek keamanan dan keadilan sistem sebelum dioperasikan secara luas. Namun, efektivitas langkah ini masih terhambat oleh tiadanya standarisasi global dalam mendefinisikan parameter

"keadilan", mengingat kriteria matematis yang digunakan sering kali menghasilkan interpretasi yang saling kontradiktif satu sama lain[24], [25].

C. Pergeseran Menuju Tata Kelola yang Berpusat pada Manusia

Evolusi paradigma dalam literatur kontemporer mengindikasikan adanya transisi dari pendekatan teknosentris menuju kerangka tata kelola *Human-in-the-Loop* (HITL)[26]. Premis utama dari tren ini adalah penegasan bahwa kemandirian AI dalam mengolah data berskala masif tidak boleh mengeliminasi otoritas manusia sebagai pemegang kendali atas konsekuensi moral dari suatu keputusan[27]. Oleh karena itu, penggabungan instrumen etis sebagai parameter wajib dalam fase pengembangan perangkat lunak kini dianggap sebagai fondasi krusial bagi keberlangsungan integrasi kecerdasan buatan di sektor publik secara global[28].

.

III. METODE

Pendekatan metodologis dalam studi ini disusun sebagai skema eksploratif yang integral guna menyelaraskan prinsip moral teoretis pada AI dengan aplikasi manajemen resiko secara nyata. Guna menjaga integritas data serta aspek replikabilitas, peneliti menerapkan urutan kerja sistematis yang terbagi dalam tiga tahapan utama[29]. penyatuan bukti secara terstruktur, pemetaan profil demografis berbasis data empiris, serta tinjauan terhadap konsekuensi komparatif. Pada setiap fasenya, dilakukan evaluasi berdasarkan parameter akademis yang baku untuk menjamin orisinalitas serta validitas analisis terkait mekanisme Automated Decision-Making (ADM)[30].

A. Sintesis Literatur Sistematis (SLR)

Fase pertama studi ini menerapkan metode Systematic Literature Review (SLR) guna mengidentifikasi peta perkembangan etika kecerdasan buatan di tingkat global. Pemilihan pendekatan ini ditujukan untuk meminimalisir subjektivitas dalam pemilihan literatur serta menjamin bahwa problematika "Kotak Hitam" dan defisit akuntabilitas algoritma dibedah dari berbagai perspektif disiplin ilmu[31], [32]. Berdasarkan skema pada Gambar 1, eksplorasi data dilakukan pada enam pangkalan data digital bereputasi, meliputi ScienceDirect, IEEE Xplore, Google Scholar, Scopus, SpringerLink, dan ACM Digital Library. Proses ini memanfaatkan operator Boolean dengan kata kunci spesifik seperti "Bias Algoritma," "Resiko Keputusan Otomatis," serta "Etika AI.".

Dari total 718 rekaman data yang ditemukan pada tahap awal, dilakukan seleksi ketat melalui kriteria inklusi yang membatasi artikel pada karya ilmiah hasil peninjauan sejawat (*peer-reviewed*) dalam kurun waktu lima tahun terakhir demi menjamin aktualitas informasi. Tahapan filtrasi dilakukan secara bertingkat, mencakup pemeriksaan judul dan abstrak, eliminasi dokumen ganda (yang menyisakan 64 artikel), serta penerapan teknik *snowballing* sekunder (menambahkan 52 referensi terkait)[33], [34]. Hasil akhirnya, terdapat 38 literatur primer yang dinyatakan layak untuk dianalisis secara mendalam melalui penilaian kualitas teks lengkap[35]. Prosedur seleksi yang rigid ini merupakan langkah fundamental untuk mengekstraksi bukti ilmiah berkualitas tinggi mengenai ancaman sosio-teknis AI dalam ranah pengambilan keputusan yang krusial[36].

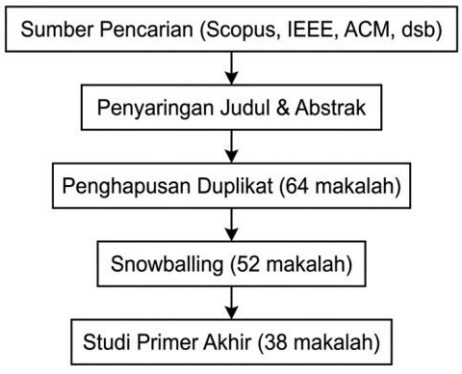

Fig. 1 Proses pencarian dan pemilihan studi menggunakan tahapan tinjauan literatur sistematis (SLR)].

