



Kompromi Etis dalam AI Generatif Memetakan Konflik Nilai Keadilan, Transparansi, dan Utilitas

Mohammad Alvi Pratama

Fakultas Hukum, Universitas Pasundan

Article Info	ABSTRACT
<p>Keywords: Etika Generative AI Konflik Nilai Kebijakan Hukum</p>	<p>Kemajuan dalam Kecerdasan Buatan Generatif (Generative AI) telah menimbulkan kekhawatiran etika yang signifikan, terutama dalam menyeimbangkan nilai-nilai utama seperti keadilan, transparansi, dan manfaat (utilitas). Semakin banyaknya implementasi sistem AI generatif dalam berbagai aplikasi, menjadi sangat penting untuk menangani dilema etika yang muncul dari <i>trade-off</i> yang inheren di antara prinsip-nilai tersebut. Sistem AI generatif sering kali menghadapi <i>trade-off</i> etika di mana kebutuhan peningkatan satu nilai, seperti keadilan, dapat menyebabkan kompromi pada nilai lain, seperti transparansi atau manfaat. Misalnya, usaha untuk membuat model AI lebih transparan dapat mengurangi efisiensi kinerjanya, sementara optimasi kemanfaatan menimbulkan masalah bias dan ketidakadilan. <i>Trade-off</i> ini bukan hanya persoalan teknis, tetapi juga merupakan dilema etika mendasar yang mempengaruhi kepercayaan, akuntabilitas, dan penerimaan masyarakat terhadap sistem AI. Oleh karena itu, tujuan utama dari penelitian ini adalah untuk mengembangkan kerangka kerja yang secara efektif menyeimbangkan <i>trade-off</i> etika dalam sistem AI generatif, dengan fokus khusus pada nilai-nilai tersebut. Penelitian ini menggunakan pendekatan multi-metode, yang menggabungkan analisis kritis dan penelitian kualitatif. Hasil penelitian menunjukkan bahwa <i>trade-off</i> etika dalam AI generatif sangat bergantung pada konteks dan sangat bervariasi di berbagai aplikasi. Sebagai contoh, dalam beberapa kasus, peningkatan transparansi dapat secara signifikan mengurangi kemanfaatan, sementara di kasus lain, upaya untuk memastikan keadilan dapat menyebabkan penurunan kinerja model. Hasil ini menekankan kebutuhan akan pendekatan yang fleksibel dan sensitif terhadap konteks dalam mengelola <i>trade-off</i> ini, daripada solusi satu ukuran untuk semua.</p>
<p>This is an open access article under the CC BY-NC license</p> 	<p>Corresponding Author: Mohammad Alvi Pratama Fakultas Hukum, Universitas Pasundan Jl. Lengkong Besar No.68, Cikawao, Kec. Lengkong, Kota Bandung, Jawa Barat 40261 alvi.pratama@unpas.ac.id</p>

INTRODUCTION

Kecerdasan Buatan Generatif (Generative Artificial Intelligence/AI) merupakan lompatan signifikan dalam kemampuan sistem AI untuk melakukan tugas-tugas yang memerlukan kreativitas, inovasi, serta kemampuan untuk menghasilkan konten baru (Epstein et al., 2023). Tidak seperti sistem AI tradisional yang mengikuti aturan atau pola yang telah

ditentukan dan dilatih sebelumnya, model AI generatif seperti Generative Adversarial Networks (GANs) dan model berbasis *transformer* (Ekman, 2021) seperti GPT-4o dirancang untuk menghasilkan output yang tidak dapat dibedakan dari konten yang dibuat oleh manusia (Bengesi et al., 2024). Kemampuan ini telah mendorong adopsi luas AI generatif di berbagai industri, mulai dari hiburan dan pendidikan hingga kesehatan dan keuangan. Namun, adopsi yang cepat ini juga menyoroti tantangan etika yang kritis, terutama dalam menyeimbangkan nilai-nilai yang bersaing seperti keadilan, transparansi, dan manfaat (utilitas) (Oniani et al., 2023; Solaiman et al., 2023)

