Semantika
Seminar Nasional Teknologi Informatika
Prosiding Seminar Nasional Teknik Informatika

Volume 3 Nomor 1 April 2026 | E-ISSN: 9999-9999|
Doi: https://semantika.polgan.ac.id/index.php
Dipublikasikan Oleh Politeknik Ganesha Medan

ANALISIS KEAMANAN DATA DALAM PENGGUNAAN ARTIFICIAL INTELLIGENCE : SISTEMATIKA LITERATURE REVIEW

Auliya Maharlika
Program Studi Manajemen Informatika Politeknik Ganesha Medan
auliyamaharlika44@gmail.com

Abstrak

Perkembangan teknologi Artificial Intelligence (AI) yang sangat pesat dalam dekade terakhir telah membawa transformasi fundamental dalam berbagai aspek kehidupan manusia, mulai dari sektor pendidikan, kesehatan, bisnis, hingga industri kreatif. Namun demikian, pesatnya adopsi teknologi AI juga memunculkan berbagai tantangan serius terkait keamanan data yang perlu mendapat perhatian khusus dari berbagai pemangku kepentingan. Penelitian ini bertujuan untuk menganalisis secara komprehensif aspek keamanan data dalam penggunaan teknologi AI, mengidentifikasi berbagai ancaman dan kerentanan yang muncul, serta merumuskan strategi mitigasi yang dapat diimplementasikan untuk melindungi privasi dan keamanan informasi pengguna. Metode yang digunakan dalam penelitian ini adalah studi literatur kualitatif dengan mengkaji berbagai sumber ilmiah berupa jurnal, artikel, dan publikasi akademik yang relevan dengan topik keamanan data dan AI. Hasil penelitian menunjukkan bahwa meskipun AI memberikan berbagai manfaat signifikan dalam meningkatkan efisiensi dan produktivitas, terdapat risiko keamanan data yang substansial meliputi kebocoran informasi sensitif, serangan adversarial, bias algoritma, dan potensi penyalahgunaan data pribadi. Kesimpulan penelitian ini menegaskan perlunya kerangka regulasi yang komprehensif, implementasi prinsip privacy by design, serta peningkatan literasi digital masyarakat untuk memastikan pemanfaatan AI yang aman dan bertanggung jawab.

Kata Kunci: Artificial Intelligence, Keamanan Data, Privasi, Machine Learning, Perlindungan Informasi

Pendahuluan, Revolusi industri keempat yang sedang berlangsung saat ini ditandai dengan konvergensi teknologi digital, fisik, dan biologis yang mengubah cara manusia hidup, bekerja, dan berinteraksi satu sama lain. Di antara berbagai teknologi yang menjadi pilar utama revolusi ini, Artificial Intelligence (AI) atau kecerdasan buatan menempati posisi sentral sebagai katalis perubahan yang paling transformatif (Bostrom, 2014). Teknologi AI telah berkembang dari sekadar konsep teoretis dalam laboratorium penelitian menjadi realitas yang menyentuh hampir setiap aspek kehidupan sehari-hari masyarakat modern, mulai dari asisten virtual di perangkat genggam hingga sistem diagnosis medis canggih di rumah sakit (Russell & Norvig, 2020).

Fenomena perkembangan AI global menunjukkan pertumbuhan yang sangat signifikan dalam beberapa tahun terakhir. Perkembangan ini didorong oleh kemajuan dalam bidang Machine Learning yang memungkinkan sistem belajar dari data dalam skala besar (Jordan & Mitchell, 2015). Negara-negara maju seperti Amerika Serikat, Tiongkok, dan negara-negara Uni Eropa berlomba-lomba mengembangkan kapabilitas AI nasional mereka, menyadari bahwa penguasaan teknologi ini akan menentukan daya saing ekonomi dan posisi geopolitik di masa depan. Indonesia sendiri tidak ketinggalan dalam arus transformasi ini melalui berbagai inisiatif adopsi AI di berbagai sektor.

Kemunculan teknologi generative AI seperti ChatGPT, Gemini, Claude, dan berbagai platform serupa dalam beberapa tahun terakhir telah memicu gelombang antusiasme sekaligus kekhawatiran. Kemampuan sistem AI generatif untuk menghasilkan teks, gambar, kode pemrograman, dan berbagai bentuk konten kreatif lainnya dengan kualitas tinggi tidak terlepas dari perkembangan arsitektur deep learning berbasis transformer (Vaswani et al., 2017) serta model bahasa besar yang mampu menghasilkan teks secara

otomatis (Brown et al., 2020).