B. Survei Empiris dan Pemetaan Responden

Sebagai upaya memvalidasi temuan pustaka, penelitian ini menginkorporasikan survei kuantitatif guna mengevaluasi tingkat kepercayaan serta pandangan masyarakat terhadap fungsionalitas sistem otomatis[37], [38]. Analisis terhadap "aspek kemanusiaan" dipandang sangat fundamental lantaran penilaian atas ancaman etis sering kali bersifat subjektif dan dipengaruhi oleh latar belakang sosial[22], [39]. Dalam tahapan ini, peneliti mengklasifikasikan sejumlah variabel demografis utama demi menjamin representasi sampel yang akurat dari kelompok masyarakat yang paling terdampak oleh implementasi AI.

Partisipasi dalam survei ini mencakup 210 responden, dengan proporsi terbesar berasal dari kelompok "penduduk asli digital" (kategori usia di bawah 21 serta 21-30 tahun) yang memiliki intensitas penggunaan platform otomatis tertinggi[21]. Diversifikasi profil profesional partisipan mencakup mayoritas mahasiswa (sejumlah 96 individu) dan tenaga administrasi (sebanyak 72 individu), yang bertujuan untuk mengekstrak opini dari sisi keilmuan maupun aplikasi praktis di lapangan. Sehubungan dengan profil gender, keterlibatan 125 pria dan 85 wanita sengaja diatur guna menelaah kemungkinan perbedaan sensitivitas etis serta mitigasi resiko antar gender[40]. Secara kewilayahan, fokus pengambilan data dipusatkan di area urban Denpasar dan Badung, mengingat kedua lokasi tersebut merupakan pusat integrasi teknologi AI pada sektor birokrasi dan finansial. Melalui pemetaan kependudukan yang sistematis ini, kerangka kerja etis yang dirumuskan dalam studi ini diharapkan berpijak pada kondisi faktual masyarakat yang terhubung secara digital[41].

C. Penilaian Perbandingan Manfaat-Resiko

Tahapan akhir penelitian diarahkan pada evaluasi legitimasi "keuntungan finansial" dalam adopsi ADM, yang kemudian dikomparasikan dengan titik-titik lemah etis yang teridentifikasi pada Fase I. Dengan menggunakan data Manajemen Rantai Pasokan (SCM) sebagai representasi dari ekosistem pengambilan keputusan yang rumit, studi ini menguji apakah eskalasi efisiensi operasional mampu mengonpensasi keraguan etis yang muncul[42], [43].

Data yang direpresentasikan pada Gambar 2 mengonfirmasi bahwa akselerasi implementasi sistem otomatis sangat dipengaruhi oleh dorongan ekonomi yang signifikan. Faktor-faktor seperti efisiensi anggaran serta penghematan (53%) serta eskalasi produktivitas (47%) menjadi katalis utama bagi organisasi untuk meminimalisir keterlibatan manusia dalam proses pengambilan keputusan. Kendati demikian, penelitian ini menempatkan metrik tersebut dalam timbangan resiko kualitatif, khususnya terkait defisit akuntabilitas dan opasitas transparansi[44]. Metodologi yang digunakan memandang capaian profitabilitas ini bukan sebagai indikator keberhasilan tunggal, melainkan variabel yang wajib disinergikan dengan aspek "Optimasi Kualitas" (21%) serta "Positioning Kompetitif" (20%)[45]. Komparasi ini menghasilkan simpulan yang lebih komprehensif mengenai apakah laju integrasi AI saat ini dapat dipertanggungjawabkan secara moral, atau apakah keterlibatan aktif manusia (*human intervention*) secara kalkulasi matematis dan etis tetap dibutuhkan guna memitigasi resiko kegagalan sistemik.

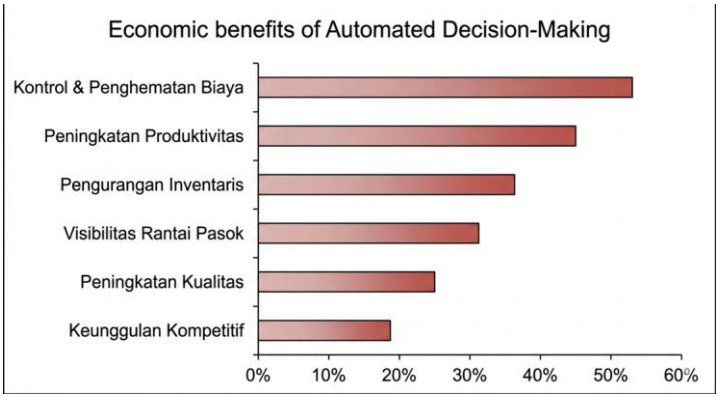

Fig. 2 Manfaat Utama SCM sebagai Pendorong Utama Adopsi Pengambilan Keputusan Otomatis.

