PUBLISHER
YAYASAN PENDIDIKAN DAN PENGEMBANGAN HARAPAN ANANDA

JURNAL
WAWASAN DAN AKSARA

Juwara | p-issn: 2797-2097 | e-issn: 2797-2119 | **Vol. 4 No. 1 (2024)**
https://doi.org/10.58740/juwara.v4i1.271

Etika dalam Pengembangan *Artificial Intelligence*: Tinjauan Pedoman dan Penerapannya

Raynaldi Nugraha Prasetya[1](✉),
Ari Kusdiyanto[2], Usman Radiana[3],
Luhur Wicaksono[4]
[1,2,3,4]Universitas Tanjungpura

[1]e-mail:
rnugraha.prasetya@gmail.com

ABSTRAK

Kemajuan dalam pengembangan kecerdasan buatan (AI) telah memunculkan berbagai diskusi terkait aspek etika teknologi ini. Banyak pedoman etika diterbitkan untuk memastikan AI dikembangkan dan digunakan dengan cara yang bertanggung jawab, dengan fokus pada privasi, keadilan, transparansi, dan keamanan. Teknologi AI yang semakin "disruptif" membuat aturan etika ini menjadi sangat penting. Penilitian ini meninjau dan membandingkan 22 pedoman etika AI. Peneliti menemukan bahwa meskipun banyak prinsip yang tumpang tindih, ada kekurangan di beberapa pedoman, terutama terkait keadilan sosial dan penerapan dalam praktik. Penilaian ini menunjukkan bahwa prinsip etika seringkali tidak sepenuhnya diterapkan di lapangan, meski pedoman-pedoman tersebut telah disusun dengan baik. Kurangnya implementasi yang tepat bisa menimbulkan masalah serius di masa depan, terutama karena AI sangat berpengaruh dalam berbagai aspek kehidupan seperti pekerjaan, pendidikan, dan kesehatan. Oleh karena itu, evaluasi mendalam diperlukan untuk memperbaiki pendekatan etika AI. Penulis menyarankan beberapa langkah perbaikan, termasuk peningkatan transparansi dan akuntabilitas dalam pengembangan AI, serta penerapan pedoman etika yang lebih konsisten. Dengan memperkuat prinsip-prinsip ini, diharapkan AI dapat dikembangkan dan digunakan dengan lebih etis, membawa manfaat maksimal bagi masyarakat.

KATA KUNCI

kecerdasan buatan; etika; pedoman

ABSTRACT

Progress in the development of artificial intelligence (AI) has given rise to various discussions regarding the ethical aspects of this technology. Many ethical guidelines are published to ensure AI is developed and used in a responsible manner, with a focus on privacy, fairness, transparency, and security. AI technology is increasingly "disruptive" making these ethical rules very important. This research reviews and compares 22 AI ethical guidelines. Researchers found that while many of the principles overlap, there are gaps in some of the guidelines, particularly regarding social justice and application in practice. This assessment shows that ethical principles are often not fully implemented in the field, even though the guidelines are well developed. Lack of proper implementation could cause serious problems in the future, especially because AI is very influential in various aspects of life such as work, education, and health. Therefore, in-depth evaluation is needed to improve AI ethical approaches. The authors suggest several steps for improvement, including increased transparency and accountability in AI development, as well as more consistent implementation of ethical guidelines. By strengthening these principles, it is hoped that AI can be developed and used more ethically, bringing maximum benefits to society.

KEYWORDS

artificial intelligence; ethics; guidelines

PENDAHULUAN

Perkembangan pesat kecerdasan buatan (AI) saat ini disertai dengan seruan yang terus menerus untuk menerapkan etika guna memanfaatkan potensi besar yang dimilikinya. Sebagai hasilnya, berbagai pedoman etika telah dirumuskan dalam beberapa tahun terakhir, yang berisi prinsip-prinsip yang diharapkan dapat diikuti oleh pengembang teknologi. Namun, muncul pertanyaan penting: Apakah pedoman etika ini benar-benar mempengaruhi keputusan manusia dalam bidang AI dan *machine learning*? Jawabannya adalah tidak. Penelitian ini menganalisis 22 pedoman etika AI utama dan memberikan saran untuk meningkatkan efektivitasnya.

