



Deepfake dan Krisis Kejujuran di Era *Artificial Intelligence*

(Analisis Dilema Moral dalam Perspektif Al-Ghazali)

Siti Rahayu^{1*}, Elvira Purnamasari²

¹⁻²UIN Fatmawati Sukarno Bengkulu, Indonesia

Email: sitisitirahayu35@gmail.com¹, elvira.purnamasari@mail.uinfabengkulu.ac.id²

*Penulis Korespondensi: sitisitirahayu35@gmail.com

Abstract. *Advances in Artificial Intelligence (AI) technology, particularly Generative Adversarial Networks (GANs), have given rise to the phenomenon of deepfakes technology capable of authentically manipulating audio-visual reality while producing content that is, in fact, fake. The August 2025 deepfake video involving Finance Minister Sri Mulyani serves as clear evidence of a digital integrity crisis that is eroding public trust in Indonesia. Although technical and regulatory solutions continue to be developed, the moral dimension of this crisis has not been thoroughly examined through the framework of classical Islamic ethics, particularly the thought of Al-Ghazali. This study aims to analyze the moral dilemmas posed by deepfake technology and to examine its relevance to the crisis of truthfulness in the digital age from the perspective of Abu Hamid al-Ghazali's thought. Using a qualitative method through library research with the primary source being the book *Ihya' Ulum al-Din*, this study applies a descriptive-analytical approach to examine the phenomenon of digital manipulation and uncover its ethical dimensions from Al-Ghazali's perspective. The results indicate that deepfakes constitute a modern form of *kadzib* that creates a false reality and triggers a crisis of public trust. From Al-Ghazali's perspective, this phenomenon contradicts the principle of *sidq* as well as the objectives of *maqasid al-shariah*. Therefore, the values of *sidq*, *amanah*, *tabayun*, and *tazkiyatun nafs* serve as an ethical framework in the production, distribution, and consumption of digital information to strengthen technical approaches to maintaining integrity and public trust.*

Keyword: *Al-Ghazali; Artificial Intelligence; Deepfake; Moral Crisis; Public Trust.*

Abstrak. Perkembangan teknologi *Artificial Intelligence* (AI), khususnya teknik *Generative Adversarial Networks* (GAN), telah melahirkan fenomena *deepfake* yang mampu memanipulasi realitas audio-visual secara autentik namun sebenarnya palsu. Kasus video *deepfake* Menteri Keuangan Sri Mulyani pada Agustus 2025 menjadi bukti nyata krisis kejujuran digital yang mengikis kepercayaan publik di Indonesia. Meskipun solusi teknis dan regulatif terus dikembangkan, dimensi moralitas dalam krisis ini belum banyak dikaji secara mendalam melalui kerangka etika Islam klasik, khususnya pemikiran Al-Ghazali. Penelitian ini bertujuan untuk menganalisis dilema moral yang ditimbulkan oleh teknologi *deepfake* dan meneliti relevansinya terhadap krisis kejujuran di era digital dalam perspektif pemikiran Abu Hamid al-Ghazali. Menggunakan metode kualitatif melalui riset kepustakaan (*library research*) dengan sumber utama kitab *Ihya' Ulum al-Din*, penelitian ini menerapkan pendekatan deskriptif-analitis untuk mengkaji fenomena manipulasi digital dan mengungkap dimensi etisnya dalam perspektif Al-Ghazali. Hasil penelitian menunjukkan bahwa *deepfake* merupakan bentuk *kadzib* modern yang menciptakan realitas semu dan memicu krisis kepercayaan publik. Dalam perspektif Al-Ghazali, fenomena ini bertentangan dengan prinsip *sidq* serta tujuan *maqasid al-shariah*. Oleh karena itu, nilai-nilai *sidq*, *amanah*, *tabayun*, dan *tazkiyatun nafs* menjadi kerangka etis dalam produksi, distribusi, dan konsumsi informasi digital guna memperkuat pendekatan teknis dalam menjaga integritas dan kepercayaan publik.

Kata Kunci: Dilema Moral; Etika Al-Ghazali; Kecerdasan Buatan; Krisis Kejujuran; Manipulasi Media Digital.

1. PENDAHULUAN

Perkembangan *Artificial Intelligence* (AI) telah membawa perubahan signifikan dalam berbagai aspek kehidupan manusia, termasuk dalam produksi dan distribusi informasi digital baik di tingkat nasional maupun internasional. Salah satu manifestasi paling kontroversial dari kemajuan ini adalah teknologi *deepfake*, yaitu teknik berbasis kecerdasan buatan yang mampu memanipulasi audio, gambar, dan video sehingga tampak autentik namun sebenarnya palsu (Putra, UMJ, and Muksin 2025). Awalnya, teknologi ini dikembangkan untuk kepentingan riset

visual dan hiburan. Namun, dalam praktiknya, *deepfake* kerap kali disalahgunakan oleh orang-orang yang tidak bertanggung jawab untuk kepentingan politik, ekonomi, bahkan pencemaran nama baik individu. Urgensi persoalan ini tercermin nyata dalam kasus pada Agustus 2025, saat video *deepfake* Menteri Keuangan Sri Mulyani Indrawati beredar luas secara viral di platform TikTok, X, dan Instagram. Video tersebut menampilkan pernyataan provokatif yang memicu kebingungan publik hingga memaksa Kementerian Komunikasi dan Informatika (Kominfo) untuk melakukan klarifikasi resmi. Bahkan, pemerintah sampai memanggil perwakilan platform besar seperti TikTok dan Meta sebagai bentuk respon serius terhadap penyebaran konten tersebut. Fenomena ini menunjukkan bahwa *deepfake* tidak lagi sekedar inovasi teknologi, melainkan telah berkembang menjadi persoalan moral yang berpotensi mengancam stabilitas kepercayaan publik. Sehingga penting untuk dikaji secara mendalam, terutama melalui konsep *Sidq* dalam pemikiran Al-Ghazali.