Prinsip keadilan dalam AI merujuk pada kemampuan sistem untuk membuat keputusan yang tidak memihak (*impartiality*) satu kelompok di atas kelompok lain, terutama dalam hal ras, jenis kelamin, atau status sosial ekonomi serta keyakinan Agama (SARA) (Ferrara, 2023). Transparansi berkaitan dengan keterbukaan dan kemampuan sistem AI untuk dapat dijelaskan, memastikan bahwa para pemangku kepentingan dapat memahami bagaimana keputusan dibuat (Li et al., 2023). Kemanfaatan merujuk pada efektivitas dan efisiensi sistem AI dalam mencapai tujuan yang diinginkan (Wu et al., 2023). Dalam AI generatif, ketiga nilai ini sering kali berbenturan. Misalnya, peningkatan transparansi model AI dapat mengurangi manfaatnya dengan membuatnya kurang efisien, atau memastikan keadilan mungkin memerlukan kompromi yang mengurangi kinerja keseluruhan sistem. *Trade-off* ini bukan hanya tantangan teknis, tetapi juga berakar kuat pada pertimbangan etika yang memengaruhi kepercayaan, akuntabilitas, dan penerimaan masyarakat terhadap teknologi AI.

Trade-off antara keadilan, transparansi, dan manfaat dalam AI generatif sangat kompleks dan bergantung pada konteks (misalnya, dalam sektor keuangan, nilai kemanfaatan sering kali diutamakan atas keadilan untuk mencapai efisiensi yang lebih besar). Salah satu masalah yang paling mendesak adalah bias algoritmik, yang dapat muncul ketika sistem AI tanpa disadari mempertahankan atau bahkan memperkuat bias sosial yang ada. Jika model AI generatif dilatih dengan data yang bias, model ini mungkin menghasilkan output yang memperkuat bias tersebut, yang pada akhirnya menyebabkan perlakuan yang tidak adil terhadap kelompok tertentu (Ferrara, 2023). Memastikan keadilan dalam sistem semacam ini sering kali memerlukan perubahan pada data pelatihan atau modifikasi model, yang dapat mengurangi manfaatnya atau mempersulit interpretasinya. Hal ini menimbulkan dilema etika yang kritis: apakah kita harus memprioritaskan keadilan meskipun itu berarti mengorbankan efisiensi dan efektivitas sistem AI?

Model AI generatif sering kali disebut sebagai "kotak hitam" karena proses pengambilan keputusan mereka sulit diinterpretasikan (Ornes, 2023; Pfeiffer et al., 2023). Meskipun ada gerakan yang berkembang menuju Explainable AI (XAI) (Arrieta et al., 2020), yang bertujuan untuk membuat sistem AI lebih transparan, hal ini sering kali datang dengan mengorbankan nilai manfaat. Model yang lebih transparan mungkin lebih lambat, kurang efisien, atau kurang efektif, menciptakan ketegangan antara keinginan untuk transparansi dan kebutuhan akan sistem AI yang praktis dan efisien.

Persoalan nilai terakhir, nilai kemanfaatan dalam sistem AI sering kali dianggap sebagai tujuan utama bahkan sebagian menyebut satu-satunya, terutama dalam aplikasi komersial di mana kinerja dan efisiensi menjadi kunci keberhasilan. Namun, jika hanya fokus pada manfaat saja, hal ini dapat menyebabkan kompromi etika, seperti mengabaikan

keadilan atau mengorbankan transparansi (Dignum, 2021). Hal ini sangat mengkhawatirkan dalam aplikasi AI generatif yang berbasis risiko tinggi, seperti dalam diagnostik kesehatan, pengambilan keputusan keuangan, atau penalaran hukum. Oleh karena itu, menyeimbangkan nilai-nilai yang saling bersaing ini merupakan tantangan signifikan bagi pengembang AI, pembuat kebijakan, dan pemangku kepentingan.

Indonesia, seperti banyak negara lainnya, dengan cepat mengadopsi teknologi AI, termasuk AI generatif, sebagai bagian dari agenda transformasi digital yang lebih luas. Pemerintah Indonesia telah mengakui pentingnya AI dalam mendorong pertumbuhan ekonomi dan meningkatkan layanan publik (Iswara, 2020). Namun, implikasi etika dari adopsi AI juga telah menjadi fokus perhatian, terutama karena negara ini berusaha untuk menyelaraskan strategi AI-nya dengan standar global.