Etika AI, atau etika secara umum, kurang memiliki mekanisme untuk memperkuat klaim normatifnya. Penegakan prinsip etika mungkin melibatkan kerugian reputasi akibat pelanggaran atau pembatasan keanggotaan di badan profesional, tetapi secara keseluruhan mekanisme ini lemah dan tidak menimbulkan ancaman besar. Peneliti, politisi, konsultan, manajer, dan aktivis harus menghadapi kelemahan etika ini. Namun, inilah mengapa banyak perusahaan dan lembaga AI tertarik dengan etika. Saat mereka merumuskan pedoman etika sendiri atau memasukkan pertimbangan etis ke dalam pekerjaan mereka, upaya untuk menciptakan kerangka hukum yang mengikat sering terhambat. Pedoman etika industri AI memberi kesan kepada pembuat undang-undang bahwa tata kelola internal sudah cukup, dan tidak diperlukan hukum khusus untuk mengurangi risiko teknologi atau penyalahgunaan (Calo 2017) . Bahkan ketika ada tuntutan untuk undang-undang yang lebih jelas, seperti yang baru-baru ini diajukan oleh Google (2019), tuntutan tersebut tetap samar dan kurang mendalam.

Pedoman etika yang dikembangkan oleh sektor sains atau industri, beserta konsep-konsep tata kelola mandiri lainnya, sering kali digunakan untuk menciptakan kesan bahwa tanggung jawab dapat dialihkan dari otoritas negara dan lembaga demokratis kepada sektor-sektor ini. Etika juga dapat berfungsi untuk meredakan kritik dari publik, sementara praktik-praktik yang dipersoalkan tetap berlangsung dalam organisasi. Contohnya, asosiasi "Partnership on AI" (2018) yang terdiri dari perusahaan-perusahaan seperti Amazon, Apple, Baidu, Facebook, Google, IBM, dan Intel, sering digunakan untuk menunjukkan komitmen terhadap regulasi hukum, meskipun sebenarnya komitmen tersebut tidak terlalu kuat.

Hal ini menimbulkan pertanyaan mengenai seberapa jauh tujuan etika benar-benar diterapkan dan diintegrasikan dalam pengembangan dan penerapan AI, atau apakah hanya niat baik yang diungkapkan. Meskipun beberapa artikel telah membahas pengajaran etika kepada ilmuwan data (Garzcarek dan Steuer 2019; Burton et al. 2017; Goldsmith dan Burton 2017; Johnson 2017), tulisan yang membahas implementasi nyata dari tujuan dan nilai etika sangat sedikit. Penelitian ini akan menawarkan gagasan strategis untuk mentransformasi pedoman etika AI dari sekadar wacana normatif menjadi panduan praktis yang dapat dioperasionalkan dalam tindakan nyata. Penelitian ini bertujuan untuk menganalisis secara kritis 22 pedoman etika utama yang berkaitan dengan pengembangan kecerdasan buatan, dengan fokus pada efektivitas penerapan prinsip-prinsip etika dalam praktik nyata. Studi ini juga mengidentifikasi kekurangan dalam mekanisme implementasi pedoman tersebut dan memberikan rekomendasi untuk memperkuat peran etika dalam tata kelola AI. Hasil penelitian ini diharapkan dapat memberikan wawasan teoretis dan praktis bagi pengembang teknologi, pembuat kebijakan, serta masyarakat umum dalam menciptakan kerangka tata kelola AI yang lebih transparan, bertanggung jawab, dan berbasis nilai-nilai etika yang dapat diukur.

METODE

Penelitian ini merupakan penelitian evaluatif dengan pendekatan deskriptif kualitatif. Fokus utama adalah menganalisis efektivitas dan relevansi 22 pedoman etika AI utama. Metode ini digunakan untuk menilai kesesuaian pedoman dengan prinsip-prinsip etika yang diterapkan dalam pengembangan dan penerapan teknologi kecerdasan buatan.

