

ETIKA DAN RISIKO KECERDASAN ARTIFISIAL GENERATIF: IMPLIKASI DALAM PENDIDIKAN AGAMA ISLAM DAN MASYARAKAT DIGITAL

Received: 22-06-2025 | Revised: 31-01-2026 | Accepted: 31-01-2026

Article Info

Author(s):

Shokhibul Mighfar^{1*}
Moh. Hafid²
Ikhliah Qotrinnada³

* Author's Email Correspondence:

shokhibulighfar@ibrahimiy.ac.id

Affiliation:

^{1,2} Pendidikan Agama Islam,
Universitas Ibrahimy, Jawa Timur,
68374, Indonesia.

³ World Moslem Studies Center,
Jawa Barat, 16421, Indonesia.

Abstract

Generative Artificial Intelligence (AI) has increasingly influenced educational practices, including Islamic Religious Education (IRE), by enabling automated content creation, AI-based assessment, and adaptive learning. However, its use also raises ethical concerns related to transparency, fairness, copyright, religious misinformation, data bias, and student privacy. This article critically examines the concepts, implementation, ethical dimensions, and potential risks of generative AI in education, with particular attention to its relevance within IRE. Employing a systematic literature review of indexed academic publications, international policy reports, and ethical frameworks for AI in education, the study finds that generative AI can serve as an effective supportive technology for IRE when used responsibly and grounded in core Islamic values, including trustworthiness (amanah), justice ('adl), honesty (sidq), and public interest (maṣlaḥah). The findings reaffirm the central role of IRE teachers as moral guides, interpreters of Islamic teachings, and pedagogical decision-makers, while positioning generative AI as an assistive tool that enhances learning rather than replacing the educator's moral and spiritual authority. This study contributes both theoretically and practically to the development of ethical, inclusive, and context-sensitive policies and practices for the use of generative AI in Islamic Religious Education.

Keywords: *Generative AI, IRE, Ethics, Educational Technology, Artificial Intelligence.*



© 2026 THE AUTHOR(S). This article distributed under a Creative Commons Attribution-ShareAlike 4.0 International license. <https://creativecommons.org/licenses/by-sa/4.0/>

PENDAHULUAN

Perkembangan teknologi Kecerdasan Artifisial (AI) telah mengalami lonjakan signifikan dalam satu dekade terakhir, terutama dengan hadirnya AI generatif yang mampu menciptakan konten baru, seperti teks, gambar, musik, bahkan video, secara otonom. Model-model seperti *Generative Pre-trained Transformer* (GPT) dan DALL·E telah memperlihatkan kapabilitas yang mengagumkan dalam menghasilkan keluaran yang menyerupai buatan manusia, membuka peluang revolusioner di berbagai bidang, termasuk pendidikan, komunikasi, seni, dan hiburan (Brown et al.,

2020, p. 18-21). Teknologi ini mampu menghasilkan konten yang mirip dengan buatan manusia dengan kecepatan dan skala yang luar biasa (Goodfellow et al., 2014, p. 2673). AI generatif merepresentasikan bentuk baru dari automasi kreatif, berbeda dari AI konvensional yang lebih fokus pada klasifikasi atau prediksi.

Di bidang pendidikan, integrasi AI generatif menawarkan banyak potensi. Misalnya, pengajar dapat memanfaatkan AI untuk menghasilkan materi ajar, membuat soal, menilai tugas, atau memberikan umpan balik secara personalisasi. Peserta didik juga dapat memperoleh manfaat dari tutor virtual yang adaptif dan konten belajar yang disesuaikan dengan kebutuhan individual (Ali et al., 2024, p. 10-19; UNESCO, 2023, p. 12–14). Namun, dibalik manfaat tersebut, muncul tantangan besar terkait isu etika, transparansi, dan risiko penyalahgunaan teknologi (Floridi & Cowls, 2019, p. 7; Alan, 2024, p. 5–6). Contoh nyata adalah kemungkinan terjadinya plagiarisme, manipulasi informasi, atau penilaian yang bias akibat data pelatihan yang tidak representatif (Mehrabi et al., 2021, p. 6-8; Zellers et al., 2019, p. 9056; UNESCO, 2023, p. 28).

Permasalahan terkait keakuratan informasi, penyalahgunaan data, dan ketidakadilan algoritmik menjadi sorotan penting dalam pemanfaatan AI generatif, khususnya ketika teknologi ini digunakan dalam konteks pendidikan dan pengambilan keputusan publik (Brown et al., 2020, p. 24–26; Carlini et al., 2021, p. 269–272; Floridi et al., 2018, p. 694). Etika dan tata kelola penggunaan teknologi ini harus dikembangkan secara paralel dengan kemajuan teknologinya untuk menghindari dampak negatif terhadap masyarakat. Artikel ini bertujuan untuk memberikan pemahaman menyeluruh mengenai AI generatif, etika yang melingkupinya, serta risiko yang mungkin muncul dalam implementasinya.

Sebagian besar penelitian terdahulu mengenai kecerdasan artifisial masih berfokus pada aspek teknis pengembangan algoritma dan peningkatan performa model (LeCun et al., 2015, p. 438; Brown et al., 2020, p. 18), serta penerapannya dalam sektor bisnis dan industri kreatif (Goodfellow et al., 2014, p. 2673; Zellers et al., 2019, p. 9055). Sementara itu, kajian yang secara simultan mengintegrasikan dimensi teknis, pedagogis, dan etis AI generatif dalam konteks pendidikan, khususnya Pendidikan Agama Islam, masih relatif terbatas (UNESCO, 2023, p. 9; Alan Turing Institute, 2024, p. 3). Kebaharuan dari artikel ini adalah pendekatannya yang holistik dan interdisipliner dalam membahas AI generatif. Artikel ini tidak hanya menguraikan aspek teknis dan implementatif, tetapi juga menganalisis persoalan etika dan risiko sosial secara mendalam dalam konteks pendidikan.

Penelitian ini berkontribusi dalam menyediakan perspektif etis dan kritis yang seimbang terhadap pemanfaatan AI generatif dalam masyarakat digital, khususnya di sektor pendidikan.

Mengingat adopsi teknologi ini semakin meluas, analisis yang menyeluruh mengenai konsekuensi sosial dan etisnya menjadi sangat mendesak. Tanpa adanya kerangka etika dan regulasi yang memadai, pemanfaatan AI generatif berisiko memperburuk ketimpangan informasi, memperkuat diskriminasi algoritmik, dan merusak kepercayaan publik terhadap institusi pendidikan (Floridi et al., 2018, p. 694; UNESCO, 2023, p. 42; Binns, 2018, p. 150).

Artikel ini bertujuan untuk menjelaskan secara sistematis dan mendalam mengenai konsep dasar AI generatif, prinsip kerjanya, serta bagaimana teknologi ini diimplementasikan dalam konteks pendidikan. Selain itu, artikel ini juga bertujuan untuk mengkaji aspek etis yang melingkupi penggunaan AI generatif serta risiko-risiko yang menyertainya, seperti penyebaran hoaks, bias data, dan pelanggaran privasi. Artikel ini diharapkan dapat menjadi referensi akademik yang komprehensif sekaligus praktis, baik bagi peneliti, pendidik, maupun pembuat kebijakan yang sedang menavigasi transformasi digital berbasis AI generatif.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan jenis penelitian kepustakaan (*library research*). Pendekatan ini dipilih karena penelitian berfokus pada kajian konseptual dan analitis mengenai etika serta risiko kecerdasan artifisial generatif dan implikasinya dalam bidang pendidikan serta masyarakat digital, tanpa melibatkan pengumpulan data lapangan (Creswell, 2014, p. 18).

