Pawiyatan: Jurnal Pendidikan Profesi Guru
Vol. 1, No. 2, Nov 2025, pp. 116~126
ISSN: 0000-0000 , DOI: 10.28918/pawiyatan.v1i2.paperID

Analisis Penalaran Moral Pada Platform *Digital Islamic Storytelling* Melalui Model *Natural Language Processing* (NLP)

Zakiya Very Ayu Suryatina [1], Ely Mufidah[2]
[1]*Sekolah Dasar Negeri 01 Mulyorejo, Kesesi, Kabupaten Pekalongan, Jawa Tengah, Indonesia*
[2,3]*Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, Jawa Tengah, Indonesia*

Article Info

Article history:

Submitted Nov 3, 2025
Accepted Nov 28, 2025
Published Dec 30, 2025

Kata Kunci:

Natural Language Processing, penalaran moral, pendidikan karakter Islami, digital storytelling, kecerdasan buatan

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi model Natural Language Processing (NLP) berbasis domain keislaman guna menganalisis penalaran moral (moral reasoning) siswa dalam konteks *digital Islamic storytelling*. Latar belakang penelitian ini berangkat dari kebutuhan akan pendekatan penilaian karakter yang objektif dan terukur, menggantikan metode kuesioner subjektif yang selama ini dominan digunakan. Penelitian menggunakan pendekatan kualitatif eksploratif dengan strategi *design-based research (DBR)* yang melibatkan 120 siswa madrasah sebagai partisipan. Data berupa 1.200 narasi digital dianotasi berdasarkan tiga tingkat moral reasoning menurut teori Kohlberg yang diadaptasi ke dalam nilai-nilai Islam. Model NLP dikembangkan melalui *fine-tuning* IndoBERT dan korpus teks keislaman, kemudian divalidasi menggunakan metrik *precision*, *recall*, dan *F1-score*. Desain penelitian ini bersifat deskriptif-interpretatif, yang bertujuan untuk menggambarkan bagaimana model NLP dapat dirancang, dilatih, dan divalidasi dalam konteks *digital Islamic storytelling*. Hasil menunjukkan bahwa model mampu mengklasifikasikan tingkat penalaran moral siswa dengan akurasi F1-score 0,83, serta menampilkan sensitivitas linguistik terhadap ekspresi moral-spiritual seperti kejujuran, keadilan, dan ihsan. Temuan ini menegaskan potensi integrasi AI dalam pembelajaran Pendidikan Agama Islam (PAI) untuk menilai dan memperkuat karakter secara lebih data-driven, reflektif, dan kontekstual. Penelitian ini sekaligus memperkenalkan paradigma baru AI for Moral Pedagogy, yaitu penggunaan kecerdasan buatan sebagai mitra pedagogis dalam pengembangan pendidikan karakter slami berbasis teknologi.

This is an open-access article under the CC BY-SA license.

Corresponding Author:

Zakiya Very Ayu Suryatini
SDN 01 Mulyorejo, Kesesi, Kabupaten Pekalongan, Jawa Tengah, Indonesia
Email: veryayu25@gmail.com

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*/ AI) telah membawa transformasi signifikan dalam dunia pendidikan, khususnya melalui penerapan Natural Language Processing (NLP) dalam menilai kemampuan kognitif peserta didik. Beragam penelitian terdahulu menunjukkan efektivitas model NLP dalam automated essay scoring, analisis sintaksis, maupun evaluasi pemahaman membaca dalam konteks pembelajaran bahasa

dan literasi umum. Namun, terlepas dari kemajuan tersebut, penerapan NLP dalam ranah penilaian afektif dan moral-spiritual masih relatif terbatas. Penilaian karakter dan nilai-nilai moral siswa umumnya masih mengandalkan instrumen self-report seperti angket atau wawancara, yang rawan bias sosial dan sulit menangkap kedalaman penalaran moral yang bersifat implisit. Kondisi ini menimbulkan kebutuhan mendesak untuk menghadirkan pendekatan baru yang mampu mengintegrasikan kecerdasan buatan dalam menilai dimensi afektif peserta didik secara lebih objektif, kontekstual, dan terukur.

Dalam konteks Pendidikan Agama Islam (PAI), tantangan ini menjadi semakin kompleks. PAI tidak hanya bertujuan membentuk pemahaman konseptual terhadap ajaran Islam, tetapi juga menginternalisasi nilai-nilai moral dan spiritual dalam perilaku sehari-hari. Berbagai inisiatif digitalisasi pembelajaran PAI telah dilakukan, seperti penggunaan media interaktif, aplikasi e-learning, hingga platform digital storytelling berbasis kisah Islami. Namun, sebagian besar penelitian di bidang ini masih berfokus pada aspek pedagogis dan teknologis semata, tanpa eksplorasi mendalam terhadap bagaimana teknologi AI dapat berperan dalam menilai perkembangan karakter dan penalaran moral siswa. Padahal, narasi dalam digital Islamic storytelling merupakan medium yang sangat potensial untuk mengekspresikan nilai, niat, dan penalaran moral yang tersembunyi dalam pilihan kata dan struktur naratif siswa.

Penelitian-penelitian terdahulu menunjukkan bahwa pendekatan story-based learning efektif dalam menstimulasi refleksi moral siswa. Akan tetapi, analisis terhadap produk naratif siswa umumnya masih dilakukan secara manual oleh guru atau peneliti menggunakan pendekatan kualitatif deskriptif. Pendekatan ini, meskipun kaya secara interpretatif, tidak efisien untuk data dalam skala besar dan bergantung pada subjektivitas penilai. Di sisi lain, model NLP yang banyak digunakan dalam konteks umum seperti BERT, GPT, atau RoBERTa—biasanya dilatih menggunakan korpus teks umum seperti Wikipedia atau berita berbahasa Inggris. Hal ini membuat model tersebut kurang sensitif terhadap nuansa linguistik, nilai moral, dan terminologi keislaman yang khas. Dengan demikian, muncul kebutuhan akan pengembangan model NLP yang bersifat domain-specific, yang mampu menangkap semantik dan konteks nilai Islam dalam teks siswa.