Kejujuran sebagai nilai moral fundamental kini menghadapi tantangan baru akibat kemampuan teknologi dalam menciptakan realitas semu (Xie et al. 2025). Dalam kajian akademik kontemporer, terdapat banyak perdebatan mengenai bagaimana seharusnya *deepfake* dipahami dan diatur. Sebagian penelitian menekankan pendekatan regulatif dan teknologi deteksi sebagai solusi utama, Westerlund (2019) sementara yang lain berargumen bahwa pendekatan etis dan literasi digital masyarakat lebih penting dalam jangka panjang (Salsabila et al. 2025). Penelitian tentang *deepfake* menunjukkan meningkatnya perhatian terhadap aspek etika. Sejumlah studi di Indonesia menegaskan bahwa *deepfake* tidak hanya menjadi persoalan teknis, tetapi juga memunculkan dilema moral yang kompleks, termasuk manipulasi identitas, penyebaran disinformasi, serta pelanggaran privasi. Misalnya, penelitian oleh Librianti (2024); Librianti (2025) serta Salsabila et al., (2025); Salsabila et al. (2025) menunjukkan bahwa *deepfake* berpotensi disalahgunakan dalam praktik penipuan, sementara Boediman (2025) menegaskan bahwa teknologi ini dapat mengikis kepercayaan publik akibat kaburnya batas antara realitas dan manipulasi (Boediman 2025).

Perbedaan temuan tersebut menunjukkan adanya kesenjangan dalam memahami dimensi moral *deepfake*, khususnya dalam konteks nilai kejujuran, serta masih terbatasnya kajian dalam perspektif etika Islam klasik. Kajian yang secara spesifik menghubungkan fenomena *deepfake* dengan pemikiran tokoh etika Islam, seperti Al-Ghazali, khususnya terkait konsep kejujuran (*Sidq*), belum ada. Oleh karena itu, penelitian ini memiliki signifikansi dalam mengisi kesenjangan tersebut dengan menawarkan perspektif alternatif yang berbasis nilai-nilai moral Islam. Dalam perspektif filsafat moral Islam, kejujuran (*sidq*) merupakan salah satu nilai utama yang dibahas secara mendalam oleh Abu Hamid Al-Ghazali. Dalam karya-

karyanya, seperti *Ihya' Ulum al-Din*. Al-Ghazali menempatkan kejujuran sebagai fondasi integritas moral yang mencakup kesesuaian antara ucapan, niat, dan tindakan. Kebohongan (*kadzib*) tidak hanya dipahami sebagai kebohongan verbal, tetapi juga sebagai bentuk manipulasi realitas yang dapat merusak tatanan sosial dan spiritual. Dengan demikian, fenomena *deepfake* dapat dianalisis sebagai bentuk baru dari ketidakjujuran yang bersifat sistemik dan berbasis teknologi.

Penelitian ini bertujuan untuk menganalisis dilema moral yang ditimbulkan oleh teknologi *deepfake* dan meneliti relevansinya terhadap krisis kejujuran di era kecerdasan buatan (AI) dalam perspektif pemikiran Al-Ghazali. Dengan menggunakan metode kualitatif, studi ini berharap dapat menunjukkan bahwa nilai-nilai etika Islam klasik tidak hanya bersifat teologis-normatif, tetapi juga memiliki ketajaman filosofis yang dapat digunakan untuk menafsirkan isu-isu modern seperti integritas digital, manipulasi realitas, dan pemulihan kepercayaan publik. Fokus studi ini adalah pada dua konsep utama yaitu kejujuran (*Sidq*) dan kebohongan (*Kadzib*).

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan jenis penelitian kaepustakaan (*library reserch*). Pendekatan ini dipilih karena penelitian bertujuan memahami fenomena secara mendalam dan kontekstual, khususnya dalam menganalisis dimensi etika dari fenomena *deepfake* dalam perspektif pemikiran Abu Hamid Al-Ghazali. Penelitian kualitatif sendiri menekankan pada pemahaman makna, proses, serta interpretasi terhadap realitas sosial secara holistik.(Moleong 2004).

Sumber data dalam penelitian ini terdiri dari data primer dan sekunder. Data primer diperoleh dari karya utama Al-Ghazali, yaitu *ihya' Ulum al-Din*, yang menjadi rujukan utama dalam memahami konsep kejujuran (*sidq*) dan kebohongan (*kadzib*). Adapun data sekunder diperoleh dari buku, artikel jurnal, serta penelitian ilmiah yang relevan dengan tema kecerdasan buatan (*Artificial Intelligence*) dan fenomena *deepfake*, khususnya yang berkaitan dengan aspek etika dan disinformasi digital.