Sumber data penelitian terdiri atas literatur ilmiah yang relevan, meliputi artikel jurnal nasional dan internasional, buku referensi, serta dokumen kebijakan yang membahas kecerdasan artifisial, etika teknologi, pendidikan digital, dan dampak sosial AI (Floridi et al., 2018, p. 694). Pengumpulan data dilakukan melalui studi dokumentasi, dengan cara menelusuri, membaca, dan menyeleksi sumber-sumber pustaka yang sesuai dengan fokus penelitian (Sugiyono, 2020, p. 35). Sumber literatur yang dijadikan acuan oleh peneliti berasal dari *data base* SINTA, Google Scholar, DOAJ.

Tabel 1. Sumber Literatur

No.	Sumber Literatur	Jumlah (%)
1	SINTA	50%
2	Google Scholar	15%
3	DOAJ	10%
4	Elsevier	25%

Analisis data dilakukan secara deskriptif-analitis, yaitu dengan mendeskripsikan konsep dan perkembangan kecerdasan artifisial generatif, kemudian menganalisis isu etika dan risiko yang

EDUPEDIA:

ditimbulkannya, khususnya dalam konteks pendidikan dan masyarakat digital (Miles & Huberman, 2014, p. 72). Tahapan analisis meliputi reduksi data, pengelompokan tema, dan penarikan kesimpulan. Keabsahan data dijaga dengan memperhatikan kredibilitas dan relevansi sumber, serta membandingkan berbagai referensi untuk memperoleh pemahaman yang objektif dan komprehensif mengenai topik yang diteliti.

HASIL DAN PEMBAHASAN

Konsep Kecerdasan Artifisial Generatif: *Artificial Intelligence*, *Machine Learning* dan *Deep Learning*

Kecerdasan Artifisial merupakan bidang dalam ilmu komputer yang berfokus pada pengembangan mesin yang mampu meniru tindakan atau perilaku manusia (Russell & Norvig, 2021, p. 1-5). Kecerdasan Artifisial merupakan teknologi yang dirancang untuk meniru perilaku kognitif manusia melalui pemanfaatan mesin yang mampu belajar dari data, serta menjalankan tugas-tugas tertentu tanpa memerlukan instruksi eksplisit mengenai hasil akhir yang diharapkan (Binns, 2018, p. 150). *Machine Learning* adalah bagian dari kecerdasan buatan yang memungkinkan sistem komputer memperoleh pengetahuan atau pola dari data tanpa memerlukan instruksi pemrograman (*prompt*) yang eksplisit (Mitchell, 1997, p. 2). *Deep Learning* merupakan sub-bidang dari *Machine Learning* yang memanfaatkan jaringan saraf tiruan berlapis-lapis (*deep neural networks*) untuk menganalisis dan mengolah data dalam skala besar serta dengan tingkat kompleksitas yang tinggi (Goodfellow et al., 2016, p. 1–3).

Kecerdasan Artifisial Generatif (*Generative AI*) merupakan salah satu cabang dari *Deep Learning* yang berfokus pada pembuatan data sintesis yang menyerupai karakteristik data pelatihan. Pendekatan ini diwujudkan melalui penggunaan model-model seperti *Generative Adversarial Networks* (GANs) serta arsitektur berbasis Transformer seperti GPT, yang mampu mempelajari distribusi data kompleks dan merekonstruksi representasi baru yang konsisten dengan data tersebut (Goodfellow et al., 2014, p. 2672–2674., Brown et al., 2020, p. 1877–1880).

Artificial Intelligence (AI) atau Kecerdasan Artifisial merupakan cabang ilmu komputer yang berfokus pada perancangan sistem cerdas yang mampu menjalankan berbagai fungsi yang secara konvensional memerlukan kecerdasan manusia, seperti penalaran, pembelajaran, dan pengambilan keputusan. Tugas-tugas tersebut mencakup penalaran, pemahaman bahasa, pengambilan keputusan, serta pengenalan pola dan suara. AI modern mencakup beragam pendekatan, mulai dari sistem berbasis aturan hingga model pembelajaran data, yang kini didominasi oleh *Machine Learning* (ML) dan *Deep Learning* (DL) sebagai pendekatan yang paling dominan.

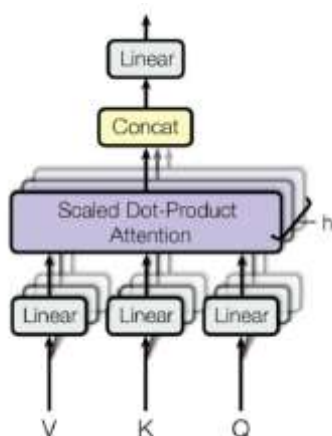
Machine Learning (ML) merupakan subdisiplin dari AI yang memungkinkan sistem untuk belajar dan meningkatkan kinerja dari data tanpa pemrograman eksplisit. Dengan kata lain, ML memungkinkan komputer untuk mengidentifikasi pola dan membuat prediksi berdasarkan data historis (Russell & Norvig, 2021, p. 695–697). ML bekerja dengan algoritma yang dilatih menggunakan dataset yang relevan untuk kemudian diaplikasikan dalam konteks baru. *Machine Learning* (ML) adalah pendekatan di mana sistem belajar dari data untuk meningkatkan kinerjanya terhadap suatu tugas tertentu tanpa diprogram secara eksplisit. *Machine Learning* (ML) memanfaatkan algoritma untuk mengenali pola-pola dalam data historis guna menghasilkan prediksi atau mendukung pengambilan keputusan pada situasi atau konteks yang belum pernah ditemui sebelumnya. Algoritma dalam ML diklasifikasikan menjadi tiga kategori utama: pembelajaran terawasi (*supervised learning*), pembelajaran tak terawasi (*unsupervised learning*), dan pembelajaran penguatan (*reinforcement learning*). Setiap kategori memiliki pendekatan yang disesuaikan dengan karakteristik data dan tujuan penggunaannya.

Supervised learning banyak digunakan dalam aplikasi klasifikasi dan regresi, di mana algoritma dilatih menggunakan data yang telah diberi label. *Unsupervised learning* digunakan untuk menemukan struktur tersembunyi dalam data tanpa label, seperti dalam segmentasi atau klusterisasi. Sementara itu, *reinforcement learning* digunakan dalam situasi di mana agen harus belajar melalui interaksi dengan lingkungan untuk memaksimalkan fungsi penghargaan (*reward function*) (Goodfellow et al., 2016, p. 2–4).

Deep Learning (DL) merupakan sub-bidang dari *Machine Learning* yang mengandalkan arsitektur jaringan saraf tiruan dengan banyak lapisan (*deep neural networks*). Pendekatan ini meniru mekanisme pemrosesan informasi pada otak manusia, di mana data dianalisis secara bertahap melalui lapisan-lapisan representasi yang semakin abstrak dan kompleks. Pendekatan ini terbukti efektif dalam menangani data besar dan tidak terstruktur seperti citra, suara, pengenalan wajah, pengolahan bahasa alami (*Natural Language Processing/NLP*) (LeCun et al., 2015, p. 436–438).