Salah satu inisiatif utama dalam hal ini adalah Strategi Nasional AI Indonesia (Stranas AI) (Strategi Nasional Kecerdasan Artifisial Indonesia, 2020) yang diluncurkan pada tahun 2020. Strategi ini menguraikan visi pemerintah untuk pengembangan AI, menekankan perlunya sistem AI yang etis, adil, dan transparan. Stranas AI mengidentifikasi lima area prioritas untuk penerapan AI: kesehatan, pendidikan, ketahanan pangan, mobilitas, dan layanan publik. Di masing-masing area ini, strategi menyoroti pentingnya pertimbangan etika, terutama dalam memastikan bahwa sistem AI tidak memperburuk ketidaksetaraan sosial atau mengorbankan hak-hak individu.

Stranas AI juga menangani masalah transparansi, menyadari bahwa esensi "kotak hitam" dari banyak sistem AI dapat merusak kepercayaan publik (Strategi Nasional Kecerdasan Artifisial Indonesia, 2020). Strategi ini menyerukan pengembangan sistem AI yang tidak hanya efektif tetapi juga dapat dijelaskan dan akuntabel. Hal ini sangat penting dalam konteks layanan publik, di mana AI semakin banyak digunakan untuk membuat keputusan yang memengaruhi kehidupan orang, seperti dalam alokasi manfaat/kesejahteraan sosial atau penyelesaian sengketa hukum.

Lebih lanjut, strategi ini menekankan peran keadilan dalam AI, terutama dalam memastikan bahwa sistem AI tidak mempertahankan bias yang ada atau menciptakan bentuk diskriminasi baru. Ini adalah kekhawatiran yang signifikan di negara yang sangat beragam suku, budaya dan agama seperti Indonesia, di mana isu-isu ras, agama, dan status sosial ekonomi sangat sensitif. Stranas AI merekomendasikan pengembangan sistem AI yang inklusif dan mencerminkan komitmen negara terhadap keadilan sosial dan kesetaraan.

Namun, masih ada tantangan besar dalam menerapkan prinsip-prinsip etika ini dalam praktiknya. Salah satu tantangan utama adalah kurangnya kerangka regulasi yang komprehensif yang secara khusus menangani etika AI. Meskipun ada undang-undang dan peraturan yang mencakup isu-isu seperti privasi data dan perlindungan konsumen, belum ada undang-undang khusus yang mengatur penggunaan etis AI, kecuali Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial (Surat Edaran Menteri Komunikasi Dan Informatika Republik Indonesia Nomor 9 Tahun 2023 Tentang Etika Kecerdasan Artifisial, 2023) yang menjadi acuan nilai etika pengembangan teknologi AI yang mencakup inklusivitas, keamanan, aksesibilitas, perlindungan data pribadi, serta pembangunan dan lingkungan berkelanjutan. Hal ini mendorong pemerintah untuk meregulasi pengembangan undang-undang etika AI yang khusus yang akan memberikan pedoman yang jelas bagi pengembang AI, pengguna, dan pembuat kebijakan.

Tantangan lain adalah kebutuhan akan kesadaran dan keterlibatan publik yang lebih besar tentang etika AI. Sementara pemerintah dan industri mendorong adopsi AI, pemahaman di kalangan masyarakat umum tentang implikasi etis dari teknologi ini masih terbatas. Hal ini sangat mengkhawatirkan mengingat potensi AI untuk memengaruhi kehidupan orang dalam cara yang signifikan. Ada kebutuhan untuk lebih banyak dialog publik dan pendidikan tentang etika AI, serta mekanisme yang memungkinkan masyarakat untuk berpartisipasi dalam bagaimana AI digunakan dan diatur.

Pendekatan Indonesia terhadap AI dibentuk oleh konteks hukum dan regulasi yang lebih luas, yang mencakup berbagai undang-undang dan kebijakan yang memiliki implikasi untuk etika AI. Salah satu dokumen hukum kunci dalam hal ini adalah Undang-Undang No. 11/2008 tentang Informasi dan Transaksi Elektronik (UU ITE), yang memberikan kerangka hukum untuk transaksi elektronik dan aktivitas siber di Indonesia. Meskipun UU ITE tidak secara khusus membahas AI, undang-undang ini mencakup ketentuan tentang perlindungan data, hak konsumen, dan pencegahan kejahatan siber, yang semuanya relevan dengan etika AI.

