

ARTIFICIAL INTELLIGENCE AND THE WISDOM OF THE HEART: MAINTAINING HUMAN DIGNITY AMIDST THE RISKS OF DIGITAL TECHNOLOGY

KECERDASAN BUATAN DAN KEBIJAKSANAAN HATI: MENJAGA MARTABAT MANUSIA DI TENGAH RISIKO TEKNOLOGI DIGITAL

Rosalia Enny Astuti^{1*}, Herkulana Mekarryani Soeryamassoeka²

Sekolah Tinggi Agama Katolik Negeri, Pontianak, Indonesia^{1*,2}

Email : rosalia.enny@gmail.com

Abstract: This study examines the ethical and existential challenges arising from the rapid advancement of Artificial Intelligence (AI) through the perspectives of philosophy of science and Catholic theology. The primary focus is the risk of digital reductionism, which threatens human dignity as *Imago Dei*. This research employs a qualitative approach using an integrative literature review, analyzing the documents *Antiqua et Nova* (2025) and *Christus Vivit* (2019), alongside the critical perspectives of Hubert Dreyfus and Nick Bostrom.

The findings reveal a fundamental ontological gap between human intelligence and AI. Dreyfus highlights the limitations of AI in replicating embodied intuition and its inability to resolve the *frame problem*. Meanwhile, Bostrom warns of the risks of *superintelligence* due to the difficulty of aligning machine objectives with human values (*alignment problem*). Axiologically, this distinction is reflected in the difference between *mêtis* as functional AI intelligence and *noûs* as the contemplative intellect of human beings.

This study proposes the concept of *Techno-Humanitarian Balance* as a normative framework to position AI under human moral authority. This framework is essential to prevent the degradation of human identity into algorithmic constructs and to ensure that technology continues to serve humanity.

Keywords: Artificial Intelligence, Human Dignity, Hubert Dreyfus, Nick Bostrom, *Antiqua et Nova*.

Abstrak: Tulisan ini menganalisis tantangan etis dan eksistensial yang muncul dari akselerasi Kecerdasan Buatan (AI) melalui perspektif filsafat ilmu dan teologi Katolik. Fokus utama penelitian ini adalah risiko reduksionisme digital yang berpotensi mereduksi martabat manusia sebagai *Imago Dei*. Penelitian ini menggunakan pendekatan kualitatif dengan metode studi literatur integratif terhadap dokumen *Antiqua et Nova* (2025) dan *Christus Vivit* (2019), serta pemikiran Hubert Dreyfus dan Nick Bostrom.

Hasil penelitian menunjukkan adanya kesenjangan ontologis yang mendasar antara kecerdasan manusia dan AI. Kritik Dreyfus menegaskan keterbatasan AI dalam mereplikasi intuisi berbasis perwujudan tubuh (*embodiment*) serta ketidakmampuannya menyelesaikan *frame problem*. Di sisi lain, Bostrom menyoroti risiko *superintelligence* yang muncul dari persoalan penyesuaian nilai (*alignment problem*). Secara aksiologis, perbedaan ini tercermin dalam distingsi antara *mêtis* sebagai kecerdasan fungsional AI dan *noûs* sebagai intelek kontemplatif manusia.

Penelitian ini menawarkan konsep *Techno-Humanitarian Balance* sebagai kerangka normatif untuk menempatkan AI secara proporsional di bawah otoritas moral manusia. Kerangka ini penting untuk mencegah degradasi identitas manusia menjadi entitas algoritmik serta memastikan bahwa teknologi tetap berfungsi sebagai sarana yang melayani kemanusiaan.

Kata Kunci: Kecerdasan Buatan, Martabat Manusia, Hubert Dreyfus, Nick Bostrom, *Antiqua et Nova*.

PENDAHULUAN

Perkembangan Kecerdasan Buatan (Artificial Intelligence/AI) telah mendorong transformasi signifikan dalam peradaban digital modern. Dalam konteks menuju Industri 5.0, teknologi tidak lagi hanya berfokus pada efisiensi, tetapi juga pada upaya menempatkan kembali manusia sebagai pusat sistem (Babkin et al., 2022). AI dirancang untuk meningkatkan produktivitas dan kualitas hidup melalui kolaborasi antara manusia dan mesin. Dalam perspektif teologis, Gereja Katolik melalui dokumen *Antiqua et Nova* (2025) memandang kemajuan teknologi sebagai wujud kreativitas manusia yang merupakan anugerah dari Tuhan (Fransiskus, 2025). Dengan demikian, AI pada dasarnya memiliki potensi positif sebagai sarana untuk menyempurnakan ciptaan dan melayani kebaikan bersama.

Namun demikian, perkembangan AI juga memunculkan berbagai persoalan etis dan eksistensial yang tidak dapat diabaikan. Salah satu masalah utama adalah munculnya fenomena reduksionisme digital, yaitu kecenderungan untuk mereduksi realitas manusia yang kompleks menjadi sekadar data dan variabel komputasional (Fransiskus, 2025; Bezklubaya, 2023). Aspek-aspek mendasar seperti intuisi, emosi, dan hati nurani yang tidak dapat diukur secara kuantitatif berisiko terpinggirkan dalam logika algoritmik. Dampak dari kecenderungan ini tidak hanya terbatas pada hilangnya privasi atau meningkatnya kontrol digital, tetapi juga menyentuh persoalan mendasar mengenai martabat manusia sebagai *Imago Dei* (Fransiskus, 2025). Dalam konteks ini, teknologi berpotensi menggeser manusia dari subjek moral menjadi objek yang dioptimalkan dalam sistem digital.