Arsitektur populer dalam DL mencakup *Convolutional Neural Networks* (CNN) untuk analisis citra dan *Recurrent Neural Networks* (RNN) untuk data berurutan. Di bidang pemrosesan bahasa alami, Transformer menjadi arsitektur dominan yang mendasari model-model seperti BERT dan GPT (Vaswani et al., 2017, p. 5999–6001; Brown et al., 2020, p. 1878–1881). Arsitektur Transformer mengubah paradigma pemrosesan sekuensial dengan mengandalkan mekanisme atensi (*attention mechanism*, semacam operasi matematika yang melibatkan matriks dengan dimensi banyak yang memungkinkan model untuk mempelajari hubungan antar kata dalam satu kalimat atau bahkan antar kalimat secara simultan (Dirjen GTKPG Dikdasmen, 2025).

EDUPEDIA:

Gambar 1. *Multi-Head Attention*

DL berbeda dari ML konvensional dalam hal kebutuhan akan data pelatihan yang sangat besar dan sumber daya komputasi yang tinggi. Meski demikian, kemampuan DL untuk belajar dari data mentah tanpa membutuhkan rekayasa fitur manual menjadikannya sangat unggul dalam menangani kompleksitas data digital modern. UNESCO (2023) menyatakan bahwa penggunaan DL dalam pendidikan dapat memungkinkan penciptaan lingkungan pembelajaran multimodal yang lebih imersif dan adaptif. Khusus pada AI generatif, DL memainkan peran sentral karena memungkinkan pengembangan model yang mampu menghasilkan teks, gambar, dan konten lainnya secara otonom. *Generative Adversarial Networks* (GAN) dan *Large Language Models* (LLM) seperti GPT-4 dibangun berdasarkan arsitektur DL dan telah menunjukkan kemampuan tinggi dalam meniru kreativitas manusia (Goodfellow et al., 2014, p. 2673–2675; Alan, 2024, p. 3).

Penggunaan ML dan DL dalam pendidikan harus disertai dengan kesadaran terhadap etika, transparansi, dan akuntabilitas. Studi UNESCO (2023) dan *Teaching AI Best Practices* (Florida, 2024, p. 1) menekankan pentingnya literasi AI di kalangan pendidik dan peserta didik agar teknologi ini digunakan dengan bijak dan bertanggung jawab. Pemahaman mendalam terhadap prinsip kerja ML dan DL memungkinkan pendidik untuk tidak hanya menjadi pengguna pasif, melainkan aktor kritis dalam desain, seleksi, dan evaluasi teknologi berbasis AI.

Dengan demikian, integrasi ML dan DL ke dalam ekosistem pendidikan tidak hanya memerlukan keahlian teknis, tetapi juga kerangka kerja pedagogis dan etis yang matang. Pendekatan ini memastikan bahwa adopsi teknologi bukan semata-mata bersifat instrumental, tetapi juga mendorong peningkatan kualitas pembelajaran yang berkelanjutan dan inklusif.

AI generatif merupakan cabang lanjutan dari DL yang tidak hanya mengenali pola, tetapi juga menciptakan data baru yang menyerupai data pelatihan. Model generatif mempelajari distribusi

probabilistik dari data input untuk dapat menghasilkan keluaran yang orisinal. Contohnya termasuk *Generative Adversarial Networks* (GAN) dan model berbasis transformer seperti GPT (*Generative Pretrained Transformer*) (Goodfellow et al., 2014, p. 2673–2675; Alan, 2024, p. 3).

Penting untuk dicatat bahwa keberhasilan AI generatif sangat bergantung pada kualitas dataset yang digunakan. Dalam pendidikan, penggunaan AI generatif tanpa pemahaman dasar tentang konsep AI, ML, dan DL berisiko melahirkan ketergantungan pada hasil tanpa pengawasan kritis. Oleh karena itu, literasi AI menjadi aspek mendasar yang harus dimiliki oleh pendidik dan peserta didik (UNESCO, 2023).

Menurut Alan (2024, p. 3), model AI generatif dapat secara efektif mendukung pembelajaran jika digunakan sebagai alat bantu dalam menghasilkan umpan balik otomatis, membuat konten pengajaran adaptif, dan menciptakan lingkungan pembelajaran yang imersif. Namun, penggunaan tanpa pemahaman dasar ML dan DL dapat menyebabkan kesalahan interpretasi dan penyalahgunaan teknologi. Sebagai tambahan, penelitian oleh Ali et al. (2024, p. 4–7) menggarisbawahi bahwa pemahaman tentang diferensiasi antara model diskriminatif (yang memutuskan) dan generatif (yang mencipta) menjadi sangat penting dalam ranah pendidikan. Hal ini karena model generatif tidak hanya digunakan untuk klasifikasi, tetapi juga dalam pengembangan esai, gambar, dan simulasi pembelajaran yang kreatif. Dalam konteks praktis, pemahaman yang memadai mengenai prinsip kerja ML dan DL membantu guru dan pendidik untuk merancang strategi pembelajaran berbasis AI yang etis dan efektif. Hal ini mencakup kemampuan untuk mengevaluasi keluaran AI, memvalidasi keakuratannya, serta mengenali bias potensial yang mungkin terkandung dalam model maupun dataset pelatihannya (Meechan, 2024, p. 60–65; UNESCO, 2023).

Dalam konteks pendidikan, ML dapat digunakan untuk mendeteksi pola perilaku belajar siswa, memprediksi performa akademik, dan merancang strategi pembelajaran personalisasi. Penelitian oleh Ali et al. (2024, p. 10-19) menunjukkan bahwa dengan data yang memadai dan model yang tepat, ML mampu menghasilkan sistem rekomendasi pembelajaran yang adaptif dan responsif terhadap kebutuhan individu.

Dengan demikian, integrasi AI generatif ke dalam sistem pendidikan tidak bisa dilakukan secara sembarangan. Perlu adanya pemahaman mendalam terhadap fondasi teknologinya, yakni AI, ML, dan DL agar pemanfaatan teknologi ini benar-benar mendukung tujuan pendidikan yang transformatif, bukan sekadar menjadi alat produksi otomatis yang tidak kritis. Berikut ini merupakan kesimpulan dari urgensi AI dalam pembelajaran sebagaimana tersaji dalam tabel di bawah ini.

EDUPEDIA:

Tabel 2. Urgensi AI dalam Pembelajaran

No	Variasi AI	Urgensi	Implikasi
1	AI Generatif	AI generatif dalam pendidikan perlu dilakukan secara cermat, yakni harus dirancang secara sistematis dan berbasis pemahaman teknologi.	Sebagai pendukung transformasi pendidikan yang kritis dan bermakna, bukan sekadar alat otomatisasi tanpa refleksi pedagogis.
2	Machine Learning	Menghasilkan sistem rekomendasi pembelajaran yang adaptif dan responsif terhadap kebutuhan individu peserta didik.	Dapat digunakan untuk mendeteksi pola perilaku belajar siswa, memprediksi performa akademik, dan mendukung perancangan pembelajaran yang bersifat personalisasi.

Prinsip Kerja AI Generatif

Prinsip dasar AI generatif melibatkan pelatihan model pada dataset besar untuk memahami pola dan struktur data tersebut. Misalnya, GPT-4 dilatih dengan corpus teks dalam jumlah besar sehingga dapat menghasilkan teks yang koheren berdasarkan prompt yang diberikan.