Penelitian ini hadir untuk menjembatani kesenjangan tersebut dengan mengembangkan dan memvalidasi model NLP yang dirancang khusus untuk menganalisis penalaran moral siswa dalam platform digital Islamic storytelling. Melalui pendekatan kualitatif, penelitian ini berfokus pada desain, pengujian, dan interpretasi model secara mendalam guna memahami bagaimana AI dapat mengidentifikasi tingkat penalaran moral (pra-konvensional, konvensional, pasca-konvensional) berdasarkan teori moral Kohlberg yang diadaptasi dalam konteks pendidikan Islam. Penelitian ini tidak hanya menghasilkan model yang terotomasi, tetapi juga menghadirkan kerangka metodologis yang mengintegrasikan tiga disiplin ilmu: Computer Science, Psikologi Moral, dan Pendidikan Islam.

Kebaruan penelitian ini terletak pada upaya membangun sistem penilaian karakter berbasis NLP yang mampu beradaptasi dengan konteks nilai dan bahasa keislaman. Berbeda dengan pendekatan konvensional yang menilai karakter melalui kuesioner, model ini menilai melalui ekspresi linguistik yang muncul secara alami dalam karya naratif siswa.

Dengan melatih model pada korpus teks keislaman seperti Al-Qur'an, Hadis, kisah Nabi, serta ribuan narasi siswa penelitian ini berpotensi menghasilkan model yang peka terhadap indikator moral reasoning Islami. Selain itu, hasil penelitian ini diharapkan dapat berkontribusi terhadap pengembangan enabling technology dalam bidang EdTech PAI, yang tidak hanya menghasilkan satu produk media, tetapi juga model dan metodologi yang dapat diterapkan secara luas pada berbagai platform pembelajaran karakter berbasis digital.

Dengan demikian, penelitian ini bukan sekadar penerapan teknologi baru, melainkan langkah konseptual untuk merekonstruksi paradigma penilaian karakter dalam pendidikan Islam di era digital. Pendekatan ini berupaya menggabungkan keunggulan analitik kecerdasan buatan dengan kedalaman nilai moral Islam, guna membangun sistem pembelajaran yang tidak hanya cerdas secara kognitif, tetapi juga berakhlak dan bermakna secara spiritual.

ISSN: 0000-0000

2. METODE

Penelitian ini menggunakan pendekatan kualitatif eksploratif dengan dukungan analisis komputasional berbasis *Natural Language Processing* (NLP). Pendekatan ini dipilih karena fokus penelitian bukan hanya pada pengembangan teknis model AI, tetapi juga pada pemahaman mendalam terhadap proses penalaran moral (moral reasoning) yang terekspresikan dalam teks naratif siswa. Dengan demikian, penelitian ini tidak sekadar membangun sistem otomatisasi, tetapi juga berupaya memahami makna, konteks, dan struktur nilai yang melandasi produksi teks dalam pembelajaran digital Islami.

Desain penelitian ini bersifat deskriptif-interpretatif, yang bertujuan untuk menggambarkan bagaimana model NLP dapat dirancang, dilatih, dan divalidasi dalam konteks *digital Islamic storytelling*. Selain itu, penelitian ini juga menggunakan strategi *design-based research* (DBR) yang memungkinkan iterasi antara teori, desain teknologi, dan implementasi pendidikan. Pendekatan DBR dipilih karena selaras dengan tujuan penelitian untuk menghasilkan *enabling technology* yang tidak hanya berfungsi secara teknis, tetapi juga relevan dengan konteks pedagogis dan nilai-nilai Islam.

Penelitian dilakukan pada platform digital Islamic storytelling yang digunakan oleh siswa madrasah tsanawiyah di beberapa sekolah mitra di Jawa Barat dan Jawa Tengah. Siswa diminta menulis cerita digital bertema nilai-nilai moral Islami, seperti kejujuran, tanggung jawab, tolong-menolong, dan kesabaran. Jumlah partisipan sebanyak 120 siswa, dengan rentang usia 13–15 tahun, dipilih menggunakan teknik purposive sampling.

Pemilihan konteks ini mempertimbangkan tiga alasan utama: pertama *Digital storytelling* memberikan ruang ekspresif bagi siswa untuk mengonstruksi makna moral secara alami melalui narasi. Kedua konteks pendidikan Islam kaya dengan kosakata religius dan nilai moral yang khas, sehingga relevan untuk pengembangan model NLP berbasis domain keislaman. Ketiga Madrasah sebagai lembaga pendidikan formal Islam menjadi representasi ideal untuk meneliti integrasi AI dalam penilaian karakter Islami.

Sumber data utama dalam penelitian ini adalah teks naratif digital yang dihasilkan oleh siswa melalui platform *digital Islamic storytelling*. Total terkumpul sekitar 1.200 narasi dengan panjang rata-rata 200–400 kata per teks. Selain itu, data tambahan diperoleh dari: Korpus teks keislaman, meliputi terjemahan Al-Qur'an, hadis, kisah Nabi, dan literatur akhlak Islami klasik seperti *Ihya' Ulumuddin* karya al-Ghazali; Wawancara mendalam dengan guru PAI dan ahli psikologi moral Islam untuk memvalidasi interpretasi nilai-nilai moral dalam teks siswa; Catatan reflektif guru dan *rubrik penilaian karakter* sebagai pembanding hasil analisis model.

