



## Agensi Moral AI dan Tanggung Jawab Manusia dalam Pemikiran Islam

Matahari Ramadhan<sup>1</sup>, Damanhuri Fattah<sup>2</sup>, Fisman Bedi<sup>3</sup>

Universitas Islam Negeri Raden Intan Lampung

Email: [ramadhan.matahari24@gmail.com](mailto:ramadhan.matahari24@gmail.com)<sup>1</sup>  
[damanhurifattah@gmail.com](mailto:damanhurifattah@gmail.com)<sup>2</sup>  
[bangbedi@gmail.com](mailto:bangbedi@gmail.com)<sup>3</sup>

P-ISSN : 2745-7796

E-ISSN : 2809-7459

**Abstrak.** Perkembangan kecerdasan buatan membuat istilah agen, otonomi, rasionalitas, keputusan, dan tanggung jawab semakin sering dipakai untuk menjelaskan sistem teknis. Pemakaian istilah tersebut dapat menimbulkan kekaburan ketika kemampuan AI mencapai tujuan operasional disamakan dengan kapasitas memikul tanggung jawab moral. Artikel ini menjelaskan batas konseptual antara AI sebagai agen rasional dan AI sebagai agen moral, lalu menempatkannya dalam kerangka tanggung jawab manusia dalam pemikiran Islam. Dengan metode kepustakaan-konseptual, artikel menelaah literatur tentang agen rasional, agensi moral, intensionalitas, kesadaran, antropomorfisme, akuntabilitas, kesenjangan tanggung jawab, dan penggunaan AI dalam ranah pengetahuan keislaman. Artikel ini berargumen bahwa AI dapat disebut agen rasional dalam arti teknis-operasional karena mampu memproses persepsi, tujuan, ukuran kinerja, dan tindakan. Akan tetapi, rasionalitas instrumental belum cukup untuk menjadikan AI sebagai agen moral. Agensi moral memerlukan syarat tambahan, seperti pemahaman normatif, intensionalitas, kapasitas memberi pertanggungjawaban, dan keterlibatan dalam relasi moral. Dalam pemikiran Islam, AI lebih tepat ditempatkan sebagai alat bantu yang berada di bawah amanah, pengawasan, dan pertanggungjawaban manusia. Dengan demikian, diskusi AI tidak perlu menolak seluruh bentuk agensi teknis, tetapi juga tidak boleh memindahkan tanggung jawab moral dari manusia dan institusi kepada sistem komputasional.

**Kata Kunci:** kecerdasan buatan; agen rasional; agensi moral; tanggung jawab, agensi moral islam.

<http://jurnal.staiddimakassar.ac.id/index.php/aujpsi>

DOI : <https://doi.org/10.55623>

### PENDAHULUAN

Kecerdasan buatan atau *artificial intelligence (AI)* tidak lagi hanya dipahami sebagai perangkat komputasi yang menjalankan instruksi tertutup. Dalam banyak konteks, AI diperlakukan sebagai sistem yang dapat memilih respons, menyesuaikan tindakan, memprediksi kebutuhan pengguna, memberi rekomendasi, bahkan berinteraksi melalui bahasa alami.

Perkembangan ini membuat istilah agen, otonomi, rasionalitas, keputusan, dan tanggung jawab masuk ke dalam percakapan umum tentang teknologi. Dalam wacana populer, AI sering disebut mengambil keputusan, memahami pengguna, membantu manusia, atau bertindak secara mandiri. Ungkapan semacam itu berguna untuk menjelaskan fungsi sistem secara praktis, tetapi dapat menjadi problematis ketika

bahasa operasional tersebut berubah menjadi klaim normatif tentang status moral *AI*.

Masalah pokoknya bukan sekadar apakah *AI* canggih, cepat, atau berguna. Masalah yang lebih mendasar adalah apakah kemampuan *AI* untuk bertindak sesuai tujuan tertentu cukup untuk menyebutnya sebagai agen moral. Dalam ilmu *AI*, istilah agen rasional memiliki makna teknis. Russell dan Norvig (2021) menjelaskan agen melalui hubungan antara persepsi, tindakan, tujuan, lingkungan tugas, dan ukuran kinerja. (Russell & Norvig, 2021) Dalam arti ini, suatu sistem disebut rasional bila tindakannya diorientasikan pada capaian kinerja yang diharapkan berdasarkan informasi dan kemampuan yang tersedia. Definisi ini kuat untuk menjelaskan *AI* sebagai sistem teknis, tetapi belum otomatis menyelesaikan pertanyaan moral. Ukuran kinerja, tujuan, dan lingkungan tugas dapat ditentukan oleh manusia; keberhasilan sistem mencapai tujuan belum berarti sistem tersebut memahami alasan moral di balik tujuan itu.

Kekaburan ini semakin penting karena *AI* semakin sering masuk ke wilayah yang memiliki konsekuensi etis. Sistem *AI* dipakai dalam pendidikan, kesehatan, rekrutmen, administrasi, keamanan, layanan publik, produksi pengetahuan, dan pendampingan sosial. Dalam kasus *AI companion* (chatbot pendamping sosial), misalnya, pengguna dapat membangun relasi afektif dan merasakan dukungan dari sistem yang merespons secara personal (Pan & de Graaf, 2025). Dalam konteks lain, keluaran *AI* dapat memengaruhi keputusan manusia tentang siapa yang diberi akses, siapa yang dipercaya, siapa yang dinilai berisiko, atau siapa yang dianggap layak mendapat layanan tertentu. Oleh sebab itu, *AI* dapat menimbulkan dampak moral sekalipun statusnya sebagai subjek moral masih diperdebatkan.