Keamanan data dalam konteks penggunaan AI menjadi isu yang semakin krusial mengingat karakteristik teknologi ini yang sangat bergantung pada ketersediaan data dalam jumlah besar. Sistem AI, khususnya yang berbasis Deep Learning, memerlukan dataset yang sangat besar untuk dapat belajar dan meningkatkan performanya (LeCun et al., 2015). Data tersebut sering kali mencakup informasi sensitif seperti data pribadi, riwayat kesehatan, hingga pola perilaku pengguna. Tanpa mekanisme perlindungan yang memadai, pengolahan data dalam skala besar ini berpotensi menimbulkan risiko kebocoran dan pelanggaran privasi.

Berbagai penelitian telah menunjukkan adanya ancaman nyata terhadap keamanan data dalam sistem AI. Serangan seperti adversarial attack memungkinkan manipulasi sistem AI melalui input tertentu sehingga menghasilkan keputusan yang salah (Goodfellow et al., 2015). Selain itu, teknik membership inference attack menunjukkan bahwa data pelatihan dapat diekstraksi dari model yang telah dilatih (Shokri et al., 2017). Hal ini membuktikan bahwa sistem AI tidak hanya rentan terhadap serangan eksternal, tetapi juga berpotensi membocorkan informasi dari dalam model itu sendiri.

Meskipun perkembangan AI telah banyak diteliti, sebagian besar penelitian masih berfokus pada peningkatan performa dan efisiensi sistem, sementara aspek keamanan data belum dikaji secara komprehensif dan terintegrasi (Zhang et al., 2021). Penelitian lain juga lebih menekankan pada aspek teknis tanpa mengaitkan implikasinya terhadap pengguna dan sektor nyata. Hal ini menunjukkan adanya kesenjangan penelitian (research gap) dalam mengkaji hubungan antara perkembangan AI, risiko keamanan data, serta strategi mitigasinya secara menyeluruh. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut melalui pendekatan Systematic Literature Review (SLR).

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini merumuskan beberapa permasalahan utama yang akan dianalisis secara mendalam, yaitu terkait dengan kondisi terkini perkembangan teknologi Artificial Intelligence serta implikasinya terhadap keamanan data, berbagai ancaman dan kerentanan keamanan data yang muncul dari penggunaan teknologi AI di berbagai sektor, strategi serta langkah-langkah mitigasi yang dapat diterapkan untuk melindungi keamanan data dalam ekosistem AI, serta peran berbagai pemangku kepentingan dalam memastikan penggunaan AI yang aman, etis, dan bertanggung jawab.

Penelitian ini memiliki beberapa tujuan yang hendak dicapai melalui analisis komprehensif terhadap berbagai sumber literatur ilmiah. Tujuan pertama adalah menganalisis dan memetakan perkembangan terkini teknologi AI serta mengidentifikasi implikasinya terhadap aspek keamanan data. Tujuan kedua adalah mengidentifikasi dan mengklasifikasikan berbagai bentuk ancaman dan kerentanan keamanan data yang terkait dengan penggunaan teknologi AI. Tujuan ketiga adalah merumuskan strategi dan rekomendasi praktis yang dapat diterapkan oleh berbagai pihak untuk memitigasi risiko keamanan data dalam pemanfaatan AI. Tujuan keempat adalah memberikan kontribusi pemikiran bagi pengembangan kerangka kebijakan dan regulasi yang mengatur penggunaan AI secara aman dan etis.

Hasil penelitian ini diharapkan dapat memberikan manfaat baik secara teoretis maupun praktis. Secara teoretis, penelitian ini berkontribusi dalam memperkaya khazanah pengetahuan akademik mengenai interelasi antara teknologi AI dan keamanan data, serta menyediakan landasan konseptual bagi penelitian-penelitian selanjutnya di bidang yang sama. Secara praktis, temuan penelitian ini dapat menjadi rujukan bagi para pengembang teknologi, pembuat kebijakan, pelaku bisnis, dan pengguna umum dalam mengadopsi dan memanfaatkan teknologi AI dengan cara yang meminimalkan risiko keamanan dan melindungi privasi data.

Urgensi penelitian mengenai keamanan data dalam penggunaan AI dilatarbelakangi oleh beberapa faktor kritis yang memerlukan perhatian serius dari komunitas akademik dan praktisi. Pertama, insiden kebocoran data dan pelanggaran privasi yang melibatkan sistem AI semakin sering terjadi dengan dampak yang semakin luas dan serius. Kedua, regulasi dan kerangka hukum yang mengatur penggunaan AI dan perlindungan data di banyak negara, termasuk Indonesia, masih dalam tahap pengembangan dan belum sepenuhnya mampu mengantisipasi kompleksitas teknologi yang terus berkembang. Ketiga, kesenjangan literasi

digital di masyarakat membuat banyak pengguna tidak sepenuhnya memahami risiko keamanan yang mereka hadapi ketika berinteraksi dengan sistem AI.