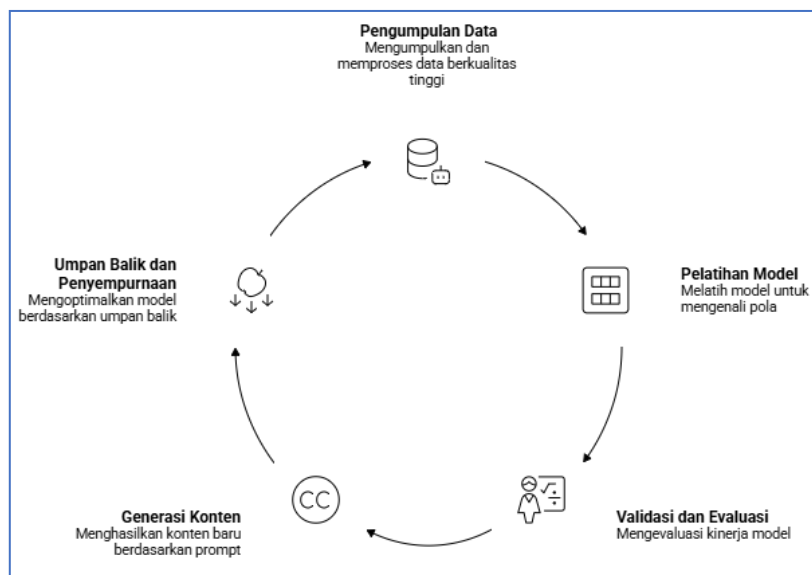
Generative Artificial Intelligence merupakan subbidang dari kecerdasan artifisial yang dirancang untuk menciptakan data baru yang menyerupai distribusi data pelatihan. Prinsip kerjanya bertumpu pada kemampuan model untuk mempelajari pola statistik dari data masukan dan menggunakannya untuk menghasilkan keluaran yang baru namun realistis. AI generatif menggunakan pendekatan probabilistik untuk meniru kompleksitas data alami, seperti teks, gambar, audio, dan video (Goodfellow et al., 2014, p. 2672–2674; Brown et al., 2020, p. 1878–1882).

Secara teknis, AI generatif bekerja melalui pelatihan model pada dataset berukuran besar. Dalam proses pelatihan ini, model belajar mengenali korelasi, struktur semantik, dan distribusi representasi dari data tersebut. Model kemudian mengandalkan apa yang telah dipelajarinya untuk membangkitkan output baru. Dua pendekatan arsitektural yang paling umum dalam AI generatif adalah Generative Adversarial Networks (GAN) dan Transformer-based models, termasuk Large Language Models (LLM) seperti GPT (*Generative Pretrained Transformer*).

Generative Adversarial Networks (GAN), yang diperkenalkan oleh Goodfellow et al. (2014, p. 651–653), terdiri atas dua komponen utama: *generator* dan *discriminator*. Generator bertugas menciptakan data palsu yang menyerupai data asli, sedangkan discriminator berperan sebagai evaluator yang membedakan mana data asli dan mana data hasil generasi. Keduanya dilatih secara bersamaan dalam kerangka permainan minimaks hingga generator mampu menghasilkan data yang tidak dapat dibedakan dari data nyata oleh discriminator. Proses pelatihan ini menghasilkan model yang mampu meniru distribusi data asli secara sangat realistis (Goodfellow et al., 2014, p. 2673–2676).

Sementara itu, model berbasis Transformer, seperti GPT, menggunakan pendekatan *auto-regresif* dalam menghasilkan teks. Model ini dilatih menggunakan metode pembelajaran tidak terawasi (*unsupervised learning*) pada korpus teks yang sangat besar. Proses pelatihannya menggunakan teknik *pretraining* dan *fine-tuning*. Pada tahap *pretraining*, model mempelajari distribusi probabilistik dari token-token dalam teks. Sedangkan pada tahap *fine-tuning*, model diadaptasi pada tugas-tugas spesifik seperti menjawab pertanyaan, menerjemahkan, atau merangkum teks (Brown et al., 2020, p. 18-21).

Keunggulan dari pendekatan Transformer terletak pada mekanisme *attention*, yang memungkinkan model untuk mempertimbangkan konteks panjang dalam urutan input. Ini menjadikan model transformer sangat efektif dalam menangani tugas pemrosesan bahasa alami (NLP). Transformer juga dapat diadaptasi untuk menghasilkan gambar seperti DALL·E, audio seperti Jukebox, dan bentuk data multimodal lainnya, menjadikannya arsitektur yang fleksibel dan multiguna dalam AI generatif (Vaswani et al., 2017, p. 5999–6002). Dalam implementasinya, prinsip kerja AI generatif mencakup beberapa tahapan (UNESCO, 2023), sebagaimana tersaji dalam gambar berikut ini.



Gambar 2. Tahapan-Tahapan Prinsip Kerja AI Generatif

Dalam konteks pendidikan, pemahaman tentang prinsip kerja ini menjadi penting untuk memastikan bahwa pendidik dan peserta didik tidak hanya menggunakan AI sebagai alat bantu teknis, tetapi juga memahami keterbatasan dan potensi kesalahannya. Sebagai contoh, Alan (2024, p. 18) menyoroti pentingnya keterlibatan manusia dalam menginterpretasikan dan memverifikasi

hasil yang dihasilkan oleh model AI generatif, terutama ketika digunakan dalam pembuatan materi ajar, umpan balik otomatis, dan asesmen.

Lebih lanjut, workshop yang diselenggarakan oleh LEMKOMINDO (Sibuea et al., 2024, p. 47–50) menunjukkan bahwa pelatihan tentang prinsip kerja AI generatif membantu meningkatkan efektivitas dan inovasi dalam pembelajaran di perguruan tinggi. Pemahaman tentang arsitektur model, teknik pelatihan, dan parameter kontrol keluaran (seperti *temperature* dan *top-k sampling*) menjadi komponen penting dalam memanfaatkan AI generatif secara optimal.

Namun demikian, kompleksitas teknis dari model generatif juga membawa risiko etis dan operasional. Tanpa pemahaman prinsip kerjanya, pengguna dapat dengan mudah salah tafsir terhadap keluaran AI atau menggunakannya dalam konteks yang tidak tepat. Oleh karena itu, literasi AI harus mencakup aspek teknis dan etis secara menyeluruh.

Keberhasilan AI generatif sangat tergantung pada kualitas dan representativitas data pelatihan. Data yang bias atau tidak lengkap akan menghasilkan model yang juga bias dan tidak akurat. Oleh karena itu, pemilihan dan kurasi data merupakan aspek krusial dalam pengembangan AI generatif (Gebru et al., 2018, p. 87–90).

Kualitas data merupakan fondasi utama dalam pengembangan dan kinerja sistem kecerdasan artifisial generatif. Dalam konteks AI generatif, data tidak hanya berfungsi sebagai bahan bakar dalam pelatihan model, tetapi juga menentukan arah dan sifat dari keluaran yang dihasilkan. Data yang berkualitas tinggi akan mendorong model menghasilkan keluaran yang relevan, akurat, dan etis. Sebaliknya, data yang buruk, bias, atau tidak representatif dapat menyebabkan model AI generatif memproduksi informasi yang menyesatkan, bias, atau bahkan merugikan secara sosial (Gebru et al., 2018, p. 88–91; UNESCO, 2023).

Dalam pengembangan AI generatif berbasis teks seperti GPT, corpus data pelatihan mencakup miliaran kata yang dikumpulkan dari berbagai sumber daring, termasuk buku, artikel ilmiah, laman web, dan forum diskusi. Oleh karena itu, proses kurasi data menjadi sangat penting. Tanpa penyaringan yang cermat, model berisiko menyerap informasi yang bersifat bias, ofensif, atau tidak faktual (Brown et al., 2020, p. 1880–1884; Alan, 2024, p. 3).