Data yang diperoleh melalui studi dokumentasi dengan cara menelaah, mengkaji, dan mengklasifikasi berbagai sumber yang relevan,(Moleong 2004) kemudian dianalisis menggunakan pendekatan deskriptif-analitis, yaitu dengan mengkaji fenomena manipulasi digital melalui kerangka moral AL-Ghazali guna mengungkap dimensi etis yang terkandung didalamnya. Pendekatan ini memungkinkan pembacaan teks secara kontekstual serta penarikan kesimpulan secara induktif mengenai kontribusi pemikiran Al-Ghazali dalam diskursus etika

teknologi. Dengan demikian, metode ini digunakan untuk menelaah relevansi filsafat moral Islam dalam memberikan alternatif solusi atas krisis kejujuran sistemik sekaligus sebagai landasan dalam pembentukan karakter manusia yang berintegritas di tengah gempuran disinformasi digital. Dengan metode ini, penelitian diharapkan mampu memberikan kontribusi konseptual dalam menghubungkan nilai-nilai Islam Klasik dengan tantangan moral modern, khususnya dalam menghadapi fenomena manipulasi realitas berbasis teknologi.

3. HASIL DAN PEMBAHASAN

Konseptualisasi *Deepfake* dalam Era *Artificial Intelligence*

Deepfake adalah salah satu teknologi kecerdasan buatan (AI) yang berasal dari penggabungan kata "*deep learning*" (pembelajaran mendalam) dan "*fake*" (palsu) (Habib 2025). Ini adalah teknik computer yang mampu memanipulasi audio, gambar, dan video yang tampak autentik namun sebenarnya palsu. Teknologi ini memanfaatkan algoritma kecerdasan buatan (AI), terutama melalui pendekatan *Generative Adversarial Networks* (GAN). Dalam sistem GAN, terdapat dua jaringan saraf utama: *generator* dan *discriminator*. Peran *generator* adalah menciptakan gambar atau video yang tampak sangat meyakinkan, sementara *discriminator* bertugas mengevaluasi apakah konten tersebut tampak asli atau hasil rekayasa. Kedua jaringan ini saling berkompetisi dalam proses pelatihan (*adversarial process*), sehingga kualitas hasil manipulasi menjadi semakin realistis dari waktu ke waktu.

Secara teknis, poses *deepfake* dimulai dengan pengumpulan data set wajah atau suara target, kemudian model GAN dilatih untuk mengenali pola-pola visual seperti ekspresi wajah, gerakan bibir, dan pencahayaan. *Generator* akan mencoba mereplikasi pola tersebut, dan *discriminator* akan menguji keaslian hasilnya. Proses ini berlangsung berulang sehingga model mampu menghasilkan video yang hampir tidak dapat dibedakan dari aslinya. Bahkan, penelitian terbaru menunjukkan bahwa *deepfake* berbasis GAN mampu menghasilkan konten dengan tingkat realisme yang sangat tinggi sehingga sulit untuk dideteksi tanpa alat khusus.

Dalam perspektif epistemologi, perkembangan *deepfake* menunjukkan bahwa kecerdasan buatan telah mengalami progresivitas yang tidak lagi hanya berfungsi sebagai alat bantu manusia, tetapi juga mampu memproduksi pengetahuan dan realitas digital baru secara mandiri. *Artificial Intelligence* melalui *deep learning* dapat mempelajari, meniru, dan merekonstruksi berbagai pola visual maupun audio manusia dengan tingkat presisi yang tinggi. Kondisi ini memperlihatkan bahwa teknologi *Artificial Intelligence* telah mengubah cara manusia dalam memahami sumber kebenaran dan validitas informasi di era perkembangan digitalisasi. Jika sebelumnya gambar dan video dipandang sebagai representasi fakta yang

objektif, maka *deepfake* justru menghadirkan keraguan epistemologis karena batas antara realitas dan manipulasi digital menjadi semakin kabur.(Surajiyo 2026). Oleh sebab itu, *deepfake* dapat dipahami bukan sekedar inovasi teknologi visual, melainkan juga sebagai manifestasi perkembangan kecerdasan buatan yang membawa implikasi filosofis terhadap konsep pengetahuan, kebenaran, dan kepercayaan publik dalam masyarakat digital modern.

***Deepfake* sebagai ancaman terhadap Kejujuran dan Kepercayaan Public**

Perkembangan *deepfake* membawa implikasi serius terhadap kejujuran dan kepercayaan publik dalam ekosistem digital, karena kepercayaan (*public trust*) sebagai fondasi utama komunikasi sosial mulai tergerus ketika konten visual yang selama ini dianggap autentik dapat dimanipulasi dengan mudah. Ketika informasi tidak lagi dapat diverifikasi secara pasti, masyarakat tidak hanya rentan terhadap disinformasi seperti pembunuhan karakter, propaganda politik, dan penipuan finansial, tetapi juga cenderung meragukan seluruh informasi yang beredar, termasuk yang sebenarnya benar.(Haida and Nuriyatman 2024) Kondisi ini memicu skeptisisme masal yang berbahaya, di mana kebenaran faktual kehilangan otoritasnya dan dipersepsikan sebagai sekadar opini, sehingga *deepfake* tidak hanya merusak kejujuran sebagai nilai moral individu, tetapi juga menciptakan krisis kepercayaan yang bersifat sistemik dalam masyarakat digital.