Di sinilah perbedaan antara dampak moral dan agensi moral perlu dijaga. Coeckelbergh (2020) menunjukkan bahwa

sistem *AI* dapat bertindak dan menghasilkan konsekuensi yang relevan secara etis, tetapi hal itu tidak identik dengan pengakuan bahwa sistem tersebut merupakan agen moral penuh (Coeckelbergh, 2020). Khasri (2024) juga menekankan pentingnya tidak mengalihkan tanggung jawab moral dari manusia kepada *AI* hanya karena *AI* tampak cerdas atau otonom (Khasri, 2024). Sabu dan Asmiyanto (2026) membuka kemungkinan pembacaan yang lebih bertingkat, yaitu *AI* dapat dipahami sebagai agen dalam level tertentu tanpa harus langsung diposisikan sebagai pemikul tanggung jawab moral reflektif (Sabu & Asmiyanto, 2026). Dengan demikian, perdebatan tidak perlu dibuat biner antara *AI* sebagai alat pasif belaka dan *AI* sebagai entitas yang disamakan dengan manusia. Yang diperlukan adalah pembedaan konseptual yang lebih presisi.

Pembedaan ini relevan bagi pemikiran Islam karena tanggung jawab moral tidak hanya dipahami sebagai persoalan efek teknis, tetapi juga berkaitan dengan kedudukan manusia sebagai subjek yang diberi amanah, akal, niat, dan kemampuan mempertanggungjawabkan tindakan. Dalam wilayah normatif-keagamaan, penggunaan *AI* tidak dapat dilepaskan dari pertanyaan tentang otoritas, validitas pengetahuan, kehati-hatian epistemik, dan tanggung jawab manusia. Azhar, Bakri, Ahmad, dan Rosele (2025) menunjukkan bahwa dalam konteks tafsir Al-Quran, *AI* dapat dipakai sebagai alat bantu sekunder, tetapi tidak dapat menggantikan otoritas epistemik manusia, ulama, sanad keilmuan, dan kerangka *maqasid* (Azhar, Bakri, Ahmad, & Rosele, 2025). Walaupun artikel ini bukan kajian tafsir digital, prinsip tersebut menegaskan bahwa fungsi operasional *AI* tidak identik dengan otoritas normatif atau tanggung jawab moral.

Kesenjangan yang hendak dijawab artikel ini bersifat terbatas. Artikel ini tidak mengklaim bahwa literatur tentang agensi moral *AI*, akuntabilitas, atau *responsibility gap* (kesenjangan tanggung jawab) belum

tersedia. Literatur tersebut sudah luas. Kebutuhan yang lebih tepat adalah penyusunan batas konseptual antara agensi rasional-operasional dan agensi moral agar diskusi *AI* tidak melompat dari performa teknis menuju pemindahan tanggung jawab moral. Dalam diskusi publik maupun kelembagaan, istilah agen, otonom, bertanggung jawab, dapat dimintai pertanggungjawaban, dan moral kerap bergerak pada level analisis yang berbeda. Ketika level-level tersebut dicampur, tanggung jawab manusia dapat menjadi kabur: pengembang dapat menyalahkan sistem, pengguna dapat menyalahkan rekomendasi mesin, institusi dapat bersembunyi di balik kompleksitas model, dan masyarakat dapat memperlakukan *AI* seolah-olah ia sendiri adalah pihak yang patut dipuji atau disalahkan.

Kontribusi artikel ini terletak pada penyusunan batas tersebut dalam bahasa konseptual yang tetap terhubung dengan pertanggungjawaban manusia. Dari sisi filsafat *AI*, artikel ini memisahkan agensi rasional dari agensi moral tanpa menolak kegunaan istilah agen dalam ilmu *AI*. Dari sisi pemikiran Islam, artikel ini menegaskan bahwa kecanggihan alat tidak menghapus amanah pengguna, pengembang, pendidik, dan institusi. Dengan demikian, kontribusi artikel bukan pada penemuan teori *AI* baru, melainkan pada klarifikasi tanggung jawab: *AI* dapat membantu tindakan manusia, tetapi manusia tetap menjadi pihak yang harus menilai tujuan, mengawasi proses, dan menjawab akibat.

Pertanyaan utama artikel ini adalah: bagaimana batas konseptual antara *AI* sebagai agen rasional dan *AI* sebagai agen moral? Pertanyaan tersebut diturunkan menjadi beberapa fokus: dalam arti apa *AI* dapat disebut agen rasional; mengapa rasionalitas instrumental tidak cukup untuk agensi moral; syarat apa saja yang biasanya diperdebatkan dalam atribusi agensi moral; dan bagaimana pembedaan tersebut memperjelas tanggung jawab manusia dalam

pemikiran Islam. Artikel ini berargumen bahwa *AI* dapat disebut agen rasional dalam arti teknis-operasional, tetapi status tersebut tidak otomatis melahirkan agensi moral. Selama syarat pemahaman normatif, intensionalitas, kesadaran, dan kemampuan memberi pertanggungjawaban belum terpenuhi secara memadai, tanggung jawab moral harus tetap ditempatkan pada manusia, organisasi, dan institusi yang mendesain, memakai, mengawasi, dan mengambil manfaat dari *AI*.