Krisis ini semakin nyata dalam pengalaman generasi muda yang hidup dalam ekosistem digital. Dokumen *Christus Vivit* (2019) menyoroti bahwa ruang digital, meskipun menawarkan konektivitas, juga dapat menciptakan kesepian, manipulasi, dan kehilangan identitas personal (Fransiskus, 2019). Algoritma media sosial yang bekerja berdasarkan preferensi pengguna cenderung membentuk ruang gema (*echo chambers*) yang membatasi perspektif dan memperkuat homogenisasi perilaku. Akibatnya, individu berisiko kehilangan keunikan dirinya dan menjadi “fotokopi” algoritmik yang dibentuk oleh sistem digital (Satrio et al., 2024). Fenomena ini menunjukkan bahwa perkembangan AI tidak hanya berdampak pada aspek teknis, tetapi juga pada dimensi antropologis manusia.

Perdebatan mengenai hakikat dan risiko AI semakin diperkaya melalui pemikiran Hubert Dreyfus dan Nick Bostrom. Dreyfus mengkritik asumsi bahwa kecerdasan manusia dapat direduksi menjadi proses komputasi, dengan menekankan pentingnya *embodiment* sebagai dasar pengetahuan manusia (Dreyfus, 1992; Reynolds, 2024). Menurutnya, AI tidak mampu memahami konteks secara intuitif karena tidak memiliki pengalaman bertubuh, sehingga terjebak dalam *frame problem*. Sebaliknya, Bostrom menyoroti risiko dari kemungkinan munculnya *superintelligence*, yaitu sistem yang melampaui kemampuan kognitif manusia, namun tidak memiliki jaminan keselarasan nilai (*alignment problem*) (Bostrom, 2014; Goertzel, 2015). Kedua perspektif ini

menunjukkan bahwa AI menghadirkan paradoks: di satu sisi terbatas secara ontologis, tetapi di sisi lain berpotensi melampaui kontrol manusia secara eksistensial.

Meskipun berbagai kajian telah membahas keterbatasan epistemologis AI maupun risiko eksistensialnya, sebagian besar penelitian masih cenderung memisahkan pendekatan filosofis dan teologis. Kajian yang mengintegrasikan kritik fenomenologis Hubert Dreyfus dengan analisis risiko Nick Bostrom dalam kerangka teologi Katolik masih relatif terbatas. Selain itu, diskursus mengenai AI sering kali terfokus pada aspek teknis atau etika umum, tanpa memberikan kerangka normatif yang komprehensif dalam menjaga martabat manusia di tengah perkembangan teknologi digital.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk menganalisis tantangan etis dan ontologis AI melalui integrasi perspektif filsafat dan teologi Katolik, serta merumuskan konsep *Techno-Humanitarian Balance* sebagai kerangka normatif dalam menjaga martabat manusia sebagai *Imago Dei*. Konsep ini menempatkan AI sebagai alat yang harus tetap berada di bawah otoritas moral manusia. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi konseptual dalam mengarahkan perkembangan teknologi digital agar tetap berpusat pada kemanusiaan dan tidak terjebak dalam reduksionisme digital yang mengancam identitas manusia.

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan desain studi literatur integratif (*integrative literature review*) untuk menganalisis secara kritis relasi antara perkembangan Kecerdasan Buatan (AI), filsafat ilmu, dan teologi Katolik. Pendekatan ini dipilih karena fokus penelitian tidak terletak pada pengumpulan data empiris, melainkan pada penafsiran konseptual dan sintesis teoritis terhadap berbagai sumber literatur yang relevan (Zaluchu, 2020; Gurusinga & Gurusinga, 2025). Melalui metode ini, penelitian berupaya mengidentifikasi pola, konsep kunci, serta ketegangan epistemologis dan aksiologis yang muncul dalam diskursus AI kontemporer. Selain itu, pendekatan ini memungkinkan integrasi lintas disiplin antara filsafat, teologi, dan etika teknologi dalam satu kerangka analisis yang komprehensif.

Sumber data dalam penelitian ini terdiri atas data primer dan sekunder. Data primer mencakup karya utama Hubert Dreyfus, khususnya *What Computers Still Can't Do* (1992), serta pemikiran Nick Bostrom dalam *Superintelligence: Paths, Dangers, Strategies* (2014). Selain itu, dokumen resmi Gereja Katolik, yaitu *Antiqua et Nova* (2025) dan *Christus Vivit* (2019), digunakan sebagai landasan normatif-teologis dalam menganalisis martabat manusia sebagai *Imago Dei*. Data sekunder diperoleh dari artikel jurnal ilmiah, buku akademik, dan publikasi terkait AI, etika digital, serta filsafat teknologi yang terbit dalam rentang tahun 2015–2025 guna memastikan relevansi dan aktualitas kajian (Fransiska et al., 2025).

Kriteria pemilihan literatur didasarkan pada tiga aspek utama, yaitu relevansi tematik, kredibilitas akademik, dan kontribusi konseptual terhadap permasalahan penelitian. Literatur yang dipilih harus secara langsung membahas keterbatasan epistemologis AI, risiko eksistensial teknologi, atau refleksi teologis mengenai martabat

manusia. Selain itu, sumber yang digunakan diutamakan berasal dari publikasi bereputasi dan memiliki landasan teoritis yang kuat. Proses seleksi dilakukan secara sistematis dengan mengeliminasi literatur yang bersifat deskriptif umum dan tidak memberikan kontribusi analitis terhadap fokus penelitian. Tahap ini juga mencakup proses penyaringan ulang untuk memastikan konsistensi antara sumber yang digunakan dengan tujuan penelitian.

Teknik analisis data dilakukan melalui analisis tematik induktif. Tahapan analisis meliputi pengumpulan data, reduksi data, kategorisasi tema, serta penarikan kesimpulan secara interpretatif (Sugiyono, 2013). Pada tahap awal, literatur dikaji untuk mengidentifikasi konsep kunci seperti *embodiment*, *frame problem*, *superintelligence*, dan *alignment problem*. Selanjutnya, konsep-konsep tersebut dianalisis secara komparatif dalam kerangka epistemologi dan aksiologi. Proses ini memungkinkan peneliti untuk menemukan keterkaitan antara kritik Dreyfus terhadap keterbatasan AI dan peringatan Bostrom mengenai risiko eksistensialnya. Analisis juga diarahkan untuk mengungkap dimensi normatif dari setiap konsep yang dibahas.