Sampel dalam penelitian ini adalah 22 dokumen pedoman etika AI yang dipilih secara *Purposive sampling*. Pemilihan didasarkan pada kriteria relevansi, kemutakhiran (maksimal lima tahun terakhir), dan cakupan global. Sampel meliputi dokumen dari organisasi internasional, pemerintah, perusahaan teknologi besar, dan asosiasi industri. Analisis data dilakukan dengan metode content analysis untuk mengevaluasi kesesuaian, kesenjangan, dan implementasi prinsip-prinsip dalam setiap pedoman. Data dikumpulkan melalui tinjauan literatur pada basis data seperti Google Scholar, ACM Digital Library, dan Algorithm Watch. Analisis dilakukan dalam dua tahap: 1) Identifikasi dan kategorisasi prinsip-prinsip etika dalam setiap dokumen; 2) Perbandingan prinsip-prinsip tersebut dengan praktik nyata dalam penelitian dan pengembangan AI.

HASIL DAN PEMBAHASAN

Penelitian di bidang etika AI mencakup berbagai refleksi tentang bagaimana prinsip-prinsip etika dapat diimplementasikan dalam rutinitas pengambilan keputusan mesin otonom (Anderson dan Anderson 2015; Etzioni dan Etzioni 2017; Yu et al. 2018), studi meta tentang etika AI (Vakkuri dan Abrahamsson 2018; Prates et al. 2018; Boddington 2017; Greene et al. 2019; Goldsmith dan Burton 2017), serta analisis empiris tentang bagaimana masalah kereta (trolley problems) diselesaikan (Awad et al. 2018). Selain itu, ada juga refleksi mengenai masalah spesifik (Eckersley 2018) dan pedoman AI yang komprehensif (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2019). Makalah ini terutama membahas isu terakhir tersebut.

Daftar pedoman etika yang dipertimbangkan dalam artikel ini mencakup kompilasi yang mencakup bidang etika AI secara komprehensif. Sejauh yang saya ketahui, ada beberapa preprint dan makalah yang saat ini tersedia yang juga membahas perbandingan berbagai pedoman etika (Zeng et al. 2018; Fjeld et al. 2019; Jobin et al. 2019). Meskipun terutama makalah dari Jobin et al. (2019) merupakan tinjauan sistematis terhadap semua literatur yang ada tentang etika AI, makalah ini tidak bertujuan untuk melakukan analisis penuh terhadap setiap dokumen norma hukum atau non-hukum yang tersedia mengenai AI, algoritma, robot, atau etika data. Sebaliknya, ini lebih merupakan gambaran semi-sistematis mengenai isu-isu dan sikap normatif dalam bidang ini, yang menunjukkan bagaimana rincian etika AI berkaitan dengan gambaran yang lebih besar.

Pemilihan dan pengumpulan 22 pedoman etika utama dilakukan melalui analisis literatur. Proses seleksi ini berlangsung dalam dua tahap. Pada tahap pertama, saya melakukan pencarian di berbagai basis data, termasuk Google, Google Scholar, Web of Science, ACM Digital Library, arXiv, dan SSRN, dengan menggunakan kata kunci seperti “etika AI”, “etika kecerdasan buatan”, “prinsip AI”, “prinsip kecerdasan buatan”, “pedoman AI”, dan “pedoman kecerdasan buatan.” Saya menelusuri setiap tautan dalam 25 hasil pencarian pertama dan mengabaikan duplikasi yang muncul. Dalam menganalisis hasil pencarian, saya juga mengeksplorasi referensi untuk menemukan pedoman yang relevan secara manual. Selain itu, saya menggunakan Inventarisasi Global Pedoman Etika AI dari Algorithm Watch, yang merupakan daftar komprehensif pedoman etika yang disusun secara crowdsourcing, untuk memastikan tidak ada pedoman penting yang terlewatkan. Melalui daftar tersebut, saya menemukan tiga pedoman tambahan yang

memenuhi kriteria seleksi. Perlu dicatat bahwa pilihan saya cenderung bias terhadap dokumen yang bersifat Barat/Utara, sehingga pedoman yang tidak ditulis dalam bahasa Inggris terabaikan.