Definisi dan konsep Articificial Intelligence

Artificial Intelligence atau kecerdasan buatan merupakan salah satu cabang ilmu komputer yang berfokus pada pengembangan sistem yang mampu melakukan tugas-tugas yang secara tradisional memerlukan kecerdasan manusia. Berbagai ahli telah mengemukakan definisi AI dari perspektif yang berbeda-beda, merefleksikan kompleksitas dan multidimensionalitas dari bidang ilmu ini.

John McCarthy, yang sering disebut sebagai bapak AI, mendefinisikan kecerdasan buatan sebagai ilmu dan teknik untuk membuat mesin-mesin cerdas, khususnya program komputer yang cerdas (McCarthy, 2007). Sementara itu, Russell dan Norvig dalam karya seminal mereka mendefinisikan AI sebagai studi tentang agen-agen yang menerima persepsi dari lingkungan dan melakukan tindakan (Russell & Norvig, 2020).

Dalam perkembangannya, AI dapat diklasifikasikan ke dalam beberapa kategori berdasarkan kemampuan dan karakteristiknya. Narrow AI atau Artificial Narrow Intelligence (ANI) merujuk pada sistem AI yang dirancang untuk melakukan tugas spesifik tertentu dengan sangat baik, seperti pengenalan wajah, terjemahan bahasa, atau rekomendasi produk. Jenis AI ini merupakan bentuk yang paling umum ditemui dalam aplikasi-aplikasi komersial saat ini. Artificial General Intelligence (AGI) merujuk pada konsep AI yang memiliki kemampuan kognitif setara dengan manusia sementara Artificial Super Intelligence (ASI) merepresentasikan tahap hipotetis di mana AI melampaui kecerdasan manusia dalam semua aspek (Bostrom, 2014).

Machine Learning sebagai Fondasi AI Modern

Machine learning merupakan subdisiplin dari AI yang berfokus pada pengembangan algoritma dan model statistik yang memungkinkan sistem komputer untuk meningkatkan performanya dalam melalui pembelajaran dari data, tanpa diprogram secara eksplisit (Jordan & Mitchell, 2015). Terdapat beberapa paradigma utama dalam machine learning. Unsupervised learning melibatkan pelatihan model pada data yang tidak dilabeli untuk menemukan pola dan struktur tersembunyi dalam data tersebut, seperti dalam aplikasi pengelompokan pelanggan atau deteksi anomali. Reinforcement learning merupakan paradigma di mana agen belajar melalui interaksi dengan lingkungan (Sutton & Barto, 2018).

Deep Learning dan Revoluci Neural Network

Deep learning merupakan subset dari machine learning yang memanfaatkan arsitektur neural network dengan banyak lapisan untuk mempelajari representasi hierarkis dari data (LeCun, Bengio, & Hinton, 2015). Arsitektur-arsitektur populer mencakup Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), serta arsitektur Transformer yang menjadi fondasi bagi model-model bahasa besar seperti GPT dan BERT (Vaswani et al., 2017).

Natural Language Processing (NLP)

NLP menggabungkan linguistik komputasional, machine learning, dan AI untuk memungkinkan komputer memahami, dan menghasilkan bahasa manusia (Jurafsky & Martin, 2023). Model-model seperti GPT (Generative Pre-trained Transformer), BERT (Bidirectional Encoder Representations from Transformers), menunjukkan kemampuan yang mengesankan. Namun demikian, kemampuan model-model ini juga membawa risiko keamanan yang unik, termasuk untuk mengekstrak informasi sensitif dari model.

Perkembangan AI Terbaru: Generate AI dan Automation

Gelombang terbaru dalam perkembangan AI ditandai dengan generative AI yang mampu menghasilkan konten baru yang orisinal, termasuk teks, gambar, audio, video, dan kode program. Teknologi seperti ChatGPT dari OpenAI, Gemini dari Google, Claude dari Anthropic, DALL-E, Midjourney, dan Stable Diffusion telah mengubah lanskap interaksi manusia-komputer secara fundamental (Ramesh et al., 2022). Bersamaan dengan generative AI, Perkembangan ini membawa implikasi keamanan data multidimensi. termasuk potensi mengingat dan mereproduksi informasi sensitif dari data pelatihan.