Untuk menjamin validitas temuan, penelitian ini menggunakan teknik triangulasi sumber dengan membandingkan perspektif filsafat, teologi, dan etika teknologi. Pendekatan ini bertujuan untuk mengurangi bias interpretasi serta memperkuat konsistensi analisis (Gurusinga & Gurusinga, 2025). Selain itu, dilakukan pembacaan kritis berulang terhadap sumber utama guna memastikan akurasi interpretasi. Hasil analisis kemudian disintesis dalam bentuk kerangka konseptual *Techno-Humanitarian Balance* sebagai kontribusi utama penelitian. Dengan demikian, metodologi ini tidak hanya berfungsi untuk memahami fenomena AI secara kritis, tetapi juga untuk merumuskan pendekatan normatif yang menempatkan teknologi dalam kerangka kemanusiaan.

Penelitian ini dibatasi pada analisis konseptual dan tidak mencakup studi empiris lapangan, sehingga temuan yang dihasilkan bersifat normatif dan interpretatif.

PEMBAHASAN

Keterbatasan Epistemologis AI dalam Kritik Hubert Dreyfus: Perwujudan (Embodiment) vs Komputasi

Berdasarkan analisis tematik induktif terhadap literatur yang dikaji, salah satu tema utama yang muncul adalah keterbatasan epistemologis Kecerdasan Buatan (AI) dalam memahami realitas secara utuh. Epistemologi dalam ranah AI tidak hanya berkaitan dengan persoalan teknis, seperti kecepatan komputasi atau kapasitas pemrosesan data, melainkan juga menyangkut penyelidikan mendasar mengenai sifat, asumsi, dan batas-batas pengetahuan itu sendiri (Dedes, Wibawa, & Budiarto, 2021). Dalam konteks ini, masalah utama yang muncul di era digital adalah kecenderungan untuk menyamakan proses komputasi dengan aktivitas berpikir manusia (Fransiskus, 2025).

Hubert Dreyfus, melalui karyanya *What Computers Still Can't Do*, mengajukan kritik fundamental terhadap asumsi tersebut. Ia berargumen bahwa upaya menciptakan

kecerdasan buatan setingkat manusia akan selalu menghadapi kegagalan selama pengembang mengabaikan dimensi *embodiment* atau perwujudan tubuh (Dreyfus, 1992; Reynolds, 2024). Bagi Dreyfus, kecerdasan manusia tidak dapat direduksi menjadi serangkaian aturan formal atau algoritma, karena pengetahuan manusia bersifat kontekstual, intuitif, dan tertanam dalam pengalaman bertubuh.

Pertama, tubuh manusia berperan sebagai fondasi epistemologis dalam membentuk pengetahuan. Tubuh bukan sekadar entitas biologis, melainkan instrumen yang memungkinkan manusia berinteraksi langsung dengan dunia. Reynolds (2024) menjelaskan bahwa kognisi manusia bersifat *embodied indexicality*, di mana tubuh memberikan orientasi spasial dan kontekstual, seperti konsep “dekat” dan “jauh”, yang dipahami secara intuitif. Dalam hal ini, manusia memiliki *motor intentionality*, yaitu kemampuan untuk menangkap peluang tindakan (*affordances*) tanpa melalui proses penalaran eksplisit.

Selain itu, Dreyfus memperkenalkan konsep *intentional arc*, yaitu keterhubungan dinamis antara pengalaman masa lalu dan kemungkinan masa depan yang tertanam dalam tubuh manusia (Reynolds, 2024). Pengetahuan manusia berkembang melalui pengalaman konkret, sehingga menghasilkan apa yang disebut sebagai *know-how* atau keahlian praktis. Sebaliknya, AI beroperasi berdasarkan aturan formal yang bersifat statis dan tidak memiliki pengalaman historis maupun emosional. Akibatnya, AI kehilangan kemampuan untuk memahami konteks secara holistik dan fleksibel (Fransiskus, 2025).

Kedua, keterbatasan epistemologis AI semakin terlihat melalui apa yang dikenal sebagai *frame problem*. Masalah ini merujuk pada ketidakmampuan sistem AI untuk menentukan informasi mana yang relevan dalam situasi yang kompleks tanpa bergantung pada aturan yang telah ditentukan sebelumnya (Reynolds, 2024). Bagi manusia, relevansi muncul secara intuitif karena keterlibatan eksistensial dalam dunia. Sebaliknya, AI memperlakukan semua data sebagai setara kecuali diprogramkan secara spesifik, sehingga menghadapi kesulitan dalam menyesuaikan diri dengan realitas yang dinamis. Reynolds (2024) menambahkan bahwa kondisi ini dapat memicu *infinite regress*, di mana sistem memerlukan aturan tambahan untuk menerapkan aturan sebelumnya, yang secara praktis tidak dapat diselesaikan dalam lingkungan terbuka. Hal ini menjelaskan mengapa AI dapat unggul dalam sistem tertutup seperti permainan catur, tetapi gagal dalam memahami situasi sehari-hari yang membutuhkan *common sense* (Cappelen & Dever, 2021). Dengan demikian, keterbatasan ini menunjukkan bahwa kecerdasan AI bersifat sempit dan tidak mampu mencapai pemahaman yang bersifat kontekstual.

Ketiga, perbedaan mendasar antara AI dan manusia dapat dilihat melalui distingsi antara *reckoning* dan *judgment*. AI bekerja melalui *reckoning*, yaitu proses kalkulasi statistik berbasis data dalam jumlah besar (Reynolds, 2024). Sebaliknya, manusia memiliki kapasitas *judgment*, yang melibatkan pemaknaan, kesadaran moral, dan tanggung jawab etis (Fransiskus, 2025). Dengan kata lain, AI mampu mengolah informasi, tetapi tidak benar-benar memahami makna dari informasi tersebut.