Selain UU ITE, Indonesia juga telah memperkenalkan peraturan yang menangani aspek-aspek spesifik AI, seperti privasi data. Peraturan Pemerintah No. 71/2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik (PP 71) menetapkan persyaratan untuk perlindungan data pribadi dalam sistem elektronik, yang sangat penting untuk memastikan penggunaan AI yang etis, terutama dalam hal transparansi dan akuntabilitas.

Namun, seperti yang telah disebutkan sebelumnya, ada pengakuan yang semakin meningkat bahwa undang-undang dan peraturan yang ada tidak cukup untuk menangani tantangan etika yang otentik yang ditimbulkan oleh AI. Sehingga, mendorong pemerintah untuk pengembangan undang-undang etika AI yang khusus, yang akan menyediakan kerangka kerja yang lebih komprehensif dan koheren untuk tata kelola AI di Indonesia.

Mengingat sifat kompleks dan kontekstual dari *trade-off* etika dalam AI generatif, penelitian ini bertujuan untuk mengidentifikasi dan pemetaan *trade-off* etika sebagai dasar kerangka kerja yang dapat menyeimbangkan nilai-nilai yang bersaing ini dengan cara yang secara teori dapat dipertanggungjawabkan dan dapat diterapkan dalam praktik.

METHODS

Metodologi penelitian ini berfokus pada identifikasi dan kategorisasi *trade-off* etika dalam sistem AI generatif, dengan penekanan khusus pada keadilan, transparansi, dan manfaat. Pendekatan ini dirancang untuk memberikan pemahaman yang komprehensif tentang bagaimana *trade-off* ini muncul dalam sistem AI dan implikasinya terhadap penerapan AI yang etis. Penelitian akan dilakukan melalui tinjauan literatur yang menyeluruh dan analisis mendalam terhadap sistem AI yang ada serta aplikasinya, terutama dalam konteks Indonesia.

Tahap pertama penelitian ini melibatkan tinjauan literatur yang komprehensif. Tinjauan ini akan mengeksplorasi literatur yang ada tentang etika AI, dengan fokus khusus pada *trade-off* etika yang melekat dalam sistem AI generatif. Tinjauan literatur ini akan mencakup area kunci seperti bias algoritmik, keadilan dalam AI, transparansi dan keterjelasan (*explainable AI*), serta keseimbangan antara manfaat dan pertimbangan etis dalam penerapan AI.

Tahap kedua penelitian ini melibatkan analisis sistem AI yang ada dan aplikasinya, terutama yang diterapkan di Indonesia. Analisis ini akan berfokus pada identifikasi *trade-off* etika yang muncul dalam penerapan AI di dunia nyata. Penelitian ini akan mengkaji berbagai studi kasus tentang sistem AI generatif di sektor-sektor seperti kesehatan, keuangan, dan layanan publik. Sektor-sektor ini sangat relevan di Indonesia, di mana AI digunakan untuk mengatasi masalah-masalah kritis seperti aksesibilitas layanan kesehatan, inklusi keuangan, dan transparansi pemerintah.

Analisis ini akan melibatkan penilaian kualitatif secara deskriptif-analitis dengan pendekatan filosofis-etis tentang bagaimana sistem-sistem AI ini menyeimbangkan nilai-nilai yang bersaing, yaitu keadilan, transparansi, dan manfaat.

RESULTS AND DISCUSSION

Identifikasi dan Kategorisasi *Trade-off* Etika

Dalam menganalisis sistem AI generatif, muncul pola *trade-off* etika yang berulang antara keadilan, transparansi, dan manfaat. *Trade-off* ini sangat terlihat dalam aplikasi-aplikasi berisiko tinggi seperti di bidang kesehatan, keuangan, dan layanan publik, di mana keputusan AI memiliki dampak signifikan pada kehidupan individu dan sosial.