Data diklasifikasikan menjadi dua kategori utama: (1) teks latih (*training corpus*) untuk model NLP, dan (2) teks uji (*testing corpus*) untuk validasi hasil. Seluruh data diproses dengan memperhatikan prinsip etika penelitian, termasuk izin tertulis dari sekolah dan anonimisasi identitas siswa.

Penelitian dilakukan dalam empat tahapan utama, yaitu: (1) pengumpulan dan anotasi data, (2) perancangan model NLP, (3) validasi hasil analisis, dan (4) interpretasi kualitatif.

Tahap pertama yaitu Pengumpulan dan Anotasi Data Setiap narasi siswa dikumpulkan dari platform digital dan dianalisis untuk mengidentifikasi indikator moral reasoning. Proses anotasi dilakukan oleh tiga expert annotators (guru PAI, dosen pendidikan Islam, dan ahli psikologi moral). Mereka mengkategorikan teks ke dalam tiga tingkat penalaran moral menurut teori Kohlberg yang diadaptasi dalam konteks Islam yaitu Pra-konvensional bahwa moralitas didasarkan pada kepatuhan terhadap hukuman atau imbalan. Kedua Konvensional, memandang moralitas didasarkan pada norma sosial dan ketaatan terhadap otoritas. Ketiga Pasca-konvensional, dimana moralitas didasarkan pada prinsip keadilan dan nilai ilahiah. Setiap teks diberi label kategori moral reasoning menggunakan kesepakatan antar-penilai (inter-rater agreement) dengan nilai reliabilitas Cohen's Kappa minimal 0,80 sebagai indikator konsistensi tinggi.

Tahap kedua yaitu perancangan dan pelatihan Model NLP. Model NLP dikembangkan menggunakan pendekatan transfer learning dengan *fine-tuning* terhadap model bahasa pra-latih (*pre-trained language model*) seperti IndoBERT dan GPT-Bahasa Indonesia. Dataset latih terdiri dari 80% korpus narasi siswa dan teks keislaman, sedangkan 20% sisanya digunakan

sebagai data validasi. Proses pelatihan melibatkan langkah-langkah berikut yaitu pertama *Preprocessing*: normalisasi teks, penghapusan tanda baca, stemming, dan konversi kata Arab-Latin. Kedua *Tokenization*: pembagian teks menjadi unit linguistik (*tokens*) yang dapat diproses model. Ketiga *Embedding generation*: representasi semantik menggunakan vektor berdimensi tinggi. Keempat *Fine-tuning*: pelatihan model untuk mengenali pola linguistik yang terkait dengan tingkat moral reasoning. Meskipun analisis dilakukan secara komputasional, proses interpretasi model tetap dilakukan secara kualitatif, yaitu dengan menelusuri potongan teks yang paling berkontribusi terhadap klasifikasi moral tertentu (*explainable AI approach*).

Tahap akhir adalah Analisis kualitatif dilakukan terhadap hasil keluaran model untuk memahami pola linguistik yang merefleksikan penalaran moral siswa. Misalnya, penggunaan kata "karena Allah", "berbuat baik meski tidak dilihat orang", atau "adil meskipun sulit" diinterpretasikan sebagai indikator moral reasoning tingkat pasca-konvensional. Sementara itu, ekspresi seperti "supaya tidak dimarahi guru" atau "agar dapat hadiah" menunjukkan moral reasoning tingkat pra-konvensional. Proses interpretasi ini menggunakan teknik thematic content analysis, dengan fokus pada tema-tema moral dominan yang muncul dalam teks. Hasil analisis kemudian dihubungkan kembali dengan teori Kohlberg serta prinsip akhlak Islam (adab, ihsan, dan takwa).

Data dianalisis dengan pendekatan campuran kualitatif-komputasional. Secara umum, analisis melibatkan dua level: pertama Analisis Komputasional dengan cara menggunakan *NLP pipeline* untuk mengidentifikasi fitur linguistik (lexical, semantic, pragmatic) yang berhubungan dengan kategori moral reasoning. Kemudian Menggunakan *topic modeling* (LDA) untuk menemukan tema moral dominan dalam narasi siswa. Lalu menganalisis *attention weights* pada model transformer guna menelusuri bagian teks yang paling memengaruhi klasifikasi moral.

Kedua Analisis Kualitatif Interpretatif diantaranya menafsirkan hasil model dalam kerangka teori moral Islam dan psikologi moral. Kemudian melakukan *cross-case analysis* untuk membandingkan perbedaan moral reasoning antar kelompok siswa berdasarkan variabel seperti jenis kelamin, sekolah, dan tema cerita. Pendekatan ini bertujuan menghasilkan pemahaman yang kaya dan mendalam tentang bagaimana siswa mengonstruksi moral reasoning dalam konteks pembelajaran digital Islami.

Untuk menjaga keabsahan (validitas) hasil, penelitian menerapkan prinsip trustworthiness menurut Lincoln & Guba (1985), meliputi: Credibility: validasi hasil model oleh ahli dan triangulasi sumber data. Transferability: deskripsi kontekstual rinci agar temuan dapat diterapkan di konteks serupa. Dependability: dokumentasi proses pelatihan model dan pengkodean kualitatif. Confirmability: pelacakan keputusan analisis dan transparansi data melalui audit trail. Aspek etika dijaga dengan memastikan seluruh data siswa digunakan secara anonim, memperoleh izin dari sekolah dan orang tua, serta menghindari interpretasi yang dapat menstigmatisasi individu. Peneliti juga memastikan bahwa penggunaan AI tidak menggantikan peran guru sebagai pembimbing moral, melainkan sebagai alat bantu analisis yang memperkaya refleksi pedagogis.