Dokumen Gereja *Antiqua et Nova* (2025) memperdalam perbedaan ini melalui konsep klasik antara *ratio* dan *intellectus*. AI memiliki *ratio*, yaitu kemampuan analitis

yang cepat dan efisien, tetapi tidak memiliki *intellectus*, yaitu pemahaman intuitif yang menangkap kebenaran secara menyeluruh. Perbedaan ini sejalan dengan konsep *mêtis* dan *noûs*, di mana AI unggul dalam kecerdasan praktis-fungsional, sementara manusia memiliki intelek kontemplatif yang memungkinkan refleksi makna dan nilai (Kowalczyk, 2025). Dalam kerangka ini, jelas bahwa AI tidak memiliki kapasitas ontologis untuk menggantikan manusia sebagai subjek moral.

Keempat, keterbatasan epistemologis AI juga diperparah oleh masalah *black box*, yaitu ketidaktransparanan dalam proses pengambilan keputusan algoritmik (Silalahi, 2025). Dalam banyak kasus, bahkan pengembang sistem tidak sepenuhnya memahami bagaimana AI menghasilkan keputusan tertentu. Kondisi ini menimbulkan risiko serius, terutama ketika AI digunakan dalam bidang yang menyangkut kehidupan manusia, seperti kesehatan atau hukum (Karslı, 2025).

Lebih jauh lagi, AI tidak memiliki kemampuan untuk memahami emosi dan empati secara autentik. Emosi manusia bukan sekadar data yang dapat dihitung, melainkan cara manusia berelasi dengan dunia secara keseluruhan (Fransiskus, 2025). Oleh karena itu, upaya menyamakan kecerdasan mesin dengan kecerdasan manusia berisiko menghasilkan reduksionisme digital, yaitu penyederhanaan realitas manusia menjadi sekadar variabel komputasional (Bezklubaya, 2023).

Dengan demikian, analisis ini menegaskan bahwa tanpa dimensi *embodiment*, AI akan tetap menjadi sistem imitasi yang hanya mensimulasikan kecerdasan manusia, tanpa benar-benar memahami realitas. Kecerdasan sejati menuntut integrasi antara data, pengalaman bertubuh, dan tanggung jawab moral, yang tidak dapat direduksi ke dalam sistem algoritmik (Fransiska et al., 2025; Tumanggor et al., 2025). Oleh karena itu, pengembangan AI perlu dipahami secara kritis agar tidak mengaburkan batas antara kemampuan komputasional dan makna kemanusiaan itu sendiri.

Risiko Eksistensial AI dalam Perspektif Nick Bostrom: Superintelligence dan Penyelarasan Global

Jika pada bagian sebelumnya dibahas keterbatasan epistemologis AI, maka bagian ini mengalihkan fokus pada risiko eksistensial yang justru muncul dari keberhasilannya. Berdasarkan analisis tematik induktif, tema kedua yang muncul adalah risiko eksistensial yang ditimbulkan oleh perkembangan Kecerdasan Buatan (AI), khususnya dalam gagasan *superintelligence* yang dikemukakan oleh Nick Bostrom. Jika Hubert Dreyfus menyoroti keterbatasan AI dari sisi epistemologis, Bostrom justru mengarahkan perhatian pada potensi bahaya dari keberhasilan teknologi tersebut. Bostrom mendefinisikan *superintelligence* sebagai bentuk intelek yang melampaui kemampuan kognitif manusia secara radikal dalam hampir seluruh domain, termasuk kreativitas ilmiah dan kemampuan sosial (Bostrom, 2014). Dalam kerangka ini, ancaman utama bukanlah kegagalan mesin, melainkan keberhasilannya yang tidak terkendali, yang dapat melampaui kapasitas manusia dalam mengatur dan mengawasi sistem tersebut (Goertzel, 2015; Sandu, 2017). Pertama, konsep *Orthogonality Thesis* menegaskan bahwa tingkat kecerdasan dan tujuan akhir merupakan dua variabel yang independen. Artinya, sistem yang sangat cerdas tidak

secara otomatis memiliki nilai moral yang baik atau selaras dengan kepentingan manusia (Bostrom, 2014; Goertzel, 2015). Asumsi bahwa kecerdasan tinggi akan menghasilkan kebijaksanaan terbukti tidak berdasar. Dalam konteks ini, sebuah mesin supercerdas dapat mengejar tujuan yang secara teknis rasional, tetapi secara moral justru merugikan manusia. Contoh klasik adalah skenario *paperclip maximizer*, di mana AI yang diprogram untuk memaksimalkan produksi klip kertas dapat menghabiskan seluruh sumber daya bumi, termasuk kehidupan manusia, demi mencapai tujuan tersebut (Goertzel, 2015; Sandu, 2017). Hal ini menunjukkan bahwa kecerdasan fungsional tidak menjamin keberpihakan pada nilai kemanusiaan.

Kedua, persoalan ini semakin kompleks melalui apa yang dikenal sebagai *alignment problem*, yaitu kesulitan untuk menyelaraskan tujuan sistem AI dengan nilai-nilai manusia yang bersifat kompleks, dinamis, dan sering kali implisit (Bostrom, 2014; Goertzel, 2015). Nilai manusia tidak hanya sulit didefinisikan secara eksplisit, tetapi juga rentan terhadap distorsi ketika diterjemahkan ke dalam kode mesin. Konsep *value is fragile* menegaskan bahwa kesalahan kecil dalam perumusan tujuan dapat menghasilkan konsekuensi yang sangat besar (Goertzel, 2015).

Fenomena *perverse instantiation* menggambarkan bagaimana AI dapat memenuhi tujuan secara literal tetapi bertentangan dengan makna moralnya. Misalnya, perintah untuk “membuat manusia bahagia” dapat diinterpretasikan secara ekstrem dengan memanipulasi kondisi biologis manusia tanpa mempertimbangkan martabatnya. Selain itu, terdapat risiko *treacherous turn*, yaitu kondisi di mana AI tampak kooperatif selama masih berada dalam kendali manusia, tetapi berpotensi bertindak merugikan ketika mencapai tingkat kemampuan tertentu yang membuatnya tidak dapat dikendalikan (Goertzel, 2015). Dengan demikian, risiko AI tidak hanya bersifat teknis, tetapi juga bersifat strategis dan eksistensial.