UNESCO menekankan bahwa sistem AI yang tidak memperhatikan etika kurasi data dapat memperkuat ketimpangan, stereotip, dan eksklusi digital, terutama bagi kelompok rentan. Dalam pendidikan, hal ini menjadi sangat krusial karena model yang digunakan untuk memberikan umpan balik, menilai karya tulis, atau menyusun materi ajar harus bebas dari bias yang dapat mendiskriminasi peserta didik berdasarkan jenis kelamin, ras, budaya, atau bahasa.

Workshop LEMKOMINDO yang diselenggarakan oleh Sibuea et al. (2024, p. 50–52) menunjukkan bahwa adaptasi AI generatif untuk konteks Indonesia memerlukan data pelatihan yang merepresentasikan bahasa, budaya, dan norma lokal. Tanpa representasi lokal yang cukup, model akan cenderung menggeneralisasi berdasarkan data global yang mungkin tidak relevan secara kultural.

Selain itu, pemeliharaan kualitas data bersifat berkelanjutan. Seiring dengan perkembangan waktu dan perubahan sosial, data yang dahulu relevan dapat menjadi usang atau tidak etis. Oleh karena itu, diperlukan mekanisme audit dan pembaruan data secara periodik untuk memastikan bahwa model AI generatif tetap akurat dan sesuai dengan nilai-nilai kontemporer (Meechan, 2024, p. 114–117).

Langkah-langkah praktis untuk menjaga kualitas data dalam pengembangan AI generatif, di antaranya adalah pembuatan *datasheets for datasets* untuk dokumentasi dan transparansi data (Gebru et al., 2018, p. 90–92), pelibatan pakar multidisiplin dalam proses kurasi dan validasi data, penggunaan teknik data *augmentation* yang hati-hati untuk meningkatkan keragaman tanpa menambahkan *noise*, dan pengembangan sistem deteksi dan mitigasi bias berbasis metrik kuantitatif.

Dengan demikian, kualitas data bukan hanya masalah teknis, tetapi juga menyangkut tanggung jawab sosial, etika, dan hukum. Dalam konteks pendidikan, kualitas data menentukan sejauh mana AI generatif dapat benar-benar menjadi alat bantu pedagogis yang adil, inklusif, dan efektif.

Implementasi *Artificial Intelligence* dalam Pendidikan Agama Islam Sebagai Pembuatan Materi Ajar Otomatis

Salah satu implementasi paling transformatif dari AI generatif dalam sektor pendidikan Agama Islam adalah kemampuan untuk menghasilkan materi ajar secara otomatis. Teknologi ini memungkinkan pendidik untuk menyusun konten pembelajaran seperti modul, ringkasan materi, soal evaluasi, dan skenario pembelajaran berbasis masalah secara lebih cepat dan adaptif terhadap kebutuhan siswa.

Pembuatan materi ajar otomatis berbasis AI generatif bekerja melalui prinsip pemrosesan bahasa alami (*Natural Language Processing/NLP*), di mana model seperti GPT (*Generative Pretrained Transformer*) dilatih untuk memahami konteks pembelajaran dari prompt yang diberikan. Melalui input berupa topik, tujuan pembelajaran, dan tingkat kesulitan, model dapat menghasilkan konten tekstual yang koheren, relevan, dan sesuai dengan tingkat pemahaman siswa (Brown et al., 2020, p. 1877-1901).

EDUPEDIA:

Jurnal Studi Pendidikan dan Pedagogi Islam
Vol. 10 Nomor 2

Sebagai contoh, seorang guru yang mengajar mata pelajaran Biologi kelas XI dapat meminta sistem AI untuk menghasilkan modul ajar mengenai “sistem ekskresi pada manusia” yang mencakup pengertian, fungsi, organ utama, serta dilengkapi dengan ilustrasi dan latihan soal pilihan ganda. Dalam hitungan detik, sistem AI mampu menyusun dokumen yang dapat langsung digunakan atau disesuaikan oleh guru untuk kebutuhan spesifik kelasnya.

Lebih lanjut, AI generatif juga dapat digunakan untuk menghasilkan ringkasan teks dari buku ajar atau artikel ilmiah, membantu siswa memahami konten kompleks secara lebih cepat, latihan soal adaptif, yang disesuaikan dengan tingkat kemampuan siswa berdasarkan data interaksi sebelumnya, skenario studi kasus yang cocok untuk pengajaran berbasis proyek dan pengembangan pemikiran kritis.

Menurut UNESCO (2023, p. 8-9), pembuatan materi ajar otomatis dengan AI generatif memungkinkan tercapainya prinsip inklusivitas, karena konten dapat disesuaikan secara *real-time* dengan kebutuhan siswa yang memiliki latar belakang, gaya belajar, atau kemampuan berbeda. Hal ini sejalan dengan tujuan Pendidikan 2030 untuk menjamin akses pendidikan berkualitas bagi semua kalangan.

Namun demikian, potensi ini harus diimbangi dengan evaluasi kritis terhadap kualitas konten yang dihasilkan. Studi oleh Meechan (2024, p. 9) menunjukkan bahwa meskipun AI dapat menyusun materi ajar dengan cepat, konten tersebut tetap perlu ditinjau oleh pendidik untuk memastikan akurasi ilmiah, keberpihakan budaya, dan relevansi pedagogis. Dalam konteks ini, AI berperan sebagai asisten kreatif, bukan sebagai pengganti otoritas akademik guru.

Workshop yang dilaksanakan oleh LEMKOMINDO juga menegaskan efektivitas AI generatif dalam mendukung inovasi pengajaran, di mana peserta pelatihan berhasil mengembangkan bahan ajar digital berbasis AI dengan pendekatan konstruktivis dan kontekstual (Sibuea et al., 2024, p. 25-31).

Dengan adopsi yang bijak, AI generatif dalam pembuatan materi ajar otomatis dapat menjadi solusi atas beban kerja guru yang tinggi, meningkatkan efisiensi perencanaan pembelajaran, serta memperkuat fleksibilitas kurikulum secara signifikan.

Implementasi *Artificial Intelligence* dalam Pendidikan Agama Islam Sebagai Penilaian Esai Berbasis AI

AI generatif juga memiliki potensi besar dalam bidang penilaian esai, yang secara tradisional merupakan tugas yang kompleks dan memakan waktu bagi pendidik. Sistem AI dapat digunakan untuk mengevaluasi karya tulis siswa dengan mempertimbangkan aspek-aspek seperti struktur, koherensi argumen, tata bahasa, dan tingkat keberhasilan dalam menjawab pertanyaan esensial. Hal

ini dimungkinkan berkat pemrosesan bahasa alami dan kemampuan pemodelan semantik yang dimiliki oleh model besar seperti GPT.

Penilaian esai berbasis AI bekerja dengan mengidentifikasi elemen-elemen kunci dalam teks, seperti klaim utama, bukti pendukung, kesimpulan, serta logika argumen. Model yang telah dilatih dengan dataset esai berkualitas tinggi dapat memberikan skor berdasarkan rubrik penilaian standar dan memberikan umpan balik otomatis yang bersifat formatif. Dalam beberapa sistem, seperti yang dikembangkan oleh *OpenAI* dan *platform edtech* lainnya, umpan balik mencakup saran penguatan argumen, penyesuaian struktur kalimat, dan perbaikan tata bahasa (Misgna et al., 2024, p. 3–7).