3. HASIL DAN PEMBAHASAN

a. Natural Language Processing (NLP) dalam Pendidikan

Perkembangan *Natural Language Processing* (NLP) telah merevolusi cara sistem komputer memahami dan menganalisis bahasa manusia. Dalam konteks pendidikan, NLP banyak digunakan untuk menilai kemampuan kognitif siswa melalui berbagai aplikasi seperti *automated essay scoring* (Shermis & Burstein, 2013), deteksi plagiarisme, dan analisis pemahaman membaca (Zhang et al., 2021). Teknologi ini memungkinkan proses penilaian menjadi lebih efisien, objektif, dan konsisten dibandingkan dengan metode manual. Berbagai model *transformer-based*, seperti *Bidirectional Encoder Representations from Transformers* (BERT) dan *Generative Pretrained Transformer* (GPT), telah terbukti efektif dalam menginterpretasi konteks semantik dan makna implisit dalam teks siswa (Devlin et al., 2019).

ISSN: 0000-0000

Namun, sebagian besar penelitian NLP dalam pendidikan masih berfokus pada domain kognitif seperti kemampuan menulis, berpikir kritis, dan pemahaman bahasa—sementara dimensi afektif dan moral-spiritual belum banyak dieksplorasi. Menurut D'Mello dan Graesser (2015), penilaian afektif memerlukan pendekatan yang lebih kompleks karena melibatkan emosi, nilai, dan motivasi yang tidak selalu eksplisit dalam teks. Di sinilah muncul kebutuhan untuk mengembangkan model NLP yang mampu mengenali indikator moral dan spiritual secara implisit dari narasi siswa, bukan sekadar menganalisis struktur linguistik atau tata bahasa.

b. Penilaian Karakter dan Moral Reasoning dalam Pendidikan

Penilaian karakter merupakan aspek penting dalam pendidikan yang berorientasi pada pembentukan pribadi utuh, bukan hanya kecerdasan intelektual. Menurut Lickona (1991), pendidikan karakter bertujuan menumbuhkan kebajikan moral (moral virtues) seperti kejujuran, tanggung jawab, dan empati. Dalam konteks teori perkembangan moral, Kohlberg (1984) menjelaskan bahwa moral reasoning berkembang melalui tiga tingkat: pra-konvensional, konvensional, dan pasca-konvensional. Pendekatan ini telah banyak digunakan dalam penelitian psikologi moral dan pendidikan, namun aplikasinya dalam konteks digital dan berbasis AI masih sangat terbatas.

Sebagian besar instrumen penilaian moral yang ada menggunakan format kuesioner seperti *Defining Issues Test* (Rest, 1979), yang mengandalkan refleksi eksplisit peserta terhadap dilema moral. Meskipun efektif, pendekatan ini memiliki keterbatasan dalam menangkap dinamika moral reasoning yang muncul secara alami dalam proses belajar, khususnya dalam konteks naratif. Oleh karena itu, beberapa peneliti seperti Walker & Hennig (2004) menyarankan pendekatan analisis wacana (discourse analysis) untuk memahami bagaimana individu mengonstruksi moral reasoning dalam bahasa sehari-hari. Pendekatan berbasis narasi memungkinkan peneliti menilai bukan hanya apa yang diyakini siswa, tetapi juga bagaimana mereka menalar dan memaknai nilai moral dalam konteks tertentu.

Dalam konteks pendidikan Islam, moral reasoning tidak hanya berlandaskan pada prinsip etika universal, tetapi juga pada nilai-nilai tauhid, adab, dan akhlak karimah yang bersumber dari Al-Qur'an dan Hadis. Menurut Azra (2012), pendidikan Islam harus menekankan integrasi antara pengetahuan (*knowledge*), iman (*faith*), dan amal (*practice*), sehingga penilaian karakter tidak dapat dilepaskan dari dimensi spiritual dan teologis. Oleh karena itu, dibutuhkan model penilaian karakter yang mampu menafsirkan moral reasoning siswa dalam kerangka nilai Islam secara autentik.

c. Digital Storytelling dalam Pembelajaran Pendidikan Agama Islam

Digital storytelling telah menjadi pendekatan pedagogis inovatif yang menggabungkan literasi digital, ekspresi kreatif, dan pembelajaran nilai. Robin (2008) mendefinisikan *digital storytelling* sebagai praktik menceritakan kisah pribadi atau tematik menggunakan media digital seperti teks, gambar, dan suara. Dalam konteks pendidikan, pendekatan ini membantu siswa mengekspresikan pengalaman, refleksi, dan nilai moral melalui narasi yang bermakna.

Dalam pembelajaran PAI, kisah Islami (qashash Islami) memiliki peran penting sebagai media internalisasi nilai dan pembentukan karakter. Kisah para Nabi, sahabat, dan tokoh-tokoh Islam sarat dengan dilema moral dan keteladanan yang dapat dijadikan bahan refleksi. Menurut Nasution (2020), *Islamic digital storytelling* berpotensi menumbuhkan kesadaran moral melalui proses kognitif dan emosional, di mana siswa tidak hanya memahami cerita tetapi juga menafsirkan nilai di baliknya. Namun, penelitian terkait umumnya masih berhenti pada evaluasi persepsi siswa terhadap media atau peningkatan motivasi belajar, bukan pada analisis mendalam terhadap kualitas moral reasoning yang tercermin dalam narasi yang dihasilkan.

Penelitian oleh Nurhasanah dan Nurdin (2021) menunjukkan bahwa siswa mampu mengekspresikan nilai-nilai Islam melalui karya naratif digital, tetapi penilaian terhadap aspek moral reasoning masih dilakukan secara manual oleh guru melalui rubrik kualitatif. Proses ini

menuntut waktu dan interpretasi subjektif yang tinggi. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi yang dapat membantu guru melakukan analisis secara lebih cepat, objektif, dan konsisten, tanpa kehilangan konteks nilai Islam yang terkandung dalam teks.

d. Penerapan NLP dalam Analisis Moral dan Nilai

Penerapan NLP untuk analisis moral dan nilai merupakan bidang yang relatif baru dan terus berkembang. Beberapa penelitian internasional telah mengembangkan model NLP untuk mendeteksi *moral foundations* dalam teks publik, seperti yang dilakukan oleh Hoover et al. (2020) menggunakan *Moral Foundations Dictionary* berbasis teori Haidt (2001). Model ini mampu mengenali pola linguistik yang merepresentasikan nilai-nilai moral seperti keadilan, kesetiaan, dan kesucian. Meski demikian, sebagian besar model tersebut dikembangkan berdasarkan konteks Barat dan korpus bahasa Inggris yang sekuler, sehingga tidak secara langsung relevan dengan konteks pendidikan Islam.

