

Algoritma Moral: Rekonstruksi Agency Etis Peserta Didik Dalam Pendidikan Agama Islam Di Era Kecerdasan Buatan

Badrus Shalih¹, Marzuqi², Zainal Abidin³

Universitas Islam Internasional Darullughah Wadda`wah

shalihbadrus@gmail.com¹, marzuqihasyim50@gmail.com², doktorzainalabidin1992@gmail.com³

Informasi Artikel

Vol: 2 No : 2 2026

Halaman : 37-44

Keywords:

Ethics of AI Agency, AI and Islamic Morality, Islamic Ethical Reconstruction

Abstract

The rapid advancement of artificial intelligence (AI) carries profound implications for the moral and ethical order of students, particularly within the context of Islamic Religious Education (PAI). This study aims to analyze AI's disruption of students' moral development, conduct an epistemological critique of existing PAI approaches, and formulate a model for reconstructing ethical agency based on Islamic values. Using a conceptual qualitative approach through library research, this study integrates perspectives from Islamic moral philosophy, contemporary moral agency theory, and critical analysis of digital education literature. The findings reveal that AI has created a condition of "algorithmic moral passivity," in which students tend to delegate ethical deliberation to automated systems, resulting in the erosion of moral *ijtihad* capacity. Grounded in Islamic concepts of 'aql, nafs, and ikhtiyar, this research proposes a Model of Ethical Agency Reconstruction (MRAE) comprising three dimensions: digital tazkiyatun nafs, technology fiqh, and algorithmic musyawarah. This model is intended to serve as an operational framework for PAI teachers in responding to AI-era challenges in a transformative and contextual manner.

Abstrak

Kemajuan kecerdasan buatan (AI) yang pesat membawa implikasi mendalam terhadap tatanan moral dan etika peserta didik, khususnya dalam konteks Pendidikan Agama Islam (PAI). Penelitian ini bertujuan untuk menganalisis disrupsi AI terhadap perkembangan moral siswa, melakukan kritik epistemologis terhadap pendekatan PAI yang ada, serta merumuskan model rekonstruksi agency etis berbasis nilai-nilai Islam. Menggunakan pendekatan kualitatif konseptual dengan metode library research, kajian ini mengintegrasikan perspektif filsafat moral Islam, teori moral agency kontemporer, dan analisis kritis terhadap literatur pendidikan digital. Hasil kajian menemukan bahwa AI telah menciptakan kondisi "algorithmic moral passivity" di mana peserta didik cenderung mendelegasikan pertimbangan etis kepada sistem otomatis, mengakibatkan erosi kapasitas *ijtihad* moral. Berdasarkan konsep 'aql, nafs, dan ikhtiyar dalam Islam, penelitian ini menawarkan Model Rekonstruksi Agency Etis (MRAE) yang terdiri dari tiga dimensi: dimensi tazkiyatun nafs digital, dimensi fiqh teknologi, dan dimensi musyawarah algoritmik. Model ini diharapkan dapat menjadi kerangka operasional bagi guru PAI dalam merespons tantangan era AI secara transformatif dan kontekstual.

Kata Kunci : Agency Etis AI, Kecerdasan Buatan Etis, Moral Islam Digital

PENDAHULUAN

Abad ke-21 ditandai oleh transformasi teknologi yang belum pernah terjadi sebelumnya. Kecerdasan buatan AI (Artificial Intelligence) bukan lagi sekadar instrumen bantu, melainkan telah menjadi ekosistem yang membentuk cara manusia berpikir, memilih, dan bertindak. Algoritma rekomendasi, asisten virtual, dan sistem pengambilan keputusan otomatis kini hadir dalam kehidupan sehari-hari peserta didik, mulai dari konten yang mereka konsumsi hingga relasi sosial yang mereka bangun. Kondisi ini menciptakan apa yang oleh sebagian pemikir disebut sebagai "masyarakat algoritmik" sebuah tatanan sosial di mana keputusan-keputusan fundamental justru semakin banyak ditentukan oleh sistem yang tidak kasat mata (Baker & Hawn, 2022a; Bojic, 2024; Mahmud dkk., 2022).