Kerangka Konseptual Keamanan Data dalam AI

Keamanan data dalam konteks AI mencakup dari kerahasiaan (confidentiality), integritas (integrity), hingga ketersediaan (availability) data—triad CIA. Kerahasiaan data

tidak hanya mencakup perlindungan terhadap informasi yang mungkin tertanam dalam model AI. Serangan seperti membership inference attack dan model inversion attack dapat mengekstrak informasi data pelatihan (Shokri et al., 2017). Integritas mencakup integritas data pelatihan perilaku model itu sendiri. Dimana Serangan adversarial merupakan ancaman unik untuk sistem AI (Goodfellow et al., 2015)

Metodologi Penelitian

Penelitian ini menggunakan pendekatan kualitatif dengan metode Systematic Literature Review (SLR) sebagai desain penelitian utama. Metode SLR yang digunakan dalam penelitian ini mengacu pada framework yang dikembangkan oleh Barbara Kitchenham, yang terdiri dari tiga tahapan utama, yaitu: Planning (Perencanaan), Conducting (Pelaksanaan), Reporting (Pelaporan)

Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari jurnal ilmiah IEEE Transactions on Information Forensics and Security, ACM Computing Surveys, dan Journal of Artificial Intelligence Research.

Teknik Pengumpulan Data

penelusuran dilakukan pada basis data IEEE Xplore, ACM Digital Library, Google Scholar, Scopus, dan Web of Science. menggunakan kata kunci “artificial intelligence security”, “machine learning privacy”, “data protection AI”, “adversarial attacks”, dan “AI ethics”, seleksi literatur mengacu pada pendekatan PRISMA Statement yang meliputi tahap identifikasi, screening, eligibility, dan inklusi akhir.

Teknik Analisis Data

Data dianalisis menggunakan teknik analisis konten kualitatif (qualitative content analysis) dan sintesis naratif sebagai bagian dari tahap reporting dalam SLR berdasarkan framework Barbara Kitchenham.

Proses analisis dimulai dengan pembacaan menyeluruh terhadap setiap literatur untuk memahami konteks, tujuan penelitian, metodologi, serta temuan utama yang dikemukakan. Selanjutnya dilakukan proses pengkodean (coding) terhadap informasi yang relevan dengan fokus penelitian, khususnya terkait aspek keamanan data dalam penggunaan AI.

Hasil Penelitian dan Analisis Literatur, Berdasarkan hasil kajian terhadap berbagai literatur ilmiah yang relevan, ditemukan bahwa penelitian terkait Artificial Intelligence (AI) secara umum masih didominasi oleh pembahasan mengenai peningkatan performa, efisiensi, dan inovasi teknologi, sementara aspek keamanan data belum mendapatkan perhatian yang seimbang dan komprehensif

No.	Penulis & Tahun	Judul Penelitian	Metode	Temuan Utama	Ulasan / Kelemahan
1.	(Nick Bostrom, 2014)	Superintelligence: Paths, Dangers, Strategies	Studi konseptual	AI berpotensi melampaui kecerdasan manusia dan membawa risiko besar	Tidak membahas keamanan data secara teknis
2.	(Tom B. Brown et al., 2020)	Language Models are Few-Shot Learners	Eksperimen (Deep Learning)	Model bahasa mampu menghasilkan teks berkualitas tinggi	Tidak membahas risiko privasi dan kebocoran data
3.	(Ian Goodfellow et al., 2015)	Explaining and Harnessing Adversarial Examples	Eksperimen	AI rentan terhadap serangan adversarial	Fokus pada aspek teknis, belum membahas dampak luas
4.	(Daniel Jurafsky & James H. Martin, 2023)	Speech and Language Processing	Buku (konseptual)	NLP memungkinkan pemrosesan bahasa alami oleh mesin	Tidak membahas keamanan data secara mendalam

5.	(John McCarthy, 2007)	What is Artificial Intelligence?	Konseptual	Mendefinisikan AI sebagai mesin cerdas	Tidak relevan dengan isu keamanan data modern
6.	(Aditya Ramesh et al., 2022)	Hierarchical Text-Conditional Image Generation	Eksperimen	AI mampu menghasilkan gambar dari teks	Tidak membahas aspek keamanan data
7.	(Stuart Russell & Peter Norvig, 2020)	Artificial Intelligence: A Modern Approach	Buku (teori)	AI sebagai sistem agen cerdas	Fokus pada konsep, belum bahas keamanan data
8.	(Reza Shokri et al., 2017)	Membership Inference Attacks	Eksperimen keamanan	Data training dapat diekstrak dari model AI	Terbatas pada skenario eksperimen
9.	(Richard S. Sutton & Andrew G. Barto, 2018)	Reinforcement Learning: An Introduction	Buku (teori)	RL memungkinkan AI belajar dari interaksi	Tidak membahas keamanan data
10.	Ashish Vaswani et al., 2017)	Attention is All You Need	Eksperimen (Deep Learning)	Transformer meningkatkan performa NLP	Tidak membahas risiko keamanan
11.	(Chiyuan Zhang et al., 2021)	Understanding Deep Learning	Analisis teoretis	Generalisasi deep learning kompleks	Tidak membahas keamanan data

sintesis dan Kontribusi Penelitian: Terdapat kesenjangan signifikan antara perkembangan AI dan pembahasan keamanan data. Penelitian ini berkontribusi mengintegrasikan perkembangan AI, risiko keamanan data, dan implikasinya di sektor pendidikan, kesehatan, bisnis, dan industri kreatif..