Dalam beberapa tahun terakhir, penelitian NLP di bidang etika dan agama mulai muncul. Misalnya, penelitian oleh Rizvi et al. (2022) mengkaji penerapan *text classification* untuk mendeteksi nilai-nilai etis dalam teks keagamaan, sedangkan Ahmad et al. (2023) menyoroti pentingnya *domain adaptation* pada model bahasa agar lebih sensitif terhadap konteks teologis dan kultural. Hal ini menunjukkan peluang besar untuk mengembangkan model NLP yang tidak hanya memahami struktur bahasa, tetapi juga semantik nilai spiritual dan moral yang terkandung di dalamnya.

e. Integrasi AI dan Pendidikan Islam

Integrasi kecerdasan buatan dalam pendidikan Islam menghadirkan tantangan epistemologis sekaligus peluang transformasional. Di satu sisi, AI dapat mempercepat proses pembelajaran, personalisasi konten, dan penilaian berbasis data. Namun di sisi lain, terdapat kekhawatiran akan reduksi makna spiritual ketika dimensi moral diukur oleh algoritma yang bersifat mekanistik. Menurut Al-Khalifa (2021), penerapan AI dalam konteks Islam harus memperhatikan prinsip *ethics by design*, yaitu memastikan bahwa algoritma mencerminkan nilai-nilai Islam dalam proses dan hasilnya.

Dalam konteks penilaian karakter, tantangan utamanya adalah bagaimana mentransformasikan indikator moral yang bersifat kualitatif ke dalam representasi linguistik yang dapat dipahami oleh model AI. Untuk itu, diperlukan kerangka konseptual yang menghubungkan teori psikologi moral (seperti Kohlberg) dengan epistemologi Islam. Misalnya, tingkat moral reasoning dapat dipadankan dengan tingkat kesadaran iman dan akhlak sebagaimana dijelaskan oleh al-Ghazali (Ihya' Ulumuddin), yang memandang moralitas sebagai hasil penyatuan antara akal, hati, dan iman. Dengan pendekatan ini, sistem AI tidak sekadar menilai benar atau salah secara normatif, tetapi menafsirkan kedalaman makna moral yang diungkapkan dalam teks.

f. Kesenjangan dan Kontribusi Penelitian

Berdasarkan telaah literatur di atas, dapat disimpulkan bahwa terdapat lima kesenjangan utama (research gaps) yang mendasari penelitian ini. Pertama, penelitian AI dalam pendidikan masih didominasi oleh fokus kognitif, dengan penilaian afektif dan moral-spiritual yang belum terotomasi. Kedua, penerapan NLP dalam konteks pendidikan Islam masih sangat terbatas, padahal konteks ini memiliki kekayaan semantik dan nilai yang unik. Ketiga, metode penilaian karakter dalam pendidikan agama masih mengandalkan analisis manual yang tidak efisien. Keempat, model NLP yang ada belum diadaptasi secara spesifik terhadap bahasa dan kosakata keislaman. Kelima, penelitian di bidang EdTech PAI cenderung berorientasi produk, bukan pengembangan metodologi penilaian karakter berbasis AI yang sistematis.

ISSN: 0000-0000

Penelitian ini menawarkan kebaruan (*novelty*) dengan mengembangkan model NLP domain-spesifik yang dirancang untuk menganalisis moral reasoning siswa berdasarkan teks naratif Islami. Model ini dilatih menggunakan korpus teks keislaman dan ribuan narasi siswa, sehingga memiliki kemampuan untuk menangkap nuansa moral reasoning yang khas dalam konteks Islam. Selain itu, penelitian ini mengintegrasikan kerangka teori dari tiga disiplin—*Computer Science*, Psikologi Moral, dan Pendidikan Islam untuk menghasilkan pendekatan penilaian karakter yang komprehensif, valid, dan aplikatif.

Penelitian ini dibangun atas dasar bahwa moral reasoning dapat direpresentasikan secara linguistik dan dianalisis menggunakan pendekatan *computational linguistics*. Dengan memanfaatkan *fine-tuning* model bahasa terhadap korpus Islami, sistem AI dapat mengenali indikator moral reasoning dalam teks siswa. Indikator tersebut kemudian dipetakan ke dalam kategori tingkat penalaran moral berdasarkan teori Kohlberg yang diadaptasi dengan nilai Islam, seperti:

Pra-konvensional: orientasi pada konsekuensi pribadi,
Konvensional: orientasi pada kepatuhan dan norma sosial,
Paska-konvensional: orientasi pada prinsip universal dan nilai ilahiah.

Kerangka ini memungkinkan model AI bertindak sebagai "asisten moral evaluator" yang membantu guru memahami perkembangan karakter siswa secara objektif dan berbasis data, tanpa menghilangkan peran reflektif dan bimbingan spiritual dari pendidik.

g. Klasifikasi Model NLP terhadap Moral Reasoning

Dari total 1.200 narasi digital yang dikumpulkan melalui platform *Digital Islamic Storytelling*, diperoleh variasi tema yang cukup luas, meliputi kejujuran (27%), tanggung jawab (21%), tolong-menolong (18%), keadilan (16%), dan kesabaran (18%). Panjang narasi berkisar antara 150 hingga 500 kata dengan gaya bahasa yang sederhana dan khas remaja madrasah. Analisis awal menunjukkan bahwa sebagian besar narasi memiliki struktur tiga bagian utama: pengenalan konteks, konflik moral, dan penyelesaian berbasis nilai religius.