Saya menolak semua dokumen yang lebih dari lima tahun agar hanya mempertimbangkan pedoman yang relatif baru. Dokumen yang hanya merujuk pada konteks nasional—seperti makalah posisi dari kelompok kepentingan nasional (Smart Dubai 2018), laporan dari British House of Lords (Bakewell et al. 2018), atau pendirian insinyur Nordik mengenai Kecerdasan Buatan dan Etika (Podgaiska dan Shklovski)—tidak termasuk dalam kompilasi ini. Namun, saya menyertakan "Ethics Guidelines for Trustworthy AI" dari Komisi Eropa (Pekka et al. 2018), "Report on the Future of Artificial Intelligence" dari pemerintahan Obama (Holdren et al. 2016), serta "Beijing AI Principles" (Beijing Academy of Artificial Intelligence 2019) yang didukung oleh Kementerian Sains dan Teknologi Tiongkok. Ketiga pedoman ini dimasukkan karena mewakili tiga kekuatan besar dalam bidang AI. Selain itu, saya juga memasukkan "OECD Principles on AI" (Organisation for Economic Co-operation and Development 2019) karena sifatnya yang supranasional. Makalah ilmiah atau teks yang termasuk dalam kategori etika AI tetapi fokus pada satu atau lebih aspek spesifik tidak diperhitungkan, begitu pula dengan pedoman atau alat bantu yang tidak secara khusus membahas AI tetapi lebih kepada big data, algoritma, atau robotika (Anderson et al. 2018; Anderson dan Anderson 2011). Kebijakan perusahaan juga dikecualikan, kecuali "Information Technology Industry AI Policy Principles" (2017), prinsip dari "Partnership on AI" (2018), serta versi pertama dan kedua dari dokumen "Ethically Aligned Design" oleh IEEE (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2016, 2019), serta daftar prinsip singkat dari Google (2018), Microsoft (2019), DeepMind (DeepMind), OpenAI (2018), dan IBM (Cutler et al. 2018) yang telah dikenal melalui media. Perusahaan besar seperti Facebook atau Twitter belum menerbitkan pedoman AI yang sistematis, hanya pernyataan perilaku baik yang terisolasi. Buku Paula Boddington tentang pedoman etika (2017) yang didanai oleh Future of Life Institute juga tidak dipertimbangkan karena hanya mengulangi prinsip Asilomar (2017). Faktor utama dalam pemilihan pedoman etika bukanlah kedalaman detail dari dokumen, tetapi niat yang jelas untuk melakukan pemetaan dan kategorisasi yang komprehensif terhadap klaim normatif dalam bidang etika AI.

Sejumlah isu muncul kembali di berbagai pedoman. Terutama, aspek akuntabilitas, privasi, dan keadilan tercantum dalam sekitar 80% dari semua pedoman, memberikan persyaratan dasar untuk membangun dan menggunakan sistem AI yang "etis." Yang menarik adalah bahwa aspek-aspek yang paling sering disebutkan adalah yang mana solusi teknisnya sudah ada atau sedang dikembangkan. Usaha teknis yang signifikan dilakukan untuk mencapai tujuan etis dalam akuntabilitas dan AI yang dapat dijelaskan (Mittelstadt et al. 2019), keadilan serta penambangan data yang peka terhadap diskriminasi (Gebru et al. 2018), dan privasi (Baron dan Musolesi 2017). Banyak inisiatif ini dikoordinasikan di bawah komunitas FAT ML atau XAI (Veale dan Binns 2017; Selbst et al. 2018). Beberapa perusahaan teknologi, seperti Google, Microsoft, dan Facebook, telah meluncurkan alat untuk mengatasi bias dan memastikan keadilan dalam *machine learning*, seperti toolkit "AI Fairness 360," "What-If Tool," "Facets," "fairlern.py," dan "Fairness Flow" (Whittaker et al. 2018). Nilai-nilai seperti akuntabilitas, keterjelasan, privasi, dan keadilan, serta nilai-nilai lainnya seperti ketahanan dan keselamatan, lebih mudah dioperasikan secara matematis, sehingga cenderung diimplementasikan sebagai solusi teknis. Mengacu pada penelitian psikolog Carol Gilligan, bisa dikatakan bahwa cara etika AI dipraktikkan dan disusun mencerminkan dominasi laki-laki dalam etika keadilan (Gilligan 1982). Gilligan menunjukkan melalui studi empiris di tahun 1980-an bahwa perempuan tidak menyelesaikan masalah moral dengan pendekatan "rasional" dan "logis" seperti yang umum dilakukan pria, melainkan melalui kerangka kerja yang lebih luas dengan pendekatan "empati" dan "emosi". Dalam hal ini, analisis distribusi penulis di pedoman menunjukkan bahwa 41,7% penulisnya adalah perempuan. Namun, angka ini menjadi 31,3% jika laporan dari organisasi yang dipimpin oleh perempuan diabaikan. Proporsi penulis perempuan dalam pedoman komunitas FAT ML, yang lebih berfokus pada solusi teknis, paling rendah, yakni 7,7% (Dia- kopoulos et al.). Dengan demikian, cara berpikir laki-laki tentang masalah etika tercermin dalam hampir semua pedoman etika dengan penekanan pada aspek seperti akuntabilitas dan privasi. Sementara itu, sangat sedikit pedoman yang membahas tentang AI dalam konteks perawatan, kesejahteraan, atau tanggung jawab sosial. Dalam etika AI, artefak teknis sering dipandang sebagai entitas terpisah yang dapat dioptimalkan oleh para ahli untuk mencari solusi bagi masalah teknis.