Di tengah arus perubahan ini, Pendidikan Agama Islam (PAI) menghadapi tantangan eksistensial yang serius. Jika selama ini PAI berperan sebagai wahana pembentukan akhlak dan internalisasi nilai-nilai ilahiah, maka kini ia harus bersanding dengan sistem AI yang secara diam-diam menawarkan “etika algoritmik” seperangkat aturan perilaku yang diprogramkan berdasarkan kalkulasi data dan optimasi kinerja, bukan berdasarkan hikmah, tazkiyah, atau kesadaran transendental. Ironisnya, peserta didik yang lahir di era digital ini kerap lebih terpapar oleh logika algoritma ketimbang logika wahyu (Elmahjub, 2023a; Holmes dkk., 2022a; Mahmud dkk., 2022).

Krisis moral yang melanda generasi muda saat ini bukan sekadar persoalan kenakalan remaja konvensional. Lebih dari itu, ia mencerminkan pergeseran mendasar dalam cara subjek moral dibentuk. Studi-studi terkini menunjukkan bahwa paparan intensif terhadap konten yang dikurasi algoritma berkontribusi pada peningkatan perilaku impulsif, penurunan empati, dan kecenderungan outsourcing keputusan etis kepada platform digital. Fenomena ini menandai apa yang penulis sebut sebagai “deagentifikasi moral” kondisi di mana individu kehilangan kapasitasnya sebagai pelaku moral yang otonom dan bertanggung jawab (Alon-Barkat & Busuioc, 2023; Baumann dkk., 2026; Lewin dkk., 2023).

Dalam perspektif Islam, agency etis kemampuan manusia untuk memilih dan bertindak berdasarkan kesadaran moral yang otonom merupakan salah satu esensi kemanusiaan yang paling fundamental. Konsep khalifah fi al-ardh mengandaikan bahwa manusia adalah subjek moral aktif yang mampu membedakan antara haq dan bathil, antara maslahat dan mudarat. Kemampuan ini tidak dapat didelegasikan kepada mesin, sebab ia melekat pada dimensi ruh, aql, dan nafs yang menjadi karakteristik unik manusia sebagai ciptaan Allah SWT (Elmahjub, 2023b; Hasibuan dkk., 2024; Taha dkk., 2025).

Berangkat dari latar belakang tersebut, penelitian ini berupaya menjawab tiga pertanyaan mendasar: (1) Bagaimana AI mendisrupsi perkembangan moral dan agency etis peserta didik? (2) Apakah pendekatan PAI yang ada saat ini secara epistemologis memadai untuk merespons tantangan tersebut? (3) Bagaimana model rekonstruksi agency etis yang relevan dapat dirumuskan dalam bingkai PAI? Melalui analisis konseptual yang mendalam, penelitian ini diharapkan dapat memberikan kontribusi teoretis sekaligus praktis bagi pengembangan PAI di era kecerdasan buatan.

METODE

Penelitian ini menggunakan pendekatan kualitatif konseptual dengan metode library research (kajian kepustakaan). Pilihan metodologis ini didasarkan pada beberapa pertimbangan epistemologis. Pertama, pertanyaan penelitian yang diajukan bersifat konseptual-normatif, yaitu berkaitan dengan bagaimana seharusnya PAI direkonstruksi secara konseptual, bukan hanya mendeskripsi apa yang ada. Kedua, objek kajian ini yakni relasi antara AI, moral agency, dan PAI masih merupakan wilayah yang relatif baru dan membutuhkan penyiapan kerangka konseptual yang matang sebelum kajian empiris dapat dilakukan secara bermakna (Aspers & Corte, 2021; Holmes dkk., 2022a; A. Nguyen dkk., 2023; T. N. M. Nguyen dkk., 2022).

Sumber data penelitian ini meliputi: (1) literatur primer berupa karya-karya klasik dan kontemporer dalam etika Islam, filsafat pendidikan Islam, dan filsafat moral; (2) literatur sekunder berupa artikel jurnal, buku teks, dan laporan penelitian tentang AI, pendidikan digital, dan pendidikan karakter; (3) dokumen kebijakan nasional dan internasional terkait pendidikan di era digital dan pengembangan AI yang bertanggung jawab (Baker & Hawn, 2022b; Holmes dkk., 2022b; Kasneci dkk., 2023).