Menariknya, sekitar 78% siswa menggunakan ungkapan-ungkapan religius seperti "karena Allah", "semoga mendapat pahala", atau "agar tidak berdosa" dalam argumentasi moral mereka. Ini menandakan bahwa dimensi teologis dan spiritual menjadi dasar penalaran moral yang signifikan dalam konteks pendidikan Islam. Akan tetapi, tingkat kedalaman refleksi moral masih bervariasi: sebagian siswa menunjukkan moral reasoning berbasis kepatuhan eksternal (*obedience-based*), sementara sebagian lainnya mampu menampilkan penalaran yang lebih reflektif dan prinsipil (*principle-based*).

Model NLP yang dikembangkan berhasil mencapai performa komputasional yang cukup baik. Berdasarkan data uji, model menghasilkan skor precision 0,84, recall 0,82, dan F1-score 0,83, yang menunjukkan bahwa model cukup mampu membedakan teks berdasarkan tingkat moral reasoning. Adapun distribusi klasifikasi moral reasoning dari 1.200 teks adalah sebagai berikut:

Kategori Moral Reasoning	Frekuensi	Persentase (%)
Pra-konvensional	312	26.0
Konvensional	573	47.8
Pasca-konvensional	315	26.2

Distribusi tersebut menunjukkan bahwa sebagian besar siswa masih berada pada tingkat konvensional, yakni menilai moralitas berdasarkan kepatuhan terhadap norma sosial dan religius. Ini konsisten dengan teori perkembangan moral Kohlberg (1984) yang menyatakan bahwa usia remaja umumnya masih berorientasi pada penerimaan sosial dan otoritas eksternal.

Model juga berhasil mengidentifikasi beberapa *lexical patterns* yang menjadi ciri khas tiap tingkat penalaran moral. Pra-konvensional ditandai oleh frasa seperti "supaya tidak dimarahi guru", "takut dosa", atau "agar tidak dihukum". Konvensional menampilkan

ungkapan seperti "karena Allah menyukai orang jujur", "sebagai murid yang baik", atau "demi menjaga nama baik". Pasca-konvensional ditandai oleh refleksi yang lebih abstrak, misalnya "keadilan harus ditegakkan meskipun sulit", "kebaikan sejati bukan karena imbalan", dan "manusia diuji agar menjadi lebih sabar dan ikhlas".

Temuan ini menunjukkan bahwa model tidak hanya mengidentifikasi pola linguistik, tetapi juga menangkap struktur semantik dan nilai-nilai yang melandasi penalaran moral Islami. Analisis lebih lanjut menggunakan *attention visualization* pada model transformer menunjukkan bahwa bagian teks yang paling memengaruhi klasifikasi adalah klausa alasan (causal clauses) dan pernyataan nilai (value statements). Sebagai contoh: Narasi pra-konvensional: "*Aku tidak mau menyontek karena nanti dapat dosadanbisadihukum."*Model fokus pada kata "karena", "dosa", dan "dihukum" sebagai indikator orientasi pada konsekuensi. Lalu Narasi konvensional: *"Aku jujur karena ingin menjadi murid yang baik seperti yang diajarkan ustaz."* Model menyoroti frasa "ingin menjadi murid yang baik" dan "diajarkan ustaz" sebagai indikator norma sosial-religius. Kemudian Narasi pasca-konvensional: *"Meskipun teman-temanku menertawakan, aku tetap membantu karena yakin Allah melihat keikhlasan."* Model mengidentifikasi frasa "meskipun", "tetap membantu", dan "Allah melihat keikhlasan" sebagai indikator prinsip internal dan keimanan reflektif.

Hasil ini memperlihatkan bahwa struktur linguistik dalam teks siswa berfungsi sebagai jendela untuk memahami cara berpikir moral mereka. Secara teoretis, temuan ini mendukung pandangan Walker & Hennig (2004) bahwa analisis wacana dapat digunakan untuk mengeksplorasi dimensi penalaran moral secara lebih autentik dibandingkan tes moral konvensional.

Proses validasi hasil model oleh tiga ahli (guru PAI, psikolog pendidikan Islam, dan dosen teknologi pendidikan) menunjukkan tingkat kesesuaian klasifikasi sebesar 87% dengan interpretasi manual. Ahli menilai bahwa model cukup akurat dalam mengenali indikator moral reasoning berbasis nilai Islam. Namun, ditemukan beberapa kasus ambiguitas, khususnya ketika siswa menggunakan bahasa metaforis atau idiomatik seperti "cahaya kebenaran" atau "jalan kesabaran". Dalam kasus tersebut, model cenderung mengklasifikasikan ke kategori pasca-konvensional, sementara ahli menilainya masih berada pada tingkat konvensional.

Triangulasi data dari hasil model, wawancara guru, dan penilaian rubrik menunjukkan bahwa AI dapat berfungsi sebagai *co-assessor* yang membantu guru menilai karakter siswa secara cepat dan konsisten. Guru menyatakan bahwa sistem ini membantu mereka menemukan pola umum perkembangan moral siswa, misalnya kecenderungan sebagian siswa untuk memahami moralitas sebagai kepatuhan ritual semata, bukan refleksi etis yang mendalam. Dengan demikian, AI berperan bukan menggantikan guru, melainkan memperkuat basis data untuk refleksi pedagogis.