### METODE

Artikel ini menggunakan metode penelitian kepustakaan dengan pendekatan filosofis-konseptual dan hermeneutik-interpretatif. Data penelitian berupa literatur primer dan sekunder yang membahas *AI* sebagai agen rasional, agensi moral, intensionalitas, pemahaman, kesadaran, antropomorfisme, akuntabilitas, kesenjangan tanggung jawab, tata kelola *AI* yang bertanggung jawab, serta penggunaan *AI* dalam kerangka epistemologi Islam. Literatur teknis tentang agen rasional dipakai untuk membangun definisi awal *AI* sebagai sistem yang mempersepsi, bertindak, dan diorientasikan oleh ukuran kinerja. Literatur filsafat pikiran dan etika *AI* dipakai untuk menguji apakah struktur operasional tersebut cukup untuk agensi moral. Literatur akuntabilitas dan kesenjangan tanggung jawab dipakai untuk menurunkan implikasi tanggung jawab manusia-institusional. Literatur pemikiran Islam dipakai untuk membaca hubungan antara teknologi, amanah, otoritas keilmuan, dan pertanggungjawaban manusia.

Proses analisis dilakukan melalui beberapa tahap. *Pertama*, sumber dibaca untuk memahami argumen internal masing-masing. *Kedua*, konsep-konsep utama diklasifikasikan ke dalam klaster: agensi rasional, rasionalitas instrumental, pemahaman dan intensionalitas, kesadaran dan status moral, agensi moral, antropomorfisme, akuntabilitas, kesenjangan

tanggung jawab, serta tanggung jawab manusia dalam pemikiran Islam. *Ketiga*, relasi antarkonsep dibandingkan untuk menentukan apakah suatu klaim berada pada level teknis, psikologis, moral, atau institusional. *Keempat*, hasil pembacaan disintesis menjadi arsitektur argumentatif yang membedakan tiga level: *AI* sebagai agen operasional, *AI* sebagai objek atribusi mental atau moral, dan manusia-institusi sebagai letak pertanggungjawaban.

Keabsahan kajian dijaga dengan membedakan klaim langsung dari sumber, inferensi berbasis sumber, dan sintesis konseptual. Misalnya, definisi teknis agen rasional dirujuk pada Russell dan Norvig (2021), sedangkan kesimpulan bahwa definisi tersebut belum cukup untuk agensi moral dibangun melalui perbandingan dengan Coeckelbergh (2020), Khasri (2024), Searle (1980), Wittgenstein (1953), Novelli, Taddeo, dan Floridi (2024), serta Santoni de Sio dan Mecacci (2021). Dengan cara ini, artikel tidak menjadikan satu sumber sebagai pemutus seluruh persoalan, melainkan memetakan batas masing-masing konsep dan menempatkannya dalam kerangka tanggung jawab manusia.

Kajian ini tidak dimaksudkan sebagai pembahasan hukum Islam yang komprehensif atau kajian tafsir digital. Sumber Azhar et al. (2025) digunakan untuk memperjelas satu prinsip epistemologis: dalam domain normatif-keagamaan, *AI* dapat membantu pekerjaan manusia, tetapi tidak menggantikan tanggung jawab, otoritas, dan pertimbangan manusia. Prinsip ini kemudian dipakai untuk membaca persoalan agensi moral *AI* secara lebih luas. Kemampuan teknis *AI* perlu ditempatkan di bawah amanah manusia, bukan dijadikan alasan untuk mengosongkan tanggung jawab moral.

### **HASIL DAN PEMBAHASAN** **AI sebagai Agen Rasional**

Langkah pertama untuk menjernihkan persoalan adalah mengakui bahwa istilah agen tidak selalu bermakna moral. Dalam

ilmu *AI*, agen dapat dipahami secara teknis sebagai sistem yang menerima masukan dari lingkungan, memproses informasi, dan menghasilkan tindakan atau keluaran. Russell dan Norvig (2021) menjelaskan bahwa rasionalitas agen berkaitan dengan tindakan yang diharapkan memaksimalkan ukuran kinerja berdasarkan persepsi, pengetahuan, tindakan yang tersedia, dan lingkungan tugas (Russell & Norvig, 2021). Kerangka ini menjelaskan mengapa *AI* dapat disebut agen rasional tanpa perlu menganggapnya sebagai subjek moral. Ia rasional dalam arti diarahkan pada capaian tertentu, bukan dalam arti memahami nilai moral dari tujuan tersebut.

Rasionalitas teknis selalu bergantung pada desain tujuan dan ukuran kinerja. Jika sistem rekomendasi dioptimalkan untuk meningkatkan keterlibatan pengguna, sistem dapat memilih keluaran yang efektif bagi tujuan itu. Jika sistem deteksi risiko dioptimalkan untuk meminimalkan kesalahan tertentu, ia dapat menyesuaikan prediksi berdasarkan data. Namun ukuran kinerja tersebut bukan hasil refleksi moral sistem; ia merupakan struktur tujuan yang diberikan, dirancang, dipilih, atau disetujui manusia dan institusi. Di sinilah rasionalitas *AI* bersifat instrumental. Sistem dapat sangat efektif memilih sarana, tetapi efektivitas sarana belum menjawab apakah tujuan itu baik, adil, sah, atau bertanggung jawab.