Proses analisis data dilakukan melalui empat tahap yang saling terkait. Tahap pertama adalah inventarisasi dan seleksi sumber, di mana penulis mengidentifikasi, mengklasifikasi, dan mengevaluasi relevansi serta kualitas berbagai sumber yang tersedia. Tahap kedua adalah analisis konseptual, yang meliputi identifikasi konsep-konsep kunci, klarifikasi maknanya, dan pemetaan hubungan antar-konsep. Tahap ketiga adalah sintesis integratif, di mana berbagai perspektif dari disiplin yang berbeda filsafat moral, etika Islam, ilmu pendidikan, dan studi AI disintesis ke dalam kerangka analisis yang koheren. Tahap keempat adalah rekonstruksi teoritis, di mana

penulis merumuskan model konseptual baru berdasarkan temuan analisis (Bowen, 2009; Jaakkola, 2020; Torraco, 2005).

Keabsahan penelitian ini dijamin melalui beberapa strategi. Pertama, triangulasi sumber dengan memastikan bahwa setiap klaim konseptual didukung oleh minimal tiga sumber yang berbeda. Kedua, coherence checking untuk memastikan konsistensi internal argumen yang dikembangkan. Ketiga, member checking melalui diskusi dengan para pakar di bidang PAI, filsafat Islam, dan teknologi pendidikan (De Loyola González-Salgado dkk., 2024; Lloyd dkk., 2024; Nowell dkk., 2017).

HASIL DAN PEMBAHASAN

1. Disrupsi AI terhadap Moralitas Peserta Didik

Temuan analisis menunjukkan bahwa disrupsi AI terhadap moralitas peserta didik berlangsung melalui setidaknya empat mekanisme yang berbeda namun saling memperkuat. Memahami mekanisme-mekanisme ini secara mendalam merupakan prasyarat untuk merumuskan respons pendidikan yang tepat.

a. Algorithmic Moral Passivity

Mekanisme pertama dan paling mendasar adalah apa yang penulis konseptualisasikan sebagai "*algorithmic moral passivity*." Fenomena ini mengacu pada kecenderungan peserta didik untuk mengosongkan kapasitas deliberasi moralnya dan menyerahkan keputusan etis kepada sistem AI. Manifestasinya beragam: mulai dari menanyakan kepada chatbot AI tentang "apakah tindakan ini etis" tanpa melakukan refleksi personal, hingga mengikuti rekomendasi algoritma tanpa mempertanyakan nilai-nilai yang tertanam di dalamnya.

Algorithmic moral passivity berbeda secara kualitatif dari ketaatan normatif biasa. Ketika seseorang mengikuti fatwa ulama atau nasihat orang tua, masih ada proses relasional yang melibatkan kepercayaan, otoritas moral, dan konteks komunal. Namun, ketika keputusan moral diserahkan kepada algoritma, seluruh infrastruktur sosial-relasional dari pertimbangan moral tersebut lenyap. Yang tersisa hanyalah output berupa rekomendasi yang muncul dari kalkulasi matematis yang tidak transparan.

Kondisi ini, jika dibiarkan berlanjut, dapat menghasilkan generasi yang secara teknologis melek tetapi secara etis pasif generasi yang mampu mengoperasikan sistem AI dengan mahir namun tidak mampu mengajukan pertanyaan kritis tentang nilai-nilai yang dikodekan di dalamnya. Dalam terminologi Islam, ini adalah kondisi di mana fungsi 'aql sebagai alat tamyiz (pembeda) mengalami atrofi akibat tidak terlatih.

b. Filter Bubble dan Krisis Perspektif Moral

Mekanisme kedua adalah efek filter bubble yang diciptakan oleh algoritma kurasi konten. Platform media sosial dan mesin pencari modern dirancang untuk memaksimalkan keterlibatan pengguna melalui penyajian konten yang sesuai dengan preferensi dan keyakinan yang sudah ada. Akibatnya, peserta didik terperangkap dalam gelembung epistemis di mana mereka hanya terpapar pada pandangan yang mengkonfirmasi keyakinan yang sudah mereka miliki.

Dalam konteks perkembangan moral, kemampuan untuk memahami dan menghargai perspektif yang berbeda apa yang oleh Kohlberg disebut sebagai "*perspective-taking*" merupakan komponen krusial dari pematangan moral (Walker, 1980). Penelitian menunjukkan bahwa individu yang mampu melihat suatu isu dari sudut pandang berbeda cenderung membuat keputusan moral yang lebih matang dan adil. Filter bubble secara sistematis menghambat perkembangan kapasitas ini.