Sejumlah penelitian terbaru menunjukkan bahwa *deepfake* tidak hanya berdampak pada aspek teknis komunikasi, tetapi juga memiliki implikasi etis dan sosial yang luas. Studi oleh Robert Chesney dan Danielle Citron menegaskan bahwa *deepfake* berpotensi merusak kepercayaan public terhadap bukti visual, yang selama ini dianggap sebagai representasi paling objektif dari realitas.(Chesney and Citron 2019) Ketika video tidak lagi dapat dipercaya, maka masyarakat akan memasuki kondisi epistemologis yang rapuh, dimana batas antara kebenaran dan kepalasuan menjadi kabur. Hal ini kemudian akan memunculkan fenomena *liar's dividend*, yaitu kondisi dimana individu dapat menyangkal fakta dengan alasan bahwa bukti yang ada mungkin merupakan hasil dari manipulasi digital.

Kemampuan *deepfake* dalam meniru ekspresi wajah, gerakan bibir, dan suara manusia telah mencapai tingkat realisme yang sangat tinggi, sehingga sulit dideteksi tanpa adanya bantuan teknologi forensik digital. Kondisi ini memperkuat potensi penyebaran disinformasi secara masif, karena konten yang tampak meyakinkan cenderung lebih mudah diterima dan disebarkan oleh masyarakat. Dengan demikian, *deepfake* tidak hanya menjadi alat manipulasi, tetapi juga instrumen produksi realitas semu yang mampu memengaruhi persepsi publik secara luas.

Fenomena ini memberikan dampak sosial yang sangat nyata dalam berbagai kasus di Indonesia, khususnya terkait penipuan identitas digital dan pelanggaran privasi. Studi hukum menunjukkan bahwa *deepfake* telah digunakan dalam berbagai bentuk kejahatan siber, termasuk manipulasi identitas dan penyebaran konten palsu yang merugikan individu maupun kelompok. Bahkan, penyalahgunaan teknologi ini juga berpotensi mengancam keamanan data pribadi, terutama data biometrik yang dapat direplikasi melalui teknologi AI. Dalam hukum Indonesia, kecerdasan buatan AI dipandang sebagai alat atau agen elektronik yang bekerja berdasarkan perintah manusia, sehingga tanggung jawab hukum tetap berada pada pengembangan atau penyelenggara sistem, sebagaimana diatur dalam UU ITE dan peraturan Pemerintah Nomor 71 Tahun 2019 yang menegaskan bahwa setiap tindakan sistem elektronik, termasuk AI dapat bekerja secara otomatis, manusia tetap menjadi pihak yang bertanggung jawab jika terjadi pelanggaran, misalnya dalam pengumpulan dan pengolahan data pribadi tanpa izin yang dapat melanggar hukum perlindungan data (Pane and Permana 2025). Namun dalam kajian teori hukum, terdapat beberapa model pertanggungjawaban seperti yang dijelaskan oleh Hallevy.

Pertama, model yang menganggap AI hanya sebagai alat sehingga tanggung jawab ada pada pengguna atau pembuatnya. Kedua, model yang melihat bahwa tindakan AI dapat menimbulkan tanggung jawab bagi pengembang karena merupakan akibat yang bisa diperkirakan. Dan yang ketiga, model yang lebih maju yang menganggap AI dapat bertanggung jawab secara langsung layaknya manusia, sehingga perbedaan pandangan ini menunjukkan bahwa semakin canggih dan mandiri AI, semakin kompleks pula penentuan siapa yang harus bertanggung jawab atas pelanggaran yang terjadi.

Konsep Kejujuran (*Sidq*) dan Kebohongan (*Kadzib*) dalam Pemikiran Al-Ghazali

Secara etimologis, kata *sidq* (الصدق) merupakan Masdar (kata kerja benda) dari kata *shadaqa-yashdiq-shidqan* (صدق-يصدق-صدقا). (Hasiah 2014) secara literal term, *al-shidq* berarti "benar, nyata, berkata benar". Trem *al-shidq* asalnya adalah perkataan (*al-qaul*) yang benar dalam berita tanpa merubah substansi kalimatnya, baik pada masa lampau maupun masa yang akan datang. *Al-Shidq* dalam pengertian ini diperlawankan dengan *al-kidzb* (الكذب). *Al-Kadzib* berasal dari kata *kadzaba-yaakdzibu* yang berarti dusta, tidak benar, atau menyimpang dari fakta. Secara terminologis, *as-sidq* dipahami sebagai kesesuaian antara ucapan dengan realitas, sedangkan *kadzib* merupakan ketidaksesuaian antara keduanya. Namun, dalam perspektif Islam klasik, pengertian ini tidak hanya terhenti pada aspek verbal, melainkan mencakup dimensi yang lebih luas, yaitu keselarasan antara hati, ucapan, dan tindakan.