LeCun, Bengio, dan Hinton (2015) menjelaskan keunggulan deep learning dalam meningkatkan akurasi sistem AI melalui pemanfaatan neural network yang kompleks. Meskipun memberikan kontribusi signifikan dalam perkembangan AI, penelitian ini belum membahas secara mendalam risiko kebocoran data yang berpotensi berdampak pada sektor sensitif seperti kesehatan.

Penelitian oleh (Brown et al., 2020) menunjukkan bahwa model bahasa berbasis transformer mampu menghasilkan teks secara otomatis dengan tingkat akurasi yang tinggi. Namun demikian, penelitian ini belum secara eksplisit mengkaji potensi penyalahgunaan teknologi tersebut, termasuk risiko penyebaran informasi palsu (misinformation) serta pelanggaran privasi dalam konteks penggunaan data pengguna pada platform digital.

Di sisi lain, (Shokri et al., 2017) mengungkap adanya ancaman serius berupa membership inference attack yang memungkinkan pihak tertentu mengekstrak informasi dari model AI. Penelitian ini memberikan kontribusi penting dalam aspek keamanan, namun masih terbatas pada pengujian eksperimental dan belum mengkaji implikasi praktisnya dalam sektor nyata seperti layanan kesehatan dan keuangan.

(Goodfellow et al., 2015) memperkenalkan konsep adversarial attack yang menunjukkan bahwa sistem AI dapat dimanipulasi melalui input yang dimodifikasi secara halus. Meskipun penelitian ini menjadi dasar penting dalam keamanan AI, pembahasannya masih terbatas pada aspek teknis dan belum mengaitkan dampaknya terhadap pengguna akhir, seperti risiko keselamatan pada kendaraan otonom atau sistem diagnosis medis.

Penelitian (Davenport dan Ronanki, 2018) lebih menyoroti implementasi AI dalam sektor bisnis yang mampu meningkatkan efisiensi operasional. Namun, penelitian ini belum membahas secara mendalam risiko keamanan data pelanggan yang diolah dalam sistem AI, padahal sektor bisnis merupakan salah satu sektor dengan tingkat kerentanan kebocoran data yang tinggi.

(Zhang et al., 2021) membahas tantangan generalisasi dalam deep learning, tetapi tidak secara spesifik mengaitkannya dengan keamanan data. Hal ini menunjukkan adanya kesenjangan penelitian dalam menghubungkan performa model dengan risiko keamanan yang ditimbulkan.

Dengan demikian, penelitian ini tidak hanya merangkum literatur yang ada, tetapi juga mengidentifikasi kelemahan utama dalam penelitian sebelumnya, yaitu kurangnya pendekatan holistik dalam mengkaji keamanan data dalam ekosistem AI.

Pembahasan

LeCun, Bengio, dan Hinton (2015) menjelaskan keunggulan deep learning dalam meningkatkan akurasi sistem AI melalui pemanfaatan neural network yang kompleks. Meskipun memberikan kontribusi signifikan dalam perkembangan AI, penelitian ini belum membahas secara mendalam risiko kebocoran data yang dapat terjadi akibat sifat model yang menyimpan representasi data pelatihan, yang berpotensi berdampak pada sektor sensitif seperti kesehatan.

Penelitian oleh (Brown et al., 2020) menunjukkan bahwa model bahasa berbasis transformer mampu menghasilkan teks secara otomatis dengan tingkat akurasi yang tinggi. Namun demikian, penelitian ini belum secara eksplisit mengkaji potensi penyalahgunaan teknologi tersebut, termasuk risiko penyebaran informasi palsu (misinformation) serta pelanggaran privasi dalam konteks penggunaan data pengguna pada platform digital.

Di sisi lain, (Shokri et al., 2017) mengungkap adanya ancaman serius berupa membership inference attack yang memungkinkan pihak tertentu mengekstrak informasi dari model AI. Penelitian ini memberikan kontribusi penting dalam aspek keamanan, namun masih terbatas pada pengujian eksperimental dan belum mengkaji implikasi praktisnya dalam sektor nyata seperti layanan kesehatan dan keuangan.

(Goodfellow et al., 2015) memperkenalkan konsep adversarial attack yang menunjukkan bahwa sistem AI dapat dimanipulasi melalui input yang dimodifikasi secara halus. Meskipun penelitian ini menjadi dasar penting dalam keamanan AI, pembahasannya masih terbatas pada aspek teknis dan belum mengaitkan dampaknya terhadap pengguna akhir, seperti risiko keselamatan pada kendaraan otonom atau sistem diagnosis medis.