Keterkaitan antara bisnis dan sains terlihat jelas, terutama karena semua konferensi AI besar disponsori oleh mitra dari industri. Hal ini juga didukung oleh data dalam AI Index 2018 (Shoham et al. 2018), yang menunjukkan peningkatan signifikan jumlah publikasi AI yang berasal dari perusahaan dalam beberapa tahun terakhir. Selain itu, pertumbuhan pesat dalam jumlah startup AI yang aktif didorong oleh investasi besar dari perusahaan modal ventura. Setiap tahun, puluhan ribu paten terkait AI terdaftar. Berbagai sektor industri kini menerapkan aplikasi AI di banyak bidang, seperti manufaktur, manajemen rantai pasokan, pengembangan layanan, pemasaran, dan penilaian risiko. Secara keseluruhan, nilai pasar AI global diperkirakan lebih dari 7 miliar dolar (Wiggers 2019). Namun, saat kita mengamati pasar AI global dan penerapan sistem AI dalam ekonomi serta sistem sosial lainnya, terlihat adanya efek samping yang tidak diinginkan dan konteks penggunaan yang merugikan. Penggunaan militer AI dalam perang siber dan kendaraan tak berawak bersenjata atau drone menjadi sorotan utama (Ernest dan Carroll 2016; Anderson dan Waxman 2013). Menurut laporan media, pemerintah AS berencana untuk menginvestasikan dua miliar dolar dalam proyek AI militer dalam lima tahun ke depan (Fryer-Biggs 2018). Selain itu, pemerintah dapat memanfaatkan aplikasi AI untuk propaganda otomatis dan kampanye disinformasi (Lazer et al. 2018), pengawasan sosial (Engelmann et al. 2019), serta teknik interogasi yang lebih baik (McAllister 2017). Di sisi lain, perusahaan dapat menyebabkan pemutusan hubungan kerja yang masif akibat penerapan AI (Frey dan Osborne 2013), melakukan eksperimen AI tanpa pengawasan pada masyarakat tanpa mendapatkan persetujuan (Kramer et al. 2014), mengalami pelanggaran data (Schneier 2018), dan mengandalkan algoritma yang bias (Eubanks 2018). Mereka juga berisiko menyediakan produk AI yang tidak aman (Sitawarin et al. 2018), menyembunyikan fungsi AI yang berbahaya (Whittaker et al. 2018), dan terburu-buru memasarkan aplikasi AI yang belum teruji. Selain itu, peretas dengan niat jahat dapat memanfaatkan AI untuk melancarkan serangan siber, mencuri data, atau menyebarkan informasi yang salah melalui teknologi seperti deepfake (Bendel 2017). Meskipun demikian, hanya sedikit publikasi yang membahas penyalahgunaan sistem AI, meskipun contoh-contoh yang ada menunjukkan potensi kerusakan besar yang dapat ditimbulkan oleh sistem tersebut (Brundage et al. 2018; King et al. 2019; O'Neil 2016).