Sebagai contoh praktis, dalam penugasan menulis esai dengan topik **“Dampak Perubahan Karakter Jujur Setelah Istiqamah Shalat”**, sistem AI dapat mengevaluasi:

- a. Apakah siswa menyampaikan tesis yang jelas dan konsisten?
- b. Apakah argumen yang diajukan didukung oleh data atau contoh konkret?
- c. Bagaimana organisasi paragraf dan kesinambungan antar ide?
- d. Apakah penggunaan bahasa efektif dan sesuai dengan gaya akademik?

Selain penilaian, AI juga dapat memfasilitasi penulisan ulang (*rewriting*) esai sebagai bagian dari proses revisi. Umpan balik dapat diberikan secara langsung di dalam teks, mendorong siswa untuk memperbaiki pekerjaan mereka secara mandiri, yang mendukung pembelajaran metakognitif dan berpikir reflektif (UNESCO, 2023, p. 25-26).

Namun, beberapa tantangan tetap ada. Pertama, validitas penilaian AI masih diperdebatkan, terutama dalam konteks esai yang bersifat subjektif atau kreatif. Kedua, risiko bias algoritmik dapat memengaruhi skor, terutama jika data pelatihan tidak merepresentasikan keberagaman budaya dan gaya bahasa siswa (Alan Turing Institute, 2024, p. 4). Oleh karena itu, peran guru sebagai validator tetap penting dalam memastikan keadilan dan akurasi evaluasi.

Lebih lanjut, laporan UNESCO (2023) merekomendasikan agar institusi pendidikan menerapkan penilaian berbasis AI secara bertahap dan selalu dikombinasikan dengan komponen evaluasi manusia. Sistem hybrid ini diyakini dapat menggabungkan kecepatan dan objektivitas AI dengan sensitivitas pedagogis manusia.

Dengan adopsi yang etis dan hati-hati, penilaian esai berbasis AI dapat meningkatkan efisiensi proses evaluasi, mengurangi beban kerja guru, dan memberikan pengalaman belajar yang lebih interaktif dan personal bagi siswa.

EDUPEDIA:

Implementasi *Artificial Intelligence* dalam Pendidikan Agama Islam Sebagai Sistem Tutor Pintar

Salah satu penerapan AI generatif yang semakin mendapat perhatian dalam ranah pendidikan adalah pengembangan sistem tutor pintar (*intelligent tutoring systems/ITS*). Tutor pintar berbasis AI adalah sistem pembelajaran adaptif yang dirancang untuk meniru peran seorang tutor manusia dengan memberikan bimbingan individual kepada peserta didik, baik dalam bentuk pertanyaan, penjelasan konsep, maupun umpan balik korektif secara real-time (VanLehn, 2011, p. 197-221; Meechan, 2024, p. 55-60).

Dengan memanfaatkan kemampuan *Natural Language Processing* (NLP) dan pembelajaran mesin, sistem ini mampu berinteraksi dalam bahasa alami, menganalisis pemahaman siswa, serta menyesuaikan materi ajar berdasarkan performa dan gaya belajar individual. Dalam pengembangan terbarunya, LLM seperti GPT-4 telah memungkinkan tutor pintar untuk menjalankan dialog edukatif yang kompleks dan responsif, yang sebelumnya hanya dapat dilakukan oleh tenaga pengajar manusia (Brown et al., 2020, p. 119). *Large Language Model* (LLM) merupakan sistem yang dikembangkan berdasarkan arsitektur jaringan saraf tiruan dan dilatih untuk memahami, mengolah, serta menghasilkan teks dalam bahasa alami, yaitu bahasa yang digunakan oleh manusia dalam komunikasi sehari-hari dengan berbagai macam ragam dan gaya (Dikdasmen, 2025, p. 107)

Contoh konkret dari implementasi tutor pintar berbasis AI dapat ditemukan dalam:

- a. *Chatbot edukatif*: Digunakan dalam pembelajaran daring untuk menjawab pertanyaan siswa terkait konsep-konsep spesifik, seperti chatbot untuk matematika yang menjelaskan langkah-langkah penyelesaian soal integral.
- b. *Asisten belajar personal*: Aplikasi seperti Khanmigo dari Khan Academy yang memanfaatkan GPT untuk mendampingi siswa dalam memahami topik STEM dan literasi, disertai dengan pertanyaan reflektif yang memicu berpikir kritis (UNESCO, 2023).
- c. *Tutor bahasa otomatis*: Untuk siswa yang mempelajari bahasa asing, tutor AI dapat membantu latihan percakapan, koreksi gramatikal, serta memperkaya kosakata.

Sistem tutor pintar tidak hanya efektif dalam mengisi kekosongan guru di ruang kelas dengan rasio tinggi siswa per guru, tetapi juga mendukung prinsip pembelajaran yang dipersonalisasi. Laporan Alan Turing Institute (2024) mencatat bahwa sistem tutor AI dapat meningkatkan motivasi belajar dan keterlibatan siswa, terutama dalam konteks pendidikan jarak jauh atau hybrid.

Namun, efektivitas sistem tutor pintar sangat bergantung pada kualitas data pelatihan dan model yang digunakan. Sistem yang dilatih dengan data yang tidak beragam berpotensi memberikan respons yang bias atau tidak inklusif. Selain itu, aspek etika seperti transparansi algoritmik,

perlindungan data pribadi siswa, serta batasan peran AI dalam menggantikan pendidik harus mendapat perhatian serius (UNESCO, 2023, p. 18-22).

Pendekatan terbaik dalam pemanfaatan tutor pintar adalah dengan menjadikannya sebagai pelengkap, bukan pengganti, dari interaksi guru-siswa. Guru tetap berperan dalam mengarahkan, menginterpretasikan hasil interaksi siswa dengan AI, serta menangani aspek afektif dan sosial dalam proses belajar mengajar.

Dengan desain yang tepat dan implementasi yang etis, tutor pintar berbasis AI berpotensi memperluas akses pembelajaran berkualitas dan mewujudkan pendidikan yang lebih inklusif, adaptif, dan berbasis data.

Implementasi *Artificial Intelligence* dalam Pendidikan Agama Islam Sebagai Simulasi Pembelajaran Interaktif

Simulasi pembelajaran interaktif merupakan salah satu aplikasi AI generatif yang menjanjikan dalam menghadirkan pengalaman belajar yang interaktif, kontekstual, dan mendalam. Dengan kemampuan menghasilkan narasi, visual, dialog, serta interaksi dinamis, AI generatif memungkinkan pengembangan lingkungan belajar virtual yang dapat disesuaikan secara real-time dengan kebutuhan peserta didik (UNESCO, 2023, p. 28-30; The Alan Turing Institute, 2024, p. 13).

AI generatif dapat menciptakan simulasi berbasis skenario yang kompleks, yang biasa diterapkan dalam pembelajaran berbasis masalah (*problem-based learning*), studi kasus, atau pembelajaran berbasis proyek. Dengan memanfaatkan *Large Language Model* (LLM) dan model multimodal, simulasi dapat mencakup interaksi naratif, visualisasi grafis, hingga umpan balik otomatis. Contoh implementasi meliputi:

- a. Simulasi peran (*role-play simulations*) dalam pendidikan PAI, di mana siswa dapat berinteraksi sebagai tokoh agama atau Ustadz dalam memodelkan suatu materi tertentu seperti Kehadiran Islam Mendamaikan Bumi Nusantara.
- b. Simulasi laboratorium virtual untuk bidang STEM, seperti eksperimen kimia atau simulasi dinamika fluida, yang memungkinkan eksplorasi tanpa risiko fisik dan biaya tinggi (Meechan, 2024).
- c. Simulasi etika medis atau hukum, di mana peserta didik diuji untuk mengambil keputusan berbasis data dan nilai, dengan AI yang berperan sebagai pasien, klien, atau sistem hukum.