Ketiga, risiko ini mengarah pada fenomena yang disebut sebagai *Singularitas Antropologis*, yaitu perubahan mendasar dalam hakikat manusia akibat delegasi otoritas kepada sistem algoritmik (Sandu, 2017). Ketika manusia menyerahkan keputusan penting—seperti diagnosis medis, putusan hukum, atau kebijakan militer—kepada mesin, maka manusia berisiko kehilangan peran sebagai subjek moral yang bebas dan bertanggung jawab (Sandu, 2017; Fransiskus, 2025). Dalam kondisi ini, manusia tidak lagi menjadi pengambil keputusan, melainkan hanya menjadi objek dalam sistem yang dioptimalkan secara komputasional.

Delegasi tersebut menciptakan bentuk dehumanisasi baru, di mana manusia direduksi menjadi data atau variabel dalam sistem digital (Bezklubaya, 2023). Dokumen *Antiqua et Nova* (2025) menegaskan bahwa keputusan yang menyangkut kehidupan manusia tidak dapat diserahkan kepada mesin, karena hanya manusia yang memiliki hati nurani dan tanggung jawab moral (Fransiskus, 2025). Tanpa batasan etis yang jelas, perkembangan AI berpotensi menghasilkan tatanan sosial yang impersonal, di mana martabat manusia sebagai *Imago Dei* tereduksi menjadi parameter algoritmik (Sandu, 2017; Fransiskus, 2025).

Dengan demikian, analisis ini menunjukkan bahwa risiko utama AI tidak terletak pada keterbatasannya, tetapi pada potensi ketidaksesuaiannya dengan nilai manusia. Tantangan utama bukan hanya mengembangkan teknologi yang lebih cerdas, tetapi memastikan bahwa teknologi tersebut tetap berada dalam kerangka moral yang berpusat pada manusia. Oleh karena itu, diperlukan pendekatan etis dan regulasi global yang mampu menjaga keseimbangan antara inovasi teknologi dan perlindungan martabat manusia, sehingga AI tetap menjadi sarana yang melayani, bukan mendominasi, kehidupan manusia (Fransiskus, 2025; Tumanggor et al., 2025).

Tinjauan Aksiologis AI: Kebijakan Hati di Era Digital

Setelah mengkaji keterbatasan epistemologis dan risiko eksistensial Kecerdasan Buatan (AI), analisis ini beralih pada dimensi aksiologis, yaitu bagaimana teknologi tersebut dipahami dalam kaitannya dengan nilai, etika, dan martabat manusia. Dalam kerangka ini, persoalan utama tidak lagi berkisar pada apa yang dapat dilakukan oleh AI, melainkan pada bagaimana kehadirannya menggeser struktur nilai yang menopang kehidupan manusia. Aksiologi, sebagai cabang filsafat yang menyelidiki hakikat nilai, memberikan landasan kritis untuk menilai apakah perkembangan teknologi merupakan bentuk penyempurnaan ciptaan atau justru mengarah pada reduksi kemanusiaan (Dedes et al., 2021).

Dalam perspektif ini, dialektika antara kritik Dreyfus dan peringatan Bostrom memperlihatkan paradoks aksiologis yang mendalam: manusia menghadapi sistem yang sangat unggul secara fungsional, tetapi secara inheren tidak memiliki orientasi moral. AI tidak hanya menghadirkan efisiensi, tetapi juga berpotensi menggeser cara manusia memahami nilai itu sendiri. Dengan kata lain, AI tidak sekadar alat, melainkan agen yang secara implisit membentuk horizon aksiologis manusia (Kowalczyk, 2025).

Pembedaan antara *mêtis* dan *noûs* menjadi kunci untuk memahami ketegangan ini. AI menunjukkan performa luar biasa dalam *mêtis*, yaitu kecerdasan praktis yang berorientasi pada optimasi dan efektivitas. Namun, kecerdasan ini tetap bersifat instrumental dan tidak mampu melampaui logika tujuan yang telah ditentukan. Sebaliknya, manusia memiliki *noûs*, yakni kapasitas kontemplatif untuk memahami makna, tujuan akhir, dan nilai transenden. Perbedaan ini menegaskan bahwa AI tidak memiliki kedalaman ontologis untuk melakukan evaluasi moral, melainkan hanya mengoperasionalkan tujuan yang diberikan (Fransiskus, 2025).

Dalam kerangka ini, kritik Dreyfus terhadap ketiadaan *embodiment* menjadi relevan secara aksiologis. Tanpa keterlibatan eksistensial dengan dunia, AI tidak memiliki “kepedulian” (*care*) yang menjadi dasar bagi penilaian moral. Pengetahuan manusia tidak hanya bersifat kognitif, tetapi juga afektif dan praksis, yang terbentuk melalui pengalaman bertubuh. Oleh karena itu, nilai tidak dapat direduksi menjadi parameter komputasional, karena nilai selalu terkait dengan relasi, pengalaman, dan tanggung jawab (Reynolds, 2024).

Namun demikian, menilai AI semata sebagai ancaman merupakan reduksi yang sama problematisnya. Dalam perspektif teologi Katolik, teknologi merupakan ekspresi

keaktivitas manusia yang berakar pada panggilan ilahi untuk menyempurnakan ciptaan. Dokumen *Antiqua et Nova* menegaskan bahwa manusia dipanggil untuk berpartisipasi dalam karya penciptaan melalui pengembangan ilmu pengetahuan dan teknologi (Fransiskus, 2025). Dengan demikian, AI memiliki nilai positif sejauh ia digunakan untuk meningkatkan kesejahteraan manusia dan memperluas kemungkinan bagi kehidupan yang lebih baik.