Apakah pedoman etika mampu mempengaruhi pengambilan keputusan individu tanpa mempertimbangkan konteks sosial yang lebih luas? Dalam sebuah studi terkendali

terbaru, para peneliti menilai secara kritis anggapan bahwa pedoman etika dapat berfungsi sebagai dasar bagi pengambilan keputusan etis para insinyur perangkat lunak (McNamara et al. 2018). Singkatnya, temuan utama mereka menunjukkan bahwa efektivitas pedoman atau kode etik tersebut nyaris nol dan tidak berdampak pada perilaku profesional dalam komunitas teknologi. Survei ini melibatkan 63 mahasiswa teknik perangkat lunak dan 105 pengembang perangkat lunak profesional. Mereka diberikan sebelas skenario pengambilan keputusan etis yang berkaitan dengan perangkat lunak, untuk menguji apakah pedoman etika dari Association for Computing Machinery (ACM) (Gotterbarn et al. 2018) berpengaruh pada pengambilan keputusan etis dalam enam vignette, yang meliputi tanggung jawab pelaporan, pengumpulan data pengguna, hak kekayaan intelektual, kualitas kode, kejujuran kepada pelanggan, serta manajemen waktu dan personel. Hasilnya cukup mengecewakan: "Tidak ditemukan perbedaan yang signifikan secara statistik dalam respons untuk vignette mana pun antara individu yang melihat dan tidak melihat kode etik, baik untuk mahasiswa maupun profesional." (McNamara et al. 2018, 4).

SIMPULAN

Hasil penelitian menunjukkan bahwa, meskipun pedoman-pedoman ini mencakup prinsip-prinsip dasar seperti akuntabilitas, transparansi, dan keadilan, implementasinya dalam praktik nyata masih lemah. Penelitian ini berhasil mengungkap bahwa pedoman etika sering kali hanya menjadi dokumen normatif tanpa mekanisme penguatan yang efektif. Dampaknya, banyak risiko etika dalam pengembangan AI tetap tidak terselesaikan. Dibandingkan dengan penelitian sebelumnya (Jobin et al., 2019; Zeng et al., 2018), penelitian ini memberikan kontribusi baru berupa rekomendasi strategis untuk mentransformasi etika AI dari wacana menjadi panduan tindakan yang konkret. Ke depannya, diharapkan penelitian ini dapat mendorong pengembangan kerangka kerja etika AI yang lebih transparan, inklusif, dan berorientasi pada nilai-nilai sosial. Implikasinya adalah perlunya kerjasama lintas sektor untuk memastikan penguatan mekanisme implementasi etika AI yang berdaya guna.

DAFTAR PUSTAKA

Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge: Cambridge University Press.

Anderson, M., Anderson, S. L. (2015). Towards ensuring ethical behavior from autonomous systems: A case-supported principle-based paradigm. *In Artificial intelligence and ethics: Papers from the 2015 AAAI Workshop (pp. 1–10).*

Anderson, D., Bonaguro, J., McKinney, M., Nicklin, A., Wiseman, J. (2018). *Ethics & algorithms toolkit.* dalam https://ethicstoolkit.ai/. Diakses 1 February 2023.

Anderson, K., Waxman, M. C. (2013). Law and ethics for autonomous weapon systems: Why a ban won't work and how the laws of WAR can. *SSRN Journal*, 1–32.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., et al. (2018). The moral machine experiment. *Nature, 563*(7729), 59–64.

Bakewell, J. D., Clement-Jones, T. F., Giddens, A., Grender, R. M., Hollick, C. R., Holmes, C., Levene, P. K. et al. (2018). AI in the UK: Ready, willing and able?. *Select committee on artificial intel- ligence (pp. 1–183).*

Baron, B., Musolesi, M. (2017). Interpretable machine learning for privacy-preserving pervasive systems. *arXiv (pp. 1–10).*

Beijing Academy of Artificial Intelligence. (2019). *Beijing AI principles* dalam https://www.baai.ac.cn/blog/beijing-ai-principles. Diakses 18 Juni 2023.