Pembedaan ini penting karena *AI* kontemporer sering tampak lebih dari sekadar alat pasif. Sistem generatif dapat menjawab pertanyaan, menyusun argumen, memberi saran, dan menyesuaikan gaya komunikasi. Dalam interaksi sosial, pengguna dapat mengalami *AI* sebagai lawan bicara yang responsif. Placani (2024) menunjukkan bahwa antropomorfisme terhadap *AI* dapat diperkuat oleh cara sistem dirancang dan dipresentasikan (Placani, 2024). Ladak, Loughnan, dan Wilks (2023) juga memperlihatkan bahwa manusia dapat mengatribusikan kapasitas moral atau mental kepada *AI* dalam cara yang tidak selalu

sejalan dengan status normatif sistem itu sendiri. Dengan kata lain, tampilan agensi dapat memperkuat kesan bahwa *AI* bermaksud atau bertanggung jawab, padahal yang terlihat mungkin hanya keberhasilan operasional dalam memproduksi respons yang sesuai.

Karena itu, artikel ini membedakan agensi operasional dari agensi moral. Agensi operasional menunjuk pada kemampuan sistem bertindak menurut struktur masukan-keluaran, tujuan, model, atau ukuran kinerja. Agensi moral menunjuk pada kapasitas yang lebih kuat: memahami alasan normatif, mempertimbangkan baik-buruk, memberi jawaban atas tindakan, dan menjadi pihak yang layak dipuji, dikritik, atau dimintai pertanggungjawaban. *AI* dapat memenuhi level pertama tanpa memenuhi level kedua. Perbedaan ini memungkinkan pengakuan terhadap kecanggihan *AI* tanpa personifikasi moral yang berlebihan.

### **Rasionalitas Instrumental dan Batas Pemahaman Normatif**

Rasionalitas instrumental *AI* perlu dibedakan dari pemahaman normatif. Suatu sistem dapat menghasilkan jawaban yang tampak tepat secara bahasa atau efisien secara tugas, tetapi hal itu belum membuktikan bahwa sistem memahami makna dan alasan moral sebagaimana manusia memahami tindakan dalam ruang sosial. Searle (1980) mengajukan kritik klasik terhadap penyamaan manipulasi simbol dengan pemahaman (Searle, 1980). Walaupun konteks *AI* saat ini berbeda dari program simbolik klasik, kritik tersebut tetap berguna sebagai pagar: keberhasilan menghasilkan keluaran yang benar secara pola tidak otomatis menunjukkan pemahaman intrinsik.

Wittgenstein (1953) juga membantu melihat bahwa mengikuti aturan bukan sekadar menghasilkan urutan tindakan yang tampak sesuai, melainkan terkait dengan praktik, penggunaan, dan kemampuan bertindak dalam bentuk kehidupan tertentu.

Jika gagasan ini diterapkan secara hati-hati pada *AI*, maka kemampuan sistem mengikuti pola bahasa atau instruksi belum cukup untuk menunjukkan bahwa ia menguasai norma dalam arti manusiawi. *AI* dapat menghasilkan respons yang sejalan dengan aturan sosial karena dilatih pada data dan diarahkan oleh prosedur optimisasi, tetapi pertanyaan tentang apakah ia memahami alasan normatif di balik aturan itu tetap terbuka.

Avramides (2001) mengingatkan bahwa persoalan pikiran orang lain tidak dapat direduksi menjadi akses langsung ke pengalaman batin, tetapi juga tidak boleh disederhanakan menjadi keluaran lahiriah belaka (Avramides, 2001). Posisi ini penting agar artikel tidak jatuh pada dua ekstrem. Ekstrem pertama adalah behaviorisme kasar: jika *AI* berperilaku seperti memahami, maka ia pasti memahami. Ekstrem kedua adalah skeptisisme privatistis: karena pengalaman batin *AI* tidak dapat diakses secara langsung, maka semua bentuk atribusi mental dianggap keliru. Artikel ini mengambil posisi yang lebih hati-hati: perilaku dan performa *AI* relevan untuk analisis, tetapi belum cukup untuk atribusi agensi moral penuh tanpa kriteria tambahan tentang pemahaman, intensionalitas, dan pertanggungjawaban.

Dalam konteks moral, pemahaman normatif tidak hanya berarti dapat menyebutkan aturan. Seseorang dapat mengulang prinsip etis tanpa sungguh-sungguh memikul tanggung jawab atas tindakannya. Agensi moral membutuhkan hubungan antara alasan, keputusan, tindakan, dan kemampuan memberi jawaban ketika tindakan dipertanyakan. Novelli et al. (2024) menempatkan akuntabilitas dalam relasi antara agen dan forum, yakni adanya pihak yang dapat diminta memberi penjelasan dan justifikasi (Novelli, Taddeo, & Floridi, 2024). Jika sistem *AI* tidak memiliki kapasitas menjawab dalam arti normatif, maka mekanisme akuntabilitas harus diarahkan pada manusia dan institusi yang dapat

memberi alasan, memperbaiki prosedur, menerima sanksi, dan mengubah praktik.