Penelitian (Davenport dan Ronanki, 2018) lebih menyoroti implementasi AI dalam sektor bisnis yang mampu meningkatkan efisiensi operasional. Namun, penelitian ini belum membahas secara mendalam risiko keamanan data pelanggan yang diolah dalam sistem AI, padahal sektor bisnis merupakan salah satu sektor dengan tingkat kerentanan kebocoran data yang tinggi.

(Zhang et al., 2021) membahas tantangan generalisasi dalam deep learning, tetapi tidak secara spesifik mengaitkannya dengan keamanan data. Hal ini menunjukkan adanya kesenjangan penelitian dalam menghubungkan performa model dengan risiko keamanan yang ditimbulkan.

Sektor bisnis dan keuangan AI digunakan untuk deteksi fraud, penilaian kredit, manajemen risiko, trading algoritmik, dan personalisasi layanan pelanggan. Risiko keamanan meliputi pencurian identitas, kerugian reputasional, serta serangan siber yang memanfaatkan AI sebagai alat serangan.

Penerapan AI dalam sektor kesehatan AI digunakan untuk diagnosis penyakit dari citra medis seperti X-ray, CT scan, dan MRI; prediksi risiko penyakit berdasarkan data genetik dan riwayat kesehatan; personalisasi pengobatan berdasarkan karakteristik individual pasien; serta akselerasi penemuan dan pengembangan obat kualitas. Data kesehatan adalah kategori paling sensitif. Model AI dapat "mengingat" data pasien dan berpotensi mengekspos informasi melalui berbagai teknik serangan. Kebocoran data pelanggan dapat mengakibatkan kerugian finansial langsung melalui pencurian identitas dan fraud, serta kerugian reputasional yang dapat menghancurkan nilai bisnis yang dibangun selama bertahun-tahun.

Industri Kreatif

Generative AI mengubah cara produksi konten. Isu meliputi hak kekayaan intelektual, pencurian identitas kreatif, dan deepfake yang mengancam kepercayaan terhadap bukti digital izin untuk melatih sistem yang kemudian dapat menghasilkan karya yang bersaing dengan mereka. kemampuan AI untuk menghasilkan konten yang menyerupai gaya individu tertentu menimbulkan risiko pencurian identitas kreatif dan pemalsuan.

Deepfake, yang merupakan aplikasi AI untuk menghasilkan video atau audio palsu yang sangat realistis, merepresentasikan salah satu ancaman keamanan paling serius yang muncul dari kemajuan AI generatif.

Dampak positif penggunaan AI

teknologi AI membawa dampak positif yang sangat signifikan bagi berbagai aspek kehidupan manusia. Pemahaman yang seimbang mengenai manfaat dan risiko diperlukan untuk memformulasikan pendekatan yang proporsional terhadap adopsi dan regulasi AI. Hal ini dapat membantu mengatasi kesenjangan pendidikan dan memberikan kesempatan yang lebih setara bagi semua peserta didik. Dalam konteks lingkungan hidup, AI digunakan untuk mengoptimalkan konsumsi energi, memprediksi dan memitigasi dampak perubahan iklim, serta mendukung pengelolaan sumber daya alam yang lebih berkelanjutan. Efisiensi ekonomi yang dimungkinkan oleh AI juga memiliki potensi untuk meningkatkan produktivitas dan kesejahteraan secara agregat.

Dampak Negatif dan Risiko Kemanan AI

penggunaan AI juga membawa berbagai dampak negatif dan risiko yang tidak dapat diabaikan. Dari perspektif keamanan teknis, sistem AI menghadapi berbagai vektor serangan yang terus berkembang. Adversarial attacks merupakan bentuk serangan di mana input yang dimodifikasi secara halus—tidak terdeteksi oleh mata manusia—dapat menyebabkan sistem AI membuat kesalahan yang parah. Dalam konteks kendaraan otonom, misalnya, serangan adversarial terhadap sistem pengenalan rambu lalu lintas dapat memiliki konsekuensi fatal.

Dalam bidang kesehatan, AI telah menunjukkan kemampuan untuk meningkatkan akurasi diagnosis penyakit, mempercepat pengembangan obat baru, dan memperluas akses layanan kesehatan ke daerah-daerah yang kurang terlayani melalui telemedicine berbasis AI. Sistem AI mampu menganalisis citra medis dengan akurasi yang menyamai atau bahkan melampaui dokter spesialis dalam beberapa kasus, memungkinkan deteksi dini penyakit yang dapat menyelamatkan nyawa.

Dalam bidang pendidikan, AI berpotensi mewujudkan visi pembelajaran yang benar-benar Kebocoran data melalui model AI merupakan risiko yang semakin dipahami dan dikhawatirkan oleh komunitas riset. Model machine learning, terutama model-model besar dengan kapasitas tinggi, memiliki kecenderungan untuk "mengingat" contoh-contoh spesifik dari data pelatihannya.