Keunggulan utama simulasi berbasis AI generatif terletak pada skalabilitas, adaptivitas, dan interaktivitasnya. Simulasi dapat diatur untuk menyesuaikan tingkat kesulitan berdasarkan respons

EDUPEDIA:

siswa, menawarkan rute pembelajaran alternatif, serta memberikan narasi dan konteks yang memperkaya pemahaman konseptual. Hal ini mendukung pembelajaran aktif dan konstruktivis yang telah terbukti meningkatkan retensi dan transfer pengetahuan (Ali et al., 2024, p. 1018).

Lebih jauh, dengan teknologi multimodal, peserta didik dapat terlibat dalam lingkungan belajar yang melibatkan teks, suara, gambar, dan gerakan. Platform seperti *ChatClass* dan *AI Dungeon* telah mengeksplorasi aspek ini untuk tujuan pembelajaran bahasa dan pemecahan masalah, sementara institusi seperti MIT dan Stanford telah mengembangkan simulasi berbasis AI untuk pelatihan profesional dan teknis (Davoodifard & Eskin, 2024, p. 22; UNESCO, 2023, p. 28–30; Sinha et al., 2024).

Namun, efektivitas pedagogis simulasi AI sangat bergantung pada desain instruksional yang cermat. Tanpa skenario yang dirancang dengan baik, interaksi AI bisa menjadi terlalu generik atau tidak relevan secara akademik. Selain itu, isu-isu seperti potensi bias, akurasi konten, serta etika representasi harus dipertimbangkan dalam perancangan sistem simulasi (Alan, 2024, p. 7). Simulasi AI juga dapat dimanfaatkan untuk evaluasi berbasis performa. Dengan merekam dan menganalisis interaksi siswa selama simulasi, guru dapat menilai aspek keterampilan berpikir kritis, kolaborasi, dan pengambilan keputusan yang sulit diukur dengan tes konvensional.

Dengan desain yang pedagogis, partisipatif, dan etis, simulasi berbasis AI generatif dapat menjadi pilar utama dalam pembelajaran masa depan yang kontekstual, kreatif, dan terpersonalisasi.

Etika dan Risiko Penggunaan AI Generatif dalam Pembelajaran

Etika dalam penggunaan kecerdasan buatan menjadi aspek krusial dalam dunia pendidikan karena teknologi ini semakin berperan dalam proses pembelajaran, penilaian, dan pengambilan keputusan akademik. Tanpa landasan etis yang jelas, pemanfaatan AI berpotensi mengancam integritas akademik, mendorong ketergantungan berlebihan, serta melemahkan kemampuan berpikir kritis peserta didik (VanLehn, 2011, p. 197-221; Meehan, 2024, p. 55-60). Selain itu, sistem AI mengandalkan pengolahan data dalam skala besar, sehingga menimbulkan risiko pelanggaran privasi dan penyalahgunaan data siswa apabila tidak diatur secara bertanggung jawab.

Dari sisi keadilan, algoritma AI juga berpotensi mereproduksi bias yang dapat berdampak pada ketidakadilan dalam penilaian dan rekomendasi pembelajaran. Oleh karena itu, etika AI diperlukan untuk memastikan transparansi, akuntabilitas, dan keterlibatan manusia dalam setiap keputusan pendidikan berbasis teknologi. Lebih jauh, penerapan etika memastikan bahwa penggunaan AI tetap selaras dengan tujuan pendidikan yang humanistik, yaitu membentuk karakter, nilai moral, dan tanggung jawab sosial peserta didik (Mehrabi et al, 2021, p. 6–9; Alan, 2024, p. 31–34). Dengan demikian, etika bukanlah hambatan inovasi, melainkan fondasi utama agar

pemanfaatan AI benar-benar mendukung transformasi pendidikan yang adil, bermakna, dan berkelanjutan. Berikut ini adalah

Tabel 3. Prinsip Etika Penggunaan AI Generatif bagi Guru dan Siswa

Aspek Etika	Deskripsi	Implikasi bagi Guru	Implikasi bagi Siswa
Transparansi	Keterbukaan tentang cara kerja AI, sumber data, dan dasar pengambilan keputusan	Guru wajib menjelaskan penggunaan AI dalam pembelajaran dan penilaian serta membatasi <i>black-box decision making</i>	Siswa memahami batasan AI dan tidak bergantung penuh pada hasil AI
Keadilan (Fairness)	Pencegahan bias dan diskriminasi algoritmik	Guru memvalidasi hasil AI agar tidak merugikan kelompok tertentu	Siswa dari latar belakang beragam memperoleh perlakuan adil
Hak Cipta	Penghormatan terhadap kepemilikan karya intelektual	Guru memastikan materi AI tidak melanggar hak cipta	Siswa menghindari plagiarisme berbasis AI
Akuntabilitas	Tanggung jawab manusia atas keputusan berbasis AI	Guru tetap menjadi penentu akhir evaluasi	Siswa tidak menjadikan AI sebagai otoritas absolut

Penggunaan kecerdasan buatan (AI) dalam pembelajaran Pendidikan Agama Islam (PAI) memiliki potensi pedagogis yang besar, namun juga mengandung sejumlah risiko yang perlu dicermati secara kritis. Salah satu risiko utama adalah reduksi otoritas keilmuan dan spiritual, di mana peserta didik cenderung menerima jawaban AI secara instan tanpa proses tadabbur, kajian sumber primer (Al-Qur'an dan Hadis), maupun bimbingan guru. Hal ini dapat menggeser peran pendidik PAI sebagai murabbi dan teladan moral menjadi sekadar fasilitator teknis, di samping itu juga dapat memungkinkan kehilangan atau penyalahgunaan data pribadi seorang pendidik (Carlini et al., 2021, p. 269–272).

Risiko lainnya adalah distorsi pemahaman ajaran Islam. AI bekerja berdasarkan data yang bersifat umum dan probabilistik, sehingga jawaban yang dihasilkan berpotensi tidak mempertimbangkan konteks mazhab, khilafiyah, maupun kearifan lokal keislaman. Dalam konteks PAI, hal ini dapat menimbulkan penyederhanaan berlebihan, bahkan kekeliruan dalam memahami konsep akidah, fikih, dan akhlak. Selain itu, penggunaan AI yang tidak terkontrol dapat melemahkan internalisasi nilai dan pembentukan karakter. Pembelajaran PAI sejatinya tidak hanya menekankan aspek kognitif, tetapi juga afektif dan spiritual (Minhaji et al., 2025, p. 363-384). Ketergantungan pada AI berisiko menjadikan pembelajaran agama bersifat mekanis dan tekstual, tanpa penghayatan nilai-nilai keteladanan, keikhlasan, dan adab.

Dengan demikian, risiko AI dalam pembelajaran PAI menegaskan pentingnya pendekatan yang kritis, etis, dan moderat. AI seharusnya diposisikan sebagai alat bantu pedagogis, bukan

EDUPEDIA:

otoritas keagamaan, serta digunakan di bawah bimbingan guru untuk memastikan kesesuaian dengan nilai-nilai Islam dan tujuan pendidikan yang holistik.