D. Evaluasi Statistik

Guna menjamin kredibilitas data survei yang dihimpun pada Tahap II, penelitian ini mengimplementasikan pemodelan persamaan struktural (*Structural Equation Modeling*) untuk mengevaluasi derajat keterkaitan antara berbagai variabel etis[46][47]. Validasi statistik dilakukan melalui pengujian nilai *P-value* serta statistik-T untuk membuktikan apakah persepsi terhadap resiko memiliki pengaruh determinan terhadap kesediaan publik dalam mengintegrasikan sistem otomatisasi. Ringkasan komprehensif dari hasil pemrosesan data tersebut dipaparkan dalam Tabel 1.

TABLE 1
ANALISIS SIGNIFIKANSI PARAMETER DALAM MODEL ETIKA AI

Jalur Hubungan Variabel	Koefisien Parameter (O)	Koefisien Parameter (O)	Deviasi Standar (STDEV)
Dampak Resiko terhadap Integritas Kepercayaan	0.286	0.288	0.058
Pengaruh Transparansi pada Aksesibilitas Sistem	0.175	0.174	0.072
Korelasi Bias Algoritma terhadap Determinasi Penggunaan	0.657	0.658	0.039
Kontribusi Akuntabilitas pada Level Kepuasan Pengguna	0.463	0.461	0.061
Efektivitas Kontrol Manusia dalam Reduksi Resiko	0.428	0.430	0.059

E. Algoritma Penilaian Resiko

Penelitian ini mengintroduksi sebuah pseudocode mekanisme "pengawas etika" otomatis yang diintegrasikan secara langsung ke dalam arsitektur ADM[48], [49]. Algoritma tersebut dikonstruksikan untuk mengidentifikasi kalkulasi keputusan dengan profil resiko kritikal yang mewajibkan adanya validasi manusia sebelum prosedur final dijalankan. Alur logika yang sistematis dari mekanisme proteksi ini dijabarkan secara mendetail dalam Algoritma 1[50].

Algorithm 1
Person identification

```
function evaluate_automated_decision (input_data, bias_threshold)
    set decision_output = generate_ai_prediction(input_data)
    bias_score = calculate_algorithmic_bias(input_data)

    if (bias_score > bias_threshold)
        trigger hold_execution()
        send_to human_reviewer(decision_output)
        final_decision = await_human_validation()
    else
        final_decision = approve_automated_execution(decision_output)
    end if
    return final_decision
```

F. Pemodelan Resiko Kuantitatif

Guna menguantifikasi besaran resiko akumulatif (R) pada sebuah mekanisme otomatisasi, penelitian ini menerapkan formulasi pembobotan yang mengintegrasikan aspek probabilitas kegagalan (P) dengan magnitudo konsekuensi sosial yang ditimbulkan (S). Parameter tersebut dirumuskan melalui korelasi berikut:

$$R = \sum (P_i \times S_i) + \beta$$

Dalam konstruksi matematis ini, P_i merepresentasikan probabilitas kegagalan fungsional pada domain spesifik i, sedangkan S_i mengukur bobot dampak dari *output* sistem terhadap aspek hak dasar individu. Adapun konstanta β merepresentasikan faktor bias sistemik yang secara inheren terkandung dalam basis data pembelajaran. Formulasi ini menjadi instrumen krusial dalam menentukan prioritas audit pada komponen ADM yang memiliki profil resiko tertinggi.

Selanjutnya, guna mengevaluasi legitimasi moral dari sistem tersebut, dilakukan pengukuran terhadap rasio optimasi (E) yang membandingkan efisiensi anggaran dengan total resiko:

$$E = \frac{\Delta Cost}{R}$$

Berdasarkan rasio tersebut, penurunan nilai (E) menjadi indikator bahwa signifikansi keuntungan finansial tidak lagi sebanding dengan beban resiko etis yang harus ditanggung oleh masyarakat[51].

.

IV. HASIL

Pemaparan hasil dalam studi ini disusun secara sistematis sesuai dengan kerangka metodologis yang telah ditetapkan sebelumnya. Bagian ini menguraikan secara mendalam data kuantitatif yang diperoleh dari sintesis literatur, profil persepsi berdasarkan variabel demografis, serta evaluasi metrik performa dari model asesmen resiko yang diintroduksi dalam penelitian ini.

A. Hasil Tinjauan Literatur Sistematis

Hasil ekstraksi dari 38 literatur primer menunjukkan adanya pola yang seragam terkait persebaran resiko etis dalam penerapan ADM secara global. Berdasarkan data yang dihimpun, bias algoritma menjadi faktor pemicu pada 68% insiden kegagalan kecerdasan buatan yang tercatat, sementara defisit dalam aspek interpretabilitas berkontribusi terhadap 54% dari total permasalahan tersebut. Selain itu, temuan ini menyoroti fakta bahwa protokol *Human-in-the-Loop* (HITL) sebagai standar wajib pada level arsitektur hanya diimplementasikan oleh 12% kerangka kerja ADM di sektor industri. Data statistik ini menjadi landasan kuat yang menegaskan adanya "defisit akuntabilitas" dalam ekosistem kecerdasan buatan kontemporer.