Lebih dari sekadar masalah epistemis, filter bubble juga memiliki implikasi langsung terhadap pembentukan karakter islami. Nilai-nilai seperti tasamuh (toleransi), tawadhu' (kerendahhatian epistemis), dan hikmah (kebijaksanaan yang muncul dari keluasan wawasan) mensyaratkan pengalaman berjumpa dengan keberagaman perspektif. Ketika algoritma secara otomatis menyaring keberagaman tersebut, proses pembudayaan nilai-nilai ini terhambat.

c. Gamifikasi Perilaku dan Reduksi Moralitas

Mekanisme ketiga adalah gamifikasi perilaku yang dipraktikkan oleh berbagai platform digital. Sistem likes, shares, followers, dan berbagai reward mechanism lainnya menciptakan struktur insentif yang mereduksi moralitas menjadi kalkulasi popularitas. Peserta didik belajar

untuk mengevaluasi tindakannya bukan berdasarkan nilai intrinsiknya, melainkan berdasarkan respons algoritmik yang dihasilkannya.

Fenomena ini berimplikasi serius terhadap pembentukan niat yang dalam etika Islam merupakan inti dari nilai moral suatu perbuatan. Hadis Nabi yang sangat terkenal "*innamal a'malu binniyyat*" menegaskan bahwa kualitas moral suatu tindakan ditentukan oleh niat yang mendasarinya, bukan oleh respons eksternal yang ditimbulkannya. Namun, ekosistem digital mendorong peserta didik untuk terus menerus memonitor dan merespons validasi eksternal, yang secara tidak langsung melemahkan orientasi niat yang ikhlas.

Penelitian neurosains terkini mendukung kekhawatiran ini. Dopaminergic feedback loop yang diciptakan oleh sistem notifikasi media sosial secara struktural menyerupai mekanisme adiksi, dan dapat melemahkan kemampuan pengaturan diri (self-regulation) yang merupakan komponen penting dari kematangan moral (Morris dkk., 2023).

d. Dehumanisasi Relasional dan Krisis Empati

Mekanisme keempat adalah dehumanisasi relasional yang terjadi ketika interaksi antar-manusia semakin digantikan oleh interaksi antara manusia dan mesin. Empati kemampuan untuk merasakan dan memahami pengalaman orang lain merupakan fondasi bagi banyak kebajikan moral, termasuk belas kasih (*rahmah*), keadilan (*'adl*), dan solidaritas (*ukhuwwah*). Namun, empati hanya dapat berkembang melalui pengalaman relasional autentik dengan sesama manusia.

Ketika porsi signifikan dari kehidupan sosial peserta didik berlangsung di dunia digital yang dimediasi oleh algoritma termasuk dalam interaksi dengan chatbot AI yang mensimulasikan empati kapasitas empatik mereka berisiko mengalami stunting. Mereka mungkin menjadi mahir dalam mengelola interaksi digital, tetapi kehilangan kemampuan untuk terlibat secara empatik dalam hubungan interpersonal yang kompleks dan penuh ketidakpastian.

Dalam perspektif pendidikan Islam, krisis empati ini merupakan ancaman langsung terhadap cita-cita pembentukan insan kamil manusia paripurna yang tidak hanya unggul secara intelektual dan spiritual, tetapi juga mampu hadir secara penuh dalam relasi sosialnya sebagai cerminan sifat-sifat Rahman dan Rahim Allah SWT.

2. Kritik Epistemologis terhadap Pendekatan PAI Konvensional

Setelah mengidentifikasi mekanisme disrupsi AI, langkah selanjutnya adalah melakukan evaluasi kritis terhadap kapasitas PAI konvensional untuk merespons tantangan tersebut. Evaluasi ini bersifat konstruktif bukan untuk mendiskreditkan PAI, melainkan untuk mengidentifikasi titik-titik kelemahan yang perlu diperkuat secara sistematis.

a. Dualisme Epistemis: Dikotomi Ilmu Agama dan Ilmu Umum

Kritik epistemologis pertama menyangkut warisan dualisme epistemis yang masih kuat dalam praktik PAI. Meskipun wacana integrasi ilmu telah berkembang pesat dalam dekade terakhir, pada level praktik pedagogis, PAI masih sering diperlakukan sebagai domain yang terpisah dari sains dan teknologi. Akibatnya, ketika peserta didik menghadapi pertanyaan etis yang terkait dengan teknologi AI, mereka tidak memiliki kerangka konseptual yang integratif untuk meresponsnya.