Dari perspektif sosial-ekonomi, otomatisasi berbasis AI mengancam lapangan pekerjaan di berbagai sektor, Konsentrasi kapabilitas AI di tangan segelintir perusahaan teknologi besar juga menimbulkan kekhawatiran mengenai monopoli data dan kekuasaan yang tidak seimbang.

Tantangan Etika, Privasi, dan Masa Depan Pekerjaan

Penggunaan AI dalam skala luas memunculkan tantangan etis yang fundamental mengenai nilai-nilai yang harus diutamakan dalam desain dan deployment sistem AI. Pertanyaan-pertanyaan seperti siapa yang bertanggung jawab ketika sistem AI membuat kesalahan, bagaimana memastikan keadilan dan non-diskriminasi dalam sistem AI, serta bagaimana menyeimbangkan manfaat agregat dengan perlindungan hak-hak individual, merupakan pertanyaan-pertanyaan yang belum memiliki jawaban yang jelas dan konsensual.

Kemampuan sistem AI untuk mengumpulkan, mengintegrasikan, dan menganalisis data dari berbagai sumber memungkinkan pembangunan profil individu yang sangat komprehensif dan intrusif. Dalam konteks ini, privasi tidak hanya terancam oleh pelanggaran keamanan dan kebocoran data, tetapi juga oleh praktik-praktik pengumpulan dan penggunaan data yang legal namun etis dipertanyakan.

Tantangan terhadap privasi semakin kompleks dengan kemajuan dalam teknologi biometrik berbasis AI. Pengenalan wajah, pengenalan suara, analisis gait, dan berbagai bentuk identifikasi biometrik lainnya memungkinkan pelacakan dan identifikasi individu dalam skala yang belum pernah terjadi sebelumnya. Beberapa negara telah mengimplementasikan sistem surveillance berbasis AI yang sangat ekstensif, menimbulkan kekhawatiran serius mengenai kebebasan sipil dan potensi penyalahgunaan oleh pemerintahan otoriter.

Dampak AI terhadap pasar tenaga kerja merupakan salah satu tantangan sosial-ekonomi yang paling signifikan. Sementara setiap revolusi teknologi dalam sejarah pada akhirnya menciptakan lebih banyak pekerjaan daripada yang dihilangkannya,

transisi jangka pendek dapat sangat menyakitkan bagi pekerja yang terdampak. Pekerjaan-pekerjaan yang bersifat rutin dan dapat diprediksi, baik di bidang kerah putih maupun kerah biru, paling rentan terhadap otomatisasi. Kecepatan perubahan yang diinduksi oleh AI juga menyulitkan adaptasi melalui pendidikan dan pelatihan ulang, yang secara tradisional merupakan mekanisme utama untuk transisi tenaga kerja antar sektor.

Analisis Kritis Terhadap Strategi Mitigasi Risiko

Menghadapi berbagai risiko keamanan data yang telah diidentifikasi, diperlukan strategi mitigasi yang komprehensif dan multi-dimensi. Pendekatan yang efektif harus mengintegrasikan intervensi teknis, organisasional, regulatoris, dan edukatif dalam kerangka yang koheren.

Dari perspektif teknis, prinsip privacy by design harus menjadi fondasi dalam pengembangan sistem AI. Pendekatan ini mengadvokasi integrasi pertimbangan privasi dan keamanan sejak tahap awal desain sistem, bukan sebagai afterthought atau add-on setelah sistem dikembangkan. Teknik-teknik seperti differential privacy, yang menambahkan noise terstruktur pada data untuk melindungi privasi individu sambil tetap memungkinkan analisis agregat yang bermakna, federated learning yang memungkinkan pelatihan model tanpa mengumpulkan data mentah secara terpusat, dan secure multi-party computation yang memungkinkan komputasi pada data terenkripsi, merepresentasikan kemajuan penting dalam rekonsiliasi antara utilitas AI dan perlindungan privasi.

Namun demikian, solusi teknis semata tidak cukup untuk mengatasi kompleksitas tantangan keamanan data dalam AI. Diperlukan kerangka tata kelola organisasional yang menetapkan akuntabilitas yang jelas, prosedur penilaian risiko yang sistematis, dan mekanisme audit yang independen. Organisasi yang mengadopsi AI perlu mengembangkan kapabilitas internal untuk memahami dan mengelola risiko yang terkait, termasuk pembentukan fungsi-fungsi seperti AI ethics board dan data protection officer yang memiliki otoritas dan sumber daya yang memadai.