Dalam perspektif Abu Hamid Al-Ghazali, kejujuran (*sidq*) merupakan fondasi utama dalam pembentukan akhlak dan integritas manusia. Dalam kitabnya *Ihya' Ulum al-Din*, Al-Ghazali menjelaskan bahwa kejujuran tidak hanya terbatas pada ucapan, tetapi juga mencakup keseluruhan dimensi eksistensi manusia, yaitu kesesuaian antara hati, niat, dan tindakan. Ia menyatakan bahwa :

الصِّدْقُ هُوَ مُطَابَقَةُ الْقَوْلِ لِلْوَاقِعِ

(kejujuran Adalah kesesuaian antara ucapan dengan kenyataan). Dalam kitabnya *al-Arba'in fi Ushul al-Din*, Al-Ghazali menyatakan bahwa kejujuran (*Sidq*) terdiri dari enam peringkat. Peringkat pertama *al-sidq fi al-qawl*, yakni kejujuran dalam perkataan. Kedua, *al-sidq fi al-niyyah wa al-iradah* kejujuran dalam niat dan keinginan. Ketiga, *al-sidq fi al-'azm*, yaitu kejujuran dalam tekad (*azm*). Keempat, *al-sidq fi al wafa' bi al-'azm* yakni sifat benar pada menepati janji dengan penuh keazaman. Kelima, *al-sidq fi al 'amal* yakni sifat benar pada amal perbuatan. Dan yang keenam adalah *al-sidq fi tahqiq maqamat al-din* yakni sifat benar pada merealisasikan makam-makam keagamaan.

Hal ini menunjukkan bahwa kejujuran merupakan integritas menyeluruh antara dimensi internal dan eksternal manusia. Sebaliknya, kebohongan (*kadzib*) menurut Al-Ghazali tidak hanya berupa ucapan yang tidak sesuai dengan fakta, tetapi juga mencakup segala bentuk penyimpangan dari kebenaran yang dilakukan secara sadar. Ia menyatakan:

الكذبُ هو الإخبارُ عن الشيءِ بخلافِ ما هو عليه

(kebohongan adalah menyampaikan sesuatu yang tidak sesuai dengan kenyataannya). Al-Ghazali juga menekankan bahwa kebohongan memiliki dampak yang luas, baik secara spiritual maupun sosial. Secara spiritual, kebohongan merusak integritas hati dan menjauhkan manusia dari kebenaran (*al-haq*). Secara sosial, kebohongan menghancurkan kepercayaan (*thiqah*) yang menjadi dasar hubungan antara individu dalam masyarakat.

Dengan demikian, *al-sidq* dalam perspektif al-Ghazali dapat dipahami sebagai kesesuaian total antara manusia dengan kebenaran (*al-haq*), baik dalam aspek internal maupun eksternal. Sebaliknya, setiap bentuk ketidaksesuaian dalam salah satu dimensi tersebut dapat dikategorikan sebagai *al-kadzib*, yang tidak hanya terbatas pada kebohongan verbal, tetapi juga mencakup bentuk-bentuk penyimpangan moral seperti *riya'*, *nifaq*, dan perilaku manipulatif lainnya. Oleh karena itu, *al-sidq* tidak hanya berfungsi sebagai prinsip etika individual, tetapi juga sebagai fondasi pembentukan integritas moral dan spiritual manusia secara holistik dalam kehidupan beragama dan sosial.

Analisis *Deepfake* dalam Perspektif Etika Al-Ghazali

Fenomena *deepfake* dapat dipahami sebagai bentuk modern dari *kadzib* yang bersifat sistemik dan berbasis teknologi. Jika dalam konteks klasik kebohongan terjadi melalui ucapan atau tulisan, maka dalam era kecerdasan buatan kebohongan dapat diwujudkan dalam bentuk visual yang tampak nyata. *Deepfake* tidak hanya menyampaikan informasi yang salah, tetapi menciptakan realitas palsu yang mampu menggantikan realitas yang sebenarnya.

Dalam perspektif Al-Ghazali, tindakan ini jelas bertentangan dengan prinsip *sidq*, karena tidak terdapat kesesuaian antara representasi yang ditampilkan dengan realitas yang sesungguhnya. Bahkan, *deepfake* dapat dikategorikan sebagai bentuk kebohongan yang lebih berbahaya, karena ia bekerja pada tingkat persepsi inderawi manusia. Ketika mata melihat sesuatu yang tampak nyata, maka kecenderungan untuk mempercayainya menjadi lebih besar. Tidak hanya itu, *deepfake* juga melibatkan dimensi moral kolektif. Tidak hanya pembuat konten yang bertanggung jawab, tetapi juga penyebar dan bahkan konsumen informasi. Dalam etika Islam, setiap individu memiliki tanggung jawab untuk memastikan kebenaran informasi (*tabayyun*). Ketika seseorang menyebarkan informasi tanpa verifikasi, maka ia turut berkontribusi dalam penyebaran kebohongan.

Melalui kerangka *Maqasid al-Shariah* dalam pemikiran Al-Ghazali, fenomena *deepfake* dapat dipahami sebagai bentuk penyimpangan etika yang bertentangan dengan tujuan dasar syariat, khususnya dalam 3 aspek, yaitu: menjaga akal (*hifz al-aql*), menjaga kehormatan (*hifz al-ird*), dan menjaga harta (*hifz al-mal*). (Habib 2025) Menjaga akal (*hifz al-aql*) karena teknologi ini mampu memanipulasi informasi sehingga tampak nyata padahal palsu, yang berpotensi menyesatkan masyarakat, menyebarkan hoaks, serta merusak kemampuan manusia dalam membedakan antara kebenaran dan kebohongan.