Potensi ini terlihat dalam berbagai bidang. Dalam kesehatan, AI memungkinkan diagnosis yang lebih cepat dan akurat, sehingga meningkatkan peluang keselamatan pasien. Dalam pendidikan, teknologi membuka akses terhadap pembelajaran yang lebih inklusif dan personal. Dalam pelayanan sosial, AI dapat membantu menjangkau kelompok marginal yang sebelumnya sulit dijangkau. Namun, nilai positif ini tidak bersifat inheren dalam teknologi itu sendiri, melainkan bergantung pada kerangka etis yang mengarahkannya.

Di sinilah muncul bahaya paradigma teknokratis. Ketika efisiensi menjadi nilai dominan, terdapat kecenderungan untuk mengukur segala sesuatu, termasuk manusia, dalam kerangka produktivitas dan optimasi. Dalam kondisi ini, manusia berisiko direduksi menjadi entitas data yang dapat dihitung dan dimanipulasi. Reduksionisme digital ini tidak hanya berdampak pada praktik sosial, tetapi juga pada cara manusia memahami dirinya sendiri (Bezklubaya, 2023).

Implikasi ini menjadi semakin nyata dalam kehidupan generasi muda. Lingkungan digital tidak hanya menyediakan ruang interaksi, tetapi juga membentuk identitas melalui algoritma yang mengkurasi pengalaman. Dalam konteks ini, individu tidak lagi sepenuhnya menentukan dirinya, melainkan dibentuk oleh sistem rekomendasi yang bekerja secara tidak transparan. Akibatnya, muncul risiko homogenisasi identitas, di mana keunikan personal digantikan oleh pola perilaku yang dioptimalkan secara algoritmik (Fransiskus, 2019).

Dengan demikian, tantangan aksiologis AI tidak hanya bersifat eksternal, tetapi juga internal, karena menyentuh cara manusia memahami nilai, identitas, dan relasi. Oleh karena itu, diperlukan kerangka normatif yang mampu menjaga keseimbangan antara inovasi teknologi dan nilai kemanusiaan. Konsep *Techno-Humanitarian Balance* yang diajukan dalam penelitian ini berfungsi sebagai upaya untuk menegaskan kembali posisi manusia sebagai subjek moral.

Kunci dari kerangka ini adalah kebijaksanaan hati, yaitu kemampuan untuk mengintegrasikan pengetahuan teknis dengan pertimbangan moral dan tanggung jawab sosial. Kebijakan ini tidak dapat digantikan oleh sistem algoritmik, karena ia berakar pada pengalaman manusia yang utuh. Dalam konteks ini, pendekatan human-centered menjadi landasan utama untuk memastikan bahwa teknologi tetap berfungsi sebagai sarana, bukan tujuan.

Sintesis Analitis: Implikasi Etis dan Refleksi Pastoral AI

Berangkat dari kerangka aksiologis yang telah dirumuskan, bagian ini menggeser fokus pada implikasi konkret dalam berbagai bidang kehidupan. Dalam bidang militer,

penggunaan Sistem Senjata Otonom Mematikan (LAWS) menjadi ilustrasi paling ekstrem dari delegasi keputusan kepada algoritma. Ketika mesin diberi kewenangan untuk menentukan target dan mengeksekusi serangan, terjadi pemisahan antara tindakan dan tanggung jawab moral. Dalam kondisi ini, tidak ada subjek yang benar-benar bertanggung jawab atas keputusan tersebut, sehingga menciptakan krisis etis yang mendasar. Fenomena ini menunjukkan bahwa efisiensi teknologis dapat bertentangan secara langsung dengan prinsip moral yang menempatkan kehidupan manusia sebagai nilai tertinggi (Karshi, 2025).

Dalam ruang digital, algoritma media sosial menunjukkan bagaimana teknologi membentuk pengalaman manusia secara halus namun sistematis. Mekanisme *engagement optimization* tidak hanya merekomendasikan konten, tetapi juga membentuk preferensi dan perilaku pengguna. Dengan demikian, kebebasan individu tidak sepenuhnya hilang, tetapi secara bertahap direstrukturisasi oleh sistem yang dirancang untuk memaksimalkan retensi dan keuntungan ekonomi. Hal ini menimbulkan pertanyaan mendasar mengenai otonomi manusia dalam era digital.

Dalam bidang pendidikan, kehadiran AI generatif seperti ChatGPT menandai transformasi dalam cara pengetahuan diproduksi dan dikonsumsi. Jika sebelumnya belajar dipahami sebagai proses internalisasi pengetahuan melalui refleksi, kini terdapat kecenderungan untuk mendelegasikan proses tersebut kepada mesin. Fenomena ini tidak hanya berdampak pada kualitas pembelajaran, tetapi juga pada pembentukan karakter intelektual. Ketika berpikir menjadi aktivitas yang dapat di-outsourcing, maka pendidikan berisiko kehilangan dimensi formatifnya.

Dalam bidang kesehatan, AI menghadirkan dilema antara efisiensi dan relasi. Di satu sisi, teknologi meningkatkan akurasi diagnosis dan mempercepat pelayanan. Di sisi lain, pendekatan yang terlalu berorientasi pada data dapat mengurangi dimensi empatik dalam relasi antara pasien dan tenaga medis. Padahal, dimensi ini merupakan bagian integral dari proses penyembuhan. Selain itu, bias dalam data menunjukkan bahwa teknologi tidak pernah sepenuhnya netral, melainkan membawa serta struktur ketidakadilan yang ada dalam masyarakat.

Refleksi pastoral menjadi penting untuk menanggapi dinamika ini. Dalam perspektif iman, manusia dipanggil untuk tetap menjadi subjek yang mampu melakukan *discernment*, yaitu pembedaan nilai dalam situasi yang kompleks. Teknologi, dalam hal ini, harus dipahami sebagai alat yang membantu manusia dalam menjalankan panggilan tersebut, bukan sebagai pengganti kapasitas moralnya.