Padahal, tradisi intelektual Islam sesungguhnya menawarkan fondasi yang kuat bagi integrasi ilmu. Konsep tauhid sebagai kesatuan fundamental realitas, tradisi tafakkur (refleksi mendalam) terhadap fenomena alam, dan metodologi usul fiqh yang bersifat holistik-kontekstual semuanya merupakan modal epistemis yang dapat menjembatani dikotomi tersebut. Tantangannya adalah bagaimana mentransformasikan kekayaan tradisi ini menjadi pedagogi yang aktual dan fungsional.

b. Orientasi Ritualistik vs. Transformasi Moral

Kritik kedua berkaitan dengan kecenderungan PAI untuk lebih menekankan dimensi ritualistik-formal (seperti hafalan ayat, tata cara ibadah, dan pengetahuan hukum Islam) ketimbang pembentukan karakter yang mendalam dan kemampuan penalaran moral yang kritis. Penelitian (Muhaimin dkk., 2021) menunjukkan bahwa sebagian besar penilaian dalam PAI masih didominasi oleh aspek kognitif-deklaratif, sementara aspek afektif dan psikomotor yang lebih relevan dengan pembentukan karakter relatif terabaikan.

Dalam konteks tantangan AI, orientasi ritualistik ini terbukti tidak memadai. Menghadapi ekosistem digital yang terus berubah, peserta didik membutuhkan kapasitas untuk melakukan istinbat (penarikan kesimpulan hukum) etis secara mandiri terhadap situasi-situasi baru yang belum pernah diantisipasi oleh fiqh klasik. Kapasitas ini hanya dapat dikembangkan melalui latihan penalaran moral yang aktif, bukan melalui hafalan ketentuan fiqh yang bersifat statis.

c. Absennya Literasi Teknologi Berbasis Nilai

Kritik ketiga adalah absennya apa yang dapat disebut sebagai "*literasi teknologi berbasis nilai*." Selama ini, literasi teknologi dipahami terutama sebagai kemampuan teknis untuk menggunakan perangkat digital secara efektif. Namun, dimensi evaluatif-etis dari literasi teknologi kemampuan untuk menilai, mempertanyakan, dan merespons secara kritis terhadap asumsi-asumsi nilai yang tertanam dalam teknologi hampir sepenuhnya absen dari kurikulum PAI.

Padahal, setiap sistem teknologi, termasuk AI, membawa serta seperangkat nilai, asumsi, dan ideologi yang dikodekan oleh para pembuatnya. Platform media sosial yang dirancang untuk memaksimalkan keterlibatan secara implisit mempromosikan nilai-nilai seperti sentralitas perhatian (*attention centrality*), eksibisionisme, dan kompetisi sosial nilai-nilai yang bertentangan dengan etika Islam tentang *qana'ah*, *tawadhu'*, dan *ukhuwwah*. Tanpa literasi teknologi berbasis nilai, peserta didik tidak memiliki perangkat untuk mengenali dan merespons inkonsistensi nilai ini.

3. Model Rekonstruksi Agency Etis (MRAE)

Berdasarkan analisis di atas, bagian ini merumuskan Model Rekonstruksi Agency Etis (MRAE) sebagai kerangka konseptual bagi transformasi PAI di era AI. MRAE dibangun di atas fondasi ontologis Islam tentang manusia sebagai makhluk yang memiliki dimensi ruh, aql, dan nafs, sekaligus mengintegrasikan wawasan dari teori pendidikan kritis dan etika teknologi kontemporer. Model ini terdiri dari tiga dimensi yang saling berhubungan dan saling menopang.

a. Dimensi Pertama: Tazkiyatun Nafs Digital

Dimensi pertama MRAE adalah *tazkiyatun nafs digital*, yang mengacu pada praktik penyucian jiwa yang dikontekstualisasikan untuk menghadapi tantangan lingkungan digital. *Tazkiyatun nafs* dalam tradisi tasawuf Islam merupakan proses pemurnian jiwa dari sifat-sifat rendah (*mazmumah*) menuju sifat-sifat terpuji (*mahmudah*) melalui latihan spiritual yang terstruktur. Dalam konteks digital, dimensi ini diterjemahkan ke dalam serangkaian praktik pedagogi yang konkret.