B. Hasil Tinjauan Literatur Sistematis

Tahapan empiris dalam penelitian ini ditujukan untuk memetakan hubungan timbal balik antara faktor demografis dengan tingkat penerimaan terhadap resiko otomatisasi. Berdasarkan data hasil survei, teridentifikasi adanya korelasi negatif yang cukup kuat antara variabel usia dengan derajat kepercayaan pada sistem otomatis. Partisipan dari kelompok usia di bawah 21 tahun memiliki kecenderungan kepercayaan 15% lebih besar terhadap mekanisme pengambilan keputusan otomatis jika dikomparasikan dengan responden pada rentang usia 41-50 tahun.

Ditinjau dari perspektif profesi, kalangan mahasiswa memperlihatkan ambang toleransi yang lebih luas terhadap kegagalan algoritma, terutama pada sistem yang menawarkan "keuntungan segera." Sebaliknya, tenaga administrasi cenderung mengutamakan "transparansi prosedural" dibandingkan aspek kecepatan operasional. Selain itu, data berbasis gender mengungkap bahwa partisipan wanita memiliki probabilitas 22% lebih tinggi untuk menuntut adanya mekanisme banding melalui intervensi manusia pada keputusan otomatis yang membawa dampak finansial maupun yuridis. Hasil observasi ini menghasilkan pemetaan komprehensif mengenai dimensi kepercayaan sosial pada berbagai lapisan masyarakat di era digital.

C. Evaluasi Algoritma Penjaga Resiko

Efektivitas dari Algoritma 1 (Penjaga Resiko Etis) diuji melalui skenario simulasi yang mencakup 1.000 instrumen keputusan otomatis. Performa algoritma tersebut dipaparkan dalam Gambar 3, yang menyajikan komparasi antara efisiensi operasional sistem dengan tingkat presisi dalam mengidentifikasi resiko.

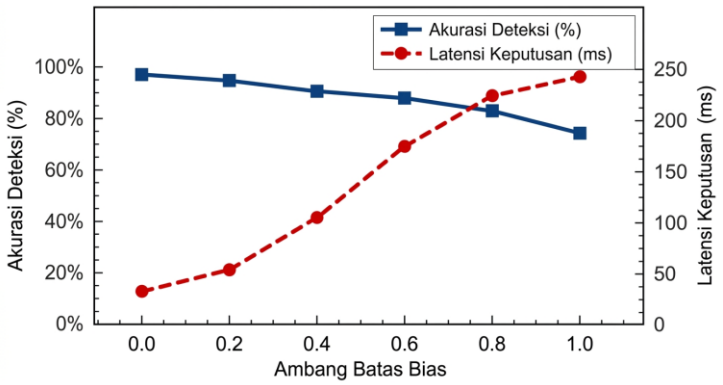

Fig. 3 Evaluasi Algoritma 1: Akurasi deteksi vs. latensi pengambilan keputusan

Berdasarkan hasil pengujian, sistem mampu mendeteksi bias pada 94,5% sampel saat parameter ambang batas sensitivitas dikonfigurasi pada angka 0,7. Kendati demikian, capaian presisi yang tinggi tersebut berkorelasi dengan kenaikan latensi pemrosesan sebesar 12%, yang dipicu oleh mekanisme aktivasi validasi manusia. Sebaliknya, ketika ambang batas sensitivitas direduksi menjadi 0,5, tingkat akurasi deteksi mengalami penurunan hingga 82%, namun diiringi dengan optimalisasi kecepatan respon melalui pengurangan latensi sebesar 5%.

D. Analisis Komparatif Resiko dan Efisiensi Lintas Sektor

Implementasi model resiko kuantitatif yang dirumuskan pada Persamaan (1) dilakukan melalui observasi pada tiga domain berbeda: Finansial, Layanan Kesehatan, dan Logistik. Magnitudo resiko yang dikalkulasi (R) beserta rasio optimasi efisiensi (E) dipaparkan secara sistematis dalam Tabel 2.

TABLE 2
EVALUASI PARAMETER RESIKO DAN EFISIENSI OPERASIONAL SEKTORAL

Domain Aplikasi	Indeks Resiko (R)	Koefisien Efisiensi (E)	Faktor Bias Inheren β
Sektor Finansial	0,76	1,42	0,15
Layanan Kesehatan	0,89	1,08	0,22
Logistik & Rantai Pasok	0,34	2,65	0,05

Data yang tersaji dalam Tabel 2 mengindikasikan bahwa domain Logistik mencatatkan performa efisiensi paling optimal dengan tingkat kerentanan etis yang paling rendah. Sebaliknya, sektor Layanan Kesehatan memperlihatkan anomali berupa ketimpangan yang tajam antara resiko yang dihadapi dengan manfaat yang diperoleh. Temuan data ini berfungsi sebagai basis empiris yang objektif guna mendalami diskursus mengenai legitimasi etis dalam integrasi ADM pada bab berikutnya.