Dalam pandangan Al-Ghazali, akal merupakan instrumen utama untuk memahami kebenaran yang harus dijaga dari segala bentuk distorsi, sehingga penyebaran konten *deepfake* jelas tidak sejalan dengan prinsip tersebut. Selain itu, praktik ini juga dapat melanggar prinsip menjaga kehormatan (*hifz al-ird*) dan menjaga harta (*hifz al-mal*), terutama ketika digunakan untuk merusak reputasi seseorang atau melakukan penipuan yang menimbulkan kerugian material maupun non-material. Dalam kerangka moral Al-Ghazali, tindakan semacam ini termasuk dalam kategori *kadzib* (kebohongan) yang dilarang karena merusak tatanan sosial dan mengikis kepercayaan publik. Dengan demikian, *deepfake* dapat dipahami sebagai bentuk “kebohongan digital” yang melampaui batas-batas tradisional etika komunikasi. Ia menciptakan krisis kejujuran yang tidak hanya bersifat individual, tetapi juga struktural dan sistemik.

Dilema Moral dalam Penggunaan Teknologi *Deepfake*

Teknologi *deepfake* hadir sebagai manifestasi kemajuan kecerdasan buatan yang bersifat ambivalen. Di satu sisi, teknologi ini menawarkan efisiensi revolusioner dalam industri kreatif, seperti rekonstruksi wajah aktor dalam sinematografi atau visualisasi tokoh sejarah untuk kepentingan edukasi. Namun, potensi kegunaannya yang bersifat positif sering kali berbenturan dengan risiko manipulasi yang destruktif. Dilema moral muncul ketika kemampuan ini dilakukan untuk menghasilkan konten yang realistis digunakan untuk mengaburkan fakta objektif. (Wahyudi 2025). Sehingga integritas informasi menjadi rapuh akibat pengaburan batasan antara realitas autentik dan simulasi artifisial. Secara etis, sebuah inovasi tidak hanya dinilai dari tujuan akhirnya, tetapi juga dari proses dan dampak sistemik yang ditimbulkan terhadap kepercayaan publik.

Penyalahgunaan *deepfake* menciptakan krisis kejujuran di ruang digital, dimana motif ekonomi dan politik sering kali menjadi penggerak utama dibalik pembuatan konten manipulatif. Ketika identitas dipalsukan tanpa adanya persetujuan untuk kepentingan propaganda atau pencemaran nama baik, maka prinsip otonomi dan martabat individu telah dilanggar secara fundamental. Kondisi ini memaksa kita untuk merumuskan ulang strategi mitigasi yang tidak hanya bertumpu pada regulasi hukum, tetapi juga pada penguatan fondasi etika digital guna melindungi hak individu dari ancaman disinformasi.

Oleh karena itu, tantangan moral dalam penggunaan *deepfake* terletak pada bagaimana manusia mampu menyeimbangkan ambisi inovasi dengan batasan etis yang ketat. (Desisca, Ramadhanti, and Khan 2025) Penegakan etika digital menjadi sangat krusial agar kemajuan kecerdasan buatan tetap berfungsi sebagai instrumen kemajuan peradaban yang bertanggung jawab, tanpa meruntuhkan tatanan kebenaran yang menjadi landasan kepercayaan sosial di era digital.

Relevansi Etika Al-Ghazali dalam Era *Artificial Intelligence*

Pemikiran etika Imam Al-Ghazali memiliki urgensi yang kuat meskipun dihadapkan pada fenomena teknologi modern seperti *Artificial Intelligence* (AI). Di era digital ini, kecerdasan buatan sering kali dipandang sebagai entitas yang netral, namun penggunaannya sangat bergantung pada integritas moral manusia di baliknya. Al-Ghazali menekankan bahwa inti dari etika adalah *tazkiyatun nafs* atau penyucian jiwa, yang berfungsi sebagai kontrol internal bagi setiap individu agar bijak dalam menggunakan teknologi. (Ali and Teh 2025) Dalam konteks AI, prinsip ini menjadi relevan sebagai fondasi bagi pengguna maupun pengembang teknologi untuk selalu mengedepankan niat yang benar (*niyyah*) dan pengawasan diri agar teknologi ini tidak digunakan untuk tujuan yang merusak atau manipulatif.

Konsep Amanah (tanggung jawab) dalam perspektif Al-Ghazali memberikan kerangka kerja yang solid dalam menghadapi tantangan etis AI, seperti penyebaran informasi palsu atau manipulasi digital. Al-Ghazali memandang ilmu pengetahuan sebagai alat untuk mencapai kebaikan, sehingga setiap inovasi teknologi harus disertai dengan tanggung jawab moral untuk menjaga kebenaran. (Serevan 2025) Ketika AI memiliki kemampuan untuk mengaburkan batasan antara fakta dan fiksi, pemikiran Al-Ghazali mengenai kejujuran menjadi kompas etis agar manusia tidak terjebak dalam krisis moral di era otomatis. (Syifa and Ridwan 2024) Dengan demikian, etika Al-Ghazali berperan untuk memastikan bahwa AI tetap berada dalam kendali manusia yang memiliki kesadaran moral tinggi.