Pertama, keadilan sering kali dikompromikan demi nilai kemanfaatan, terutama dalam sistem yang dirancang mengutamakan efisiensi. Sebagai contoh, dalam sistem penilaian kredit otomatis, model AI generatif sering kali mengandalkan dataset besar yang tanpa disadari mungkin mencerminkan bias sosial yang ada yang berujung diskriminatif (Klein, 2020; Verma, 2019). Meskipun model-model ini dioptimalkan untuk akurasi dan manfaat, mereka dapat memperkuat atau bahkan memperburuk ketidakadilan yang ada secara sistemik dan masif, yang merugikan kelompok-kelompok tertentu. *Trade-off* ini sangat terlihat dalam sektor keuangan Indonesia, di mana penilaian kredit berbasis AI semakin banyak digunakan untuk menentukan kelayakan pinjaman. Meskipun ada potensi untuk meningkatkan inklusi keuangan, kurangnya keadilan dalam sistem ini dapat memperburuk eksklusi bagi kelompok yang terpinggirkan. Otoritas Jasa Keuangan Indonesia telah merilis Panduan Kode Etik Kecerdasan Buatan (Artificial Intelligence/AI) yang Bertanggung Jawab dan Terpercaya di Industri Teknologi Finansial (Panduan Kode Etik Kecerdasan Buatan (Artificial Intelligence/AI) Yang Bertanggung Jawab Dan Terpercaya Di Industri Teknologi Finansial, 2023) yang digunakan sebagai mitigasi risiko dalam optimalisasi teknologi AI.

Kedua, transparansi adalah nilai penting lainnya yang sering berbenturan dengan manfaat. Banyak model AI generatif, terutama yang menggunakan teknik pembelajaran mendalam (deep learning), berfungsi sebagai "kotak hitam" di mana proses pengambilan keputusannya tidak transparan (Jha et al., 2023). Meskipun model-model ini unggul dalam tugas-tugas seperti generasi bahasa atau sintesis gambar, kurangnya transparansi mereka menimbulkan kekhawatiran etika yang signifikan. Hal ini sangat bermasalah di sektor kesehatan, di mana sistem AI digunakan untuk membantu diagnosis (Cadario et al., 2021; Procter et al., 2023). Di Indonesia, di mana sistem kesehatan sudah mengalami tekanan, penggunaan sistem AI yang tidak transparan dapat merusak kepercayaan di antara pasien dan penyedia layanan kesehatan, yang berpotensi menyebabkan perawatan yang kurang optimal.

Nilai kemanfaatan cenderung mendominasi dalam penerapan nilai dalam sistem AI, terutama dalam aplikasi komersial. Namun, mengutamakan nilai kemanfaatan sering kali menyebabkan kompromi dalam nilai keadilan maupun transparansi. Sebagai contoh, dalam sistem perekrutan otomatis, model AI generatif sering kali dioptimalkan untuk menyaring sejumlah besar pelamar secara efisien. Namun, fokus pada manfaat ini dapat menyebabkan hasil yang bias, di mana kelompok-kelompok tertentu secara sistematis dirugikan (Chen, 2023). Di Indonesia, di mana pasar kerja sangat kompetitif, penggunaan sistem semacam itu dapat memperburuk ketidaksetaraan dalam peluang kerja.

Analisis Perbandingan *Trade-off* Etika

Trade-off etika yang diidentifikasi dalam penelitian ini bukanlah hal yang otentik bagi Indonesia, tetapi merupakan bagian dari tantangan global yang lebih luas dalam etika AI. Namun, konteks sosial-budaya dan hukum yang spesifik di Indonesia memperkuat tantangan ini dengan cara sendiri berdasarkan kearifan lokal.

Di Indonesia, prioritas kemanfaatan dalam sistem AI sering kali menyebabkan kompromi yang signifikan dalam keadilan. Ini terlihat di sektor-sektor seperti keuangan dan kesehatan, di mana peningkatan efisiensi yang ditawarkan oleh sistem AI sering kali dicapai dengan mengorbankan kesetaraan. Tantangan ini bukan hanya masalah teknis tetapi juga etis, karena sistem ini dapat memperkuat ketidakadilan sosial yang ada secara terstruktur. Pemerintah Indonesia melalui Stranas AI (Strategi Nasional Kecerdasan Artifisial Indonesia, 2020) menekankan perlunya keadilan dalam penerapan AI, namun terdapat kesenjangan antara aspirasi kebijakan ini dan realitas implementasi AI, di mana manfaat sering kali mengungguli keadilan.