Dengan demikian, *AI* dapat memproses norma sebagai data, tetapi belum tentu memahami norma sebagai tuntutan moral. Sistem dapat menghasilkan jawaban tentang keadilan, amanah, atau tanggung jawab karena istilah-istilah itu muncul dalam korpus pelatihan atau instruksi pengguna. Akan tetapi, kemampuan memproduksi bahasa moral belum identik dengan kapasitas menjadi subjek yang diwajibkan oleh norma moral. Perbedaan ini menjadi pusat argumentasi artikel: rasionalitas instrumental adalah syarat teknis bagi agensi operasional, tetapi agensi moral menuntut lebih dari keberhasilan instrumen.

### Kesadaran, Intensionalitas, dan Syarat Agensi Moral

Literatur tentang kesadaran *AI* memperlihatkan bahwa hubungan antara kecerdasan, agensi, dan kesadaran tidak sederhana. Butlin et al. (2023) mengusulkan pendekatan berbasis indikator untuk menilai kemungkinan kesadaran pada *AI*, tetapi tidak menyamakan kecanggihan komputasional dengan kesadaran (Butlin dkk., 2023). Chalmers (2023) juga membedakan kemampuan bahasa dan kecerdasan dari kesadaran, sambil tetap membuka pertanyaan tentang kemungkinan sistem masa depan (Chalmers, 2023). Seth (2025) menekankan bahwa sistem yang tampak sadar dapat menciptakan persoalan etis bahkan ketika status kesadarannya belum pasti (Seth, 2025). Rangkaian literatur ini membantu artikel mengambil posisi non-dogmatis: tidak perlu menyatakan bahwa *AI* mustahil memiliki status mental apa pun di masa depan, tetapi juga tidak boleh menganggap *AI* saat ini sudah menjadi subjek moral hanya karena performanya mengesankan.

Kesadaran sendiri tidak identik dengan agensi moral. Nagel (1974) menunjukkan pentingnya sudut pandang subjektif dalam pengalaman (Nagel, 2024), sementara Chalmers (1996) membuka ruang bagi

diskusi filosofis tentang kesadaran nonbiologis (Chalmers, 1996). Namun agensi moral melibatkan lebih dari pertanyaan apakah ada pengalaman subjektif. Ia juga berkaitan dengan kemampuan merespons alasan, memahami norma, mengendalikan tindakan, dan menjadi pihak yang dapat dimintai pertanggungjawaban. Karena itu, status moral, kepatutan sebagai penerima pertimbangan moral, agensi moral, dan tanggung jawab moral perlu dibedakan. Sebuah entitas yang memiliki pengalaman mungkin patut dipertimbangkan secara moral, tetapi belum tentu menjadi agen yang bertanggung jawab. Sebaliknya, sistem yang bertindak secara efektif belum tentu memiliki pengalaman atau tanggung jawab.

Khasri (2024) menekankan pentingnya kesadaran dan subjektivitas dalam menilai agensi moral *AI* (Khasri, 2024). Coeckelbergh (2020) mengingatkan bahwa *AI* dapat memiliki peran moral dalam sistem tindakan manusia, namun itu tidak langsung berarti *AI* adalah agen moral penuh (Coeckelbergh, 2020). Sabu dan Asmiyanto (2026) memberi ruang bagi pemahaman agensi artifisial yang lebih bertingkat, terutama melalui pembacaan level abstraksi dan moralitas terdistribusi (Sabu & Asmiyanto, 2026). Ketiga arah ini dapat disintesis sebagai berikut: *AI* dapat dianalisis sebagai agen pada level operasional atau fungsional, tetapi agensi moral harus diuji melalui syarat tambahan yang tidak habis oleh performa teknis.

Syarat tambahan tersebut sekurang-kurangnya mencakup empat unsur. *Pertama*, intensionalitas, yakni keterarahan tindakan bukan hanya sebagai korelasi masukan-keluaran, tetapi sebagai relasi bermakna terhadap objek, alasan, atau tujuan. *Kedua*, pemahaman normatif, yakni kemampuan menangkap alasan moral sebagai alasan, bukan sekadar memprediksi kalimat moral yang sesuai. *Ketiga*, kapasitas reflektif, yakni kemampuan mengevaluasi tindakan dalam kaitan dengan norma dan konsekuensi. *Keempat*, kemampuan memberi

pertanggungjawaban, yakni posisi sebagai pihak yang dapat memberi jawaban, menerima kritik, dan memikul konsekuensi normatif. Selama unsur-unsur ini belum dapat ditunjukkan secara memadai pada AI, atribusi agensi moral penuh masih lemah.

Namun kehati-hatian ini tidak berarti AI kembali diperlakukan sebagai benda netral tanpa relevansi moral. AI dapat memengaruhi pilihan, membentuk perhatian, memperkuat bias, mengatur akses, dan memediasi hubungan sosial. Pan dan de Graaf (2025) menunjukkan bahwa hubungan manusia-chatbot dapat menghasilkan pengalaman dukungan sosial dan timbal balik (Pan & de Graaf, 2025). Walaupun hal itu tidak membuktikan kedirian moral chatbot, dampaknya nyata bagi pengguna. Karena itu, pertanyaan etis tidak hilang; hanya letak tanggung jawabnya yang perlu diarahkan dari personifikasi AI menuju jaringan manusia, desain, institusi, dan praktik penggunaan.