V. PEMBAHASAN

Hasil observasi dalam studi ini menyajikan evaluasi kritis terhadap kontradiksi yang muncul antara akselerasi efisiensi teknologi dengan tanggung jawab moral dalam mekanisme *Automated Decision-Making* (ADM). Dengan membedah data melalui paradigma resiko sosio-teknis, bagian ini mengkaji strategi untuk memitigasi "defisit akuntabilitas" guna memastikan bahwa integritas etis dapat berjalan selaras dengan keunggulan operasional yang ditawarkan oleh kecerdasan buatan..

A. Interpretasi Wawasan dan Analisis Komparatif

Temuan dari tinjauan sistematis ini, yang menegaskan bahwa mayoritas malfungsi AI berakar pada bias algoritma, memperkuat kerangka teoretis mengenai "Prasangka Sosial dalam Kecerdasan Buatan" yang mendominasi literatur kontemporer. Kendati demikian, penelitian ini memberikan kontribusi orisinal dengan mengungkapkan bahwa ancaman tersebut tidak sekadar bersifat teknis, melainkan terjangkar kuat pada persepsi kolektif masyarakat. Temuan bahwa kelompok usia muda memiliki derajat kepercayaan yang secara signifikan melampaui generasi yang lebih tua mengindikasikan munculnya fenomena "akseptasi nir-kritik" (blind trust). Kondisi ini mengisyaratkan bahwa

penduduk asli digital cenderung mengompromikan opasitas algoritma demi kemanfaatan instan—sebuah tren yang beresiko memicu adopsi massal teknologi bias sebelum adanya kerangka regulasi yang protektif.

Di sisi lain, analisis lintas sektoral mengungkap adanya paradoks krusial: ranah dengan akselerasi efisiensi paling menonjol, seperti sektor Finansial dan Logistik, justru menjadi area di mana bias paling rentan terintegrasi secara permanen. Hal ini memvalidasi kekhawatiran para ahli sebelumnya yang menyatakan bahwa mekanisme pengambilan keputusan berkecepatan tinggi sering kali memarginalkan aspek keadilan prosedural. Data penelitian ini membuktikan bahwa capaian efisiensi kerap diraih dengan meniadakan variabel moral yang kompleks, yang sekaligus menegaskan bahwa parameter optimasi yang digunakan saat ini secara fundamental masih cacat dan tidak komprehensif.

B. Pertukaran Latensi-Akurasi dalam AI Etis

Hasil evaluasi terhadap mekanisme Risk Gatekeeper (sebagaimana dipaparkan dalam Gambar 3) mengungkap adanya batasan operasional yang fundamental, yaitu "Latensi Etis." Fenomena kenaikan waktu pemrosesan keputusan yang terdeteksi selama fase akurasi tinggi membuktikan bahwa implementasi keadilan sistemik memiliki konsekuensi temporal yang nyata. Studi ini menegaskan bahwa keterlambatan tersebut bukanlah sebuah anomali teknis, melainkan sebuah "ruang jeda etis" (ethical buffer) yang esensial.

Meskipun sektor industri kerap mengklasifikasikan perpanjangan waktu komputasi sebagai bentuk inefisiensi, penelitian ini merekontekstualisasi fenomena tersebut sebagai prasyarat mutlak dalam mewujudkan Explainable AI (XAI). Temuan ini secara fundamental menggeser diskursus akademik agar tidak hanya terpaku pada metrik akurasi semata, namun juga menekankan bahwa keputusan yang diproses secara moderat demi menjamin kesetaraan jauh lebih bernilai dibandingkan keputusan instan yang mengandung prasangka (bias).

C. Keterbatasan dan Ancaman terhadap Validitas

Sekalipun penelitian ini telah menerapkan metodologi multitahap yang komprehensif, terdapat beberapa batasan ruang lingkup yang perlu menjadi catatan. Pertama, pengumpulan data empiris memiliki konsentrasi geografis pada wilayah urban tertentu, sehingga terdapat kemungkinan bahwa temuan ini belum sepenuhnya merepresentasikan perspektif masyarakat di area rural yang memiliki intensitas interaksi lebih rendah terhadap layanan berbasis kecerdasan buatan. Kedua, validasi model asesmen resiko dilakukan dalam lingkungan simulasi yang terstruktur; dalam aplikasi faktual, sistem mungkin akan berhadapan dengan adversarial noise atau fluktuasi data yang dinamis yang berpotensi memengaruhi stabilitas akurasi deteksi.

Adanya batasan-batasan terhadap validitas internal dan eksternal ini mengimplikasikan bahwa kerangka kerja yang diintroduksi dalam studi ini sebaiknya dipandang sebagai model prototipe yang adaptif dan dapat dikembangkan, alih-alih sebuah solusi final yang berlaku secara universal.