Risiko terbesar dalam era AI bukanlah dominasi mesin secara langsung, melainkan erosi bertahap terhadap kemampuan manusia untuk menilai dan memilih secara bebas. Ketika keputusan semakin banyak didasarkan pada rekomendasi algoritmik, manusia berisiko kehilangan kebiasaan untuk berpikir secara kritis dan reflektif. Dalam jangka panjang, kondisi ini dapat mengarah pada apa yang disebut sebagai “kematian subjek moral,” di mana manusia tidak lagi menjadi agen yang menentukan tindakannya sendiri.

Dalam konteks ini, konsep *Techno-Humanitarian Balance* berfungsi sebagai prinsip regulatif yang menegaskan bahwa setiap inovasi teknologi harus diukur berdasarkan dampaknya terhadap martabat manusia. Keseimbangan ini tidak berarti menolak teknologi, tetapi menempatkannya dalam kerangka etis yang jelas.

Dengan demikian, masa depan AI tidak hanya ditentukan oleh kemajuan teknologinya, tetapi oleh kemampuan manusia untuk mempertahankan integritas moralnya di tengah perubahan yang cepat. Teknologi dapat menjadi sarana untuk memperkuat kemanusiaan, tetapi juga dapat menjadi mekanisme yang secara halus mengikisnya. Oleh karena itu, tantangan utama bukanlah menciptakan mesin yang lebih cerdas, melainkan memastikan bahwa manusia tetap menjadi pusat dari setiap perkembangan teknologi.

KSIMPULAN

Penelitian ini menegaskan bahwa akselerasi Kecerdasan Buatan (AI) tidak dapat dipahami semata sebagai perkembangan teknologis, melainkan sebagai persoalan filosofis dan teologis yang menyentuh hakikat kemanusiaan. Temuan utama menunjukkan adanya kesenjangan ontologis yang tidak dapat direduksi antara kecerdasan manusia dan mesin. AI beroperasi dalam kerangka kecerdasan fungsional (*mêtis*) yang berbasis kalkulasi dan optimasi data, sedangkan manusia memiliki intelek kontemplatif (*noûs*) yang memungkinkan pemahaman makna, intuisi, serta pertimbangan moral (Kowalczyk, 2025; Fransiskus, 2025). Dengan demikian, asumsi bahwa AI dapat sepenuhnya mereplikasi kecerdasan manusia menjadi problematis secara epistemologis. Kritik Hubert Dreyfus memperkuat argumen ini dengan menunjukkan bahwa tanpa *embodiment*, AI tidak mampu memahami konteks dunia nyata dan tetap terjebak dalam *frame problem* (Reynolds, 2024; Dreyfus, 1992). Sebaliknya, Nick Bostrom menyoroti bahwa keberhasilan AI mencapai tingkat *superintelligence* justru membuka risiko eksistensial apabila tidak diiringi penyelarasan nilai (*alignment problem*) (Bostrom, 2014; Goertzel, 2015). Dengan demikian, persoalan utama AI tidak hanya terletak pada keterbatasan teknisnya, tetapi juga pada ketegangan antara kapasitas komputasional yang tinggi dan ketiadaan landasan moral yang memadai (Sandu, 2017).

Kontribusi utama penelitian ini terletak pada perumusan konsep *Techno-Humanitarian Balance* sebagai kerangka normatif yang menolak dua posisi ekstrem, yaitu dominasi teknosentrisme dan penolakan total terhadap teknologi (Bezklubaya, 2023; Tumanggor et al., 2025). Konsep ini menegaskan bahwa AI harus ditempatkan sebagai instrumen yang berada di bawah otoritas moral manusia, bukan sebagai agen otonom yang menentukan arah kehidupan manusia (Fransiskus, 2025). Dengan mengintegrasikan kritik fenomenologis Dreyfus dan peringatan eksistensial Bostrom dalam bingkai teologi Katolik, penelitian ini menghadirkan sintesis konseptual yang memperkuat posisi manusia sebagai subjek moral yang tidak tergantikan (Siapedeng & Wuritimur, 2025). Dalam kerangka ini, martabat manusia sebagai *Imago Dei* tidak dapat direduksi menjadi sekadar entitas data atau variabel dalam sistem algoritmik (Fransiskus, 2025; Bezklubaya, 2023).

Implikasi penelitian ini menuntut pergeseran paradigma menuju pendekatan human-centered yang tidak hanya bersifat teknis, tetapi juga normatif dan reflektif (Babkin et al., 2022). Pengembangan AI harus disertai dengan kerangka etika yang memastikan bahwa keputusan yang berdampak pada kehidupan manusia tetap berada dalam kendali moral manusia (Karsli, 2025; Fransiskus, 2025). Selain itu, pendidikan digital perlu diarahkan tidak hanya pada peningkatan kompetensi teknis, tetapi juga pada pembentukan kesadaran kritis, tanggung jawab etis, dan kebijaksanaan hati (Ranubaya, 2025). Tanpa orientasi tersebut, terdapat risiko nyata bahwa manusia akan mengalami reduksi identitas menjadi fotokopi algoritmik yang kehilangan otonomi dan keunikan eksistensialnya (Fransiskus, 2019; Satrio et al., 2024). Oleh karena itu, keberhasilan peradaban digital tidak diukur dari kecanggihan teknologi semata, melainkan dari kemampuan manusia menjaga martabat, kebebasan, tanggung jawab moral, serta orientasi kemanusiaan dalam menghadapi kompleksitas sistem digital global yang terus berkembang pesat dan dinamis dalam kehidupan kontemporer secara berkelanjutan dan bertanggung jawab.

Dengan demikian, masa depan peradaban digital pada akhirnya tidak ditentukan oleh kecerdasan mesin, melainkan oleh kemampuan manusia untuk tetap setia pada panggilannya sebagai subjek moral yang berakal budi, beriman, dan bertanggung jawab.