Di lain sisi, *trade-off* antara transparansi dan manfaat sangat bermasalah dalam konteks Indonesia, di mana kepercayaan publik terhadap institusi adalah isu kritis. Penggunaan sistem AI yang tidak transparan, terutama dalam layanan publik dan sektor hukum, dapat memperburuk defisit kepercayaan ini. Meskipun Stranas AI menganjurkan pengembangan sistem AI yang transparan dan akuntabel, kenyataannya adalah banyak model AI generatif yang saat ini digunakan bersifat opak dan sulit diinterpretasikan. Kurangnya transparansi ini tidak hanya merusak kepercayaan tetapi juga membatasi kemampuan pemangku kepentingan untuk mempertanggungjawabkan keputusan AI.

Implikasi Hukum dan Kebijakan

Kerangka hukum Indonesia, termasuk Stranas AI, memberikan dasar yang kuat untuk mempromosikan praktik AI yang etis. Namun, terdapat kesenjangan yang signifikan dalam implementasi dan penegakan kebijakan ini. Misalnya, meskipun Stranas AI menekankan pentingnya keadilan dan transparansi, hanya ada sedikit regulasi atau pedoman khusus yang mengoperasionalkan prinsip-prinsip ini dalam praktik. UU ITE dan PP 71 memberikan panduan umum tentang perlindungan data dan transaksi elektronik, tetapi tidak secara khusus menangani tantangan etika yang ditimbulkan oleh sistem AI.

Lebih jauh lagi, kurangnya undang-undang yang meregulasi AI yang bertanggungjawab dan etis yang khusus di Indonesia berarti tidak ada kerangka kerja yang komprehensif untuk mengatur pengembangan dan penerapan sistem AI. Hal ini mengarah pada situasi di mana pertimbangan etika sering kali menjadi prioritas sekunder dibandingkan

dengan prioritas teknis dan komersial. Ada kebutuhan mendesak untuk reformasi regulasi guna memastikan bahwa sistem AI di Indonesia dikembangkan dan diterapkan dengan cara yang menyeimbangkan nilai-nilai tersebut.

Temuan dari penelitian ini mengidentifikasi tantangan etika yang signifikan yang ditimbulkan oleh sistem AI generatif, terutama dalam konteks Indonesia. Meskipun ada pengkajian yang semakin meningkat akan pentingnya keadilan dan transparansi, nilai-nilai ini sering kali dikompromikan dalam praktiknya, terutama ketika berbenturan dengan upaya untuk mengejar kemanfaatan. Hal ini menimbulkan pertanyaan penting tentang peran AI dalam masyarakat dan tanggung jawab mereka yang mengembangkan dan menerapkan teknologi ini.

Salah satu tantangan utama yang telah diidentifikasi dalam penelitian ini adalah kurangnya kerangka kerja etika yang komprehensif untuk AI di Indonesia. Meskipun Stranas AI dan Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial (Surat Edaran Menteri Komunikasi Dan Informatika Republik Indonesia Nomor 9 Tahun 2023 Tentang Etika Kecerdasan Artifisial, 2023) menyediakan visi tinggi untuk pengembangan AI yang etis dan bertanggungjawab, ada kebutuhan akan pedoman dan regulasi yang lebih spesifik yang menangani tantangan kompleks yang ditimbulkan oleh sistem AI generatif. Ini termasuk kebutuhan akan kesadaran publik yang lebih besar dan keterlibatan dalam etika AI, serta pengembangan alat dan kerangka kerja yang dapat membantu pemangku kepentingan menavigasi *trade-off* yang kompleks antara keadilan, transparansi, dan manfaat.

Implikasi Etika yang Lebih Luas

Trade-off etika yang diidentifikasi dalam penelitian ini memiliki implikasi signifikan tidak hanya bagi Indonesia tetapi juga bagi wacana global yang lebih luas tentang etika AI. Seiring dengan semakin meluasnya sistem AI generatif, kebutuhan akan pemahaman yang lebih mendalam tentang *trade-off* ini menjadi semakin penting. Temuan dari penelitian ini berkontribusi pada pemahaman ini dengan memberikan analisis mendalam tentang bagaimana *trade-off* ini muncul dalam praktik dan tantangan yang terlibat dalam mengelolanya.

Salah satu implikasi utama dari penelitian ini adalah perlunya pendekatan yang lebih holistik terhadap etika AI, yang melampaui solusi teknis dan mencakup nilai-nilai sosial yang lebih luas yang inklusif. Ini termasuk kebutuhan akan keterlibatan publik yang lebih besar dalam etika AI, serta pengembangan alat dan kerangka kerja yang dapat membantu pemangku kepentingan menavigasi *trade-off* yang kompleks antara keadilan, transparansi, dan manfaat.