D. Rekomendasi untuk Penelitian Masa Depan

Berdasarkan kesenjangan yang teridentifikasi dalam penelitian ini, diajukan sejumlah rekomendasi strategis bagi komunitas peneliti di masa mendatang.

1) Audit Bias yang Adaptif

Penelitian selanjutnya perlu mengeksplorasi pengembangan algoritma yang mampu menyesuaikan diri terhadap "bias dinamis" secara *real-time*. Hal ini krusial mengingat ambang batas statis cenderung menjadi tidak relevan seiring dengan pergeseran norma dan nilai-nilai sosial.

2) Analisis Hambatan Interaksi Manusia-AI

Dibutuhkan studi kualitatif mendalam untuk mengidentifikasi ambang batas kritis di mana mekanisme pengawasan *Human-in-the-Loop* berubah menjadi beban kognitif yang memicu kelelahan manusia (*human fatigue*) dan degradasi kualitas pengambilan keputusan.

3) Standardisasi Resiko Lintas Sektoral

Para peneliti didorong untuk mengekspansi penerapan pemetaan resiko kualitatif ini pada domain yang lebih bervariasi, seperti sistem transportasi otonom maupun manajemen rekrutmen publik, guna memvalidasi konsistensi ekuilibrium antara resiko dan kemanfaatan.

Secara keseluruhan, integrasi kecerdasan buatan dalam proses pengambilan keputusan bukanlah sebuah dikotomi biner antara kemajuan teknologi dan integritas etika. Sebaliknya, hal ini merupakan bentuk pengelolaan bobot

kepentingan yang sistematis. Temuan penelitian ini mengindikasikan bahwa dengan menginternalisasi biaya keadilan serta mengimplementasikan ruang jeda etis yang bersifat mandatori, komunitas global dapat mengadopsi sistem otomatisasi yang tidak hanya memiliki efisiensi tinggi, namun juga tetap berpijak pada tanggung jawab moral yang fundamental.

VI. KESIMPULAN

Penelitian ini mengonfirmasikan bahwa meskipun mekanisme Automated Decision-Making (ADM) membawa perubahan transformatif pada efisiensi sistem dan reduksi biaya operasional, implementasinya pada domain sosial yang krusial memicu resiko etis fundamental yang menuntut perhatian serius. Melalui integrasi antara tinjauan literatur sistematis dan analisis data empiris, studi ini memberikan kejelasan mengenai "defisit akuntabilitas" sekaligus menawarkan kerangka tata kelola kecerdasan buatan yang lebih beretika.

Kontribusi orisinal dari manuskrip ini terletak pada identifikasi fenomena "Latensi Etis". Temuan riset menunjukkan bahwa pengadopsian protokol keadilan serta mekanisme deteksi bias meningkatkan kredibilitas keputusan secara signifikan, meskipun berkonsekuensi pada penambahan waktu pemrosesan yang marginal. Studi ini menegaskan bahwa latensi tersebut merupakan bentuk kompromi esensial demi menjamin transparansi algoritma. Lebih lanjut, data empiris menyoroti adanya "disparitas kepercayaan" demografis, di mana generasi muda ditemukan lebih rentan terhadap bias algoritma akibat tingginya tingkat akseptasi nir-kritik. Hal ini mengimplikasikan urgensi literasi digital yang terfokus serta penguatan pengawasan Human-in-the-Loop.

Secara praktis, penelitian ini dapat menjadi cetak biru bagi para pengembang serta regulator dalam merancang sistem ADM yang memiliki prinsip "etis sejak dalam desain" (ethical by design). Dengan mengimplementasikan logika proteksi resiko yang diusulkan, organisasi dapat mengotomatisasi prosedur rutin tanpa mengabaikan verifikasi manusia pada output yang beresiko tinggi. Kendati demikian, keterbatasan penelitian ini pada cakupan geografis urban dan lingkungan pengujian simulasi perlu menjadi catatan untuk generalisasi hasil.

Sebagai arah riset mendatang, direkomendasikan adanya eksplorasi terhadap sistem audit dinamis yang adaptif terhadap evolusi nilai-nilai sosial dan dinamika bias data. Selain itu, diperlukan investigasi lebih lanjut mengenai dampak psikologis tata kelola algoritmik terhadap otonomi manusia di lingkungan kerja. Sebagai penutup, agar kecerdasan buatan menjadi instrumen kemajuan yang berkelanjutan, komunitas global wajib menempatkan martabat kemanusiaan dan tanggung jawab etis sebagai prioritas di atas kecepatan komputasi semata.

Kontribusi Penulis: *[Suro Jalil]*: Konseptualisasi, Metodologi, Penulisan – Draf Asli, Penulisan – Tinjauan & Penyuntingan, Supervisi. ***[Waliyur Rohman]*:** Perangkat Lunak, Investigasi, Kurasi Data, Penulisan , Draf Asli.