Bendel, O. (2017). The synthetization of human voices. *AI & SOCIETY - Journal of Knowledge, Culture and Communication, 82*, 737.

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Cham: Springer.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A. et al. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv (pp. 1–101).*

Burton, E., Goldsmith, J., Koening, S., Kuipers, B., Mattei, N., & Walsh, T. (2017). Ethical considera- tions in artificial intelligence courses. *Artificial Intelligence Magazine*, 38(2), 22–36.

Calo, R. (2017). Artificial intelligence policy: a primer and roadmap. *SSRN Journal*, 1–28.

Cutler, A., Pribić, M., Humphrey, L. (2018). *Everyday ethics for artificial intelligence: A practical guide for designers & developers* dalam https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf. Diakses 4 February 2023.

DeepMind. *DeepMind ethics & society principles*. dalam https://deepmind.com/applied/deepmind-ethics-society/principles/. Diakses 17 Juli 17 2023.

Eckersley, P. (2018). Impossibility and uncertainty theorems in AI value alignment or why your AGI should not have a utility function. *arXiv (pp. 1–13).*

Engelmann, S., Chen, M., Fischer, F., Kao, C., Grossklags, J. (2019). Clear sanctions, vague rewards: How China's social credit system currently defines "Good" and "Bad" behavior. *In Proceedings of the conference on fairness, accountability, and transparency—FAT* '19 (pp. 69–78).*

Ernest, N., & Carroll, D. (2016). Genetic fuzzy based artificial intelligence for unmanned combat aerial vehicle control in simulated air combat missions. *Journal of Defense Management.*

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics, 21*(4), 403–418.

Fjeld, J., Hilligoss, H., Achten, N., Daniel, M. L., Feldman, J., Kagay, S. (2019). *Principled artificial intelligence: A map of ethical and rights-based approaches* dalam https://ai-hr. cyber.harvard.edu/primp-viz.html. Diakses 17 Juli 2023.

Frey, C. B., Osborne, M. A. (2013). The future of employment: *How susceptible are jobs to computerisa- tion: Oxford Martin Programme on Technology and Employment (pp. 1–78).*

Fryer-Biggs, Z. (2018). *The pentagon plans to spend $2 billion to put more artificial intelligence into its weaponry* dalam https://www.theverge.com/2018/9/8/17833160/pentagon- darpa-artificial-intelligence-ai-investment. Diakses 25 Januari 2023.

Garzcarek, U., Steuer, D. (2019). Approaching ethical guidelines for data scientists. *arXiv (pp. 1–18).*

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé, III, H., Crawford, K. (2018). Datasheets for datasets. *arXiv (pp. 1–17).*

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge: Harvard University Press.

Goldsmith, J., Burton, E. (2017). Why teaching ethics to AI practitioners is important. *ACM SIGCAS Computers and Society (pp. 110–114).*

Google. (2018). *Artificial intelligence at Google: Our principles* dalam https://ai.google/principles/. Diakses 24 Januari 2023.

Google. (2019). *Perspectives on issues in AI governance (pp. 1–34)* dalam https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf. Diakses 11 Februari 2023.

Gotterbarn, D., Brinkman, B., Flick, C., Kirkpatrick, M. S., Miller, K., Vazansky, K., Wolf, M. J. (2018). *ACM code of ethics and professional conduct: Affirming our obligation to use our skills to ben- efit society (pp. 1–28)* dalam https://www.acm.org/binaries/content/assets/ about/acm-code-of-ethics-booklet.pdf. Diakses 1 February 2023.

Greene, D., Hoffman, A. L., Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *In Hawaii international confer- ence on system sciences (pp. 1–10).*

Holdren, J. P., Bruce, A., Felten, E., Lyons, T., & Garris, M. (2016). *Preparing for the future of artificial intelligence (pp. 1–58).* Washington, D.C: Springer.

Howard, P. N., Kollanyi, B. (2016). Bots, #StrongerIn, and #Brexit: Computational propaganda during the UK-EU Referendum. *arXiv (pp. 1–6).*

Hursthouse, R. (2001). *On virtue ethics. Oxford: Oxford University Press. Information Technology Industry Council. (2017)* dalam https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf.. Diakses 29 Januari 2023.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1(9), 389–399.*

Johnson, D. G. (2017). Can engineering ethics be taught?. *The Bridge, 47*(1), 59–64.