### Tanggung Jawab Manusia dalam Pemikiran Islam

Dalam pemikiran Islam, persoalan AI tidak cukup dijawab dengan menyebut sistem sebagai cerdas atau otonom. Tanggung jawab moral berkaitan dengan manusia sebagai subjek yang memiliki akal, niat, amanah, dan kemampuan mempertanggungjawabkan tindakan. Karena itu, teknologi yang membantu tindakan manusia tidak otomatis menggantikan posisi manusia sebagai pemikul tanggung jawab. AI dapat menjadi alat yang memperluas kemampuan manusia, tetapi perlu tetap berada dalam kerangka pengawasan, kehati-hatian, dan orientasi nilai.

Kerangka ini juga mencegah reduksi etika menjadi sekadar kepatuhan prosedural. Dalam tradisi Islam, tindakan tidak hanya dinilai dari keluaran lahiriah, tetapi juga dari orientasi niat, pemahaman, tanggung jawab, dan akibat yang ditimbulkan. Ketika AI menghasilkan jawaban yang tampak benar, manusia masih perlu menilai apakah jawaban itu sah, adil, proporsional, dan sesuai dengan tujuan kemaslahatan. Dengan

demikian, penggunaan AI memerlukan adab epistemik: tidak tergesa-gesa menganggap keluaran mesin sebagai otoritas, tidak menyembunyikan keterlibatan AI ketika transparansi diperlukan, dan tidak memakai kompleksitas teknologi sebagai alasan untuk menghindari pertanggungjawaban.

Azhar et al. (2025) memberi contoh penting dari wilayah tafsir Al-Quran (Azhar dkk., 2025). Dalam konteks itu, AI dapat membantu pencarian, pengolahan teks, dan analisis awal, tetapi tidak dapat menggantikan otoritas tafsir yang berakar pada epistemologi Islam, *sanad*, penguasaan ilmu, *maqasid*, dan pertimbangan ulama. Walaupun artikel ini tidak membahas tafsir secara khusus, prinsipnya dapat dipakai untuk membaca AI dalam ranah normatif yang lebih luas: kemampuan komputasional tidak identik dengan otoritas normatif. AI boleh berfungsi sebagai asisten, tetapi manusia tetap harus menilai, menyaring, mempertanggungjawabkan, dan memastikan bahwa penggunaan AI tidak merusak tujuan etis.

Dalam pendidikan Islam, kajian keagamaan, layanan konsultasi, atau produksi pengetahuan, persoalan utamanya bukan hanya apakah AI akurat, tetapi siapa yang bertanggung jawab atas otoritas jawaban, dampak kesalahan, bias, penyederhanaan, dan kemungkinan hilangnya adab keilmuan. Sistem AI tidak memiliki kedudukan sebagai mukallaf dalam pengertian moral-keagamaan; ia tidak memikul niat, dosa, pahala, atau pertanggungjawaban spiritual. Karena itu, pengalihan tanggung jawab moral kepada AI sulit dipertahankan dalam kerangka pemikiran Islam yang menempatkan manusia sebagai subjek amanah.

Poin ini tidak berarti setiap penggunaan AI dalam ranah Islam harus ditolak. Sebaliknya, AI dapat dipakai secara bertanggung jawab bila fungsinya dibatasi sebagai alat bantu. Ia dapat membantu pencarian literatur, transliterasi, pemetaan tema, perbandingan teks, atau penyusunan bahan ajar. Namun hasilnya harus diperiksa

oleh manusia yang memiliki kompetensi. Dalam bahasa artikel ini, *AI* dapat menjadi agen operasional dalam sistem kerja manusia, tetapi bukan agen moral yang menggantikan kewajiban manusia untuk menimbang, mengoreksi, dan bertanggung jawab.

Dengan demikian, pemikiran Islam memperkuat perbedaan antara agensi rasional dan agensi moral. *AI* dapat rasional secara instrumental karena mampu memilih cara efektif menuju tujuan tertentu. Akan tetapi, pertanyaan tentang tujuan baik, niat, amanah, keadilan, dan pertanggungjawaban tetap berada pada manusia. Jika *AI* digunakan untuk tujuan yang merugikan, diskriminatif, menyesatkan, atau merendahkan martabat manusia, sistem tidak dapat menjadi tempat terakhir penyaluran tanggung jawab. Manusia dan institusi yang mendesain, memilih data, menentukan tujuan, menerapkan sistem, dan menggunakan hasilnya tetap harus menjawab.