Ketika hasil klasifikasi model diinterpretasikan dengan pendekatan Islam, ditemukan korelasi menarik antara tingkat moral reasoning dan konsep akhlak. Pada tingkat pra-konvensional, perilaku moral lebih banyak dipengaruhi oleh rasa takut terhadap dosa atau hukuman sejajar dengan tahap *taubat* awal dalam tasawuf yang berorientasi pada *al-khauf* (rasa takut kepada Allah). Pada tingkat konvensional, moral reasoning siswa berpusat pada kepatuhan terhadap norma sosial dan agama, mirip dengan konsep *ta'dib* (pembiasaan adab). Sedangkan tingkat pasca-konvensional mencerminkan internalisasi nilai spiritual tertinggi, yaitu *ihsan* berbuat baik karena kesadaran akan pengawasan Allah (*muraqabah*).

Korelasi ini memperkuat argumentasi bahwa teori moral Barat (Kohlberg) dapat diadaptasi ke dalam kerangka Islam dengan menambahkan dimensi teologis. Siswa yang mampu mengekspresikan prinsip moral berdasarkan keimanan, bukan sekadar ketaatan sosial, menunjukkan *moral maturity* yang lebih tinggi. Dengan demikian, model AI ini tidak hanya mendeteksi struktur linguistik, tetapi juga berpotensi mengukur kedalaman spiritual reasoning yang terefleksi dalam teks Islami.

Melalui *topic modeling* (LDA) terhadap 1.200 narasi, ditemukan lima tema utama yang sering muncul: Keadilan dan Kejujuran (24%), Kesabaran dan Keteguhan Hati (20%), Tolong-menolong dan Empati Sosial (19%), Ketaatan terhadap Orang Tua dan Guru (18%), Pengendalian Diri dan Istiqamah (19%)

ISSN: 0000-0000

Tema keadilan dan kejujuran menjadi yang paling dominan. Siswa banyak menulis kisah tentang kejujuran dalam ujian, berbagi makanan dengan teman, atau menolak perilaku curang. Tema ini menunjukkan bahwa nilai-nilai moral yang diajarkan di madrasah berhasil diinternalisasi dalam konteks naratif. Namun, kedalaman argumentasi moral masih bervariasi: sebagian siswa menulis sekadar perbuatan benar-salah tanpa refleksi, sementara yang lain menautkan tindakan pada keyakinan iman dan tanggung jawab spiritual.

Sebagai contoh, berikut perbandingan dua bentuk moral reasoning dalam tema yang sama yaitu pertama dalam narasi konvensional: *"Aku tidak mau menyontek karena guru bilang itu dosa dan membuat nilai tidak berkah."* Dan kedua narasi pasca-konvensional:*"Aku menolak menyontek karena ingin jujur kepada diriku sendiri dan* yakin Allah menilai usaha, bukan hasil." Perbedaan kedalaman refleksi ini menunjukkan bagaimana sistem NLP dapat membantu guru membedakan tingkat internalisasi nilai, bukan sekadar perilaku permukaan.

Meskipun hasil model tergolong baik, terdapat beberapa keterbatasan yang perlu dicatat. Pertama, model masih mengalami kesulitan dalam menafsirkan bahasa simbolik dan implikatur moral yang tidak eksplisit. Misalnya, kalimat seperti *"aku ingin menjadi lentera di tengah kegelapan"* sering diklasifikasikan secara salah karena konteks metaforisnya. Kedua, beberapa istilah religius yang multifungsi, seperti "ujian" atau "takdir", sulit dikontekstualisasi secara otomatis tanpa pemahaman pragmatik yang lebih dalam.

Selain itu, model kadang tidak mampu membedakan antara ekspresi normatif (menyatakan nilai) dan reflektif (menalar nilai). Sebagai contoh, kalimat *"kita harus sabar karena sabar itu baik"* dianggap moral reasoning tinggi oleh model, padahal lebih tepat dikategorikan sebagai peniruan normatif, bukan penalaran moral reflektif. Temuan ini menunjukkan perlunya integrasi antara model semantik dengan analisis pragmatik berbasis konteks wacana.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi Natural Language Processing (NLP) dengan kerangka pendidikan karakter Islam dapat menjadi terobosan metodologis dalam penilaian moral siswa secara lebih objektif, kontekstual, dan skalabel. Melalui analisis 1.200 narasi digital siswa, model AI yang dikembangkan mampu mengidentifikasi dan mengklasifikasikan tingkat moral reasoning ke dalam tiga kategori utama: pra-konvensional, konvensional, dan pasca-konvensional, dengan tingkat akurasi rata-rata F1-score 0,83. Hasil ini membuktikan bahwa model tidak hanya mengenali pola linguistik permukaan, tetapi juga menangkap struktur semantik dan refleksi moral yang tertanam dalam bahasa religius siswa.

Secara konseptual, penelitian ini memperluas paradigma *AI for Education* dengan menghadirkan pendekatan "AI for Moral Pedagogy", yaitu penggunaan kecerdasan buatan untuk memahami kedalaman nilai dan spiritualitas yang termanifestasi dalam bahasa. Pendekatan ini menegaskan bahwa teknologi, ketika dirancang dengan basis epistemologi Islam dan teori psikologi moral, dapat berfungsi bukan sekadar sebagai alat otomatisasi, tetapi juga sebagai asisten reflektif bagi guru dalam memahami perkembangan karakter siswa.

Implikasi teoretis dari penelitian ini terletak pada kontribusinya terhadap tiga ranah utama: (1) memperluas teori perkembangan moral Kohlberg dengan menambahkan dimensi *spiritual reasoning* berbasis iman dan akhlak; (2) memperkaya konsep pendidikan karakter Islam (ta'dib dan ihsan) melalui pendekatan berbasis data linguistik; dan (3) memperkenalkan kerangka integratif AI-Islamic Pedagogy, di mana nilai, bahasa, dan teknologi berpadu dalam satu ekosistem etis pembelajaran digital.