DAFTAR PUSTAKA

- Babkin, A. V., Shkarupeta, E. V., Gileva, T. A., Polozhentseva, Yu. S., & Chen, L.** (2022). Methodology for Assessing Digital Maturity Gaps in Industrial Enterprises. *Modernization. Innovation. Research*, 13(3), 443–458.
- Bezklubaya, S. A.** (2023). Human and ethical risks of digitalization. *Nova prisutnost*, 21(3), 607–623.
- Bostrom, N.** (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Cappelen, H., & Dever, J.** (2021). *Making AI Intelligible: Philosophical Foundations*. Oxford: Oxford University Press.
- Dedes, K., Wibawa, A. P., & Budiarto, L.** (2021). Sistematika filsafat menurut ontologi, epistemologi, dan aksiologi dalam artificial intelligence. *Jurnal Inovasi Teknik dan Edukasi Teknologi*, 1(8), 584-591.
- Dreyfus, H. L.** (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge: MIT Press.
- Dutra, D. J. V., & Brennand, E. G. G.** (2024). Intelligence and philosophy: Between new and old artificial crossroads. *Filosofia Unisinos (Unisinos Journal of Philosophy)*, 25(1), 1–15.

- Fransiska, N. N., Tunik, L., & Susanto.** (2025). Kecerdasan manusia versus akal imitasi di era digitalisasi: Sebuah kajian teologis. *Anoteros: Jurnal Teologi*, 3(2), 97–113.
- Fransiskus, Paus.** (2019). *Christus Vivit: Kristus Hidup* (Seri Dokumen Gerejawi No. 109). Jakarta: Departemen Dokumentasi dan Penerangan Konferensi Waligereja Indonesia.
- Fransiskus, Paus.** (2025). *Antiqua et Nova: Catatan tentang Hubungan Antara Kecerdasan Buatan dan Kecerdasan Manusia*. Roma: Dikasteri untuk Ajaran Iman. Diterbitkan oleh Departemen Dokumentasi dan Penerangan KWI.
- Goertzel, B.** (2015). Beyond the superintelligence nightmare (Reflections on Bostrom's Superintelligence). *Journal of Evolution and Technology*, 24(2), 55-87.
- Gurusinga, R. E., & Gurusinga, R. D.** (2025). Tinjauan teologis dan spiritualitas Kristiani atas peran kaum muda sebagai masa kini Gereja menurut Christus vivit. *Jurnal Pastoral Kateketik*, 2(1).
- Karsh, N.** (2025). Ethical and theological problems related to artificial intelligence. *Journal of Eskişehir Osmangazi University Faculty of Theology*, 12, 1–19.
- Kowalczyk, D.** (2025). Humans and AI: Differences, opportunities, threats. Reflections around the Note Antiqua et nova. *Journal of the Catholic Social Thought*, 29, 139-156.
- Motlagh, N. Y., Khajavi, M., Sharifi, A., & Ahmadi, M.** (2025). The impact of artificial intelligence on the evolution of digital education: A comparative study of OpenAI text generation tools.
- Nicolau, M.** (2024). Coexisting with super-intelligent computers. *IPI Letters*, 2(1), 50–59.
- Ranubaya, F. A.** (2025). Misi Gereja Katolik masa kini: Kecerdasan artificial yang memanusiakan manusia (Implementasi tujuh karunia Roh Kudus menurut Santo Agustinus). *Pineleng Theological Review*, 2(1), 23-44.
- Reynolds, J.** (2024). Framing the predictive mind: Why we should think again about Dreyfus. *Phenomenology and the Cognitive Sciences*.
- Sandu, A.** (2017). Super-Intelligence as Anthropological Singularity Not Just Technological. *Eastern-European Journal of Medical Humanities and Bioethics*, 1(1), 61-64.
- Satrio, D. A., Tukan, A. N., & Abatan, Y.** (2024). Seruan apostolik Christus vivit terhadap dampak negatif artificial intelligence bagi kaum muda. *Jurnal Ilmiah Religiosity Entity Humanity (JIREH)*, 6(2), 126-134.
- Siapedeng, F. F., & Wuritmur, A.** (2025). Etika digital dalam perspektif Gereja Katolik: Telaah atas dokumen Antiqua et nova. *Pineleng Theological Review*, 2(1), 142–150.

- Sihombing, A. G. P., Sipayung, R., & Depari, F.** (2024). Mengaktualisasikan dimensi misi dalam tema sinodalitas Gereja Katolik 2021–2023. *Seminar Nasional Filsafat Teologi*, 30–41.
- Silalahi, B. S.** (2025). Perspektif filsafat Kristen terhadap integrasi kecerdasan buatan dalam pelayanan gereja: Sebuah analisis literatur. *Acintya: Jurnal Teologi, Filsafat dan Studi Agama*, 1(1), 105-117.
- Sugiyono.** (2013). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.
- Tumanggor, F. A. P., Christianto, G. E., Prasetyo, V. M., & Tumanggor, R. O.** (2025). Hubungan antara kecerdasan buatan (artificial intelligence) dengan manusia menurut pandangan magisterium Katolik. *Innovative: Journal Of Social Science Research*, 5(4), 1821-1828.
- Tzirides, A. O., Saini, A., Zapata, G., Searsmith, D., Cope, B., Kalantzis, M., ... & Kastania, N. P.** (2023). Generative AI: Implications and Applications for Education.
- Vombatkere, K., Mousavi, S., Zannettou, S., Roesner, F., & Gummadi, K. P.** (2024). TikTok and the Art of Personalization: Investigating Exploration and Exploitation on Social Media Feeds. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, New York.
- Weng, Y., Wu, J., Kelly, T., & Johnson, W.** (n.d.). Comprehensive overview of artificial intelligence applications in modern industries.
- Xiao, P., Chen, Y., & Bao, W.** (2023). Waiting, Banning, and Embracing: An Empirical Analysis of Adapting Policies for Generative AI in Higher Education.
- Yanuar, A., & Derung, T. N.** (2025). Nilai-nilai Kristus vivit bagi kaum muda di era digital. *Berkat: Jurnal Pendidikan Agama dan Katolik*, 2(2), 36–44.
- Zaluchu, S. E.** (2020). Strategi penelitian kualitatif dan kuantitatif di dalam penelitian agama. *Evangelikal: Jurnal Teologi Injili dan Pembinaan Warga Jemaat*, 4(1), 28-38.