Praktik utama dalam dimensi ini meliputi: pertama, digital mindfulness atau *muraqabah digital* latihan untuk hadir secara sadar dan reflektif dalam setiap interaksi digital, alih-alih bersikap reaktif dan terbawa arus algoritma. Kedua, digital fasting atau *shaumul-digital* praktik pembatasan konsumsi digital yang disengaja sebagai bentuk latihan kendali diri (*mujahadah*). Ketiga, *contemplative technology use* penggunaan teknologi yang diorientasikan bukan untuk konsumsi pasif, melainkan untuk refleksi, penciptaan, dan kontribusi positif.

Implementasi pedagogis dari dimensi ini dapat mencakup: sesi refleksi mingguan tentang penggunaan teknologi masing-masing peserta didik, proyek pembuatan konten digital yang bermakna dan berdampak sosial positif, serta latihan-latihan mindfulness yang mengintegrasikan praktik dzikir dengan kesadaran terhadap penggunaan teknologi.

b. Dimensi Kedua: Fiqh Teknologi

Dimensi kedua MRAE adalah *fiqh teknologi*, yang merupakan pengembangan dan pengaplikasian metodologi fiqh Islam untuk menganalisis dan mengevaluasi fenomena teknologi secara kritis. *Fiqh teknologi* tidak dipahami sebagai sekadar kumpulan fatwa tentang halal haram penggunaan teknologi tertentu, melainkan sebagai metodologi berpikir dinamis yang memungkinkan peserta didik untuk melakukan istinbat etis terhadap situasi-situasi baru yang terus berkembang.

Komponen utama *fiqh teknologi* dalam konteks PAI meliputi: pemahaman tentang *maqashid al-syariah* (tujuan hukum Islam) sebagai bingkai evaluasi apakah suatu teknologi melindungi atau mengancam jiwa, akal, agama, keturunan, dan harta? Prinsip-prinsip *ushul fiqh* seperti *la dharara wa la dharir* (tidak boleh ada bahaya yang ditimbulkan), *dar al-mafasid muqaddam 'ala jalb al-mashalih* (mencegah kerusakan didahulukan dari mendatangkan

kemaslahatan), dan sadd al-dzara'i (menutup jalan menuju kerusakan) menyediakan kerangka metodologis yang kaya untuk evaluasi etis terhadap teknologi AI.

Secara pedagogis, dimensi ini dapat diimplementasikan melalui studi kasus etis tentang penggunaan AI dalam berbagai konteks, latihan analisis menggunakan kerangka maqashid al-syariah, dan proyek penelitian tentang dampak sosial-moral dari teknologi spesifik. Metode pembelajaran yang dianjurkan adalah deliberasi moral berbasis kasus (case-based moral deliberation) yang mendorong peserta didik untuk aktif mengaplikasikan kerangka etis Islam pada situasi nyata.

c. Dimensi Ketiga: Musyawarah Algoritmik

Dimensi ketiga MRAE adalah musyawarah algoritmik, yang mengacu pada praktik deliberasi kolektif yang kritis terhadap sistem dan keputusan algoritmik. Musyawarah dalam Islam bukan sekadar prosedur pengambilan keputusan, melainkan juga praktik etis yang mencerminkan nilai-nilai seperti kerendahan hati epistemis, penghargaan terhadap keragaman perspektif, dan komitmen terhadap keadilan. Dalam konteks AI, musyawarah algoritmik diterjemahkan ke dalam praktik pedagogi yang mengembangkan kemampuan peserta didik untuk bersama-sama mengidentifikasi, menganalisis, dan merespons bias serta dampak etis dari sistem AI.