Secara praktis, model NLP ini dapat diimplementasikan sebagai *assessment assistant* dalam platform pembelajaran PAI untuk mendukung guru melakukan penilaian karakter berbasis narasi siswa. Sistem ini dapat mempercepat proses evaluasi, mengurangi bias subjektif, dan memberikan data diagnostik bagi pengembangan kurikulum yang lebih berorientasi nilai. Selain itu, hasil klasifikasi moral reasoning dapat menjadi dasar bagi guru untuk merancang kegiatan reflektif, diskusi etika, dan pembelajaran berbasis kasus yang lebih mendalam.

Namun demikian, penelitian ini juga menegaskan pentingnya menjaga keseimbangan antara kecerdasan buatan dan kecerdasan spiritual manusia. AI sebaiknya tidak menggantikan peran guru sebagai pembimbing moral, melainkan berfungsi sebagai mitra epistemik yang memperkaya pemahaman tentang perkembangan karakter siswa.

Ke depan, penelitian ini membuka peluang untuk mengembangkan model NLP multibahasa berbasis teks keislaman yang lebih luas, serta integrasi dengan analisis emosi dan narasi multimodal. Dengan demikian, pengembangan AI dalam pendidikan Islam tidak hanya berorientasi pada efisiensi teknologi, tetapi juga menjadi bagian dari gerakan besar humanisasi teknologi dalam bingkai nilai-nilai Ilahi.

FUNDING INFORMATION *(mandatory)*

"This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors."

DAFTAR PUSTAKA

[1] Al-Attas, S. M. N. (1993). Islam and Secularism. Kuala Lumpur: ISTAC.
[2] Al-Ghazali. (1998). Ihya' Ulumuddin (The Revival of the Religious Sciences). Beirut: Dar al-Fikr.
[3] Alimardani, M., & Hashemi, M. (2019). A review of text classification and its applications. Journal of Information Science and Engineering, 35(1), 1–20.
[4] Allen, C., & Wallach, W. (2009). Moral machines: Teaching robots right from wrong. AI & Society, 24(1), 15–26.
[5] Arifin, M. (2021). Analisis naratif dalam pengembangan karakter siswa melalui media digital Islami. Jurnal Pendidikan Karakter, 11(1), 32–47.
[6] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. International Conference on Learning Representations (ICLR), 1–15.
[7] Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.). Thousand Oaks, CA: Sage Publications.
[8] Darmawan, D. (2020). Pemanfaatan kecerdasan buatan dalam dunia pendidikan: Peluang dan tantangan. Jurnal Teknologi Pendidikan, 22(3), 187–198.
[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.
[10] Fadli, A., & Mustaqim, R. (2021). Pendekatan integratif AI dan PAI dalam pembelajaran abad 21. Jurnal Tarbiyah dan Ilmu Pendidikan Islam, 9(1), 75–89.
[11] Fancourt, D., & Finn, S. (2019). What is the evidence on the role of the arts in improving health and well-being? World Health Organization Regional Office for Europe.
[12] Fitriyah, S. (2020). Pembelajaran berbasis digital storytelling untuk pendidikan karakter Islami. Jurnal Edukasi Islam, 5(2), 101–118.
[13] Hamid, N., & Latifah, S. (2019). Moral reasoning dalam perspektif psikologi dan pendidikan Islam. Jurnal Ilmiah Pendidikan Agama Islam, 17(2), 201–214.
[14] Hidayat, A. (2021). Analisis nilai moral dalam karya sastra digital berbasis Islam. Jurnal Bahasa dan Sastra, 10(1), 51–63.
[15] Kaur, J., & Arora, P. (2021). Sentiment analysis using deep learning techniques: A review. Journal of King Saud University – Computer and Information Sciences, 33(4), 436–447.
[16] Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the EMNLP, 1746–1751.
[17] Kohlberg, L. (1981). The Philosophy of Moral Development: Moral Stages and the Idea of Justice. San Francisco: Harper & Row.
[18] Lickona, T. (1991). Educating for Character: How Our Schools Can Teach Respect and Responsibility. New York: Bantam Books.
[19] Loke, S. H., & Choudhury, T. (2022). Towards responsible AI in education: Ethics, equity, and explainability. Computers & Education: Artificial Intelligence, 3(1), 100–112.
[20] Maulana, R., & Zaini, A. (2023). Desain model penilaian karakter berbasis AI pada pembelajaran PAI. Jurnal Teknologi Pendidikan Islam, 4(1), 12–27.
[21] Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). Qualitative Data Analysis: A Methods Sourcebook (4th ed.). Thousand Oaks, CA: Sage Publications.
[22] Moustakas, C. (1994). Phenomenological Research Methods. Thousand Oaks, CA: Sage.

ISSN: 0000-0000

[23] Qardhawi, Y. (1999). Al-Khuluq al-Islami (Etika dalam Islam). Kairo: Maktabah Wahbah.
[24] Rai, R., & Singh, V. (2023). Assessing moral reasoning in digital narratives: A computational linguistics approach. Journal of Educational Computing Research, 61(2), 189–210.
[25] Rahmawati, N. (2020). Penggunaan Natural Language Processing dalam analisis teks keagamaan. Jurnal Informatika dan Komputasi, 9(2), 99–108.
[26] Saldaña, J. (2016). The Coding Manual for Qualitative Researchers (3rd ed.). London: Sage Publications.
[27] Walker, L. J., & Hennig, K. H. (2004). Moral Development and Moral Education. New York: Psychology Press.
[28] Wahyudi, T., & Anshori, A. (2022). Kecerdasan buatan dalam perspektif etika Islam. Jurnal Filsafat Islam, 6(2), 201–218.*
[29] Wilson, R. A., & Keil, F. C. (2020). The moral mind in AI systems: A review of psychological and computational perspectives. Frontiers in Psychology, 11, 3114