Dalam konteks Indonesia, penelitian ini memetakan persoalan guna menyelaraskan pengembangan AI dengan tujuan sosial dan ekonomi yang lebih luas. Ini termasuk memastikan bahwa sistem AI tidak hanya canggih secara teknologi tetapi juga inklusif secara sosial dan selaras dengan nilai-nilai dan kebutuhan masyarakat Indonesia yang berbasis kearifan lokal. Temuan dari penelitian ini memberikan dasar yang kuat untuk mengembangkan kebijakan dan strategi yang dapat membantu mencapai kesejajaran ini.

Temuan dari penelitian ini juga memiliki implikasi penting bagi standar etika AI global. Meskipun ada konsensus yang semakin berkembang tentang pentingnya keadilan,

transparansi, dan manfaat dalam sistem AI, masih ada perdebatan signifikan tentang bagaimana nilai-nilai ini harus diseimbangkan dalam praktiknya.

Maka dari itu, beberapa rekomendasi strategis dapat diberikan untuk mengelola *trade-off* etika dalam sistem AI generatif. Rekomendasi ini sangat relevan bagi pembuat kebijakan, pengembang AI, dan pemangku kepentingan lain yang terlibat dalam penerapan AI di Indonesia:

1. Ada kebutuhan mendesak untuk kerangka etika AI yang komprehensif di Indonesia yang memberikan pedoman yang jelas untuk menyeimbangkan keadilan, transparansi, dan manfaat dalam sistem AI. Kerangka ini harus dikembangkan dalam konsultasi dengan berbagai pemangku kepentingan, termasuk pengembang AI, pembuat kebijakan, organisasi masyarakat sipil, dan masyarakat umum sebagai *blueprint* bagi kemajuan AI yang bertanggungjawab.
2. Kesadaran dan keterlibatan publik dalam etika AI harus diperkuat untuk memastikan bahwa sistem AI dikembangkan dan diterapkan dengan cara yang mencerminkan nilai-nilai dan kebutuhan masyarakat Indonesia. Ini termasuk pengembangan program pendidikan publik tentang etika AI yang demokratis, serta pembentukan mekanisme untuk masukan dan *feedback* publik tentang kebijakan dan praktik AI yang akan diimplementasikan.
3. Transparansi dan akuntabilitas yang lebih besar dalam sistem AI harus dipromosikan melalui pengembangan teknik AI yang dapat dijelaskan (*Explainable AI*) dan pembentukan mekanisme pengawasan yang memastikan sistem AI diaudit dan ditinjau secara berkala.
4. Upaya harus dilakukan untuk memastikan bahwa sistem AI diterapkan dengan cara yang mempromosikan nilai keadilan dan menghindari memperparah ketidakadilan sosial yang ada. Ini termasuk mengembangkan sistem mitigasi bias, serta implementasi kebijakan yang memastikan sistem AI inklusif dan dapat diakses oleh semua anggota masyarakat agar tidak terciptanya ketidakadilan yang tersistematis dan terstruktur.
5. Pengembangan AI di Indonesia harus diselaraskan dengan tujuan sosial dan ekonomi negara yang lebih luas, sebagaimana diuraikan dalam Stranas AI dan dokumen kebijakan terkait lainnya. Ini termasuk memastikan bahwa sistem AI berkontribusi pada inklusi sosial, pembangunan ekonomi, dan kesejahteraan seluruh rakyat Indonesia.

CONCLUSION

Menyeimbangkan *trade-off* etika dalam sistem AI generatif, terutama dalam konteks keadilan, transparansi, dan manfaat adalah hal utama yang harus segera dilakukan. Hal itu karena *trade-off* ini merupakan hal kompleks dan sangat kontekstual, dengan implikasi signifikan bagi pengembang AI dan pembuat kebijakan serta dampak langsung kepada masyarakat luas. Dalam konteks Indonesia, penelitian ini menggarisbawahi kebutuhan mendesak akan kerangka etika AI yang komprehensif yang mencerminkan lanskap sosial, budaya, dan hukum yang spesifik di negara ini. Meskipun Stranas AI menyediakan fondasi yang kuat untuk pengembangan AI yang etis, masih ada kebutuhan untuk pedoman dan regulasi yang lebih spesifik yang menangani tantangan unik yang ditimbulkan oleh sistem AI