Implementasi dimensi ini dapat mencakup: pertama, *"algorithm audit"* projects di mana kelompok peserta didik menyelidiki cara kerja dan dampak algoritma dari platform yang mereka gunakan sehari-hari. Kedua, ethical scenario deliberation di mana kelompok mendiskusikan skenario etis yang melibatkan AI dari berbagai perspektif nilai personal, sosial, dan islami. Ketiga, participatory design workshops di mana peserta didik diajak untuk merancang "AI yang ideal dari perspektif Islam," sebuah latihan yang mendorong mereka untuk mengeksternalisasi dan mengkritisi nilai-nilai yang mereka harapkan terwujud dalam teknologi.

Tabel berikut merangkum komponen utama MRAE beserta target kompetensi dan contoh implementasi pedagogisnya:

| Tazkiyatun Nafs Digital | Nafs, Tazkiyah, Mujahadah, Muraqabah | Kendali diri digital, kesadaran reflektif, ikhlas dalam berekspresi | Digital journaling, digital fasting, refleksi muraqabah |
|--------------------------------|---|--|---|
| Fiqh Teknologi | Maqashid Syariah, Ushul Fiqh, Ijtihad, Istinbat | Penalaran moral kritis, evaluasi etis berbasis nilai, ijtihad kontekstual | Studi kasus etis, analisis maqashid, deliberasi moral berbasis fiqh |
| Musyawarah Algoritmik | Syura, Ukhuwwah, 'Adl, Tawasuth | Deliberasi kolektif, literasi algoritmik kritis, partisipasi etis | Algorithm audit, skenario etis kolaboratif, participatory AI design |

Tabel 1. Dimensi Model Rekonstruksi Agency Etis (MRAE)

Ketiga dimensi MRAE tersebut tidak bersifat hierarkis atau terpisah, melainkan saling melengkapi dalam satu sistem yang organik. Dimensi tazkiyatun nafs digital menyediakan fondasi spiritual-subjektif; dimensi fiqh teknologi menyediakan kerangka metodologis-intelektual; dan dimensi musyawarah algoritmik menyediakan arena sosial-deliberatif di mana agency etis peserta didik diekspresikan dan diuji dalam konteks komunal. Bersama-sama, ketiganya menciptakan ekosistem pembelajaran moral yang komprehensif dan relevan untuk era AI.

MRAE juga berbeda secara fundamental dari pendekatan literasi digital yang ada. Sementara literasi digital konvensional berfokus pada penguasaan keterampilan teknis dan keamanan digital, MRAE menempatkan pembentukan subjek moral sebagai tujuan utama. Teknologi dalam MRAE dipahami bukan sebagai tujuan atau bahkan sebagai alat netral, melainkan sebagai arena etis di mana nilai-nilai diperjuangkan, dipertahankan, dan diwujudkan.

KESIMPULAN

Penelitian ini telah mencapai tiga tujuannya secara sistematis. Pertama, analisis menunjukkan bahwa disrupsi AI terhadap moralitas peserta didik berlangsung melalui empat mekanisme utama: algorithmic moral passivity, filter bubble, gamifikasi perilaku, dan dehumanisasi relasional. Keempat mekanisme ini bersifat saling memperkuat dan mengancam integritas agency etis peserta didik secara holistik.

Kedua, kritik epistemologis terhadap PAI konvensional mengidentifikasi tiga kelemahan mendasar: warisan dualisme epistemis, orientasi ritualistik yang mengabaikan transformasi moral yang mendalam, dan absennya literasi teknologi berbasis nilai. Kelemahan-kelemahan ini menjadikan PAI konvensional tidak cukup responsif terhadap tantangan yang ditimbulkan oleh ekosistem AI.

Ketiga, penelitian ini merumuskan Model Rekonstruksi Agency Etis (MRAE) yang terdiri dari tiga dimensi yang saling melengkapi: tazkiyatun nafs digital, fiqh teknologi, dan musyawarah algoritmik. Model ini berakar pada tradisi intelektual dan spiritual Islam yang kaya, sekaligus responsif terhadap tantangan etis kontemporer yang ditimbulkan oleh AI. MRAE menawarkan bukan hanya kritik, tetapi sebuah visi afirmatif tentang bagaimana PAI dapat bertransformasi menjadi wahana pembentukan manusia yang benar-benar merdeka bebas bukan hanya dari penjajahan fisik, tetapi juga dari kolonisasi algoritmik terhadap kesadaran moral.