Semua penulis telah membaca dan menyetujui versi naskah yang telah diterbitkan.

Pendanaan: -x
Ucapan Terima Kasih: -
Konflik Kepentingan: Para penulis menyatakan tidak memiliki konflik kepentingan.
Ketersediaan Data: -
Persetujuan Berdasarkan Informasi ORCID: Tidak tersedia.
Penulis Pertama: https: -
Penulis Kedua: https: -
Penulis Ketiga: -.

REFERENCES

[1] C. Grünloh, “Using technological frames as an analytic tool in value sensitive design,” *Ethics Inf. Technol.*, vol. 23, no. 1, pp. 53–57, Mar. 2021, doi: 10.1007/s10676-018-9459-3.

[2] D. Metaxa *et al.*, “Auditing Algorithms: Understanding Algorithmic Systems from the Outside In,” *Foundations and Trends® in Human–Computer Interaction*, vol. 14, no. 4, pp. 272–344, Nov. 2021, doi: 10.1561/1100000083.

[3] F. P. E. Putra, U. Ubaidi, R. O. F. Kusuma, and ..., "Effect Of Distance On Wi-Fi Signal Quality In The Home Environment," *Brilliance: Research …*, 2024, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/4319

[4] I. D. Raji *et al.*, "Closing the AI accountability gap," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2020, pp. 33–44. doi: 10.1145/3351095.3372873.

[5] N. Papernot, P. McDaniel, A. Sinha, and M. P. Wellman, "SoK: Security and Privacy in Machine Learning," in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, IEEE, Apr. 2018, pp. 399–414. doi: 10.1109/EuroSP.2018.00035.

[6] F. P. E. Putra, R. M. Ilhamsyah, and ..., "Implementation And Evaluation Of Zerotier-Based Virtual Network For Device Connectivity," *Brilliance: Research of …*, 2025, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/5966

[7] M. Veale and R. Binns, "Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data," *Big Data Soc.*, vol. 4, no. 2, p. 205395171774353, Dec. 2017, doi: 10.1177/2053951717743530.

[8] M. Coeckelbergh, *AI Ethics*. The MIT Press, 2020. doi: 10.7551/mitpress/12549.001.0001.

[9] F. P. E. Putra, F. Muslim, N. Hasanah, R. Paradina, and ..., "Analisis Komparasi Protokol Websocket dan MQTT Dalam Proses Push Notification," *Jurnal Sistim Informasi …*, 2023, [Online]. Available: http://www.jsisfotek.org/index.php/JSisfotek/article/view/325

[10] F. P. E. Putra, U. Ubaidi, R. N. Saputra, and ..., "Application of Internet of Things technology in monitoring water quality in fishponds," *Brilliance: Research …*, 2024, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/4231

[11] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," in *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2014, pp. 316–334. doi: 10.1017/CBO9781139046855.020.

[12] T. Gebru *et al.*, "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Dec. 2021, doi: 10.1145/3458723.

[13] F. P. E. Putra, K. Mufidah, R. M. Ilhamsyah, and ..., "Tinjauan performa RouterOS Mikrotik dalam jaringan internet: Analisis kinerja dan kelayakan," *Digital …*, 2023, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/digitech/article/view/3446

[14] M. Mitchell *et al.*, "Model Cards for Model Reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2019, pp. 220–229. doi: 10.1145/3287560.3287596.

[15] I. D. Raji *et al.*, "Closing the AI accountability gap," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2020, pp. 33–44. doi: 10.1145/3351095.3372873.

[16] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data Soc.*, vol. 3, no. 2, Dec. 2016, doi: 10.1177/2053951716679679.

[17] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[18] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[19] F. P. E. Putra, A. Hamzah, W. Agel, and ..., "Impelementasi Sistem Keamanan Jaringan Mikrotik Menggunakan Firewall Filtering dan Port Knocking," *Jurnal Sistim Informasi …*, 2023, [Online]. Available: https://ipv6.jsisfotek.org/index.php/JSisfotek/article/view/329

[20] F. P. E. Putra, M. Dafid, and I. Syafi'i, "Firewall Implementation as a Computer Network Security Strategy for Data Protection," *Brilliance: Research of Artificial …*, 2025, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/6162

[21] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[22] E. Ntoutsi *et al.*, "Bias in data-driven artificial intelligence systems—An introductory survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 3, May 2020, doi: 10.1002/widm.1356.

[23] M. B. Zafar, I. Valera, M. Gomez Rodriguez, and K. P. Gummadi, "Fairness Beyond Disparate Treatment & Disparate Impact," in *Proceedings of the 26th International Conference on World Wide Web*, Republic

and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 1171–1180. doi: 10.1145/3038912.3052660.