### **Akuntabilitas dan Kesenjangan Tanggung Jawab**

Kesenjangan tanggung jawab sering muncul ketika tindakan atau keluaran *AI* sulit dilacak kepada satu aktor manusia. Santoni de Sio dan Mecacci (2021) menunjukkan bahwa kesenjangan tersebut tidak tunggal; ia dapat muncul dalam bentuk kesenjangan kesalahan moral, akuntabilitas, tanggung jawab hukum, atau tanggung jawab aktif (de Sio & Mecacci, 2021). Königs (2022) mengingatkan bahwa tidak semua kesenjangan harus dipahami sebagai kekosongan metafisik (Königs, 2022). Sebagian kesulitan atribusi dapat bersifat epistemik atau institusional: penjelasan sistem belum memadai, pembagian peran belum jelas, atau prosedur pertanggungjawaban belum dirancang dengan baik. Dengan demikian, kompleksitas *AI* tidak boleh langsung dipakai sebagai alasan bahwa tidak ada pihak manusia yang bertanggung jawab.

Novelli et al. (2024) membantu memperjelas akuntabilitas sebagai relasi

kemampuan memberi pertanggungjawaban (Novelli dkk., 2024). Dalam sistem *AI*, yang perlu ditata bukan hanya kemampuan model menjelaskan keluaran, tetapi juga siapa yang wajib memberi alasan kepada siapa, berdasarkan standar apa, dan dengan konsekuensi apa. Penjelasan teknis dapat membantu akuntabilitas, tetapi tidak menggantikannya. Penjelasan model tidak sama dengan pertanggungjawaban moral atau institusional. Jika sebuah keputusan berbasis *AI* merugikan seseorang, jawaban yang dibutuhkan bukan hanya penjelasan bahwa model memberi skor tertentu, melainkan siapa yang memilih model itu, siapa yang memvalidasinya, siapa yang mengawasi dampaknya, siapa yang memperbaikinya, dan siapa yang bertanggung jawab atas kerugian.

Tata kelola *AI* yang bertanggung jawab menempatkan pertanggungjawaban dalam siklus hidup teknologi. Papagiannidis, Mikalef, dan Conboy (2025) menunjukkan pentingnya tata kelola yang mencakup desain, penerapan, pengawasan, pembagian peran, dan evaluasi (Papagiannidis, Mikalef, & Conboy, 2025). Dalam perspektif artikel ini, tata kelola tersebut perlu dibaca sebagai konsekuensi praktis dari batas agensi moral *AI*. Karena *AI* belum dapat diposisikan sebagai agen moral penuh, tanggung jawab harus didistribusikan pada aktor manusia dan institusi, bukan dibiarkan mengambang di antara sistem teknis dan pengguna akhir. Distribusi tanggung jawab ini tidak berarti semua orang bertanggung jawab secara sama; sebaliknya, perlu ada perbedaan peran antara pengembang, penyedia, institusi pengguna, operator, pendidik, regulator, dan pengguna.

Pemikiran Islam dapat memperkuat arah ini melalui gagasan amanah. Amanah teknologi bukan hanya kewajiban memakai alat secara efisien, tetapi juga kewajiban menjaga akibat sosial, epistemik, dan moral dari penggunaannya. Dalam pendidikan dan studi Islam, misalnya, institusi tidak cukup menyatakan bahwa *AI* hanya alat bantu. Institusi perlu menetapkan batas penggunaan,

standar verifikasi, transparansi kepada peserta didik atau pembaca, dan mekanisme koreksi ketika *AI* menghasilkan kekeliruan. Pengguna juga perlu memahami bahwa menyerahkan penilaian moral kepada *AI* merupakan bentuk pelepasan tanggung jawab yang tidak dapat dibenarkan.

Dengan demikian, artikel ini menawarkan arsitektur tiga tingkat. Tingkat pertama adalah agensi operasional: *AI* dapat bertindak sebagai agen rasional dalam kerangka teknis karena ia merespons lingkungan dan mengejar ukuran kinerja. Tingkat kedua adalah atribusi mental-moral: manusia dapat menganggap *AI* memahami, berniat, atau bertanggung jawab karena sistem tampil responsif dan menyerupai perilaku sosial. Tingkat ketiga adalah pertanggungjawaban normatif: selama syarat agensi moral belum terpenuhi, tanggung jawab harus tetap diletakkan pada manusia dan institusi. Arsitektur ini menahan dua kesalahan sekaligus. Kesalahan pertama adalah merendahkan *AI* seolah-olah tidak memiliki dampak moral karena dianggap hanya alat. Kesalahan kedua adalah mempersonifikasikan *AI* seolah-olah dampak moralnya membuat ia sendiri menjadi penanggung jawab moral.

Implikasinya bagi kajian Islam adalah perlunya etika penggunaan *AI* yang menggabungkan kecermatan konseptual dan tanggung jawab praktis. *AI* tidak harus ditolak sebagai inovasi, tetapi juga tidak boleh diberi otoritas moral yang tidak dimilikinya. Dalam ranah pendidikan dan studi Islam, *AI* dapat mendukung pembelajaran, penelitian, dan pengelolaan informasi, tetapi manusia tetap harus menjaga niat, amanah, akurasi, keadilan, dan pertanggungjawaban. Dengan kata lain, *AI* dapat menjadi instrumen rasional, tetapi bukan pengganti subjek moral.