Regulasi memainkan peran penting dalam menetapkan standar minimum dan menciptakan level playing field bagi semua aktor. Berbagai yurisdiksi telah mulai mengembangkan kerangka regulasi untuk AI, dengan European Union's AI Act menjadi salah satu yang paling komprehensif. Regulasi yang efektif harus menyeimbangkan antara perlindungan terhadap risiko dengan fleksibilitas untuk inovasi, serta harus cukup adaptif untuk mengakomodasi perkembangan teknologi yang sangat cepat. Di Indonesia, pengembangan regulasi AI masih dalam tahap awal, dan terdapat kebutuhan mendesak untuk mempercepat proses ini sambil belajar dari pengalaman yurisdiksi lain.

Edukasi dan peningkatan kesadaran masyarakat merupakan komponen penting dari strategi mitigasi yang holistik. Pengguna yang terinformasi lebih mampu membuat keputusan yang bijak mengenai penggunaan teknologi AI dan lebih waspada terhadap risiko yang mungkin mereka hadapi. Program-program literasi digital dan AI perlu dikembangkan dan diintegrasikan ke dalam kurikulum pendidikan formal maupun program-program pelatihan untuk masyarakat umum.

Kesimpulan

Berdasarkan analisis komprehensif yang telah dilakukan terhadap berbagai sumber literatur ilmiah, penelitian ini menghasilkan beberapa kesimpulan substantif mengenai keamanan data dalam penggunaan Artificial Intelligence. Pertama, perkembangan teknologi AI yang sangat pesat telah membawa transformasi fundamental dalam berbagai sektor kehidupan, namun bersamaan dengan itu juga memunculkan tantangan keamanan data yang kompleks dan multidimensi. Karakteristik inherent dari sistem AI yang sangat bergantung pada data dalam jumlah besar, dikombinasikan dengan kemampuannya untuk mempelajari dan menyimpan pola-pola dari data tersebut, menciptakan risiko-risiko

keamanan yang unik dan berbeda dari sistem teknologi informasi konvensional. Kedua, ancaman keamanan data dalam ekosistem AI mencakup spektrum yang luas, mulai dari serangan adversarial dan data poisoning di tingkat teknis, hingga kebocoran informasi melalui model, pencurian model, dan penyalahgunaan data di tingkat aplikasi. Berbagai sektor menghadapi manifestasi risiko yang berbeda sesuai dengan karakteristik data dan konteks penggunaannya. Ketiga, mitigasi risiko keamanan data dalam AI memerlukan pendekatan holistik yang mengintegrasikan intervensi teknis, tata kelola organisasional, kerangka regulasi, dan edukasi

masyarakat. Regulasi yang adaptif dan berbasis risiko diperlukan untuk memberikan perlindungan yang memadai tanpa menghambat inovasi. Keempat, keberhasilan dalam mengamankan ekosistem AI memerlukan kolaborasi aktif dari seluruh pemangku kepentingan, termasuk pengembang teknologi, organisasi pengguna, regulator, akademisi, dan masyarakat. Pembangunan kepercayaan publik terhadap AI sangat bergantung pada komitmen terhadap praktik yang aman dan bertanggung jawab. Kelima, meskipun AI menghadirkan risiko keamanan data yang substansial, risiko tersebut dapat dikelola melalui kombinasi pendekatan teknis, organisasional, dan regulatoris yang tepat. Oleh karena itu, diperlukan peningkatan penelitian lanjutan, penerapan prinsip *privacy by design* oleh pengembang, penguatan tata kelola oleh organisasi, percepatan regulasi oleh pemerintah, serta peningkatan literasi masyarakat agar penggunaan AI dapat berlangsung secara aman, etis, dan berkelanjutan.

Daftar Pustaka

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

https://global.oup.com/academic/product/superintelligence-9780199678112

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.

https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *Proceedings of the International Conference on Learning Representations (ICLR)*.

https://www.deeplearningbook.org/

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed. draft). Stanford University. web.stanford.edu

McCarthy, J. (2007). What is artificial intelligence? Stanford University.

https://cicerocq.wordpress.com/wpcontent/uploads/2018/08/trabalho1_ia_dinter_ufrgs_ufma_uema_cicero.pdf

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents.

https://3dvar.com/Ramesh2022Hierarchical.pdf

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

https://drive.usercontent.google.com/download?id=1PNxatm8ir2MI81iiLBJg1g44BvJOhFoz&export=download&authuser=0

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. *2017 IEEE Symposium on Security and Privacy (SP)*, 3-18.

https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=7958568&ref=

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

https://drive.usercontent.google.com/download?id=1ZnM1D8arEFtCnZGGd84DS600e7jdKOl5&export=download&authuser=0

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998-6008.

https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845 aa-Paper.pdf

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM, 64*(3), 107-115.

https://dl.acm.org/doi/pdf/10.1145/3446776