generatif. Penelitian ini juga menekankan pentingnya keterlibatan publik dalam etika AI, serta perlunya transparansi dan akuntabilitas yang lebih besar dalam sistem AI. Dengan mengatasi tantangan ini, Indonesia dapat memastikan bahwa sistem AI-nya tidak hanya canggih secara teknologi tetapi juga etis dan selaras dengan nilai-nilai dan kebutuhan masyarakatnya. Hal itu juga didukung dengan wawasan global dan terlibat dalam wacana global tentang etika AI dengan memberikan pemahaman dan analisis mendalam tentang *trade-off* etika dalam sistem AI generatif dan menawarkan rekomendasi strategis untuk mengelola *trade-off* ini dalam praktiknya. Seiring dengan terus berkembangnya AI generatif, sangat penting untuk memastikan bahwa pertimbangan etika tetap menjadi prioritas utama dalam pengembangan dan penerapan AI.

REFERENCE

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Benges, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access*.
- Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*, 5(12), 1636–1642.
- Dignum, V. (2021). The myth of complete AI-fairness. *Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15–18, 2021, Proceedings*, 3–8.
- Ekman, M. (2021). *Learning deep learning: Theory and practice of neural networks, computer vision, natural language processing, and transformers using TensorFlow*. Addison-Wesley Professional.
- Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M. R., Groh, M., Schroeder, H., Smith, A., Akten, M., & Fjeld, J. (2023). Art and the science of generative AI: A deeper dive. *ArXiv Preprint ArXiv:2306.04141*.
- Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 3.
- Strategi Nasional Kecerdasan Artifisial Indonesia (2020). <https://ai-innovation.id/strategi>
- Iswara, M. A. (2020). Indonesia sets sights on artificial intelligence in new national strategy. *The Jakarta Post*, 14.
- Jha, D., Rauniar, A., Srivastava, A., Hagos, D. H., Tomar, N. K., Sharma, V., Keles, E., Zhang, Z., Demir, U., & Topcu, A. (2023). Ensuring trustworthy medical artificial intelligence through ethical and philosophical principles. *ArXiv Preprint ArXiv:2304.11530*.
- Klein, A. (2020). *Reducing bias in AI-based financial services*. *Brookings Report*.
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1–46.
- Surat Edaran Menteri Komunikasi Dan Informatika Republik Indonesia Nomor 9 Tahun 2023 Tentang Etika Kecerdasan Artifisial (2023).

- Oniani, D., Hilsman, J., Peng, Y., Poropatich, R. K., Pamplin, J. C., Legault, G. L., & Wang, Y. (2023). Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *NPJ Digital Medicine*, 6(1), 225.
- Ornes, S. (2023). Peering inside the black box of AI. *Proceedings of the National Academy of Sciences*, 120(22), e2307432120.
- Panduan Kode Etik Kecerdasan Buatan (Artificial Intelligence/Ai) Yang Bertanggung Jawab Dan Terpercaya Di Industri Teknologi Finansial (2023).
- Pfeiffer, J., Gutschow, J., Haas, C., Möslein, F., Maspfuhl, O., Borgers, F., & Alpsancar, S. (2023). Algorithmic fairness in AI: an interdisciplinary view. *Business & Information Systems Engineering*, 65(2), 209–222.
- Procter, R., Tolmie, P., & Rouncefield, M. (2023). Holding AI to account: challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction*, 30(2), 1–34.
- Solaiman, I., Talat, Z., Agnew, W., Ahmad, L., Baker, D., Blodgett, S. L., Chen, C., Daumé III, H., Dodge, J., & Duan, I. (2023). Evaluating the social impact of generative ai systems in systems and society. *ArXiv Preprint ArXiv:2306.05949*.
- Verma, S. (2019). Weapons of math destruction: how big data increases inequality and threatens democracy. *Vikalpa*, 44(2), 97–98.
- Wu, C., Lib, Y.-F., & Bouvry, P. (2023). Survey of Trustworthy AI: A Meta Decision of AI. *ArXiv Preprint ArXiv:2306.00380*.