### KESIMPULAN

Artikel ini menyimpulkan bahwa *AI* dapat disebut agen rasional dalam arti teknis-operasional, tetapi tidak otomatis menjadi

agen moral. Sebagai agen rasional, *AI* dapat mempersepsi masukan, mengolah informasi, merespons lingkungan, dan diarahkan oleh tujuan atau ukuran kinerja. Namun rasionalitas semacam itu bersifat instrumental. Ia menjelaskan kemampuan sistem memilih sarana yang efektif, bukan kemampuan memahami alasan moral, memikul niat, memberi pertanggungjawaban, atau menjadi subjek amanah. Batas ini perlu dijaga terutama ketika *AI* digunakan dalam ruang pendidikan dan keagamaan.

Agensi moral membutuhkan syarat tambahan seperti intensionalitas, pemahaman normatif, kesadaran atau subjektivitas dalam arti yang relevan, kapasitas reflektif, dan kemampuan memberi pertanggungjawaban. Literatur yang ditelaah menunjukkan bahwa performa linguistik, otonomi teknis, respons sosial, atau keberhasilan optimisasi belum cukup untuk memenuhi syarat tersebut. Karena itu, atribusi moral kepada *AI* perlu dibatasi. *AI* dapat memiliki dampak moral dan dapat menjadi bagian dari sistem tindakan manusia, tetapi tanggung jawab moral tidak boleh dipindahkan begitu saja kepada sistem komputasional.

Dalam pemikiran Islam, perbedaan ini semakin penting karena tanggung jawab terkait dengan manusia sebagai pemikul amanah. *AI* dapat digunakan sebagai alat bantu dalam pendidikan, studi Islam, atau kegiatan intelektual lain, tetapi tidak menggantikan otoritas, kehati-hatian, niat, dan pertanggungjawaban manusia. Penggunaan *AI* harus diarahkan oleh manusia dan institusi yang mampu menilai tujuan, dampak, dan batasnya. Dengan demikian, posisi yang lebih tepat bukan menolak *AI* secara total atau mempersonifikasikannya sebagai subjek moral, melainkan menempatkannya sebagai agen operasional yang berada di bawah tanggung jawab manusia.

Saran penelitian berikutnya adalah memperluas kajian pemikiran Islam tentang *AI* melalui konsep *taklif*, amanah, *mas'uliyah*, niat, akal, dan *maqasid* secara

lebih mendalam. Penelitian lanjutan juga dapat menguji bagaimana lembaga pendidikan Islam merumuskan pedoman penggunaan AI yang menjaga akurasi, adab keilmuan, transparansi, dan akuntabilitas. Di sisi lain, kajian filsafat AI perlu terus memeriksa kemungkinan bentuk agensi moral terbatas pada sistem masa depan tanpa tergesa-gesa menyamakan performa teknis dengan tanggung jawab moral.

### DAFTAR PUSTAKA

- Avramides, A. (2001). *Other Minds*. London and New York: Routledge.
- Azhar, M. H. M., Bakri, M. F. M., Ahmad, K., & Rosele, M. I. (2025). Etika dan Had Penggunaan Kecerdasan Buatan (AI) dalam Tafsir Al-Quran Menurut Kerangka Epistemologi Ilmu Islam. *QURANICA: International Journal of Quranic Research*, 14, 97–124.
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., ... VanRullen, R. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. <https://doi.org/10.48550/arXiv.2308.08708>
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, D. J. (2023). *Could a Large Language Model be Conscious?* Boston Review.
- Coeckelbergh, M. (2020). *AI Ethics*. Cambridge, MA: The MIT Press. Diambil dari <https://lcn.loc.gov/2019018827>
- de Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them. *Philosophy & Technology*, 34, 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>
- Khasri, M. R. K. (2024). Etika Kecerdasan Artifisial (AI): Sebuah Eksplorasi Metaetika atas Kebijakan Regulasi AI di Indonesia dan Internasional. *Refleksi: Jurnal Filsafat dan Pemikiran Keislaman*, 24(1). <https://doi.org/10.14421/ref.v24i1.6047>
- Königs, P. (2022). Artificial Intelligence and Responsibility Gaps: What Is the Problem? *Ethics and Information Technology*, 24, 36. <https://doi.org/10.1007/s10676-022-09643-0>
- Nagel, T. (2024). *What Is It Like to Be a Bat?* New York, NY: Oxford University Press. <https://doi.org/10.1093/oso/9780197752791.001.0001>
- Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in Artificial Intelligence: What It Is and How It Works. *AI & Society*, 39, 1871–1882. <https://doi.org/10.1007/s00146-023-01635-y>
- Pan, S., & de Graaf, M. M. A. (2025). Developing a Social Support Framework: Understanding the Reciprocity in Human-Chatbot Relationship. *CHI Conference on Human Factors in Computing Systems*, 1–13. Yokohama, Japan: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713503>
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible Artificial Intelligence Governance: A Review and Research Framework. *Journal of Strategic Information Systems*, 34, 101885. <https://doi.org/10.1016/j.jsis.2024.101885>
- Placani, A. (2024). Anthropomorphism in AI: Hype and Fallacy. *AI and Ethics*, 4, 691–698. <https://doi.org/10.1007/s43681-024-00419-4>
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4 ed.). Pearson.
- Sabu, F. X., & Asmiyanto, T. (2026). Rethinking Moral Agency: Luciano Floridi and the Ethics of AI in the Age of ChatGPT. *Khizanah al-Hikmah: Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan*, 14(1). <https://doi.org/10.24252/kah.v14i1a5>