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial intelligence crime: An interdisci- plinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics, 26, 89–120.*

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emo- tional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America, 111(24), 8788–8790.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018).The science of fake news. *Science, 359(6380), 1094–1096.*

McAllister, A. (2017). Stranger than science fiction: The rise of A.I. interrogation in the dawn of autono- mous robots and the need for an additional protocol to the U.N. convention against torture. *Minne- sota Law Review, 101, 2527–2573*.

McNamara, A., Smith, J., Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development?" In G. T. Leavens, A. Garcia, C. S. Păsăreanu (Eds.) *Proceed- ings of the 2018 26th ACM joint meeting on european software engineering conference and sym- posium on the foundations of software engineering—ESEC/FSE 2018 (pp. 1–7).* New York: ACM Press.

Microsoft Corporation. (2019). *Microsoft AI principles* dalam https://www.micro soft.com/en-us/ai/our-approach-to-ai. Diakses 1 February 2023.

Mittelstadt, B., Russell, C., Wachter, S. (2019). *Explaining explanations in AI*. In Proceedings of the con- ference on fairness, accountability, and transparency—FAT* '19 (pp. 1–10).

O'Neil, C. (2016*). Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishers.

OpenAI. (2018). *OpenAI Charter* dalam https://openai.com/charter/. Diakses 17 Juli 2023.

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence (pp. 1–12)* dalam https://legalinstruments.oecd.org/en/ instruments/OECD-LEGAL-0449. Diakses 18 Juni 2023.

Partnership on AI. (2018). *About us* dalam https://www.partnershiponai.org/about/. Diakses 25 Januari 2023.

Pekka, A.-P., Bauer, W., Bergmann, U., Bieliková, M., Bonefeld-Dahl, C., Bonnet, Y., Bouarfa, L. et al. (2018). The European Commission's high-level expert group on artificial intelligence: Ethics guidelines for trustworthy ai. Working Document for stakeholders' consultation. *Brussels (pp.1–37).*

Prates, M., Avelar, P., Lamb, L. C. (2018). On quantifying and understanding the role of ethics in AI research: A historical account of flagship conferences and journals. *arXiv (pp. 1–13).*

Schneier, B. (2018). Click here to kill everybody. New York: W. W. *Norton & Company*.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., Vertesi, J. (2018). Fairness and abstrac- tion in Sociotechnical Systems. *In ACT conference on fairness, accountability, and transparency (FAT) (vol. 1, No. 1, pp. 1–17).*

Sitawarin, C., Bhagoji, A. N., Mosenia, A., Chiang, M., Mittal, P. (2018). DARTS: Deceiving autono- mous cars with toxic signs. *arXiv (pp. 1–27).*

Smart Dubai. 2018. AI ethics principles & guidelines dalam https://smartdubai.ae/docs/default-source/ai-principles-resources/aiethics.pdf. Diakses 1 February 2023.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751–752. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2016). *Ethically aligned design: A vision for prioritizing human well-being with artificial intelligence and autonomous sys- tems (pp. 1–138).*

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (pp.1–294).*

Vakkuri, V., Abrahamsson, P. (2018). The key concepts of ethics of artificial intelligence. *In Proceedings of the 2018 IEEE international conference on engineering, technology and innovation (pp. 1–6).*

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination with- out collecting sensitive data. *Big Data & Society, 4(2), 1–17.*

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., Schwartz, O. (2018). *AI now report 2018 (pp. 1–62).*

Wiggers, K. (2019). CB insights: *Here are the top 100 AI companies in the world* dalam https://venturebeat.com/2019/02/06/cb-insights-here-are-the-top-100-ai-companies-in- the-world/. Diakses 11 February 2023.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., Yang, Q. (2018). Building ethics into artificial intel- ligence. *arXiv (pp. 1–8).*

Zeng, Y., Lu, E., Huangfu, C. (2018). Linking artificial intelligence principles. *arXiv (pp. 1–4).*