REFERENCES

- Alon-Barkat, S., & Busuioc, M. (2023). Human–AI Interactions in Public Sector Decision Making: “Automation Bias” and “Selective Adherence” to Algorithmic Advice. *Journal of Public Administration Research and Theory*, 33(1), 153–169. <https://doi.org/10.1093/jopart/muac007>
- Aspers, P., & Corte, U. (2021). What is Qualitative in Research. *Qualitative Sociology*, 44(4), 599–608. <https://doi.org/10.1007/s11133-021-09497-w>
- Baker, R. S., & Hawn, A. (2022a). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Baker, R. S., & Hawn, A. (2022b). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Baumann, L., Spangenberg, P., Birkefeld, K., Dörre, M., Bollmacher, L., & Nebel, S. (2026). Social Networking Site Use and Adolescents Empathic Skills: A Systematic Literature Review. *Adolescent Research Review*. <https://doi.org/10.1007/s40894-025-00279-x>
- Bojic, L. (2024). AI alignment: Assessing the global impact of recommender systems. *Futures*, 160, 103383. <https://doi.org/10.1016/j.futures.2024.103383>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- De Loyola González-Salgado, I., Rivera-Navarro, J., Gutiérrez-Sastre, M., Conde, P., & Franco, M. (2024). Conducting member checking within a qualitative case study on health-related behaviours in a large European city: Appraising interpretations and co-constructing findings. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, 28(1), 3–21. <https://doi.org/10.1177/13634593221109682>
- Elmahjub, E. (2023a). Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI. *Philosophy & Technology*, 36(4), 73. <https://doi.org/10.1007/s13347-023-00668-x>
- Elmahjub, E. (2023b). Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI. *Philosophy & Technology*, 36(4), 73. <https://doi.org/10.1007/s13347-023-00668-x>
- Hasibuan, U. S., Utami, P. I., Novia, S., Surahman, C., & Sumarna, E. (2024). Konsep Khalifah dalam Qs. Al-Baqarah/ 2: 30 dan Implikasinya Terhadap Tujuan Pendidikan Islam di Era Society 5.0. *JOURNAL OF QURAN AND HADITH STUDIES*, 13(2), 272–285. <https://doi.org/10.15408/quhas.v13i2.42166>

- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022a). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022b). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lewin, K. M., Kaur, A., & Meshi, D. (2023). Problematic Social Media Use and Impulsivity. *Current Addiction Reports*, 10(3), 553–562. <https://doi.org/10.1007/s40429-023-00495-2>
- Lloyd, N., Hyett, N., & Kenny, A. (2024). To Member Check or not to Member Check? An Evaluation of Member Checking in an Interpretive Descriptive Study. *International Journal of Qualitative Methods*, 23, 16094069241301383. <https://doi.org/10.1177/16094069241301383>
- Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175, 121390. <https://doi.org/10.1016/j.techfore.2021.121390>
- Morris, R., Moretta, T., & Potenza, M. N. (2023). The Psychobiology of Problematic Use of Social Media. *Current Behavioral Neuroscience Reports*, 10(4), 65–74. <https://doi.org/10.1007/s40473-023-00261-8>
- Muhaimin, M., Fajrie, N., & Setiawan, D. (2021). Respon Kognitif Anak dalam Produk Permainan Remitan Desa Mayong Lor Jepara. *Jurnal Inovasi Penelitian*, 2(5), 1349–1358. <https://doi.org/10.47492/jip.v2i5.803>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Nguyen, T. N. M., Whitehead, L., Dermody, G., & Saunders, R. (2022). The use of theory in qualitative research: Challenges, development of a framework and exemplar. *Journal of Advanced Nursing*, 78(1). <https://doi.org/10.1111/jan.15053>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1), 1609406917733847. <https://doi.org/10.1177/1609406917733847>
- Taha, J. Z. A., Sulaiman, S. Z., & Ajmain@Jima'ain, M. T. (2025). The Use of Artificial Intelligence (ChatGPT) in Islamic Education: An Analysis of Challenges, Opportunities, and Ethical Considerations Based on Maqasid Shariah. *Journal of Research, Innovation, and Strategies for Education (RISE)*, 2(3), 35–46. <https://doi.org/10.70148/rise.v2i3.3>
- Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- Walker, L. J. (1980). Cognitive and perspective-taking prerequisites for moral development. *Child Development*, 131–139.