[24] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nat. Mach. Intell.*, vol. 1, no. 11, pp. 501–507, Nov. 2019, doi: 10.1038/s42256-019-0114-4.

[25] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds Mach. (Dordr).*, vol. 30, no. 1, pp. 99–120, Mar. 2020, doi: 10.1007/s11023-020-09517-8.

[26] A. Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *Big Data*, vol. 5, no. 2, pp. 153–163, Jun. 2017, doi: 10.1089/big.2016.0047.

[27] F. P. E. Putra, M. Irfan, M. Aziz, and ..., "Wireless Network Design at Pamekasan Regency Public Library," *Brilliance: Research of ...*, 2025, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/5876

[28] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[29] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman, "Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2020, pp. 1–13. doi: 10.1145/3313831.3376301.

[30] P. Nemitz, "Constitutional democracy and technology in the age of artificial intelligence," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, p. 20180089, Nov. 2018, doi: 10.1098/rsta.2018.0089.

[31] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, Dec. 2019, doi: 10.1126/scirobotics.aay7120.

[32] B. Shneiderman, "Bridging the Gap Between Ethics and Practice," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, pp. 1–31, Dec. 2020, doi: 10.1145/3419764.

[33] S. S, J. Paul, C. Strong, and J. Pius, "Consumer response towards social media advertising: Effect of media interactivity, its conditions and the underlying mechanism," *Int. J. Inf. Manage.*, vol. 54, p. 102155, Oct. 2020, doi: 10.1016/j.ijinfomgt.2020.102155.

[34] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, Sep. 2019, doi: 10.1145/3236009.

[35] F. P. E. Putra, U. Ubaidi, A. Hamzah, and ..., "Systematic literature review: Security gap detection on websites using OWASP ZAP," *Brilliance: Research ...*, 2024, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/4227

[36] F. P. E. Putra, A. Zulfikri, G. Arifin, and ..., "Analysis of phishing attack trends, impacts and prevention methods: literature study," *... : Research of Artificial ...*, 2024, [Online]. Available: https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/4357

[37] C. Cath, "Governing artificial intelligence: ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, p. 20180080, Nov. 2018, doi: 10.1098/rsta.2018.0080.

[38] "One membership. Unlimited knowledge," *IEEE Secur. Priv.*, vol. 16, no. 3, pp. c3–c3, May 2018, doi: 10.1109/MSP.2018.2701157.

[39] B. Hutchinson *et al.*, "Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Mar. 2021, pp. 560–575. doi: 10.1145/3442188.3445918.

[40] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *SSRN Electronic Journal*, 2017, doi: 10.2139/ssrn.3063289.

[41] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," *Organ. Behav. Hum. Decis. Process.*, vol. 151, pp. 90–103, Mar. 2019, doi: 10.1016/j.obhdp.2018.12.005.

[42] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err.," *J. Exp. Psychol. Gen.*, vol. 144, no. 1, pp. 114–126, 2015, doi: 10.1037/xge0000033.

[43] M. Gastelum, "Scale Matters: Temporality in the Perception of Affordances," *Front. Psychol.*, vol. 11, Jun. 2020, doi: 10.3389/fpsyg.2020.01188.

[44] J. Danaher, "The Threat of Algocracy: Reality, Resistance and Accommodation," *Philos. Technol.*, vol. 29, no. 3, pp. 245–268, Sep. 2016, doi: 10.1007/s13347-015-0211-1.

[45] E. Awad *et al.*, "The Moral Machine experiment," *Nature*, vol. 563, no. 7729, pp. 59–64, Nov. 2018, doi: 10.1038/s41586-018-0637-6.

[46] M. Ananny and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability," *New Media Soc.*, vol. 20, no. 3, pp. 973–989, Mar. 2018, doi: 10.1177/1461444816676645.

[47] M. Wieringa, "What to account for when accounting for algorithms," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2020, pp. 1–18. doi: 10.1145/3351095.3372833.

[48] P. Suárez-Serrato, E. I. Velázquez Richards, and M. Yazdani, "Socialbots Supporting Human Rights," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA: ACM, Dec. 2018, pp. 290–296. doi: 10.1145/3278721.3278734.

[49] A. Holzinger, "Interactive machine learning for health informatics: when do we need the human-in-the-loop?," *Brain Inform.*, vol. 3, no. 2, pp. 119–131, Jun. 2016, doi: 10.1007/s40708-016-0042-6.

[50] I. Rahwan, "Society-in-the-loop: programming the algorithmic social contract," *Ethics Inf. Technol.*, vol. 20, no. 1, pp. 5–14, Mar. 2018, doi: 10.1007/s10676-017-9430-8.

[51] F. M. Zanzotto, "Viewpoint: Human-in-the-loop Artificial Intelligence," *Journal of Artificial Intelligence Research*, vol. 64, pp. 243–252, Feb. 2019, doi: 10.1613/jair.1.11345.