Relevansi pemikiran Al-Ghazali juga terlihat dalam upayanya menjaga martabat manusia di tengah dominasi mesin. Meskipun AI mampu mengolah data dengan kecepatan luar biasa, Al-Ghazali mengingatkan bahwa manusia memiliki dimensi spiritual dan jiwa rasional (*Nafs Nathiqah*) yang memiliki kesadaran serta *fitrah*, sesuatu yang tidak akan pernah bisa ditiru oleh algoritma mesin. (Ningrum et al. 2024) Hal ini memberikan batasan yang jelas bahwa AI hanyalah alat bantu, sedangkan keputusan moral dan kebijakan tetap berada sepenuhnya di tangan manusia. Dengan mengintegrasikan etika Al-Ghazali ke dalam pengembangan AI, kita dapat menciptakan ekosistem teknologi yang tidak hanya cerdas secara fungsional, namun juga mulia secara moral dan beradab.

Selain aspek kontrol diri dan tanggung jawab, etika Al-Ghazali juga menawarkan solusi preventif melalui konsep keseimbangan antara ilmu (*ilm*) dan amal. Dalam pandangan Al-Ghazali, ilmu yang bermanfaat adalah ilmu yang mendekatkan manusia pada kebenaran dan kemaslahatan, bukan sekadar kecanggihan teknis tanpa arah moral. Relevansi ini sangat krusial dalam pengembangan AI saat ini, di mana kecepatan inovasi sering kali melampaui kesiapan etika masyarakat. Dengan merujuk pada prinsip metafisika Al-Ghazali, pengembangan kecerdasan buatan seharusnya tidak hanya mengejar efisiensi materi, tetapi juga harus selaras dengan nilai-nilai kemanusiaan yang luhur demi menjaga martabat manusia sebagai subjek moral utama. (Neonnub and Kosat 2026). Integrasi antara nilai-nilai sufistik Al-Ghazali dengan kemajuan teknologi ini pada akhirnya akan melahirkan pola hubungan yang harmonis antara manusia dan mesin, di mana teknologi berfungsi sebagai pelayan bagi kesejahteraan spiritual dan jasmani manusia.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa fenomena *deepfake* merupakan bentuk perkembangan teknologi *Artificial Intelligence* yang tidak hanya berdampak pada aspek teknis, tetapi juga menimbulkan dilema moral yang serius, khususnya terkait krisis kejujuran di era digital. *Deepfake* mampu menciptakan realitas semu yang berpotensi merusak kepercayaan publik, menyebarkan disinformasi, serta melanggar privasi dan kehormatan individu.

Dalam perspektif Al-Ghazali, fenomena *deepfake* dapat dikategorikan sebagai bentuk modern dari *kadzib* (kebohongan) yang bertentangan dengan prinsip *sidq* (kejujuran). Al-Ghazali menegaskan bahwa kejujuran tidak hanya terbatas pada ucapan, tetapi mencakup keselarasan antara niat, perkataan, dan tindakan. Oleh karena itu, praktik *deepfake* yang menyajikan sesuatu yang tidak sesuai dengan kenyataan secara jelas bertentangan dengan nilai-nilai etika Islam dan berpotensi merusak integritas moral individu maupun tatanan sosial. Melalui kerangka *maqāṣid al-sharī'ah*, *deepfake* terbukti bertentangan dengan tujuan dasar syariat, khususnya dalam menjaga akal (*hifz al-'aql*), menjaga kehormatan (*hifz al-'ird*), dan menjaga harta (*hifz al-māl*). Hal ini menunjukkan bahwa dampak *deepfake* tidak hanya bersifat individual, tetapi juga meluas ke ranah sosial dan struktural, sehingga memerlukan pendekatan yang komprehensif dalam penanganannya. Oleh karena itu, nilai-nilai seperti *sidq*, *amanah*, dan *tazkiyyatun nafs* relevan untuk dijadikan landasan etis dalam penggunaan dan pengembangan teknologi digital.

Penelitian ini masih memiliki keterbatasan karena bersifat kajian kepustakaan yang bertumpu pada analisis sumber-sumber tertulis, sehingga kajian yang dihasilkan lebih bersifat konseptual dan belum didukung oleh data empiris di lapangan. Selain itu, penelitian ini hanya berfokus pada pemikiran Al-Ghazali dan belum mengkaji perbandingan dengan tokoh atau pendekatan etika lainnya. Oleh karena itu, hasil penelitian ini perlu dipahami sebagai kajian awal yang membuka ruang bagi pengembangan analisis yang lebih komprehensif.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya diharapkan dapat mengembangkan kajian ini ke arah yang lebih luas aplikatif. Penggunaan pendekatan empiris, seperti studi kasus, observasi, atau survei, menjadi penting untuk melihat secara langsung bagaimana nilai-nilai etika seperti *sidq*, *amanah*, dan *tabayyun* diimplementasikan dalam praktik bermedia digital. Selain itu, penelitian lanjutan juga dapat mengintegrasikan perspektif etika Islam dengan pendekatan teknologi dan kebijakan publik, sehingga mampu menghasilkan formulasi solusi yang lebih komprehensif dalam menghadapi fenomena *deepfake*. Kajian komparatif dengan pemikiran tokoh lain maupun teori etika modern juga diperlukan guna

memperkaya analisis serta memperluas cakrawala pemahaman tentang etika teknologi di era kecerdasan buatan. Dengan demikian, penelitian ini di masa mendatang diharapkan tidak hanya berhenti pada tataran konseptual, tetapi juga mampu memberikan kontribusi praktis dalam membangun integritas moral dan memperkuat kepercayaan publik di ruang digital.