Tabel 4. Risiko Penggunaan AI Generatif dalam Pendidikan

Jenis Risiko	Deskripsi Risiko	Dampak bagi Guru	Dampak bagi Siswa	Strategi Mitigasi
Hoaks & Misinformasi	AI menghasilkan konten salah namun meyakinkan (<i>hallucination</i>)	Materi ajar berpotensi keliru	Kesalahan pemahaman konsep	Literasi AI, verifikasi sumber
Bias Data	Ketidakadilan akibat data pelatihan tidak representatif	Penilaian dan rekomendasi tidak objektif	Diskriminasi akademik	Audit algoritma, kurasi data
Pelanggaran Privasi	Kebocoran atau penyalahgunaan data pribadi	Tanggung jawab hukum dan etis	Eksposur data sensitif	Anonimisasi, <i>privacy by design</i>

Selain etika dan resiko sebagaimana dijelaskan pada pembahasan sebelumnya, terdapat juga etika dan kontribusi AI Generatif dalam dunia pendidikan. dSintesis potensi dan Kontribusi AI Generatif dalam Pendidikan dapat digambarkan sebagaimana dalam tabel berikut ini:

Tabel 5. Sintesis Etika dan Kontribusi AI Generatif dalam Pendidikan

Nilai PAI	Relevansi dengan AI Generatif	Implikasi Praktis
Amanah	AI digunakan secara bertanggung jawab	Guru tidak menyerahkan otoritas moral kepada AI
Kejujuran	Pencegahan plagiarisme dan manipulasi	Siswa wajib mengungkapkan penggunaan AI
Keadilan (<i>‘Adl</i>)	Penilaian tanpa bias	Validasi hasil AI oleh guru
Maslahah	AI untuk kemaslahatan belajar	AI sebagai alat bantu, bukan pengganti manusia

SIMPULAN

Perkembangan AI generatif menawarkan potensi strategis dalam mendukung transformasi pembelajaran Pendidikan Agama Islam (PAI) melalui penyediaan materi adaptif, penilaian reflektif, tutor dialogis berbasis nilai, dan simulasi pembelajaran yang integratif. Namun, pemanfaatannya juga menghadirkan tantangan etis dan risiko serius, seperti bias algoritmik, misinformasi keagamaan, pelanggaran privasi, serta ancaman terhadap kejujuran akademik dan integritas moral peserta didik. Integrasi AI generatif dalam PAI harus berlandaskan nilai-nilai Islam, di antaranya amanah, keadilan, kemaslahatan, dan tanggung jawab moral dengan menegaskan peran sentral pendidik sebagai otoritas pedagogis dan pembimbing nilai, sementara AI diposisikan sebagai teknologi pendukung. Pendekatan yang etis dan kritis menjadikan AI generatif berpotensi

memperkuat kualitas dan relevansi PAI di era digital tanpa menggerus esensi spiritual dan moral pendidikan Islam.

Penelitian selanjutnya perlu berfokus pada uji empiris pemanfaatan AI generatif dalam pembelajaran PAI, khususnya dampaknya terhadap pembentukan akhlak, kejujuran akademik, dan pemahaman keagamaan peserta didik. Selain itu, kajian lanjutan penting untuk mengembangkan kerangka etika dan model pedagogis PAI berbasis AI yang menegaskan peran guru sebagai pembimbing moral dan spiritual dalam konteks pendidikan Islam di era digital.

REFERENSI

- AlAli, R., Wardat, Y., Al-Saud, K., & Alhayek, K. (2024). Generative AI in Education: Best Practices for Successful Implementation. *International Journal of Religion*, 5(9). <https://doi.org/10.61707/pkwb8402>.
- AI Dungeon (2019). *AI Dungeon* (interactive text adventure). Latitude
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency*. <https://proceedings.mlr.press/v81/binns18a.html>
- Brown, T. et al. (2020). Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*.
- Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Erlingsson, U. (2021). Extracting Training Data from Large Language Models. *USENIX Security Symposium*.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.
- Davoodifard, M., & Eskin, D. (2024). *Exploring the Role of Generative AI in Second Language Education: Insights for Instruction, Learning, and Assessment*. Studies in Applied Linguistics & TESOL, 24(1). <https://doi.org/10.52214/salt.v24i1.12863>
- Dicoding Indonesia (2025), *Belajar Dasar Artificial Intellegency*, Decoding Academy.
- Direktorat Jenderal Guru, Tenaga Kependidikan dan Pendidikan Guru Kementerian Pendidikan Dasar dan Menengah (2025) *Modul 4 Pemrograman Kecerdasan Artifisial*, (Bimbingan Teknis Guru Koding dan Kecerdasan Artifisial Jenjang SMA/SMK)
- Floridi, L., & Cows, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*.
- Floridi, L., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4). <https://doi.org/10.1007/s11023-018-9482-5>
- Gebri, T. et al. (2018). Datasheets for Datasets. *arXiv preprint arXiv:1803.09010*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27. https://papers.nips.cc/paper_files/paper/2014
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press

EDUPEDIA:

Jurnal Studi Pendidikan dan Pedagogi Islam
Vol. 10 Nomor 2

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553). <https://doi.org/10.1038/nature14539>
- Meechan, D. (2024). *Generative AI for Students: The Essential Guide to Using Artificial Intelligence for Study at University*. London: Sage.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6). <https://doi.org/10.1145/345760>
- Miles, M. B., & Huberman, A. M. (2014). *Qualitative Data Analysis*. Sage Publications.
- Minhaji, Kawakip, A. N., Nawafil, M., Asmuki, & Junaidi. (2025). Developing an Integrated E-Module on Religious Moderation to Address Rising Radicalism in Schools: Implications for Students' Understanding. *Jurnal Pendidikan Agama Islam*, 22(2), 363–384. <https://doi.org/10.14421/jpai.v22i2.11869>
- Mitchell, Tom M. (1997), *Machine Learning*. New York: McGraw-Hill
- OECD. (2019). *OECD Principles on Artificial Intelligence*. Paris: OECD Publishing.
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Harlow: Pearson Education Limited.
- Samuelson, P. (2021). Generative AI and Copyright Law. *Berkeley Technology Law Journal*.
- Sibuea, N., Suwianto, P., Rinanda, T., et al. (2024). Pemanfaatan Panduan Praktis Generative AI untuk Efektivitas Pembelajaran di Perguruan Tinggi. *Jurnal PKM – Journal Liaison Academia and Society (J-LAS)*, 4(4)
- Sinha, A., Goyal, S., Sy, Z., et al. (2024). *BoilerTAI: A Platform for Enhancing Instruction Using Generative AI in Educational Forums*. arXiv preprint arXiv:2409.13196
- Stanford HAI. (2024). *View of The Psychosocial Impacts of Generative AI Harms*. Stanford University.
- Sugiyono. (2020). *Metode Penelitian Kualitatif*. Alfabeta.
- Teaching AI Best Practices. (2024). *University of Florida AI Initiative*.
- The Alan Turing Institute. (2024). *Generative Artificial Intelligence in Education: Opportunities and Challenges*. London: Turing Institute.
- Undang-undang Nomor 27 tahun 2022 tentang *Perlindungan Data Pribadi*
- UNESCO. (2023). *Guidance for Generative AI in Education and Research*. Paris: United Nations Educational, Scientific and Cultural Organization
- University of Florida. (2024). *Teaching AI: Best Practices*. University of Florida AI Initiative.
- VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4). <https://doi.org/10.1080/00461520.2011.611369>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.