DAFTAR REFERENSI

- Ali, M. F., & Mat Teh, K. S. (2025). Etika dan adab penggunaan kecerdasan buatan (AI) dalam pembelajaran bahasa Arab menurut perspektif Imam Al-Ghazali dan Ibn Miskawayh. *Jurnal Al-Haady*, 7(1), 22–31.
- Boediman, E. P. (2025). *Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review* (Vol. 19). <https://doi.org/10.20885/komunikasi.vol19.iss2.art8>
- Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98, 147.
- Desisca, A., Ramadhanti, E., & Khan, A. (2025). Etika kecerdasan buatan: Tantangan baru dalam filsafat moral. *Journal of Development Economics and Digitalization, Tourism Economics*, 2(1), 59–69. <https://doi.org/10.70248/jdedte.v2i1.1746>
- Habib, Z. (2025). Ethics of artificial intelligence in Maqāsid Al-Sharīa's perspective. *KARSA: Journal of Social and Islamic Culture*, 33(1), 105–134. <https://doi.org/10.19105/karsa.v33i1.19617>
- Haida, R. S. N., & Nuriyatman, E. (2024). Urgensi pengaturan perlindungan hukum terhadap korban deepfake melalui artificial intelligence (AI) dari perspektif hukum pidana Indonesia. *Jurnal Hukum Respublica*, 24(1).
- Hasiah. (2014). Konsep kejujuran (Al-Sidq) dalam perspektif hadis. *Studi Multidisipliner*, 1.
- Librianti, E. O. I. (2025). Deepfake dalam komunikasi politik: Tantangan etika dan aspek hukum dalam era artificial intelligence. *Siyasah*, 5(2), 221–238. <https://doi.org/10.32332/siyasah.v4i1>
- Moleong, L. J. (2004). *Metodologi penelitian kualitatif*. Remaja Rosdakarya.
- Neonnub, P., & Kosat, O. (2026). Metaphysical foundations of AI ethics: A Thomistic-Islamic dialogue on human dignity and moral agency in the digital age. In *Proceeding International Conference on Islam and Civilization (ICONIC)* (Vol. 1, pp. 467–475).
- Ningrum, H. M. C., Mualim, A. N., Sudarman, S., & Bedi, F. (2024). Artificial intelligence in Al-Ghazali's perspective: Bridging the concept of nafs and machine. *Islamic Thought Review*, 2(2), 140–150. <https://doi.org/10.30983/itr.v2i2.8849>
- Pane, M. D., & Permana, M. Z. S. (2025). Pertanggungjawaban pidana terhadap pengembang artificial intelligence pada kasus pelanggaran privasi dan data pribadi. *Judge: Jurnal Hukum*, 6(3), 467–472. <https://doi.org/10.54209/judge.v6i03.1593>
- Putra, Z. F., FISIP UMJ, I. K., & Muksin, N. N. (2025). Deepfake dan disinformasi: Krisis kebenaran di era digital kajian filsafat dan etika komunikasi. *Kompasiana*.
- Salsabila, N., Nursyamsi, S. E., Lailani, N., Safitri, U., & Azzahra, S. F. (2025). Etika AI dalam konteks disinformasi digital: Penyalahgunaan identitas pada penipuan online. *Politika*

Progresif: Jurnal Hukum, Politik dan Humaniora, 2(4), 199–209.
<https://doi.org/10.62383/progres.v2i4.2747>

- Serevan, B. (2025). Al-Ghazali's principle of trust (amanah) as a framework for ethical AI governance in organizations. *AI and Ethics*, 5(5), 5273–5282.
<https://doi.org/10.1007/s43681-025-00780-y>
- Surajiyo. (2026). Progresivitas kecerdasan buatan dalam perspektif epistemologi dan implikasinya bagi ilmu pengetahuan kontemporer. *Journal of Instructional and Development Researches*, 6(1), 102–110. <https://doi.org/10.53621/jider.v6i1.729>
- Syifa, A., & Ridwan, A. (2024). Pendidikan karakter Islami di era digital: Tantangan dan solusi berdasarkan pemikiran sosial Imam Al-Ghazali. *Social Studies in Education*, 2(2), 107–122.
- Wahyudi, B. R. (2025). Tantangan penegakan hukum terhadap kejahatan berbasis teknologi AI. *Innovative: Journal of Social Science Research*, 5(1), 3436–3450.
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11). <https://doi.org/10.22215/timreview/1282>
- Xie, Y., Fu, R., Wang, X., Wang, Z., Li, Y., & Wen, Z. (2025). Fake speech wild: Detecting deepfake speech on social media platform. *arXiv*, 1–5.
<https://doi.org/10.48550/arXiv.2508.10559>